Bulletin

de la SOCIÉTÉ MATHÉMATIQUE DE FRANCE

Tome 149
Fascicule 1

2021

SOCIÉTÉ MATHÉMATIQUE DE FRANCE

Pages 1-235

Bull. Soc. Math. France
149 (1), 2021, p. 1-235

Sommaire

Bull. Soc. Math. France
149 (1), 2021, p. 1-235

Contents

Bull. Soc. Math. France
149 (1), 2021, p. 1-13

THE BOGOMOLOV–BEAUVILLE–YAU DECOMPOSITION FOR KLT PROJECTIVE VARIETIES WITH TRIVIAL FIRST CHERN CLASS – WITHOUT TEARS

BY FRÉDÉRIC CAMPANA

ABSTRACT. — We give a simplified proof (in characteristic zero) of the decomposition theorem for connected complex projective varieties with klt singularities and a numerically trivial canonical bundle. The proof mainly consists in reorganizing some of the partial results obtained by many authors and used in the previous proof but avoids those in positive characteristic by S. Druel. The single, to some extent new, contribution is an algebraicity and bimeromorphic splitting result for generically locally trivial fibrations with fibers without holomorphic vector fields. We first give the proof in the easier smooth case, following the same steps as in the general case, treated next. The last two words of the title are plagiarized from [4].

RÉSUMÉ (*La décomposition de Bogomolov-Beauville-Yau des variétés projectives klt à première classe de Chern triviale – sans larmes*). — Nous donnons une preuve simplifiée (en caractéristique zéro) du théorème de décomposition des variétés connexes et projectives complexes à singularités klt et fibré canonique numériquement trivial.Cette preuve consiste essentiellement en une réorganisation de la preuve originale basée sur des résultats partiels obtenus par divers auteurs, mais évite d'utiliser ceux de caractéristique positive obtenus par S. Druel. Le seul résultat nouveau, dans une certaine mesure, établit l'algébricité et le scindage méromorphe pour les fibrations génériquement localement triviales dont les fibres n'ont pas de champ de vecteur holomorphe non nul. Nous donnons tout d'abord la preuve dans le cas lisse, plus simple, suivant les mêmes étapes que dans le cas général, traité ensuite. Les deux derniers mots du titre plagient [4].

Texte reçu le 17 avril 2020, modifié le 8 juillet 2020, accepté le 21 octobre 2020.

FRÉDÉRIC CAMPANA, Université de Lorraine, Institut Elie Cartan, Nancy • *E-mail :* frederic.campana@univ-lorraine.fr

Mathematical subject classification (2010). — 14J32, 14E99, 32J25, 32Q20, 32Q25.

Key words and phrases. — Kähler–Einstein metrics, First Chern class, Hyperkähler varieties, Calabi–Yau varieties, Holonomy, Algebraic foliations, Fundamental group.

0037-9484/2021/1/$ 5.00
doi:10.24033/bsmf.2823

1. Introduction

When X is smooth, connected, compact Kähler, with $c_1(X) = 0$, the classical, metric, proof of the Bogomolov–Beauville–(Yau) decomposition theorem, given in [2] (the arguments of [6] being Hodge-theoretic), starts with a Ricci-flat Kähler metric ([26]) and then decomposes the universal cover X' of X according to De Rham theorem, in its holonomy factors. The Cheeger–Gromoll theorem then distinguishes the flat Euclidian factor $\mathbb{C}^s$ of X' from the (simply-connected) product P of the others (which are compact and with holonomy either $SU(m)$ or $Sp(k)$). The compactness of P combined with Bieberbach's theorem now imply that a finite étale cover of X is the product of a complex torus $\mathbb{C}^s/\Gamma$ with P.

We shall first give a different proof, but only for X smooth projective, of this product decomposition, weaker in the sense that P is not shown to be simply connected (see Theorem 2.1 below). Indeed, the proof does not go through the universal cover and uses neither the De Rham nor the Cheeger–Gromoll theorems.

This allows for its extension (given next) to the singular case obtained in [21], which uses many other partial results, among which are those of [18] and [14] (which plays a rôle analogous to that played by the Cheeger–Gromoll theorem). Our proof makes the step involving the delicate positive characteristic arguments of [14] superfluous. We, indeed, deduce the algebraicity of the foliation given by the flat factor of the holonomy from the splitting result (see Theorem 3.4) below, instead of using the Albanese map. This splitting result can be applied once the algebraicity of the leaves of the foliations given by the nonflat factors of the holonomy have been shown to be algebraic and without nonzero vector fields.

The author thanks Benoît Claudon and Mihai Păun for their help in reading the text and several discussions. After this text was posted on arXiv, the author received useful comments by S. Druel and H. Guenancia and thanks both of them too. He also thanks the referee for his careful reading and suggestions for making some statements more precise.

2. The smooth case

We treat this case first in order to show the steps in the general case in a simpler context.

THEOREM 2.1. — *Let X be a smooth connected complex projective manifold with $c_1(X) = 0$. There exists a finite étale cover of X, which is a product of an abelian variety with projective manifolds that are either irreducible symplectic or Calabi–Yau.*

REMARK 2.2. — The notions of irreducible symplectic and Calabi–Yau manifolds are defined as in [2]: either by the values of $h^{p,0}$, or by the holonomy of any Ricci-flat Kähler metric. We need the projectivity of X, because the Kähler version of [13] is not known. Our proof also does not show the finiteness of the fundamental groups of symplectic or Calabi–Yau manifolds. A partial solution to this finiteness property is given in Proposition 2.7 below, based on more general L^2-methods. A complete solution is also given in Proposition 2.9, but it does not (in an obvious way) extend to the singular case.

Proof of Theorem 2.1. — We equip X with any Ricci-flat Kähler metric ([26]). Let Hol^0 (or Hol) be its restricted holonomy (or holonomy) representation and $T_X = F \oplus (\bigoplus_i T_i)$ be a (local near any given point of X) splitting of the tangent bundle of X into factors that are irreducible for the action of Hol^0. These local factors also correspond to a local splitting of X into a direct product of Kähler submanifolds. In particular, these local products are regular holomorphic foliations. Here, F is the "flat" factor consisting of restricted holonomy-invariant tangent vectors. Now, Hol^0 is a normal subgroup of Hol, and Hol/Hol^0 acts by permutation on the factors of the restricted holonomy decomposition. Because the action of Hol/Hol^0 is induced by a representation $\pi_1(X) \to Hol/Hol^0$, the local holonomy decomposition of T_X above holds globally on a suitable finite étale cover of X.

We now replace X by such a finite étale cover and obtain a global product decomposition $T_X = F \oplus (\bigoplus_i T_i)$ by regular holomorphic foliations, the restricted holonomy of F being trivial, while the ones of T_i are irreducible and of the form $SU(m_i)$ or $Sp(k_i)$.

LEMMA 2.3. — *Let $T_X = \bigoplus_j E_j$ be a direct sum decomposition by foliations E_j, with $c_1(X) = 0$. Then, $c_1(E_j) = 0, \forall j$.*

Proof. — Assume not and let H be a polarization on X, with $n := dim(X)$. Then, $c_1(E_j).H^{n-1} \neq 0$, for some j. Since $\sum_j c_1(E_j).H^{n-1} = 0$, we get $c_1(E_h).H^{n-1} > 0$ for some h. It then follows from [13], Lemma 4.10, that E_h contains a subfoliation G with $\mu_{H,min}(G) > 0$ and by [13], Theorem 4.1, that K_X is not pseudo-effective, contrary to the hypothesis $c_1(T_X) = 0$.

A second, shorter, proof (suggested by the referee) consists in invoking the semistability of T_X with respect to any polarization, so that $c_1(E_j).H^{n-1} \leq 0, \forall j$. □

From the preceding Lemma 2.3, if $T_X = F \oplus (\bigoplus_i T_i)$ is the holonomy decomposition of T_X considered above for X smooth projective with $c_1(X) = 0$, we get that $c_1(F) = c_1(T_i) = 0, \forall i$.

LEMMA 2.4. — *The dual $T_i^\star$ of each T_i is not pseudo-effective (which means that for any polarization H and any given $k > 0$, $h^0(X, Sym^m(T_i^\star) \otimes H^k) = \{0\}$ for $m \geq m(k)$).*

Proof. — We proceed in two steps. From [17], §15.3, and Proposition 24.22, it follows that $Sym^m(T_i), \forall i, \forall m > 0$ is an irreducible representation and, hence, stable. Next, [11], Theorem 1.3 (or alternatively [21], Theorem 1.1) implies that $T_i^\star$ is not pseudo-effective for each i. □

From [13], Theorem 4.2, Lemma 4.6, we now get the first claim of the next result[1]

LEMMA 2.5. — *Each of the foliations T_i has algebraic leaves, which are compact*[2]*, since T_i is everywhere regular, and X is smooth. Thus, T_i defines a smooth (proper) fibration $f_i : X \to B_i$ on a smooth projective base B_i. Each of these fibrations is locally trivial with fiber F_i and becomes a product $X' = F_i \times B_i'$ after a suitable finite étale base-change $B_i' \to B_i$.*

Proof. — Second claim: let $C_i := F \oplus (\bigoplus_{\neq i} T_i)$ be the complement in T_X of T_i. This defines a regular holomorphic foliation locally over B_i, which is transversal to f_i, and thus shows that f_i is locally isotrivial over B_i. Third claim: it is sufficient to know that $Aut(F_i)$ is discrete or that $h^0(F_i, T_{F_i}) = 0$. However, this is easy, since F_i is a projective manifold with $c_1 = 0$ and irreducible nontrivial holonomy, which thus does not leave any tangent vector invariant, which implies the claimed vanishing by the Bochner principle. □

Consider any one of the projections $f_i : F_i \times B_i \to B_i$ (after a suitable finite étale cover). Then, $c_1(B_i) = 0$ and its holonomy decomposition is $F \oplus (\bigoplus_{j \neq i} T_i)$. Proceeding inductively on $dim(X)$, we obtain a decomposition in a product $X = (\times_i F_i) \times B$, where B is smooth projective with $c_1(B) = 0$ and trivial holonomy F.

The next lemma then concludes the proof of Theorem 2.1. □

LEMMA 2.6. — *([5]) Let X be a connected compact Kähler manifold with $c_1(X) = 0$ and with trivial restricted holonomy representation (relative to some Ricci-flat Kähler metric). Then, X is covered by a torus.*

The symplectic and the even-dimensional Calabi–Yau manifolds can be shown to have a finite fundamental group by L^2-methods that extend to the singular case. Another approach is given right after this first proof, which works more generally, for compact Riemannian manifolds with nonnegative Ricci curvature and *maximal* b_1 vanishing, but does not extend in any obvious way to the singular case.

1. Although not explicitly stated in [13], this is a main step of the proof of 4.2 and is suggested by the proof of Lemma 4.6 there. The explicit formulation was first given in [14], §8. Since only the particular case of a polarization H^{n-1} is used here, one could even alternatively apply [7].

2. By contradiction: if not, the leaf through a regular point of the boundary of the closure of a leaf should be contained in this boundary, and of the same dimension. In the singular case, this compactness fails, and more delicate arguments are required.

PROPOSITION 2.7. — *Let X be a connected compact Kähler manifold with $c_1(X) = 0$ and $\chi(\mathcal{O}_X) \neq 0$. Then, $\pi_1(X)$ is finite.*

Proof. — We give two proofs, both relying on [1].

First proof. This is the proof given in [10], Corollary 5.3, and Remark 5.5. By [10], Theorem 4.1, it is sufficient to show that $\kappa^+(X) \leq 0$, that is, $\kappa(X, det(F)) \leq 0$, for any subsheaf $F \subset \Omega^p_X, \forall p > 0$. This follows from the semistability of $\wedge^r\Omega^p_X, \forall r, p > 0$. Indeed, since K_X is trivial, $\Omega^p_X \cong (\Omega^{n-p}_X)^*$, and so any saturated subsheaf $D := det(F)$ of rank 1 of $\wedge^r\Omega^p_X$ is numerically trivial, since both D and $D^* = det(\Omega^p/F) \subset (\wedge^{r'}\Omega^p)^* = \wedge^{r'}\Omega^{n-p}$, have nonpositive slope with respect to any polarization.

Second proof. If $X' \to X$ is the universal cover, and h an L^2-holomorphic p-form on X', then h is parallel (because the Laplacian of its squared norm equals the square norm of its covariant derivative, and so is nonnegative everywhere. Gaffney's integration trick implies that the Laplacian identically vanishes, since h is L^2, and X' is complete). Thus, h comes from X and vanishes, if X' is noncompact. By [1], one gets $0 = \sum_{p\in\{0,n\}}(-1)^p h^0_{(2)}(X', \Omega^p_{X'}) = \chi_{(2)}(X', \mathcal{O}_{X'}) = \chi(X, \mathcal{O}_X) \neq 0$, which is a contradiction. □

COROLLARY 2.8. — *If X is a compact Kähler manifold of dimension n, and irreducible symplectic (or Calabi–Yau of even dimension), then $\pi_1(X)$ is finite, of cardinality dividing $(\frac{n}{2} + 1)$ (resp. 2).*

Proof. — Let $X' \to X$ be the (compact) universal cover of X, of degree d. We then have $\chi(\mathcal{O}_{X'}) = d.\chi(\mathcal{O}_X)$. On the other hand, X' is still irreducible symplectic (or Calabi–Yau), and so we have: $\chi(\mathcal{O}_{X'}) = \sum_{p=0}^{p=\frac{n}{2}}(-1)^{2p}h^0(X', \Omega^{2p}_{X'}) = \frac{n}{2} + 1$ (or $\chi(\mathcal{O}_{X'}) = \sum_{p\in\{0,n\}}(-1)^p h^0(X', \Omega^p_{X'}) = 2$). □

The following result applies to any Calabi–Yau manifold but does not immediately extend to the singular case.

PROPOSITION 2.9. — *Let M be a compact connected Riemannian manifold with nonnegative Ricci curvature, such that $b_1(M') = 0$, for any finite étale cover M' of M. The fundamental group of M is finite.*

Proof. — By [24], the growth of $\pi_1(M)$ is polynomial (of degree bounded by the dimension of M). From [20], $\pi_1(M)$ is virtually nilpotent. Thus, $\pi_1(M')$ is nilpotent and torsion free for some finite étale cover M' of M. Thus, $\pi_1(M')$ is either trivial or has an abelianization of positive rank. Since $b_1(M') = 0$, $\pi_1(M') = \{1\}$, hence the claim. □

3. The singular version

Let X be a complex projective variety with klt singularities whose first Chern class is zero, i.e. $c_1(X) = 0$. By [25], Chap. V, Corollary 4.9, the condition $K_X \equiv 0$ implies that K_X is $\mathbb{Q}$-trivial. We may, and shall, assume, by passing to an index-one cover, that the singularities of X are canonical, and that K_X is trivial. (Instead of [25], when the singularities are canonical, one could use either [22], Thm. 8.2, or [12], Thm. 3.1 applied to a resolution of X).

Remark 3.1. — Notice that passing to the index-one cover eliminates examples of rationally connected varieties with klt singularities and torsion canonical bundle, such as the Ueno surface, the quotient of $E \times E$ by $\mathbb{Z}_4$ acting diagonally by complex multiplication by $\sqrt{-1}$ on each factor, where E is the elliptic curve with this complex multiplication.

We denote by ω the unique Ricci-flat metric of X that belongs to a given Kähler class ([16]). We will see now that the steps of the previous proof extend to the singular context, using the results from [18], §8, 9 and [14], Prop. 4.10 and Prop. 3.13. The single new input here is the algebraicity criterion for foliations in Theorem 3.4 below, which makes superfluous the characteristic $p > 0$ methods and results by several authors used in [14]. The results of [22] used in [14] are also no longer needed.

Theorem 3.2 ([21]). — *Let X be a normal complex variety with klt singularities and with $c_1(T_X) = 0$. There exists a quasi étale cover $f : \widetilde{X} \to X$ with canonical singularities, which is a product $\widetilde{X} = \Pi_j Y_j \times A$, where A is an abelian variety, and $Y_j's$ are varieties with canonical singularities, trivial canonical bundle, and irreducible restricted holonomy either $Sp(k_j)$, or $SU(m_j)$ (see §3.1 below). The Y_j' respectively are said to be irreducible symplectic (or Calabi–Yau).*

Since, by [18]*, there always exists a finite quasi étale cover with full holonomy either $Sp(k_j)$ or $SU(m_j)$, these notions coincide with the usual ones up to such a cover.*

3.1. Restricted holonomy cover. — We consider ω the "EGZ" Ricci-flat metric on X constructed in [16]. As shown[3] in [18], Prop. 7.3, after a quasi étale cover, obtained from the permutation representation of the holonomy on the factors of the restricted holonomy, the tangent sheaf T_X of X decomposes as follows:

$$T_X = \mathcal{F} \oplus \Big(\bigoplus_i \mathcal{E}_i \Big) , \tag{1}$$

3. In the first version, Prop. 7.9 was quoted, instead of Prop. 7.3, which is sufficient for our purposes, as pointed out by S. Druel and H. Guenancia, whom the author thanks for this observation.

where the restricted holonomy of $\mathcal{F}$ is trivial, and the other ones are either $SU(n_i)$ or $Sp(k_i)$. The other properties of $\mathcal{E}_i$ used here are:

(i) *The sheaf $\mathcal{E}_i$ defines a nonsingular foliation of rank either n_i or $2k_i$. on X_{reg}.*
(ii) *The first Chern classes of $\mathcal{E}_i, \mathcal{F}$ are zero.*
(iii) *All the symmetric powers of $\mathcal{E}_i$ and their duals are irreducible representations of the holonomy factors and are stable, for any polarization on X.* The first property follows from standard representation theory, given the structure of the holonomy group. The stability is [18], Theorem 8.1, see also claim 9.17.
(iv) *In particular, we have $h^0(X, \mathcal{E}_i) = 0, \forall i$, by stability.*
(v) *The preceding properties still hold for any finite quasi étale cover of X.* Indeed, the Ricci-flat metric on X lifts to such covers, and the restricted holonomy decomposition lifts there too.
(vi) *The holonomy factors and their holonomy groups do not depend on the Ricci-flat Kähler metric chosen.*

3.2. Algebraic foliations. — Recall that a foliation on X is said to be algebraic if its leaves are so.

In the decomposition (1), the foliations $\mathcal{E}_i$ are algebraic. Indeed, by either [21] (or [11], Theorem 3.1) none of the $\mathcal{E}_i's$ is pseudo-effective[4]. We can, thus, apply [13], Theorem 4.2, Lemma 4.6, which implies that they are algebraic.

Our goal now is to show that $\mathcal{F}$ too is algebraic. This is true, if $T_X = \mathcal{F}$. We, thus, assume that some nonzero factor $\mathcal{E}_i$ appears in (1), and so:

$$T_X = \mathcal{G} \oplus \mathcal{E} \ , \tag{2}$$

where $\mathcal{E}$ has positive rank, and the properties (i)–(v) are satisfied. So, here we assume implicitly that $\mathcal{E}$ is one of the factors $\mathcal{E}_i$ in (1), and $\mathcal{G}$ is the sum of the other factors. Observe that $\mathcal{G}$ is a foliation, since the decomposition (1) is induced by the local holonomy splitting of X_{reg} (in general, the sum of two foliations need not be integrable).

LEMMA 3.3. — *Let X be an algebraic variety with canonical singularities and trivial first Chern class. Let ω be the* Ricci-*flat metric in some Kähler class on X, and $T_X = \mathcal{G} \oplus \mathcal{E}$ a corresponding decomposition as in the preceding lines. Assume that the foliation $\mathcal{G}$ is algebraic. Then:*

There exists a quasi étale cover $f : \widetilde{X} \to X$, where $\widetilde{X}$ has canonical singularities, and a product decomposition $\widetilde{X} = F \times Y$, which coincides at the tangent level with the decomposition $T_{\widetilde{X}} = f^{[*]}\mathcal{E} \oplus f^{[*]}\mathcal{G}$.

4. The result of [11] can, indeed, be applied on a resolution of the singularities of X, by lifting both the foliation and an ample class, since its argument deals with the general point of X only.

Proof. — The claim follows directly from [14], Prop. 4.10 (notice that the assumption $\tilde{q}(X) = 0$ there can be weakened to $\tilde{q}(F) = 0$, if F is the closure of a generic leaf of $\mathcal{E}$. The property $\tilde{q}(X) = 0$ is, indeed, used only to apply Prop. 4.8 of loc. cit., but 4.8 requires only the vanishing of $\tilde{q}$ for the fibers of $\mathcal{E}$). Now, $\tilde{q}(F) = 0$ follows from the properties (iii) and (v) of the holonomy factors quoted above. □

The algebraicity of $\mathcal{G}$ follows from Theorem 3.4 below, which, in fact, implies more: the bimeromorphic decomposition of X as a product, birationally, after a finite cover[5]. We may, and shall, assume that X has $\mathbb{Q}$-factorial terminal singularities by step 1 of the proof of Prop. 4.10 of [14]. By Prop. 3.13 of loc. cit, there is a Zariski open subset[6] X^0 of X, and a projective morphism $\varphi^0 : X^0 \to Y^0$, which is a locally trivial fibration in the analytic topology, its fibers being isomorphic to some F with $\tilde{q}(F) = 0$,by the properties (iv) and (v) of the holonomy factors quoted in §3.1 above. The conclusion then follows from the next algebraicity criterion for foliations.

THEOREM 3.4. — *Let X and Y be two Kähler[7] normal spaces and let $f : X \to Y$ be a surjective and proper holomorphic map with connected fibers. We denote by $\mathcal{E} := T_{X/Y}$ the foliation on X induced by f. We assume that:*

1. *f is a trivial fibration, locally in the analytic topology, with fiber F over some nonempty Zariski open set Y_0 of Y.*
2. *$h^0(F, T_F) = 0$, and so the automorphism group of F is discrete.*

Then, there is a finite map $\vartheta : V \to Y$, étale over Y_0, such that base-changing $f : X \to Y$ and normalizing the fiber-product $X_V := X \times_Y V$, we have a birational decomposition $\delta : X_V \dashrightarrow F \times V$, isomorphic over Y_0.

Moreover, if $\mathcal{G}$ is any distribution on X, such that $T_X = T_{X/Y} \oplus \mathcal{G}$ over $f^{-1}(U)$, for some nonempty analytically open $U \subset Y_0$, then: $\delta_*((id_X \times \vartheta)^*(\mathcal{G})) = \mathcal{H}$, where $\mathcal{H} := T_{X_V/F} \subset T_{X_V}$ is the horizontal foliation defined by the product decomposition of $T_{F\times V}$. In particular, $\mathcal{G}$ is an algebraic foliation, and is the unique distribution on X, which is everywhere transversal to $T_{X/Y}$ over some open subset $U \subset Y_0$ as above.

REMARK 3.5. — 1. The birational splitting after a generically finite base-change $V \to Y$ (but not necessarily étale over Y_0) always exists if X is projective (or Moishezon) under the single hypothesis (1) of Theorem 3.4. However, the algebraicity of $\mathcal{G}$ requires the hypothesis (2) as seen, for example, when $f : X \to Y$ is a morphism of Abelian varieties

5. S. Druel informed the author that one could also apply his Theorem 1.5 in [15]. Since the hypothesis, scope, and proofs of both results are different, it seems worth stating and proving Theorem 3.4.

6. Up to a finite étale cover of X^0, by shrinking the open set Y^0 of the proof.

7. Or in the class $\mathcal{C}$.

with positive-dimensional fibers, which has many horizontal nonalgebraic foliations.

2. There is certainly a bimeromorphic version of Theorem 3.4, where the generic fibers of f are assumed to be bimeromorphically equivalent, by similar arguments.
3. A global Kähler condition is required for the algebraicity of $\mathcal{G}$ to hold, as the following simple example shows. Let F be a projective $K3$ surface with an infinite and finitely generated group of automorphisms $G \subset Aut(F)$. Let $C = C'/\pi_1(C)$ be a curve of genus $g \geq 1$ with universal cover C', such that $\pi_1(C)$ admits a surjective group morphism $\rho : \pi_1(C) \to G$. Such a C exists when $g \geq m$, is the cardinality of some set of generators of G. One can choose $g = 1$ if and only if G is abelian, generated by at most 2 elements. Let $X := (C' \times F)/\pi_1(C)$, where $p \in \pi_1(C)$ acts on the right on $(C' \times F)$ by: $(c', f).p := (p^{-1}.c', \rho(p^{-1}).f)$. The foliation $\mathcal{G}'$ with leaves $C' \times \{f\}$ on X' induces a foliation $\mathcal{G}$ on X, which is not algebraic. In this case, $Iso(Z, X/C) = Iso^\star(Z, X/C)$ is irreducible, noncompact, isomorphic to $C" := C'/Ker\rho$, the Galois cover of C with group G. The leaves of $\mathcal{G}$ are isomorphic to $C"$.

Proof. — Let $\varphi : Z := F \times Y \to Y, \psi : F \times Y \to F$ be the projections onto the second (resp. first) factor. As in [8], §8, we define:

$$\mathrm{Iso}^*(Z, X/Y) \subset \mathcal{C}(Z \times_Y X/Y) \tag{3}$$

to be the subset of the relative Chow variety of $Z \times_Y X$ over Y parameterizing the graphs of isomorphisms of F-seen as a fiber of φ over a point $y \in Y_0$- to X_y, the fiber of f over y. According to [8], §8, $\mathrm{Iso}^*(Z, F)$ is a Zariski open subset (with countably many components if $Aut(F)$, which is here discrete, infinite) of the relative Chow scheme of $(Z \times_Y X/Y)$, which consists of cycles contained in one of the fibers of the fiber product over Y.

Let $\mathrm{Iso}(Z, X/Y)$ be the topological (i.e., Zariski here) closure of $\mathrm{Iso}^\star(Z, X/Y)$ in $\mathcal{C}(Z \times_Y X/Y)$. It consists of the union of the closures of the components of $\mathrm{Iso}^\star(Z, X/Y)$, all of these closures being proper over Y, and irreducible components of the Chow–Barlet scheme of $(Z \times_Y X/Y)$. It is equipped with a projection to Y, by restriction of the one on $\mathcal{C}(Z \times_Y X/Y)$. Since f is locally trivial over Y_0, the projection $\mathrm{Iso}^*(Z, X/Y) \to Y$ is open over Y_0. This projection is proper on each component of $\mathrm{Iso}^\star(Z, X/Y)$, since these irreducible components of $\mathrm{Iso}(Z, X/Y)$ are compact (essentially by a general result, [23]) of D. Lieberman, based on E. Bishop's theorem). Moreover, by the assumption (2) of Theorem 3.4, the fibers of $\mathrm{Iso}(Z, X/Y)$ to Y are discrete over Y_0. If V is an irreducible component of $\mathrm{Iso}(Z, X/Y)$, its projection $\vartheta : V \to Y$ is, thus, onto, and finite étale over, Y_0. Indeed, if $Y' \subset Y_0$ is any small analytic open subset over which $X_{Y'} := f^{-1}(Y') \cong Y' \times F$ is given, $Iso^\star(Z, X/Y)$ identifies naturally with $Y' \times Aut(F)$ over Y' and shows that $\vartheta : V \to Y$ is étale over Y_0.

We, thus, get a fiber product $X_V := X \times_Y V$, with the obvious projections $f_V : X_V \to V, g : X_V \to X$. Let $V_0 := \vartheta^{-1}(Y_0)$. Any $v \in V_0$ is, thus, equipped naturally with an isomorphism $ev_v : F \cong X_y, y := \vartheta(v)$. This evaluation map extends (see [8], §8, Prop. 1) meromorphically to $ev : F \times V \to X_V$, which is, thus, bimeromorphic and isomorphic over V_0.

In order to simplify notation, we replace X, Y, f by X_V, V, f_V, respectively, and identify via ev X_V with $F \times V = F \times Y$ (recall that ev is isomorphic over $V_0 = Y_0$), and all the assumptions of Theorem 3.4 are preserved. The projections of $X = F \times Y$ onto its second (or first) factor are denoted $f = \varphi$ and ψ.

To establish the last claim of Theorem 3.4, we only have to check that $\mathcal{G}$ coincides over Y_0 with the sheaf $T_{X/F} := \mathcal{H}$, which will also prove the algebraicity of $\mathcal{G}$.

We restrict everything over the open set $U \subset Y_0$ appearing in the last assumption of Theorem 3.4, so we assume that $X_U := f^{-1}(U) = F \times U$ and, thus, have a first decomposition $T_{X_U} = \psi^*(T_F) \oplus \mathcal{H}$, where $\mathcal{H}$ is the kernel of the map $d\psi : T_{X_U} \to \psi^*(T_F)$.

The second decomposition $T_{X_U} = \psi^*(T_F) \oplus \mathcal{G}$ gives equivalently an isomorphism $df_{|\mathcal{G}} : \mathcal{G} \to f^*(T_U)$ over X_U. Let $(df_{|\mathcal{G}})^{-1} : f^*(T_U) \to \mathcal{G}$ be its inverse. Let $\gamma := d\psi \circ (df_{|\mathcal{G}})^{-1} : f^*(T_U) \to \psi^*(T_F)$ be the composite map, seen as an element $\gamma \in H^0(X_U, f^*(\Omega^1_U) \otimes \psi^*(T_F))$. We have the following equalities:

$$\begin{aligned} H^0(X_U, f^*(\Omega^1_U) \otimes \psi^*(T_F)) &= H^0(U, \Omega^1_U \otimes f_*(\psi^*(T_F))) \\ &= H^0(U, \Omega^1_U \otimes \{0\}) = \{0\}. \end{aligned}$$

The last two equalities follow from assumption (2) of Theorem 3.4, which implies that $f_*(\psi^*(T_F)) = \{0\}$. This shows that $\mathcal{G} = \mathcal{H} = T_{Z/F}$ over X_U and so everywhere by analytic continuation. □

We can now conclude the proof of Theorem 3.2 by induction on $\dim(X)$, since we now know that (up to quasi étale covers) $X = Y \times Z$, in which Y is a product of varieties with canonical singularities, $c_1 = 0$, restricted holonomy either SU or Sp, and Z is in the same class of varieties but with trivial restricted holonomy (i.e., $T_Z = \mathcal{F}$). Theorem 3.2 then follows from:

3.3. A singular Bieberbach theorem. — Assume now that only the factor $\mathcal{F}$ appears in the decomposition (1). We are reduced to showing that if Z has canonical singularities, trivial first Chern class, and a trivial restricted holonomy group, it is covered by an abelian variety. However, this is just Corollary 1.16 in [19].

3.4. The fundamental group. — Let X be a complex projective variety with klt singularities and $K_X \equiv 0$. Recall that X is said to be irreducible symplectic (or Calabi–Yau) if its restricted holonomy representation for any, or some, EGZ

Ricci-flat Kähler metric is irreducible and of the form $Sp(m)$ (or $SU(n)$). In this situation, we have the following result, which is entirely similar to the smooth case:

THEOREM 3.6. — *(See* [18]*, 13.1) If $\chi(\mathcal{O}_X) \neq 0$, then $\pi_1(X')$ is finite, if $\rho : X' \to X$ is any resolution. This applies to irreducible symplectic varieties and to even-dimensional Calabi–Yau varieties: the cardinality of $\pi_1(X')$ lies in $[1, \frac{n}{2} + 1]$ in the first case (or in $[1, 2]$ in the second case).*

Since the map $\rho_*(\pi_1(X')) \to \pi_1(X)$ is surjective (by [9]*, Proposition 1.3), this implies the same statement for $\pi_1(X')$.*

Proof. — We apply [10], which says that $\pi_1(X')$ is finite if $\kappa(X', det(\mathcal{F})) \leq 0$ for any $\mathcal{F} \subset \Omega^p_{X'}), \forall p > 0$. Since the sections of $det(\mathcal{F})^{\otimes m}$ are sections of $Sym^m(\Omega^p_{X'})$, and the restrictions of these are reflexive sections, hence parallel over the regular locus of X by [18], Theorem 8.2, these sections are determined by their value in one single point of X_{reg}. Thus, $\kappa(X', det(\mathcal{F})) \leq 0$.

One could also argue as in the first proof of Proposition 2.7.

The invariant $\chi(\mathcal{O}_X)$ behaves as in the smooth case when X has klt and, thus, rational singularities. It is, in particular, multiplicative under finite étale covers.

If X is irreducible symplectic (or even-dimensional Calabi–Yau) and n-dimensional, we have: $h^0(X, \Omega^{[p]}_X) \leq h_{n,p}$, where $h_{n,p} = 0$ for p odd, and $h_{n,p} = 1$ for $p \leq n$ even (or $h_{n,p} = 0$ for $p \neq 0, n$, and $h_{n,p} = 1$ for $p = 0, n$), and so $\chi(\mathcal{O}_X)$ lies in $[1, \frac{n}{2} + 1]$ (or in $[1, 2]$). This shows the claim, since these inequalities still hold on the universal cover X" of X, and $\chi(\mathcal{O}_{X"}) = d.\chi(\mathcal{O}_X)$, where d is the degree of X" over X and also the cardinality of $\pi_1(X)$. □

REMARK 3.7. — Our proof of Theorem 3.6 differs slightly from the one in [18], 13.1, both relying on [10], and thus on [1]. See [10], §.5 for further remarks on this topic.

BIBLIOGRAPHY

[1] M. ATIYAH – "Elliptic operators, discrete groups, and Von Neumann algebras", *Astérisque* **32–33** (1976), p. 43–72.

[2] A. BEAUVILLE – "Variétés Kählériennes dont la première classe de Chern est nulle", *J. Diff. Geom.* **18** (1983), p. 755–782.

[3] A. BESSE – *Einstein manifolds*, Ergebnisse der Mathematik und ihrer Grenzgebiete, no. 10, Springer Verlag, 1987.

[4] F. BEUKERS, A. J. VAN DER POORTEN & R. VAN DER YAGER – "Dyson's lemma without tears", *Indagationes Mathematicae* **2** (1991), no. 1, p. 19–28.

[5] L. Bieberbach – "Über die Bewegungsgruppen der Euklidischen Raüme. I, II", *Math. Ann.* **70** (1911), p. 297–336, **72** (1912), p. 400-412.
[6] F. Bogomolov – "Hamiltonian Kähler manifolds", *Sov. Math. Dokl.* **19** (1978), p. 1462–1465.
[7] F. Bogomolov & M. McQuillan – "Rational curves on foliated varieties", 2001, IHES/M/01/07.
[8] F. Campana – "Réduction d'Albanese d'un morphisme propre et faiblement kählérien, ii", *Comp. Math.* **54** (1985), p. 373–398.
[9] ———, "On twistor spaces of class $\mathcal{C}$", *J. Diff. Geom.* **33** (1991), p. 451–459.
[10] ———, "Fundamental group and positivity of the cotangent bundles of compact Kähler manifolds", *J. Alg. Geom.* **4** (1995), p. 487–502.
[11] F. Campana, J. Cao & M. Păun – "Subharmonicity of direct images and applications", ArXiv.
[12] F. Campana & T. Peternell – "Geometric stability of the cotangent bundle and the universal cover of a projective manifold", *Bull. SMF* (2011).
[13] F. Campana & M. Păun – "Foliations with positive movable slope", (2019).
[14] S. Druel – "A decomposition theorem for singular spaces with trivial canonical class of dimension at most five", *Inv. math.* **211** (2018), p. 245–296.
[15] ———, "Codimension one foliations with numerically trivial canonical class on singular spaces", mars 2020, arXiv:1809.06905 (2018).
[16] P. Eyssidieux, V. Guej & A. Zeriahi – "Singular Kähler-Einstein metrics", *J. AMS.* **22** (2009), p. 607–639.
[17] W. Fulton & J. Harris – *Representation theory. A first course*, Springer-Verlag, 1991.
[18] D. Greb, H. Guenancia & S. Kebekus – "Klt varieties with trivial canonical class, holonomy, differential forms, and fundamental groups", *Geometry and Topology* **23** (2019), no. 4, p. 2051–2124.
[19] D. Greb, S. Kebekus & T. Peternell – "Etale fundamental group of Kawamata log-terminal spaces, flat sheaves and quotients of abelian varieties", *Duke Math. J.* **165** (2016), no. 10, p. 1965–2004.
[20] M. Gromov – "Groups of polynomial growth and expanding maps", *Publ. IHES* **53** (1981), p. 53–78.
[21] A. Höring & T. Peternell – "Algebraic integrability of foliations with numerically trivial canonical bundle", *Inv. Math.* **216** (2019), p. 395–419.
[22] Y. Kawamata – "Minimal models and the kodaira dimension of algebraic fiber spaces", *Journal für die reine und angewandte Mathematik* **363** (1985), p. 1–46.

[23] D. Lieberman – "Compactness of the Chow scheme: applications to automorphisms and deformations of Kähler manifolds", in *Séminaire François Norguet III*, Lecture Notes in Mathematics, no. 670, 1978, p. 140–186.
[24] J. Milnor – "A note on curvature and fundamental group", *J. Diff. Geom.* **2** (1968), p. 1–7.
[25] N. Nakayama – *Zariski decomposition and abundance*, MSJ Memoirs, no. 14, The Mathematical Society of Japan, Tokyo, 2004.
[26] S. Yau – "On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation I", *Comm. Pure and Appl. Math.* **31** (1978), p. 339–411.

Bull. Soc. Math. France
149 (1), 2021, p. 15-54

ON THE ORBITAL STABILITY OF A FAMILY OF TRAVELLING WAVES FOR THE CUBIC SCHRÖDINGER EQUATION ON THE HEISENBERG GROUP

BY LOUISE GASSOT

ABSTRACT. — We consider the focusing energy-critical Schrödinger equation on the Heisenberg group in the radial case

$$i\partial_t u - \Delta_{\mathbb{H}^1} u = |u|^2 u, \quad \Delta_{\mathbb{H}^1} = \frac{1}{4}(\partial_x^2 + \partial_y^2) + (x^2+y^2)\partial_s^2, \quad (t,x,y,s) \in \mathbb{R}\times\mathbb{H}^1,$$

which is a model for non-dispersive evolution equations. For this equation, the existence of global smooth solutions and the uniqueness of weak solutions in the energy space are open problems. We are interested in a family of ground-state travelling waves parameterized by their speed in $(-1,1)$. We show that the travelling waves of speed close to 1 present some orbital stability in the following sense. If the initial data is radial and close enough to one traveling wave, then there exists a global weak solution that stays close to the orbit of this travelling wave at all times. A similar result is proven for the limiting system associated to this equation.

Texte reçu le 30 septembre 2019, modifié le 9 septembre 2020, accepté le 2 octobre 2020.

LOUISE GASSOT, Département de mathématiques et applications, École normale supérieure, CNRS, PSL University, 75005 Paris, France, Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405 Orsay, France • *E-mail :* louise.gassot@universite-paris-saclay.fr

Mathematical subject classification (2010). — 35B35, 35C07, 35Q55, 43A80.

Key words and phrases. — Nonlinear Schrödinger equation, Traveling wave, Orbital stability, Heisenberg group, Dispersionless equation, Bergman kernel.

The author received no specific funding for this work.

0037-9484/2021/15/$ 5.00
doi:10.24033/bsmf.2824

Résumé (*Autour de la stabilité orbitale d'une famille d'ondes progressives pour l'équation de Schrödinger cubique sur le groupe de Heisenberg*). — On considère l'équation de Schrödinger énergie-critique sur le groupe de Heisenberg dans le cas radial

$$i\partial_t u - \Delta_{\mathbb{H}^1} u = |u|^2 u, \quad \Delta_{\mathbb{H}^1} = \frac{1}{4}(\partial_x^2 + \partial_y^2) + (x^2+y^2)\partial_s^2, \quad (t,x,y,s) \in \mathbb{R}\times\mathbb{H}^1,$$

qui est un modèle d'équation d'évolution non dispersive. Pour cette équation, dans l'espace d'énergie, l'existence globale de solutions régulières et l'unicité de solutions faibles sont des problèmes ouverts. On s'intéresse à une famille d'ondes progressives minimisantes paramétrées par leur vitesse dans $]-1,1[$. On montre que les ondes progressives dont la vitesse est proche de 1 possèdent des propriétés de stabilité orbitale au sens suivant. Pour toute donnée initiale radiale suffisamment proche d'une onde progressive, alors il existe une solution faible globale associée à cette donnée initiale qui reste proche de l'orbite de l'onde progressive en tout temps. Un résultat similaire est montré pour le système limite associé à cette équation.

1. Introduction

1.1. Motivation. — We are interested in the Schrödinger equation on the Heisenberg group

$$\begin{cases} i\partial_t u - \Delta_{\mathbb{H}^1} u = |u|^2 u \\ u(t=0) = u_0 \end{cases}, \quad (t,x,y,s) \in \mathbb{R}\times\mathbb{H}^1. \tag{1}$$

The operator $\Delta_{\mathbb{H}^1}$ denotes the sub-Laplacian on the Heisenberg group. When the solution is radial, in the sense that it only depends on t, $|x+iy|$ and s, the sub-Laplacian writes as

$$\Delta_{\mathbb{H}^1} = \frac{1}{4}(\partial_x^2 + \partial_y^2) + (x^2+y^2)\partial_s^2.$$

The Heisenberg group is a typical case of sub-Riemannian geometry where dispersive properties of the Schrödinger equation disappear (see Bahouri, Gérard and Xu [3]). To take it further, Del Hierro [12] proved sharp decay estimates for the Schrödinger equation on H-type groups, depending on the dimension of the center of the group. More generally, Bahouri, Fermanian and Gallagher [2] proved optimal dispersive estimates on stratified Lie groups of step 2 under some property of the canonical skew-symmetric form. In contrast, they also gave a class of groups without this property displaying a total lack of dispersion, which includes the Heisenberg group.

Dispersion impacts the way one can address the Cauchy problem for the Schrödinger equation. Indeed (see Burq, Gérard and Tzvetkov [7], Remark 2.12), the existence of a smooth local in the time-flow map defined on some Sobolev space $H^k(M)$ for the Schrödinger equation on a Riemannian manifold M with

the Laplace–Beltrami operator Δ

$$\begin{cases} i\partial_t u - \Delta u = |u|^2 u \\ u(t=0) = u_0 \end{cases}$$

implies the following Strichartz estimate

$$\|\mathrm{e}^{it\Delta} f\|_{L^4([0,1]\times M)} \leq C \|f\|_{H^{\frac{k}{2}}(M)}.$$

The argument also applies for the Heisenberg group with the homogeneous Sobolev spaces $\dot{H}^k(\mathbb{H}^1)$, for which the inequality holds if and only if $k \geq 2$ [9]. In particular, without a conservation law controlling the $\dot{H}^2$-norm, there is no existence result of global smooth solutions. Moreover, existence and uniqueness of weak solutions in the energy space $\dot{H}^1(\mathbb{H}^1)$ is an open problem, even if constructing global weak solutions to the Schrödinger equation on the Heisenberg group would still possible in the defocusing case through a compactness argument. Note that for weak solutions the energy of the solution is only bounded above by the initial energy. Therefore, the cancellation of the energy of the solution at some time may not imply that the solution is identically zero and does not exclude the possibility of non-uniqueness of weak solutions, as in the 2D incompressible Euler equation [15].

The aim of this paper is to construct some global weak solutions with a prescribed behaviour. More precisely, given initial data close to some ground-state travelling wave solution for the Schrödinger equation on the Heisenberg group, we want to construct a global weak solution that stays close to the orbit of the traveling wave at all times. Combined with a uniqueness result, this would lead to the orbital stability of this ground-state traveling wave.

1.2. Main results. — We consider a family of traveling waves with speed $\beta \in (-1, 1)$ in the form

$$u_\beta(t, x, y, s) = \sqrt{1-\beta} Q_\beta(x, y, s + \beta t).$$

The profile Q_β satisfies the following stationary hypoelliptic equation (with $D_s = -i\partial_s$)

$$-\frac{\Delta_{\mathbb{H}^1} + \beta D_s}{1-\beta} Q_\beta = |Q_\beta|^2 Q_\beta.$$

Because of the scaling invariance, it would have been equivalent in the rest of the study to define u_β as

$$u_\beta(t, x, y, s) = Q_\beta\left(\frac{x}{\sqrt{1-\beta}}, \frac{y}{\sqrt{1-\beta}}, \frac{s+\beta t}{1-\beta}\right).$$

From [8], we know that as β tends to 1, the ground-state solutions of speed β converge up to symmetries in $\dot{H}^1(\mathbb{H}^1)$ to some profile Q. Moreover, Q is the

solution to a limiting equation

$$D_s Q = \Pi_0^+(|Q|^2 Q), \tag{2}$$

for which the ground-state solution is unique up to symmetries equal to

$$Q(x,y,s) = \frac{i\sqrt{2}}{s + i(x^2+y^2) + i}.$$

The operator Π_0^+ is an orthogonal projector onto a relevant space for our analysis denoted by V_0^+. For more details, see Section 2.

We first prove a mild form of orbital stability for the ground state Q in the limiting equation and then focus on the orbital stability of the ground states Q_β in the Schrödinger equation on the Heisenberg group when β is close to 1.

Definition 1.1. — For $u \in \dot{H}^1(\mathbb{H}^1)$ and $X = (s_0, \theta, \alpha) \in \mathbb{R} \times \mathbb{T} \times \mathbb{R}_+^*$, we denote by $T_X u$ the element of $\dot{H}^1(\mathbb{H}^1)$ satisfying

$$T_X u(x,y,s) := \mathrm{e}^{i\theta} \alpha u(\alpha x, \alpha y, \alpha^2(s - s_0)), \quad (x,y,s) \in \mathbb{H}^1.$$

Let $\mathcal{M}$ be the orbit of Q

$$\mathcal{M} = \{T_X Q \mid X \in \mathbb{R} \times \mathbb{T} \times \mathbb{R}_+^*\},$$

then the distance of u to $\mathcal{M}$ is defined as

$$d(u, \mathcal{M}) = \inf_{X \in \mathbb{R}\times\mathbb{T}\times\mathbb{R}_+^*} \|u - T_X Q\|_{\dot{H}^1(\mathbb{H}^1)}.$$

Similarly, denote by $\mathcal{Q}_\beta$ the orbit of Q_β

$$\mathcal{Q}_\beta = \{T_X Q_\beta \mid X \in \mathbb{R} \times \mathbb{T} \times \mathbb{R}_+^*\},$$

then the distance of u to $\sqrt{1-\beta}\mathcal{Q}_\beta$ is

$$d(u, \sqrt{1-\beta}\mathcal{Q}_\beta) = \inf_{X \in \mathbb{R}\times\mathbb{T}\times\mathbb{R}_+^*} \|u - \sqrt{1-\beta} T_X Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)}.$$

Our first result is an orbital stability result for the profile Q associated the evolution problem linked to the limiting equation

$$\begin{cases} i\partial_t u = \Pi_0^+(|u|^2 u) \\ u(t=0) = u_0 \end{cases}. \tag{3}$$

Theorem 1.2 (Orbital stability for Q). — *There exist $c_0 > 0$ and $r_0 > 0$, such that the following holds. Let $r \le r_0$ and $u_0 \in \dot{H}^1(\mathbb{H}^1) \cap V_0^+$, such that*

$$\|u_0 - Q\|_{\dot{H}^1(\mathbb{H}^1)} < r^2. \tag{4}$$

Then there exists a weak solution $u \in \mathcal{C}(\mathbb{R}, \dot{H}^1(\mathbb{H}^1))$ (with the weak topology) to equation (3), such that for all $t \in \mathbb{R}$,

$$d(u(t), \mathcal{M}) \le c_0 r.$$

Using the links between the limiting equation and the Schrödinger equation, we deduce our second result: an orbital stability result for the profiles Q_β for the Schrödinger equation when β is close to 1 in the radial case.

THEOREM 1.3 (Orbital stability for Q_β). — *There exist $c_0 > 0$ and $r_0 > 0$, such that the following holds. Let $r \in (0, r_0)$. Then there exists $\beta_* \in (0,1)$, such that if $\beta \in (\beta_*, 1)$, and if $u_0 \in \dot{H}^1(\mathbb{H}^1)$ is radial and satisfies*

- *if $u_0 \in \dot{H}^1(\mathbb{H}^1) \cap V_0^+$:*

$$\|u_0 - \sqrt{1-\beta} Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)} < \sqrt{1-\beta} r^2 \tag{5}$$

- *in the general case:*

$$\|u_0 - \sqrt{1-\beta} Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)} < (1-\beta) r, \tag{6}$$

then there exists a weak radial solution $u \in \mathcal{C}(\mathbb{R}, \dot{H}^1(\mathbb{H}^1))$ (with the weak topology) to the Schrödinger equation on the Heisenberg group (1)

$$\begin{cases} i\partial_t u - \Delta_{\mathbb{H}^1} u = |u|^2 u \\ u(t=0) = u_0 \end{cases}$$

such that for all $t \in \mathbb{R}$, $\frac{u(t)}{\sqrt{1-\beta}}$ is close to the orbit of Q_β:

$$d\left(u(t), \sqrt{1-\beta}\mathcal{Q}_\beta\right) \leq c_0 \sqrt{1-\beta} r.$$

Note that, unlike the weak solutions discussed in the first part 1.1, the energy of the weak solutions from Theorem 1.2 (or Theorem 1.3) is controlled, indeed, this energy is very close to the one of the ground state Q (or $\sqrt{1-\beta}Q_\beta$). Furthermore, these two theorems would imply the orbital stability of Q and Q_β in the radial case in both situations if we had a uniqueness result for the solutions.

The assumption required on a general initial condition for the Schrödinger equation (6) is stronger than the assumption on an initial data already in V_0^+ (5). Indeed, for r fixed, and general initial data u_0, we will choose β_* sufficiently close to 1, so that the projection $\Pi_0^+(u_0)$ of u_0 onto V_0^+ satisfies $\|\Pi_0^+(u_0) - \sqrt{1-\beta}Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)} \leq C\sqrt{1-\beta}r^2$ (and so that the remainder term $(\mathrm{Id} - \Pi_0^+)(u_0)$ is also bounded). Note that thanks to the convergence of Q_β to Q with a rate $o(\sqrt{1-\beta})$ (see Appendix A), assumption (5) is comparable to assumption (4), up to a change of function $u \rightsquigarrow \frac{1}{\sqrt{1-\beta}} u(x, y, s - \beta t)$.

The key point in both proofs is the following local stability estimate for Q, which comes from the invertibility of the linearized operator around Q for the limiting equation (2) on a subspace of V_0^+ of finite co-dimension.

DEFINITION 1.4. — For $u \in \dot{H}^1(\mathbb{H}^1) \cap V_0^+$, we denote by $\mathcal{P}(u)$ (or, respectively, $\mathcal{E}(u)$) the momentum (or, respectively, the energy) for the limiting equation (3)

$$\mathcal{P}(u) := \|u\|^2_{\dot{H}^1(\mathbb{H}^1)}$$

or, respectively,

$$\mathcal{E}(u) := \|u\|^4_{L^4(\mathbb{H}^1)},$$

then define

$$\delta(u) := |\mathcal{P}(u) - \mathcal{P}(Q)| + |\mathcal{E}(u) - \mathcal{E}(Q)|.$$

PROPOSITION 1.5. — [8] *There exist $\delta_0 > 0$ and $C > 0$, such that for all $u \in \dot{H}^1(\mathbb{H}^1) \cap V_0^+$, if $\delta(u) \leq \delta_0$, then*

$$d(u, \mathcal{M})^2 \leq C\delta(u).$$

In order to prove Theorem 1.2, we construct the weak solution for the limiting initial value problem (3) as a limit of smooth functions. The approximating functions solve slightly modified equations, where we have restricted frequencies, so that the Cauchy problem is globally well posed. We show that we can control their distance to the orbit of the ground state Q using Proposition 1.5. Finally, we build modulation parameters that stay bounded on finite time intervals for the approximate solutions, and, through a compactness argument, we control the distance of the weak solution to the orbit of Q when passing to the limit.

For Theorem 1.3, the idea for the construction is the same, however we only have at our disposal the information on the limiting equation from Proposition 1.5. Therefore, we need to take advantage of the fact that Q_β is close to Q when β is close to 1. In this spirit, in order to tackle Theorem 1.3 for the speed β, we first introduce Cauchy problems for the Schrödinger equation (1) with a parameter γ increasing from β to 1. We display some continuity between the Cauchy problems, therefore it is possible to show their convergence to a Cauchy problem for the limiting equation as γ tends to 1. In the proof, we combine this strategy with the above method: we approximate by smooth functions the weak solutions to the Cauchy problems with parameter γ by restricting frequencies. Finally, we are able to get back to the problem with speed β by continuity and conclude in the same way as the proof of Theorem 1.2, by constructing bounded modulation parameters for the approximate solutions.

1.3. Comparison with other equations. — Concerning the focusing energy-critical Schrödinger equation on the Euclidean plane $\mathbb{R}^N$

$$i\partial_t u - \Delta u = |u|^{p_c - 1} u,$$

where $N \geq 3$ and $p_c = \frac{N+2}{N-2}$, there exists an explicit stationary solution

$$W(x) = \frac{1}{(1 + \frac{|x|^2}{N(N-2)})^{\frac{N-2}{2}}}.$$

The orbit $\{x \mapsto CW(\frac{x+x_0}{\lambda}) \mid (C, x_0, \lambda) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}_+^*\}$ of W is the set of minimizers for the Sobolev embedding $\dot{H}^1(\mathbb{R}^N) \hookrightarrow L^{2^*}(\mathbb{R}^N)$ (see the work of Talenti[19] and Aubin [1]). The energy $E(W) = \frac{1}{2}\|W\|_{\dot{H}^1(\mathbb{R}^N)} - \frac{1}{p_c+1}\|W\|_{L^{p_c+1}(\mathbb{R}^N)}$ and the $\dot{H}^1$ norm $\|W\|_{\dot{H}^1(\mathbb{R}^N)}$ play an important role in the dynamical behaviour of the solutions. Kenig and Merle [13] proved in the radial case that if $N \in \{3, 4, 5\}$ and the initial condition $u_0 \in \dot{H}^1(\mathbb{R}^N)$ satisfies $E(u_0) < E(W)$ and $\|u_0\|_{\dot{H}^1(\mathbb{R}^N)} < \|W\|_{\dot{H}^1(\mathbb{R}^N)}$, then the solution is global and scatters in $\dot{H}^1(\mathbb{R}^N)$, whereas if $E(u_0) < E(W)$ and $\|u_0\|_{\dot{H}^1(\mathbb{R}^N)} > \|W\|_{\dot{H}^1(\mathbb{R}^N)}$, then the solution must blow up in finite time.

The situation is different for the Schrödinger equation on the Heisenberg group. Indeed, from the equation satisfied by Q_β, one can see that the traveling waves

$$u_\beta(t, x, y, s) = \sqrt{1-\beta}Q_\beta(x, y, s + \beta t)$$

have a vanishing energy as β tends to 1:

$$E(u_\beta(t)) = \frac{1}{2}\|u_\beta(t)\|^2_{\dot{H}^1(\mathbb{H}^1)} - \frac{1}{4}\|u_\beta(t)\|^4_{L^4(\mathbb{H}^1)} \sim (1-\beta)\frac{\pi^2}{2} \to 0,$$

and, therefore, there exist solutions that do not scatter with arbitrary small energy.

A better parallel would be the mass-critical focusing half-wave equation on the real line

$$i\partial_t u + |D|u = |u|^2 u, \quad (t, x) \in \mathbb{R}\times\mathbb{R}, \tag{7}$$

where $D = -i\partial_x$, $\widehat{|D|f}(\xi) = |\xi|\widehat{f}(\xi)$. The half-wave equation in one dimension also presents some lack of dispersion and admits traveling waves with speed $\beta \in (-1, 1)$ (see Krieger, Lenzmann and Raphaël [14])

$$u(t, x) = Q_\beta\Big(\frac{x+\beta t}{1-\beta}\Big)\mathrm{e}^{-it},$$

where the profile Q_β is a solution to

$$\frac{|D| - \beta D}{1-\beta}Q_\beta + Q_\beta = |Q_\beta|^2 Q_\beta.$$

The profiles Q_β in the half-wave equation converge [10] as β tends to 1 in $H^{\frac{1}{2}}(\mathbb{R})$ to a ground-state solution Q to some limiting equation

$$DQ + Q = \Pi(|Q|^2 Q),$$

where Π is the Szegő projector from $L^2(\mathbb{R})$ onto the space $L^2_+(\mathbb{R}) = \{u \in L^2(\mathbb{R}) \mid \mathrm{Supp}(\widehat{u}) \subset \mathbb{R}_+\}$ of L^2 functions with nonnegative Fourier frequencies. From Q, we recover a traveling wave solution to the cubic Szegő equation

$$i\partial_t u = \Pi(|u|^2 u) \tag{8}$$

by setting $u(t,x) = Q(x-t)\mathrm{e}^{-it}$. Moreover, the linearized operator around Q is coercive [17]. One can deduce a similar estimate to Proposition 1.5, implying that the Szegő profile is orbitally stable in the relevant space for Q

$$H^{\frac{1}{2}}_+(\mathbb{R}) = \{u \in H^{\frac{1}{2}}(\mathbb{R}) \mid \mathrm{Supp}(\widehat{u}) \subset \mathbb{R}_+\}.$$

The following theorem comes from a graduate course given by P. Gérard and F. Rousset [11], for which no online notes are available (see Pocovnicu [16], Theorem 1.3, for a non-quantitative version, and [17], Proposition 6.3 applied to a zero Toeplitz potential, for a similar result with finite times).

THEOREM 1.6 (Orbital stability of Q for the Szegő equation). — *There exist $\varepsilon_0 > 0$ and $C > 0$ such that for all solution u of the Szegő equation* (8) *with initial condition $u_0 \in H^{\frac{1}{2}}_+(\mathbb{R})$, if*

$$\|u_0 - Q\|_{H^{\frac{1}{2}}(\mathbb{R})} \leq \varepsilon_0,$$

then

$$\sup_{t\in\mathbb{R}} \inf_{(\gamma,y)\in\mathbb{T}\times\mathbb{R}} \|e^{-i\gamma}u(t,\cdot - y) - Q\|^2_{H^{\frac{1}{2}}(\mathbb{R})} \leq C\|u_0 - Q\|_{H^{\frac{1}{2}}(\mathbb{R})}.$$

Gérard, Lenzmann, Pocovnicu and Raphaël [10] deduced the invertibility of the linearized operator for the half-wave equation around the profiles Q_β when β is close enough to 1. Their estimates imply the orbital stability of these profiles [11] indeed, one can prove a result similar to Proposition 1.5 for the profile Q_β with an adapted gap δ_β. However, this strategy does not work for the Schrödinger equation, as we will see at the beginning of Section 4.

THEOREM 1.7 (Orbital stability of Q_β for the half-wave equation). — *There exists $\beta_* \in (0,1)$, such that the following holds. Let $\beta \in (\beta_*, 1)$. Then there exist $\varepsilon_0(\beta) > 0$ and $C(\beta) > 0$, such that for all solution u of the half-wave equation* (7) *with initial condition $u_0 \in H^{\frac{1}{2}}(\mathbb{R})$, if*

$$\|u_0 - Q_\beta(\frac{\cdot}{1-\beta})\|_{H^{\frac{1}{2}}(\mathbb{R})} \leq \varepsilon_0(\beta),$$

then

$$\sup_{t\in\mathbb{R}} \inf_{(\gamma,y)\in\mathbb{T}\times\mathbb{R}} \|e^{-i\gamma}u(t,\cdot - y) - Q_\beta(\frac{\cdot}{1-\beta})\|^2_{H^{\frac{1}{2}}(\mathbb{R})} \leq C\|u_0 - Q_\beta(\frac{\cdot}{1-\beta})\|_{H^{\frac{1}{2}}(\mathbb{R})}.$$

In higher dimensions $d \geq 2$, traveling waves for the half-wave equation on $\mathbb{R}^d$

$$i\partial_t u + \sqrt{-\Delta}u = |u|^{p-1}u, \quad (t,x) \in \mathbb{R}\times\mathbb{R}^d,$$

are also orbitally stable in the radial case for mass-sub-critical non-linearities $1 < p < 1 + \frac{2}{n}$, but orbitally unstable in the mass-supercritical regime $1 + \frac{2}{n} < p < 1 + \frac{2}{n-1}$ [6]. Moreover, in the energy-critical and sub-critical case, Bellazzini, Georgiev, Lenzmann and Visciglia [5] proved that there can be no small data scattering in the energy space because of the existence of traveling waves with arbitrary small energy.

As we will see in this paper, we cannot directly adapt the proofs for the half-wave equation because we lack information on the Cauchy problem. A second complication arising in comparison to the half-wave equation is the fact that only two conservation laws are available (energy and momentum), because the masses of the ground states may be infinite (this fact is easy to check for Q, for instance). The method both for the Schrödinger equation on the Heisenberg group and for its limiting system is the construction of some weak solutions as a limit of smooth functions, and show that we can pass to the limit on their stability properties.

The paper is organized as follows. We first prove the orbital stability of Q for the limiting equation in Section 3. Then, we assess how close the solutions are to the limiting equation as β tends to 1 in order to study the orbital stability of Q_β for the Schrödinger equation in Section 4.

2. Notation

2.1. The Heisenberg group. — Let us now recall some facts about the Heisenberg group. We use coordinates and identify the Heisenberg group $\mathbb{H}^1$ with $\mathbb{R}^3$. The group multiplication is given by

$$(x, y, s) \cdot (x', y', s') = (x + x', y + y', s + s' + 2(x'y - xy')).$$

The Lie algebra of left-invariant vector fields on $\mathbb{H}^1$ is spanned by the vector fields $X = \partial_x + 2y\partial_s$, $Y = \partial_y - 2x\partial_s$ and $T = \partial_s = \frac{1}{4}[Y, X]$. The sub-Laplacian is defined as

$$\mathcal{L}_0 := \frac{1}{4}(X^2 + Y^2) = \frac{1}{4}(\partial_x^2 + \partial_y^2) + (x^2 + y^2)\partial_s^2 + (y\partial_x - x\partial_y)\partial_s.$$

When the function is radial, the sub-Laplacian coincides with the operator

$$\Delta_{\mathbb{H}^1} := \frac{1}{4}(\partial_x^2 + \partial_y^2) + (x^2 + y^2)\partial_s^2.$$

The space $\mathbb{H}^1$ is endowed with a smooth left-invariant measure, the Haar measure, which in the coordinate system (x, y, s) is the Lebesgue measure $\mathrm{d}\lambda_3(x, y, s)$. Sobolev spaces of positive order can then be constructed on $\mathbb{H}^1$ from powers of the operator $-\Delta_{\mathbb{H}^1}$, for example, $\dot{H}^1(\mathbb{H}^1)$ is the completion of the Schwarz space $\mathscr{S}(\mathbb{H}^1)$ for the norm

$$\|u\|_{\dot{H}^1(\mathbb{H}^1)} := \|(-\Delta_{\mathbb{H}^1})^{\frac{1}{2}} u\|_{L^2(\mathbb{H}^1)}.$$

2.2. Decomposition along the Hermite functions. — In order to study radial functions defined on the Heisenberg group $\mathbb{H}^1$ it is convenient to use their decomposition along Hermite-type functions (see, for example, [18], Chapters 12 and 13). The Hermite functions

$$h_m(x) = \frac{1}{\pi^{\frac{1}{4}} 2^{\frac{m}{2}} (m!)^{\frac{1}{2}}} (-1)^m \mathrm{e}^{\frac{x^2}{2}} \partial_x^m (\mathrm{e}^{-x^2}), \quad x \in \mathbb{R}, m \in \mathbb{N}$$

form an orthonormal basis of $L^2(\mathbb{R})$. In $L^2(\mathbb{R}^2)$, the family of products of two Hermite functions $(h_m(x)h_p(y))_{m,p\in\mathbb{N}}$ diagonalises the two-dimensional harmonic oscillator: for all $m, p \in \mathbb{N}$,

$$(-\Delta_{x,y} + x^2 + y^2) h_m(x) h_p(y) = 2(m + p + 1) h_m(x) h_p(y).$$

Given $u \in \mathscr{S}(\mathbb{H}^1)$, we will denote by $\widehat{u}$ its usual Fourier transform under the s variable, with corresponding variable σ

$$\widehat{u}(x, y, \sigma) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-is\sigma} u(x, y, s) \, \mathrm{d}s.$$

For $m, p \in \mathbb{N}$, set $\widehat{h_{m,p}}(x, y, \sigma) := h_m(\sqrt{2|\sigma|}x) h_p(\sqrt{2|\sigma|}y)$. Then, the family $(h_{m,p})_{m,p\in\mathbb{N}}$ diagonalises the sub-Laplacian:

$$\widehat{\Delta_{\mathbb{H}^1} h_{m,p}} = -(m + p + 1)|\sigma| \widehat{h_{m,p}}.$$

Let $k \in \{-1, 0, 1\}$ and denote by $\dot{H}^k(\mathbb{H}^1) \cap V_n^{\pm}$ the subspace of $\dot{H}^k(\mathbb{H}^1)$ spanned by $\{h_{m,p} \mid m, p \in \mathbb{N}, m + p = n\}$ in the following sense.

Definition 2.1. — Some $u_n^{\pm} \in \dot{H}^k(\mathbb{H}^1)$ belongs to $\dot{H}^k(\mathbb{H}^1) \cap V_n^{\pm}$ if there exists a family $(f_{m,p}^{\pm})_{m+p=n}$, such that

$$\widehat{u_n^{\pm}}(x, y, \sigma) = \sum_{\substack{m,p\in\mathbb{N}; \\ m+p=n}} f_{m,p}^{\pm}(\sigma) \widehat{h_{m,p}}(x, y, \sigma) \mathbb{1}_{\sigma \gtrless 0}.$$

For $u_n^{\pm} \in \dot{H}^k(\mathbb{H}^1) \cap V_n^{\pm}$, the $\dot{H}^k$-norm of $u_n^{\pm}$ writes as

$$\begin{aligned} \|u_n^{\pm}\|_{\dot{H}^k(\mathbb{H}^1)}^2 &= \int_{\mathbb{R}_{\pm}} ((n+1)|\sigma|)^k \int_{\mathbb{R}^2} |\widehat{u_n^{\pm}}(x, y, \sigma)|^2 \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}\sigma \\ &= \sum_{\substack{m,p\in\mathbb{N}; \\ m+p=n}} \int_{\mathbb{R}_{\pm}} ((n+1)|\sigma|)^k |f_{m,p}^{\pm}(\sigma)|^2 \frac{\mathrm{d}\sigma}{2|\sigma|}. \end{aligned}$$

Any function $u \in \dot{H}^k(\mathbb{H}^1)$ admits a decomposition along the orthogonal sum of the subspaces $\dot{H}^k(\mathbb{H}^1) \cap V_n^{\pm}$. Let us write $u = \sum_{\pm} \sum_{n\in\mathbb{N}} u_n^{\pm}$, where $u_n^{\pm} \in \dot{H}^k(\mathbb{H}^1) \cap V_n^{\pm}$ for all $(n, \pm)$. Then

$$\|u\|_{\dot{H}^k(\mathbb{H}^1)}^2 = \sum_{\pm} \sum_{n\in\mathbb{N}} \|u_n^{\pm}\|_{\dot{H}^k(\mathbb{H}^1)}^2.$$

For $k=0$, we get an orthogonal decomposition of the space $L^2(\mathbb{H}^1)$ and denote by $\Pi_n^{\pm}$ the associated orthogonal projectors.

The particular space $\dot{H}^k(\mathbb{H}^1)\cap V_0^+$ is spanned by a unique radial function h_0^+, satisfying

$$\widehat{h_0^+}(x,y,\sigma)=\frac{1}{\sqrt{\pi}}\mathrm{e}^{-(x^2+y^2)\sigma}\mathbb{1}_{\sigma\geq 0}.$$

Set $u\in\dot{H}^k(\mathbb{H}^1)\cap V_0^+$, then there exists f, such that

$$\widehat{u}(x,y,\sigma)=f(\sigma)\widehat{h_0^+}(x,y,\sigma),$$

and

$$\|u\|^2_{\dot{H}^k(\mathbb{H}^1)}=\int_{\mathbb{R}_+}|f(\sigma)|^2\frac{\mathrm{d}\sigma}{2\sigma^{1-k}}.$$

3. Orbital stability for the ground state Q in the limiting equation

In this section, we prove Theorem 1.2 on the orbital stability for the ground state Q in the limiting equation (3)

$$\begin{cases} i\partial_t u=\Pi_0^+(|u|^2u)\\ u(t=0)=u_0\in\dot{H}^1(\mathbb{H}^1)\cap V_0^+\end{cases},\quad (t,x,y,s)\in\mathbb{R}\times\mathbb{H}^1.$$

For convenience, in this part, we replace the elements $u\in\dot{H}^k(\mathbb{H}^1)\cap V_0^+$, $k\in\{-1,0,1\}$ with the corresponding function on the complex upper half-plane F_u defined as

$$F_u(s+i(x^2+y^2)):=u(x,y,s).$$

Due to the equality of Sobolev norms (10) below, we will see that for $k\in\{-1,0,1\}$, u belongs to $\dot{H}^k(\mathbb{H}^1)\cap V_0^+$ if and only if F_u belongs to the space of holomorphic functions $\dot{H}^{\frac{k}{2}}(\mathbb{C}_+)\cap\mathrm{Hol}(\mathbb{C}_+)$. Moreover, the Paley–Wiener theorem 3.2 enables us to identify the orthogonal projector Π_0^+ from $L^2(\mathbb{H}^1)$ onto its closed subspace $L^2(\mathbb{H}^1)\cap V_0^+$ with the orthogonal projector P_0 from $L^2(\mathbb{C}_+)$ onto the Bergman space $A_1^2=L^2(\mathbb{C}_+)\cap\mathrm{Hol}(\mathbb{C}_+)$. The projector P_0 is then a Bergman projector, which will be defined in equality (15). This change of functions then transforms the Cauchy problem for u into a Cauchy problem for F_u written as

$$\begin{cases} i\partial_t u=P_0(|u|^2u)\\ u(t=0)=u_0\in\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)\cap\mathrm{Hol}(\mathbb{C}_+)\end{cases},\quad (t,z)\in\mathbb{R}\times\mathbb{C}_+. \tag{9}$$

We establish the correspondence between u and F_u in Section 3.1. Then we construct some smooth functions approximating a weak solution of equation (9) in Section 3.2, prove their weak convergence in Section 3.3, and deduce from

their distance to the orbit of Q an upper bound on the distance of the weak limit to this orbit in Section 3.4.

3.1. Weighted Bergman spaces. — Letting $u \in \dot{H}^k(\mathbb{H}^1) \cap V_0^+$ we first note that $F_u \in \dot{H}^{\frac{k}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$. Indeed, the Fourier transform of u along the s variable corresponds to a function in $L^2(\mathbb{R}_+, \sigma^{k-1}\,\mathrm{d}\sigma)$; there exists $f \in L^2(\mathbb{R}_+, \sigma^{k-1}\,\mathrm{d}\sigma)$, such that $\widehat{u}(x,y,\sigma) = f(\sigma)\widehat{h_0^+}(x,y,\sigma)$. Therefore, we have

$$F_u(z) = \frac{1}{\pi\sqrt{2}} \int_0^{+\infty} e^{iz\sigma} f(\sigma)\,\mathrm{d}\sigma,$$

so that F_u is holomorphic. Moreover, the Sobolev norms of u and F_u are linked by

(10)

$$\|u\|^2_{\dot{H}^k(\mathbb{H}^1)} = \pi\|F_u\|^2_{\dot{H}^{\frac{k}{2}}(\mathbb{C}_+)} = \pi\|(-i\partial_z)^{\frac{k}{2}} F_u\|^2_{L^2(\mathbb{C}_+)} = \frac{1}{2}\int_0^{+\infty} |f(\sigma)|^2 \sigma^{k-1}\,\mathrm{d}\sigma.$$

For $k < 1$, F_u belongs to the weighted Bergman space A^2_{1-k}.

Definition 3.1 (Weighted Bergman spaces). — Given $k < 1$, the weighted Bergman space A^2_{1-k} is the subspace of $L^2_{1-k} := L^2(\mathbb{C}_+, \mathrm{Im}(z)^{-k}\,\mathrm{d}\lambda(z))$ composed of holomorphic functions of the complex upper half-plane $\mathbb{C}_+$:

$$A^2_{1-k} := \left\{ F \in \mathrm{Hol}(\mathbb{C}_+) \mid \|F\|^2_{L^2_{1-k}} := \int_0^{+\infty}\int_{\mathbb{R}} |F(s+it)|^2\,\mathrm{d}s\frac{\mathrm{d}t}{t^k} < +\infty \right\}.$$

Indeed, recall the Paley–Wiener theorem for Bergman spaces [4].

Theorem 3.2 (Paley–Wiener). — *For every $f \in L^2(\mathbb{R}_+, \sigma^{k-1}\,\mathrm{d}\sigma)$, $k < 1$, the following integral is absolutely convergent on $\mathbb{C}_+$,*

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \mathrm{e}^{iz\sigma} f(\sigma)\,\mathrm{d}\sigma, \tag{11}$$

and defines a function $F \in A^2_{1-k}$, which satisfies

$$\|F\|^2_{L^2_{1-k}} = \frac{\Gamma(1-k)}{2^{1-k}} \int_0^{+\infty} |f(\sigma)|^2 \sigma^{k-1}\,\mathrm{d}\sigma. \tag{12}$$

Conversely, for every $F \in A^2_{1-k}$, there exists $f \in L^2(\mathbb{R}_+, \sigma^{k-1}\,\mathrm{d}\sigma)$, such that (11) *and* (12) *hold.*

For $k = 1$, F_u belongs to the Hardy space $\mathcal{H}^2(\mathbb{C}_+)$.

Definition 3.3. — The Hardy space $\mathcal{H}^2(\mathbb{C}_+)$ of holomorphic functions of the upper half-plane $\mathbb{C}_+$ such that the following norm is finite:

$$\|F\|^2_{\mathcal{H}^2(\mathbb{C}_+)} := \sup_{t>0} \int_{\mathbb{R}} |F(s+it)|^2\,\mathrm{d}s < +\infty.$$

THEOREM 3.4 (Paley–Weiner). — *For every $f \in L^2(\mathbb{R}_+)$, the following integral is absolutely convergent on $\mathbb{C}_+$*

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \mathrm{e}^{iz\sigma} f(\sigma)\, \mathrm{d}\sigma \tag{13}$$

and defines a function F in the Hardy space $\mathcal{H}^2(\mathbb{C}_+)$, which satisfies

$$\|F\|^2_{\mathcal{H}^2(\mathbb{C}_+)} = \int_0^{+\infty} |f(\sigma)|^2\, \mathrm{d}\sigma. \tag{14}$$

Conversely, for every $F \in \mathcal{H}^2(\mathbb{C}_+)$, there exists $f \in L^2(\mathbb{R}_+)$, such that (13) *and* (14) *hold.*

In the following, we will work with the holomorphic representations, the solutions being valued in the Hardy space $\mathcal{H}^2(\mathbb{C}_+) = \dot{H}^{\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$.

3.2. Construction of approximate solutions. — Given an initial data $u_0 \in \mathcal{H}^2(\mathbb{C}_+) = \dot{H}^{\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$ close enough to the ground state

$$Q(z) = \frac{i\sqrt{2}}{z+i},$$

we want to construct a global solution to the Cauchy problem (9)

$$\begin{cases} i\partial_t u = P_0(|u|^2 u), \quad (t,z) \in \mathbb{R} \times \mathbb{C}_+ \\ u(t=0) = u_0 \end{cases},$$

which stays close to Q (up to symmetries) at all times. The Bergman projection P_0 from $L^2(\mathbb{C}_+)$ to A^2_1 writes as (see, e.g. [4])

$$P_0(u)(z) = -\frac{1}{\pi} \int_{\mathbb{R}_+} \int_{\mathbb{R}} \frac{1}{(z-s+it)^2} u(s+it)\, \mathrm{d}s\, \mathrm{d}t, \quad z \in \mathbb{C}_+. \tag{15}$$

We approximate u by functions with higher regularity, satisfying equations for which we can use a classical global well-posedness result.

Construction of smoothing projectors $\widetilde{P}_{\varepsilon,M}$: For $\varepsilon, M > 0$, we define the projector $\widetilde{P}_{\varepsilon,M}$ as follows. Write $u \in \dot{H}^k(\mathbb{C}_+) \cap \mathrm{Hol}(C_+)$, $k \leq \frac{1}{2}$ (or $u \in H^k(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$, $k \geq 0$) as

$$u(z) = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} e^{iz\sigma} f(\sigma)\, \mathrm{d}\sigma,$$

and then

$$\widetilde{P}_{\varepsilon,M}(u)(z) := \frac{1}{\sqrt{2\pi}} \int_{\varepsilon}^{M} e^{iz\sigma} f(\sigma)\, \mathrm{d}\sigma.$$

This projector removes the high and low frequencies of u in order to add some regularity in the solutions. It defines a bounded projector from $\dot{H}^k(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$ to itself for $k \leq \frac{1}{2}$ and from $H^k(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$ to itself for $k \geq 0$.

Construction of a sequence of approximate solutions $(u_n)_n$*:* We consider $f \in L^2(\mathbb{R}_+)$, such that for all $z \in \mathbb{C}_+$,

$$u_0(z) = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} e^{iz\sigma} f(\sigma)\, \mathrm{d}\sigma,$$

which satisfies

$$\|u_0\|^2_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} = \frac{1}{2}\|f\|^2_{L^2(\mathbb{R}_+)}.$$

Let us fix a sequence of positive numbers $(\varepsilon_n)_n$ going to zero and consider the following initial data belonging to $H^2(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$

$$u_0^n(z) := \widetilde{P}_{\varepsilon_n, \frac{1}{\varepsilon_n}} u_0(z) = \frac{1}{\sqrt{2\pi}} \int_{\varepsilon_n}^{1/\varepsilon_n} e^{iz\sigma} f(\sigma)\, \mathrm{d}\sigma.$$

We denote by $H^2_\varepsilon(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$ the space of functions $u \in H^2(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$ satisfying $\widetilde{P}_{\varepsilon, \frac{1}{\varepsilon}}(u) = u$. On this space, the $\dot{H}^k$-norms, $k \geq 0$, are equivalent:

$$\varepsilon^{2k}\|u\|^2_{L^2(\mathbb{C}_+)} \leq \|u\|^2_{\dot{H}^k(\mathbb{C}_+)} = \frac{1}{2}\int_\varepsilon^{1/\varepsilon} \sigma^{2k-1}|f(\sigma)|^2\, \mathrm{d}\sigma \leq \frac{1}{\varepsilon^{2k}}\|u\|^2_{L^2(\mathbb{C}_+)}.$$

Define the projection P_0^n as

$$P_0^n = \widetilde{P}_{\varepsilon_n, \frac{1}{\varepsilon_n}} \circ P_0.$$

We consider the following Cauchy problem

$$\begin{cases} i\partial_t u_n = P_0^n(|u_n|^2 u_n) \\ u_n(t=0) = u_0^n \end{cases}, \tag{16}$$

which is globally well posed in $H^2_{\varepsilon_n}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$.

Proposition 3.5. — *Let* $u_0^n \in H^2_{\varepsilon_n}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$*. Then there exists a unique solution* $u_n \in \mathcal{C}^\infty(\mathbb{R}, H^2_{\varepsilon_n}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+))$ *of* (16) *in the distribution sense.*

Proof. — The local existence comes from the Cauchy–Lipschitz theory for ODEs. Indeed, P_0 defines a bounded projector from $H^2(\mathbb{C}_+)$ onto $H^2(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$, and, therefore, P_0^n defines a bounded projector from $H^2(\mathbb{C}_+)$ onto $H^2_{\varepsilon_n}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$. Let $r := \|u_0^n\|_{H^2(\mathbb{C}_+)}$ and denote by $B(u_0, r)$ the ball centered at u_0 of radius r in $H^2_{\varepsilon_n}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$. Since $H^2(\mathbb{C}_+)$ is an algebra, we deduce that there exists $T = T(r)$, such that the map

$$\begin{aligned} \mathcal{C}([-T,T], B(u_0,r)) &\to \mathcal{C}([-T,T], B(u_0,r)) \\ v &\mapsto \left(t \mapsto v_0 + \frac{1}{i}\int_0^t P_0^n(|v|^2 v)(\tau)\, \mathrm{d}\tau\right) \end{aligned}$$

defines a contraction mapping from $\mathcal{C}([-T,T], B(u_0,r))$ to itself. We conclude the local well-posedness of equation (16). Moreover, the time of existence of

the solution is bounded below by some constant which only depends on the norm of the initial data in $H^2(\mathbb{C}_+)$.

In order to prove that local solutions extend globally in time, we show that there is no blow-up of the H^2 norm in finite time. Due to the equivalence of the $\dot{H}^{\frac{1}{2}}$ norm and the H^2 norm in $H^2_{\varepsilon_n}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$, it is enough to prove that equation (16) has conserved momentum

$$\mathcal{P}(u) := (u, -iu_z)_{L^2(\mathbb{C}_+)} = \|u\|^2_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)}.$$

However, using the equation a solution u satisfies

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}(u) &= 2\,\mathrm{Re}(\partial_t u, -i\partial_z u)_{L^2(\mathbb{C}_+)} \\ &= 2\,\mathrm{Re}(P_0^n(|u|^2 u), \partial_z u)_{L^2(\mathbb{C}_+)} \\ &= 2\,\mathrm{Re}(|u|^2 u, -i\partial_z u)_{L^2(\mathbb{C}_+)}.\end{aligned}$$

By integration by parts, one knows that the complex scalar product $(|u|^2 u, -i\partial_z u)_{L^2(\mathbb{C}_+)}$ is imaginary, leading to the conservation of momentum. □

Similarly, one can show that the energy $\mathcal{E}(u) = \|u\|^4_{L^4(\mathbb{C}_+)}$ is also conserved, using the equation,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(u) &= 2\,\mathrm{Re}(\partial_t u, |u|^2 u)_{L^2(\mathbb{C}_+)} \\ &= 2\,\mathrm{Re}(-iP_0^n(|u|^2 u), |u|^2 u)_{L^2(\mathbb{C}_+)} \\ &= 2\,\mathrm{Re}(-iP_0^n(|u|^2 u), P_0^n(|u|^2 u))_{L^2(\mathbb{C}_+)} \\ &= 0.\end{aligned}$$

We now show that $u_n(t)$ is close to the orbit $\mathcal{M}$ of the ground state Q. Dues to Proposition 1.5, it is enough to focus on $\delta(u_n(t))$. However, using the conservation laws, we know that for all $t \in \mathbb{R}$,

$$\delta(u_n(t)) = \delta(u_0^n).$$

Moreover, by construction of u_0^n, we know that $\|u_0^n - u_0\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)}$ tends to 0 as n tends to $+\infty$, and, therefore, $\delta(u_0^n)$ tends to $\delta(u_0)$. Assume that $\delta(u_0) < \delta_0$, then $\delta(u_0^n) < \delta_0$ after some rank N. Due to Proposition 1.5, we deduce that for all $n \geq N$ and $t \in \mathbb{R}$,

$$d(u_n(t), \mathcal{M})^2 \leq C\delta(u_0^n). \tag{17}$$

3.3. Weak convergence. — In this section, we show that the sequence $(u_n)_n$ has a weak limit u, which is a weak solution to equation (9). In order to do so, we first prove that $t \mapsto \partial_t u_n(t)$ is uniformly bounded in $\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)$ and then use Ascoli's theorem.

Because of the conservation of the momentum and the fact that $\mathcal{P}(u_0^n) \leq \mathcal{P}(u_0)$, for all $n \in \mathbb{N}$, we know that for all $n \in \mathbb{N}$ and $t \in \mathbb{R}$,

$$\|u_n(t)\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} \leq \|u_0\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)}.$$

Using the equation satisfied by u_n, we also know that

$$\|\partial_t u_n(t)\|_{\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)} \leq \|P_0^n(|u_n|^2 u_n)(t)\|_{\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)}.$$

By the dual Sobolev embedding $L^{\frac{4}{3}}(\mathbb{C}_+) \hookrightarrow \dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)$ and the fact that P_0 extends to a bounded projector from $L^p(\mathbb{C}_+)$ to itself as soon as $1 < p < +\infty$ (see, for instance, [4], Theorem 1.34), we can estimate that

$$\begin{aligned}
\|P_0^n(|u_n|^2 u_n)(t)\|_{\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)} &\leq \|P_0(|u_n|^2 u_n)(t)\|_{\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)} \\
&\leq C\|P_0(|u_n|^2 u_n)(t)\|_{L^{\frac{4}{3}}(\mathbb{C}_+)} \\
&\leq C'\||u_n|^2 u_n(t)\|_{L^{\frac{4}{3}}(\mathbb{C}_+)} \\
&\leq C'\|u_n(t)\|^3_{L^4(\mathbb{C}_+)}.
\end{aligned}$$

Since $u_n(t)$ is uniformly bounded in $\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)$ and, therefore, in $L^4(\mathbb{C}_+)$, we conclude that the term $\|\partial_t u_n(t)\|_{\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)}$ is also uniformly bounded.

We now prove that for all $T > 0$, up to a subsequence $(u_n)_n$ converges in $\mathcal{C}([-T,T], \dot{H}^{\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+))$ (with the weak topology) to a function u.

We know that $\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$ is separable, since it is isometric to $L^2(\mathbb{R}_+)$. Moreover, by removing the high frequencies of the Fourier function f at infinity, one can see that $\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+) \cap \dot{H}^{\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$ is dense in $\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$. We can, therefore, consider a countable sequence $(\varphi_k)_k$ in $\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+) \cap \dot{H}^{\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$, such that every function in $\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$ can be approximated by a subsequence of $(\varphi_k)_k$ for the $\dot{H}^{-\frac{1}{2}}$-norm.

Fix $k \in \mathbb{N}$. Since $(t \mapsto \partial_t u_n(t))_n$ and $(t \mapsto u_n(t))_n$ are uniformly bounded in $\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)$ and in $\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)$, respectively, the sequence $\ell_n(\cdot, \varphi_k) : t \in [-T,T] \mapsto (u_n(t), \varphi_k)$ is equicontinuous and equibounded, for all n and t,

$$|\partial_t \ell_n(t, \varphi_k)| = |(\partial_t u_n(t), \varphi_k)| \leq \|\partial_t u_n(t)\|_{\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)} \|\varphi_k\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)}$$

and

$$|\ell_n(t, \varphi_k)| = |(u_n(t), \varphi_k)| \leq \|u_n(t)\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} \|\varphi_k\|_{\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)}.$$

Applying Ascoli's theorem, for every $k \in \mathbb{N}$, there is a subsequence $(n_p)_p$, such that $(\ell_{n_p}(\cdot, \varphi_k))_p$ converges in $\mathcal{C}([-T,T], \mathbb{C})$ to some continuous function $\ell(\cdot, \varphi_k)$ as p tends to $+\infty$. By a diagonal argument, we can use the same subsequence for all $k \in \mathbb{N}$. Using a second diagonal argument on a sequence of times $(T_n)_n$ going to $+\infty$, we can assume that for all k, there exists $\ell(\cdot, \varphi_k) \in \mathcal{C}(\mathbb{R}, \mathbb{C})$, such that for all $T > 0$, the sequence $(\ell_{n_p}(\cdot, \varphi_k))_p$ converges in $\mathcal{C}([-T,T], \mathbb{C})$ to $\ell(\cdot, \varphi_k)|_{[-T,T]}$.

By density, ℓ extends to a bounded linear map $\ell \in \mathcal{C}(\mathbb{R}, (\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+))^*)$ (with weak topology). Now, by duality, ℓ can be represented by $u \in \mathcal{C}(\mathbb{R}, \dot{H}^{\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+))$, for all $\varphi \in \dot{H}^{-\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$,

$$\ell(t, \varphi) = (u(t), \varphi).$$

To conclude, by construction, for all $T > 0$, the sequence $(\ell_{n_p}|_{[-T,T]})_p$ converges weakly to $\ell|_{[-T,T]}$ in the space $\mathcal{C}([-T,T], (\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+))^*)$; therefore, $(u_n)_n$ converges weakly to u in $\mathcal{C}([-T,T], \dot{H}^{\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+))$. Passing to the limit we conclude that u is a global solution to the original equation (9) in the distribution sense.

We deduce that

$$\begin{aligned} d(u(t), \mathcal{M})^2 &= \inf_{X \in \mathbb{R}\times\mathbb{T}\times\mathbb{R}_+^*} \|u(t) - T_X Q\|^2_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} \\ &\leq \inf_{X \in \mathbb{R}\times\mathbb{T}\times\mathbb{R}_+^*} \liminf_{n\to+\infty} \|u_n(t) - T_X Q\|^2_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)}. \end{aligned}$$

Since X is not compact, this inequality is not sufficient if we want to apply inequality (17) to estimate $d(u(t), \mathcal{M})$. In the following part, we construct a map $t \mapsto X_n(t)$, such that for all $t \in \mathbb{R}$, $u_n(t)$ is close to $T_{X_n(t)}Q$, and $(X_n(t))_{n\in\mathbb{N}}$ stays bounded, then use a compactness argument.

3.4. Modulation. — Recall the notations in the Introduction. Fixing $u \in \mathcal{H}^2(\mathbb{C}_+) = \dot{H}^{\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$ and $X = (s, \theta, \alpha) \in \mathbb{R} \times \mathbb{T} \times \mathbb{R}_+^*$, we denote by $T_X u$ the element of $\mathcal{H}^2(\mathbb{C}_+)$ satisfying

$$T_X u(z) := \mathrm{e}^{i\theta} \alpha u(\alpha^2(z - s)), \quad z \in \mathbb{C}_+.$$

We write $X^{-1} = (-s, -\theta, \alpha^{-1})$ and

$$|X| := |s| + |\theta| + |\log(\alpha)|.$$

We have also defined the orbit of Q as

$$\mathcal{M} = \{T_X Q \mid X \in \mathbb{R} \times \mathbb{T} \times \mathbb{R}_+^*\},$$

and the distance of u to $\mathcal{M}$ as

$$d(u, \mathcal{M}) = \inf_{X=(s,\theta,\alpha)\in\mathbb{R}\times\mathbb{T}\times\mathbb{R}_+^*} \|T_X u - Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)}.$$

We choose $0 < r < 1$ and assume that $\|u_0 - Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} < r^2$. Given $K > 0$, for $n \geq N$ large enough, the regularized initial data u_0^n satisfies $\delta(u_0^n) < Kr^2$. Using the conservation of energy and momentum and Proposition 1.5, we deduce that there exist $c_0 > 0$ and $r_0 > 0$, such that if $0 < r < r_0$, then $d(u_n(t), \mathcal{M}) < c_0 r$ for all $t \in \mathbb{R}$.

We start from the observation that around time $t = 0$, we can choose $X_n(t) = (0, 0, 1)$ for all $n \geq N$, since $\|u_0^n - Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} < c_0 r$. By continuity, we know that $\|u_n(t) - Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} \leq (1 + \varepsilon) c_0 r$ on some small time interval, which can

be taken independently of n. Indeed, using the conservation of momentum, we have

$$\begin{aligned}\|u_n(t)-Q\|^2_{\dot H^{\frac12}(\mathbb{C}_+)} &= \|u_n(t)\|^2_{\dot H^{\frac12}(\mathbb{C}_+)} + \|Q\|^2_{\dot H^{\frac12}(\mathbb{C}_+)} - 2(u_n(t),Q)_{\dot H^{\frac12}(\mathbb{C}_+)}\\ &= \|u_0^n\|^2_{\dot H^{\frac12}(\mathbb{C}_+)} + \|Q\|^2_{\dot H^{\frac12}(\mathbb{C}_+)} - 2(u_n(t),-iQ_z)_{L^2(\mathbb{C}_+)},\end{aligned}$$

and, therefore, the derivative of $\|u_n(t)-Q\|^2_{\dot H^{\frac12}(\mathbb{C}_+)}$ is bounded by

$$\begin{aligned}\left|\frac{\mathrm{d}}{\mathrm{d}t}\|u_n(t)-Q\|^2_{\dot H^{\frac12}(\mathbb{C}_+)}\right| &= \left|2(\partial_t u_n(t),-iQ_z)_{L^2(\mathbb{C}_+)}\right|\\ &\le 2\|\partial_t u_n(t)\|_{\dot H^{-\frac12}(\mathbb{C}_+)}\|-iQ_z\|_{\dot H^{\frac12}(\mathbb{C}_+)}.\end{aligned}$$

However, we have already seen that $\|\partial_t u_n(t)\|_{\dot H^{-\frac12}(\mathbb{C}_+)}$ is bounded independently of t and n, and, therefore, there exists $K>0$, such that for $n\ge N$ and $t\in\mathbb{R}$

$$\begin{aligned}\|u_n(t)-Q\|^2_{\dot H^{\frac12}(\mathbb{C}_+)} &\le \|u_0^n-Q\|^2_{\dot H^{\frac12}(\mathbb{C}_+)} + K|t|\\ &\le (c_0 r)^2 + K|t|.\end{aligned}$$

For fixed $\varepsilon>0$, we conclude that the inequality $\|u_n(t)-Q\|_{\dot H^{\frac12}(\mathbb{C}_+)}\le(1+\varepsilon)c_0 r$ holds as long as $|t|\le\frac{(1+\varepsilon)^2-1}{K}(c_0r)^2$.

Set $\varepsilon>0$ and $t_1:=\frac{(1+\varepsilon)^2-1}{K}(c_0r)^2$. Assume that at time t_0, there exists a bounded sequence $(X_n^0)_n$ in $\mathbb{R}\times\mathbb{T}\times\mathbb{R}_+^*$, such that for all n, $\|u_n(t_0)-T_{X_n^0}Q\|_{\dot H^{\frac12}(\mathbb{C}_+)}<c_0r$. By the above method, one can show that $\|u_n(t)-T_{X_n^0}Q\|_{\dot H^{\frac12}(\mathbb{C}_+)}\le(1+\varepsilon)c_0r$ on $[t_0-t_1,t_0+t_1]$. Indeed, let $v_n:=T_{(X_n^0)^{-1}}u_n$. The equation satisfied by u_n is not invariant by scaling, but we can write down the equation satisfied by v_n. Recall that if

$$u(z)=\frac{1}{\sqrt{2\pi}}\int_0^{+\infty}e^{iz\sigma}f(\sigma)\,\mathrm{d}\sigma,$$

then

$$\widetilde{P}_{\varepsilon,M}u(z)=\frac{1}{\sqrt{2\pi}}\int_\varepsilon^M e^{iz\sigma}f(\sigma)\,\mathrm{d}\sigma.$$

Write $(X_n^0)=:(s_n^0,\theta_n^0,\alpha_n^0)$ and $\widetilde{P_0^n}:=\widetilde{P}_{\frac{\varepsilon_n}{(\alpha_n^0)^2},\frac{1}{\varepsilon_n(\alpha_n^0)^2}}\circ P_0$. Then $v_n=T_{(X_n^0)^{-1}}u_n$ satisfies $\|v_n(t_0)-Q\|_{\dot H^{\frac12}(\mathbb{C}_+)}<c_0r$, and

$$i\partial_t v_n=\widetilde{P_0^n}(|v_n|^2v_n).$$

Like equation (16), this equation conserves the energy $\|v_n(t)\|^4_{L^4(\mathbb{C}_+)}$ and the momentum $\|v_n(t)\|^2_{\dot H^{\frac12}(\mathbb{C}_+)}$. However, for all ε and M, one can see from the expression (10) of the Sobolev norms that the projector $\widetilde{P}_{\varepsilon,M}$ satisfies

$\|\widetilde{P}_{\varepsilon,M}u\|_{\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)} \leq \|u\|_{\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)}$. Therefore, we have the same inequalities as above

$$\begin{aligned}\|\partial_t v_n(t)\|_{\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)} &\leq \|\widetilde{P_0^n}(|v_n|^2 v_n)(t)\|_{\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)} \\ &\leq \|P_0(|v_n|^2 v_n)(t)\|_{\dot{H}^{-\frac{1}{2}}(\mathbb{C}_+)} \\ &\leq C\|P_0(|v_n|^2 v_n)(t)\|_{L^{\frac{4}{3}}(\mathbb{C}_+)} \\ &\leq C'\||v_n|^2 v_n(t)\|_{L^{\frac{4}{3}}(\mathbb{C}_+)} \\ &\leq C'\|v_n(t)\|^3_{L^4(\mathbb{C}_+)}.\end{aligned}$$

Since $\|v_n(t)\|_{L^4(\mathbb{C}_+)} = \|u_n(t)\|_{L^4(\mathbb{C}_+)}$ is uniformly bounded by conservation of the L^4-norm, we conclude that $\|v_n(t) - Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} = \|u_n(t) - T_{X_n^0}Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} \leq (1+\varepsilon)c_0 r$, as long as $|t - t_0| \leq t_1$.

We construct X_n as a piecewise $\mathcal{C}^1$ functional on $\mathbb{R}$ as follows. For $k \in \mathbb{Z}$, X_n is constant on $[kt_1, (k+1)t_1[$, equal to some $X_n^k \in \mathbb{R}\times\mathbb{T}\times\mathbb{R}_+^*$ to be chosen. We first set $X_n^{-1} = X_n^0 = (0,0,1)$. Then, at time $t_k = kt_1$, $k \geq 1$, we use the fact that $d(u_n(t_k), \mathcal{M}) < r$ and choose X_n^k, such that $\|u_n(t_k) - T_{X_n^k}Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} < c_0 r$. Then from the above paragraph, $\|u_n(t) - T_{X_n^k}Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} \leq (1+\varepsilon)c_0 r$ on $[t_k, t_k + t_1]$. We do a similar construction for negative times. The map X_n satisfies

$$\|u_n(t) - T_{X_n(t)}Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} \leq (1+\varepsilon)c_0 r, \quad t \in \mathbb{R}.$$

It remains to show that X_n is bounded independently of n on bounded intervals. In order to do so, it is enough to control the gap between X_n^{k-1} and X_n^k. By construction, at time t_k,

$$\|u_n(t_k) - T_{X_n^{k-1}}Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} \leq (1+\varepsilon)c_0 r$$

and

$$\|u_n(t_k) - T_{X_n^k}Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} < c_0 r,$$

and, therefore,

$$\|T_{X_n^{k-1}}Q - T_{X_n^k}Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} \leq (2+\varepsilon)c_0 r.$$

Using the following Lemma, we conclude that if r is chosen small enough, then there exists a constant $c_1 > 0$, such that for all $n \geq N$ and $k \in \mathbb{Z}$,

$$|X_n^{k-1}(X_n^k)^{-1}| \leq c_1.$$

LEMMA 3.6. — *There exist $c_1 > 0$ and $r_1 > 0$, such that the following holds. Let $X \in \mathbb{R}\times\mathbb{T}\times\mathbb{R}_+^*$, such that*

$$\|T_X Q - Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} \leq r_1.$$

Then

$$|X| \leq c_1.$$

Proof. — Due to the invariance of the $\dot{H}^{\frac{1}{2}}$-norm by symmetries, one can assume that $X = (s, \theta, \alpha)$ with $\alpha \geq 1$ up to exchanging X and X^{-1}. We expand

$$\begin{aligned}\|T_X Q - Q\|^2_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} &= \|T_X Q\|^2_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} + \|Q\|^2_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} - 2(T_X Q, Q)_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} \\ &= 2\pi - 2(T_X Q, Q)_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)}.\end{aligned}$$

Now, recall that

$$Q(z) = \frac{\sqrt{2}}{z+i} = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} e^{iz\sigma} f(\sigma)\, \mathrm{d}\sigma,$$

with

$$f(\sigma) = -2i\sqrt{\pi} e^{-\sigma},$$

and

$$\|Q\|^2_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} = \frac{1}{2} \int_0^{+\infty} |f(\sigma)|^2\, \mathrm{d}\sigma = \pi.$$

With this notation, the function corresponding to $T_X Q$ is

$$g(\sigma) = -2i\sqrt{\pi} e^{i\theta} e^{-is\sigma} e^{-\frac{\sigma}{\alpha^2}} \frac{1}{\alpha^2},$$

and, therefore,

$$\begin{aligned}\|T_X Q - Q\|^2_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} &= 2\pi - 4\pi \operatorname{Re}\left(\int_0^{+\infty} e^{i\theta} e^{-is\sigma} e^{-\frac{\sigma}{\alpha^2}} \frac{1}{\alpha^2} e^{-\sigma}\, \mathrm{d}\sigma \right) \\ &= 2\pi - 4\pi \operatorname{Re}\left(\frac{e^{i\theta}}{\alpha^2 \left(is + \frac{1}{\alpha^2} + 1 \right)} \right).\end{aligned}$$

Set $\alpha = 1 + \beta$ with $\beta \geq 0$. We want to bound s and β. By assumption,

$$\left| \operatorname{Re}\left(\frac{e^{i\theta}}{is\frac{(1+\beta)^2}{2} + 1 + \beta + \frac{\beta^2}{2}} \right) - 1 \right| \leq \frac{r_1^2}{2\pi} =: \delta_1.$$

Denote $z := \frac{e^{i\theta}}{is\frac{(1+\beta)^2}{2} + 1 + \beta + \frac{\beta^2}{2}}$. The fact $|\operatorname{Re}(z) - 1| \leq \delta_1$ implies that $|z| \geq \operatorname{Re}(z) \geq 1 - \delta_1$, and if $\delta_1 < 1$, that

$$\frac{1}{|z|} = \left| is\frac{(1+\beta)^2}{2} + 1 + \beta + \frac{\beta^2}{2} \right| \leq \frac{1}{1-\delta_1}.$$

On the one hand, taking the real part,

$$1 + \beta + \frac{\beta^2}{2} \leq \frac{1}{1-\delta_1}.$$

Since $\beta \in \mathbb{R}_+ \mapsto 1+\beta+\frac{\beta^2}{2}$ is strictly increasing and going to $+\infty$ as β goes to $+\infty$, there exists some constant $c>0$, such that $\beta \leq c$, or, in other terms, $0 \leq \log \alpha \leq \log(1+c)$. On the other hand, since $\beta \geq 0$, the bound on the imaginary part implies that

$$|s| \leq \frac{2}{1-\delta_1}. \qquad \square$$

Using the Lemma, assume that $3c_0 r < r_1$ and fix $t \in \mathbb{R}$. We now know that $(X_n(t))_n$ takes values in a compact set; up to extraction, one can assume that $(X_n(t))_n$ converges to some $X(t)$. Moreover, for all $t \in \mathbb{R}$ and $n \in \mathbb{N}$, $\|u_n(t) - T_{X_n(t)}Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} \leq (1+\varepsilon)c_0 r$, and, therefore, passing to the weak limit $n \to +\infty$ we conclude that $\|u(t) - T_{X(t)}Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} \leq (1+\varepsilon)c_0 r$. Since $\varepsilon > 0$ can be taken arbitrarily small, we have proven the following reformulation of Theorem 1.2.

THEOREM 3.7. — *There exist $c_0 > 0$ and $r_0 > 0$, such that the following holds. Let $r \leq r_0$ and $u_0 \in \dot{H}^{\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+)$, such that $\|u_0 - Q\|_{\dot{H}^{\frac{1}{2}}(\mathbb{C}_+)} < r^2$. Then there exists a weak solution $u \in \mathcal{C}(\mathbb{R}, \dot{H}^{\frac{1}{2}}(\mathbb{C}_+) \cap \mathrm{Hol}(\mathbb{C}_+))$ (with the weak topology) to equation* (9)

$$\begin{cases} i\partial_t u = P_0(|u|^2 u) \\ u(t=0) = u_0 \end{cases}, \quad (t,z) \in \mathbb{R} \times \mathbb{C}_+$$

such that for all $t \in \mathbb{R}$,

$$d(u(t), \mathcal{M}) \leq c_0 r.$$

4. Orbital stability for the ground states Q_β in the Schrödinger equation

We now consider the Schrödinger equation on the Heisenberg group (1)

$$\begin{cases} i\partial_t u - \Delta_{\mathbb{H}^1} u = |u|^2 u \\ u(t=0) = u_0 \end{cases}, \quad (t,x,y,s) \in \mathbb{R} \times \mathbb{H}^1.$$

For $\beta \in (\beta_*, 1)$, we are interested in solutions with initial data $u_0 \in \dot{H}^1(\mathbb{H}^1)$ satisfying

$$\|u_0 - \sqrt{1-\beta} Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)} < (1-\beta) r.$$

Let u be an eventual solution and set

$$u(t,x,y,s) = \sqrt{1-\beta} U((1-\beta)t, x, y, s+\beta t),$$

so that U is a solution to

$$i\partial_t U - \frac{\Delta_{\mathbb{H}^1} + \beta D_s}{1-\beta} U = |U|^2 U, \quad (t,x,y,s) \in \mathbb{R} \times \mathbb{H}^1. \tag{18}$$

The initial data U_0 satisfies

$$\|U_0 - Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)} < \sqrt{1-\beta} r.$$

There are two relevant conserved quantities for this equation: the energy

$$\mathcal{E}_\beta(V) := \frac{1}{2}(-\frac{\Delta_{\mathbb{H}^1} + \beta D_s}{1-\beta} V, V)_{L^2(\mathbb{H}^1)} - \frac{1}{4}\|V\|^4_{L^4(\mathbb{H}^1)},$$

and the momentum

$$\mathcal{P}(V) := (D_s V, V)_{L^2(\mathbb{H}^1)}, \quad V \in \dot{H}^1(\mathbb{H}^1).$$

Theorem 1.3 is equivalent to prove that if β is large, then one can construct a global weak solution U to equation (18), which stays close to the orbit of Q_β at all times, which leads to the following reformulation.

THEOREM 4.1. — *There exist some constants $c_0 > 0$ and $r_0 > 0$, such that for all $r \in (0, r_0)$, there exists a parameter $\beta^*(r) \in (0,1)$, such that the following holds. Let $\beta \in (\beta^*(r), 1)$ and $U_0 \in \dot{H}^1(\mathbb{H}^1)$ satisfying*

- *if $U_0 \in \dot{H}^1(\mathbb{H}^1) \cap V_0^+$:*

$$\|U_0 - Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)} < r^2$$

- *in the general case:*

$$\|U_0 - Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)} < \sqrt{1-\beta} r.$$

Then there exists a global weak solution $U_\beta \in \mathcal{C}(\mathbb{R}, \dot{H}^1(\mathbb{H}^1))$ (with the weak topology) to equation (18)

$$\begin{cases} i\partial_t U_\beta - \frac{\Delta_{\mathbb{H}^1} + \beta D_s}{1-\beta} U_\beta = |U_\beta|^2 U_\beta \\ U_\beta(t=0) = U_0 \end{cases},$$

such that for all $t \in \mathbb{R}$, $U_\beta(t)$ is close to the orbit $\mathcal{Q}_\beta = \{T_X Q_\beta \mid X \in \mathbb{R} \times \mathbb{T} \times \mathbb{R}_+^*\}$ of Q_β:

$$d(U_\beta(t), \mathcal{Q}_\beta) \leq c_0 r.$$

Contrary to the strategy deployed for the half-wave equation [11], the gap

$$\delta_\beta(V) := |\mathcal{E}_\beta(V) - \mathcal{E}_\beta(Q_\beta)| + |\mathcal{P}(V) - \mathcal{P}(Q_\beta)|, \quad V \in \dot{H}^1(\mathbb{H}^1),$$

does not here directly control the distance of V to $\mathcal{Q}_\beta$, so Proposition 1.5 does not hold for Q_β and δ_β. Indeed, even the fact that $\delta_\beta(V) = 0$ does not imply that V belongs to $\mathcal{Q}_\beta$. This is due to the fact that we can only use two conservation laws (energy and momentum) here, whereas an additional conservation law was available for the half-wave equation: the mass of the solution.

However, using that Q_β tends to Q as β tends to 1, one can instead show that the component of the solution along the space V_0^+ is close to Q and control the rest separately. More precisely, decompose

$$U(t) = U^+(t) + W(t),$$

where $U^+(t) \in \dot{H}^1(\mathbb{H}^1) \cap V_0^+$ and $W(t) \in \bigoplus_{(k,\pm)\neq(0,+)} \dot{H}^1(\mathbb{H}^1) \cap V_k^\pm$. If we know that $W(t)$ is small enough, then $\delta_\beta(U(t)) \approx \delta(U^+(t))$. This enables us to estimate the distance $d(U^+(t), \mathcal{M})$ of $U^+(t)$ to the orbit of Q, and therefore the distance of $U(t)$ to the orbit of Q_β for β close to 1.

The plan of the proof is as follows. Fix $\beta \in (0,1)$. We approximate the initial data and the equation by global smooth functions $(U_{\gamma,n})_{\gamma\in[\beta,1),n\in\mathbb{N}}$ valued in $H^2(\mathbb{H}^1)$ in Section 4.1. We then decompose

$$U_{\gamma,n}(t) = U^+_{\gamma,n}(t) + W_{\beta,n}(t),$$

where $U^+_{\gamma,n}(t) \in \dot{H}^1(\mathbb{H}^1) \cap V_0^+$ and $W_{\gamma,n}(t) \in \bigoplus_{(k,\pm)\neq(0,+)} \dot{H}^1(\mathbb{H}^1) \cap V_k^\pm$. In Section 4.2, we fix $n \in \mathbb{N}$ and study the limit $\gamma \to 1$. We prove by using the conservation laws that $W_{\gamma,n}(t)$ stays small and that the gap $\delta(U^+_{\gamma,n}(t))$ is controlled as $\delta(U^+_{\gamma,n}(t)) \lesssim r^2$ for $\gamma \geq \beta^*(n,t)$, which leads to an upper bound

$$d(U_{\gamma,n}(t), \mathcal{M}) < c_0 r, \quad t \in \mathbb{R}, \gamma \in [\max(\beta^*(n,t), \beta), 1). \tag{19}$$

Then, we show that the lower bound $\beta^*(n,t)$ can be taken independently of n and t. Finally, in Section 4.3, we fix $\beta \geq \beta_*$ and use the same method as for the limiting equation to find an upper bound on the modulation parameters $(X_{\beta,n}(t))_{n\in\mathbb{N}}$ in order to pass to the limit $n \to +\infty$ in the above inequality (19).

4.1. Construction of approximate solutions. —

Construction of a sequence of smoothing projectors $\Pi^{(n)}$*:* We define a sequence of projectors $\Pi^{(n)}$ close to identity, mapping elements of $\dot{H}^s(\mathbb{H}^1)$ $(s = \pm 1)$ to smoother functions, by removing the high and low Fourier frequencies and the high Hermite modes in the decomposition

$$\dot{H}^s(\mathbb{H}^1) = \bigoplus_{k\in\mathbb{N}} \bigoplus_{\pm} \dot{H}^s(\mathbb{H}^1) \cap V_k^\pm.$$

Using these projectors, we consider a sequence of equations approximating (18) for which the Cauchy problem is globally well posed.

Let $u \in \dot{H}^s(\mathbb{H}^1)$, $s = \pm 1$, which we decompose as a series of elements of $\dot{H}^s(\mathbb{H}^1) \cap V_k^\pm$ for $(k,\pm) \in (\mathbb{N},\pm)$. Write

$$u = \sum_{k=0}^{+\infty} \sum_{\pm} \Pi_{k,\pm}(u),$$

where for all $(k,\pm)\in(\mathbb{N},\pm)$, $\Pi_{k,\pm}(u)\in\dot{H}^s(\mathbb{H}^1)\cap V_k^{\pm}$. Then

$$\|u\|^2_{\dot{H}^s(\mathbb{H}^1)}=\sum_{k\in\mathbb{N}}\sum_{\pm}\int_{\mathbb{R}_\pm}((k+1)|\sigma|)^s\int_{\mathbb{R}^2}|\widehat{\Pi_{k,\pm}(u)}(x,y,\sigma)|^2\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\sigma.$$

Let $n\in\mathbb{N}$. We define $\Pi^{(n)}(u)$ as follows. We take the n-th partial sum and remove the high and low frequencies $|\sigma|\to+\infty$ and $|\sigma|\to 0$:

$$\widehat{\Pi^{(n)}(u)}(x,y,\sigma):=\sum_{k=0}^{n}\sum_{\pm}\widehat{\Pi_{k,\pm}(u)}(x,y,\sigma)\mathbb{1}_{\frac{1}{n}\le|\sigma|\le n}. \tag{20}$$

Consequently,

$$\|\Pi^{(n)}(u)\|^2_{\dot{H}^s(\mathbb{H}^1)}=\sum_{k=0}^{n}\sum_{\pm}\int_{\{\sigma\in\mathbb{R}_\pm\}\cap\{\frac{1}{n}\le|\sigma|\le n\}}((k+1)|\sigma|)^s\times\int_{\mathbb{R}^2}|\widehat{\Pi_{k,\pm}(u)}(x,y,\sigma)|^2\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\sigma$$

converges to $\|u\|^2_{\dot{H}^s(\mathbb{H}^1)}$ as n goes to $+\infty$.

Moreover, if $u\in\dot{H}^1(\mathbb{H}^1)$, then $\Pi^{(n)}(u)$ belongs to $H^3(\mathbb{H}^1)$. Indeed,

$$\|\Pi^{(n)}(u)\|^2_{H^3(\mathbb{H}^1)}=\sum_{k=0}^{n}\sum_{\pm}\int_{\{\sigma\in\mathbb{R}_\pm\}\cap\{\frac{1}{n}\le|\sigma|\le n\}}(1+(k+1)^3|\sigma|^3)\times\int_{\mathbb{R}^2}|\widehat{\Pi_{k,\pm}(u)}(x,y,\sigma)|^2\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\sigma,$$

but on the set $\{\frac{1}{n}\le|\sigma|\le n\}$, and for $k\le n$,

$$(1+(k+1)^3|\sigma|^3)\le(n|\sigma|+(n+1)^2(k+1)n^2|\sigma|)\le n(1+n(n+1)^2)(k+1)|\sigma|,$$

and, therefore, $\|\Pi^{(n)}(u)\|_{H^3(\mathbb{H}^1)}$ is finite.

Construction of a sequence of approximate solutions $(U_{\gamma,n})_{\gamma\in[\beta,1),n\in\mathbb{N}}$*:* Fix $\beta\in(0,1)$, $r>0$ and $U_0\in\dot{H}^1(\mathbb{H}^1)$, such that

- either $U_0\in\dot{H}^1(\mathbb{H}^1)\cap V_0^+$ and
$$\|U_0-Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)}<r^2;$$
- either
$$\|U_0-Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)}<\sqrt{1-\beta}r.$$

We want to construct a global solution to (18)

$$\begin{cases}i\partial_t U_\beta-\frac{\Delta_{\mathbb{H}^1}+\beta D_s}{1-\beta}U_\beta=|U_\beta|^2U_\beta\\ U_\beta(t=0)=U_0\end{cases},$$

such that for all $t\in\mathbb{R}$,

$$d(U_\beta(t),\mathcal{Q}_\beta)\le c_0 r.$$

By approximation, the idea would be to consider a sequence of equations

$$\begin{cases} i\partial_t U_{\beta,n} - \frac{\Delta_{\mathbb{H}^1}+\beta D_s}{1-\beta} U_{\beta,n} = \Pi^{(n)}(|U_{\beta,n}|^2 U_{\beta,n}) \\ U_{\beta,n}(t=0) = U_0^{\beta,n} = \Pi^{(n)}(U_0) \end{cases}, \quad n \in \mathbb{N}, \tag{21}$$

for which one can show that for all n large, there exists $\beta^*(n)$, such that if $\beta \geq \beta^*(n)$, then

$$d(U_{\beta,n}(t), \mathcal{Q}_\beta) \leq c_0 r, \quad t \in \mathbb{R}.$$

In order to get a lower bound β^* independent of n, we rather construct a set of initial data $(U_0^{\gamma,n})_{\gamma\in[\beta,1),n\in\mathbb{N}}$ and equations

$$\begin{cases} i\partial_t U_{\gamma,n} - \frac{\Delta_{\mathbb{H}^1}+\gamma D_s}{1-\gamma} U_{\gamma,n} = \Pi^{(n)}(|U_{\gamma,n}|^2 U_{\gamma,n}) \\ U_{\gamma,n}(t=0) = U_0^{\gamma,n} = \Pi^{(n)}(U_0^\gamma) \end{cases}, \quad n \in \mathbb{N}, \gamma \in [\beta, 1), \tag{22}$$

and then use a continuity argument.

For $\gamma \in [\beta, 1)$, the initial data U_0^γ is defined as follows:

- if $U_0 \in \dot{H}^1(\mathbb{H}^1) \cap V_0^+$, we choose U_0^γ constant equal to U_0;
- otherwise, we choose

$$U_0^\gamma := \frac{1-\gamma}{1-\beta} U_0 + \frac{\gamma-\beta}{1-\beta} Q.$$

We make this choice in the general case because we need the initial data U_0^γ to go to $\dot{H}^1(\mathbb{H}^1) \cap V_0^+$ as γ tends to 1.

LEMMA 4.2. — *Let $r > 0$ and $U_0 \in \dot{H}^1(\mathbb{H}^1)$. There exist $C_0 > 0$, $\beta_*(r) \in (0,1)$ and $N(r, U_0) \in \mathbb{N}$, such that the following holds. Let $\beta \in (\beta_*(r), 1)$ and assume that U_0 satisfies*

- *either $U_0 \in \dot{H}^1(\mathbb{H}^1) \cap V_0^+$ and*

$$\|U_0 - Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)} < r^2;$$

- *either*

$$\|U_0 - Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)} < \sqrt{1-\beta} r.$$

Then for all $n \geq N(r, U_0)$ and for all $\gamma \in [\beta, 1)$,

$$|\mathcal{E}_\gamma(U_0^{\gamma,n}) - \mathcal{E}(Q)| + |\mathcal{P}(U_0^{\gamma,n}) - \mathcal{P}(Q)| < C_0 r^2.$$

Proof. — We use the following convergence rate of $(Q_\beta)_\beta$ to Q as β tends to 1 (proved in Appendix A):

$$\|Q_\beta - Q\|_{\dot{H}^1(\mathbb{H}^1)} = o(\sqrt{1-\beta}).$$

If $U_0 \in \dot{H}^1(\mathbb{H}^1) \cap V_0^+$ and $\|U_0 - Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)} < r^2$, we have chosen U_0^γ constant equal to U_0, and it is enough to use that $\|\Pi^{(n)}(U_0) - U_0\|_{\dot{H}^1(\mathbb{H}^1)} \to 0$ as $n \to +\infty$.

We now treat the case $\|U_0 - Q_\beta\|_{\dot H^1(\mathbb{H}^1)} < \sqrt{1-\beta} r$. By convergence of Q_β to Q, there exists $\beta_* = \beta_*(r) \in (0,1)$, such that for all $\beta \in (\beta_*, 1)$,

$$\|Q_\beta - Q\|_{\dot H^1(\mathbb{H}^1)} < \sqrt{1-\beta} r.$$

We decompose

$$Q_\beta = Q_\beta^+ + R_\beta$$

and

$$U_0 = U_0^+ + W_0,$$

where $Q_\beta^+, U_0^+ \in \dot H^1(\mathbb{H}^1) \cap V_0^+$ and $R_\beta, W_0 \in \bigoplus_{(k,\pm)\neq(0,+)} \dot H^1(\mathbb{H}^1) \cap V_k^\pm$. In the same way, we decompose U_0^γ as

$$U_0^\gamma = (U_0^\gamma)^+ + W_0^\gamma$$

and $U_0^{\gamma,n} = \Pi^{(n)}(U_0^\gamma)$ as

$$U_0^{\gamma,n} = (U_0^{\gamma,n})^+ + W_0^{\gamma,n},$$

where $(U_0^\gamma)^+, (U_0^{\gamma,n})^+ \in \dot H^1(\mathbb{H}^1) \cap V_0^+$ and $W_0^\gamma, W_0^{\gamma,n} \in \bigoplus_{(k,\pm)\neq(0,+)} \dot H^1(\mathbb{H}^1) \cap V_k^\pm$.

Since

$$\|W_0 - R_\beta\|_{\dot H^1(\mathbb{H}^1)} \leq \|U_0 - Q_\beta\|_{\dot H^1(\mathbb{H}^1)} < \sqrt{1-\beta} r,$$

W_0 satisfies

$$\begin{aligned}\|W_0\|_{\dot H^1(\mathbb{H}^1)} &\leq \|W_0 - R_\beta\|_{\dot H^1(\mathbb{H}^1)} + \|R_\beta\|_{\dot H^1(\mathbb{H}^1)} \\ &\leq 2\sqrt{1-\beta} r.\end{aligned}$$

Therefore, $W_0^\gamma = \frac{1-\gamma}{1-\beta} W_0$ satisfies

$$\|W_0^\gamma\|_{\dot H^1(\mathbb{H}^1)} \leq 2\frac{1-\gamma}{\sqrt{1-\beta}} r,$$

which implies that for all $n \in \mathbb{N}$,

$$\|W_0^{\gamma,n}\|_{\dot H^1(\mathbb{H}^1)} \leq 2\frac{1-\gamma}{\sqrt{1-\beta}} r.$$

In particular,

$$\begin{aligned}&\left|\left(-\frac{\Delta_{\mathbb{H}^1} + \gamma D_s}{1-\gamma} W_0^{\gamma,n}, W_0^{\gamma,n}\right)_{L^2(\mathbb{H}^1)}\right| + \left|(D_s W_0^{\gamma,n}, W_0^{\gamma,n})_{L^2(\mathbb{H}^1)}\right| \\ &\quad\leq 8\frac{1-\gamma}{1-\beta} r^2 + 4\frac{(1-\gamma)^2}{1-\beta} r^2 \\ &\quad\leq 12 r^2,\end{aligned}$$

and

$$\|W_0^{\gamma,n}\|_{L^4(\mathbb{H}^1)} \leq C\|W_0^{\gamma,n}\|_{\dot H^1(\mathbb{H}^1)} \leq 2C\sqrt{1-\beta} r,$$

which is bounded by r^2 if $\beta \geq \beta_*(r)$ is large enough.

Given the form of the energy

$$\mathcal{E}_\gamma(U_0^{\gamma,n}) = \frac{1}{2}\left(-\frac{\Delta_{\mathbb{H}^1}+\gamma D_s}{1-\gamma}W_0^{\gamma,n}, W_0^{\gamma,n}\right)_{L^2(\mathbb{H}^1)} + \frac{1}{2}(D_s(U_0^{\gamma,n})^+, (U_0^{\gamma,n})^+)_{L^2(\mathbb{H}^1)} - \frac{1}{4}\|(U_0^{\gamma,n})^+ + W_0^{\gamma,n}\|^4_{L^4(\mathbb{H}^1)},$$

and given that all the terms involving $W_0^{\gamma,n}$ are bounded by some C_0r^2, it is now enough to prove an estimate of the form $\|(U_0^{\gamma,n})^+ - Q\|_{\dot{H}^1(\mathbb{H}^1)} \leq C_0r^2$. However,

$$\begin{aligned}\|(U_0^{\gamma,n})^+ - (U_0^{\gamma})^+\|_{\dot{H}^1(\mathbb{H}^1)} &\leq \frac{1-\gamma}{1-\beta}\|\Pi^{(n)}((U_0)^+) - U_0^+\|_{\dot{H}^1(\mathbb{H}^1)} \\ &\quad + \frac{\gamma-\beta}{1-\beta}\|\Pi^{(n)}(Q) - Q\|_{\dot{H}^1(\mathbb{H}^1)} \\ &\leq \|\Pi^{(n)}((U_0)^+) - U_0^+\|_{\dot{H}^1(\mathbb{H}^1)} + \|\Pi^{(n)}(Q) - Q\|_{\dot{H}^1(\mathbb{H}^1)},\end{aligned}$$

which converges to zero as n goes to $+\infty$ independently of γ and β. Moreover,

$$\begin{aligned}\|(U_0^{\gamma})^+ - Q\|_{\dot{H}^1(\mathbb{H}^1)} &= \frac{1-\gamma}{1-\beta}\|U_0^+ - Q\|_{\dot{H}^1(\mathbb{H}^1)} \\ &\leq \frac{1-\gamma}{1-\beta}(\|U_0 - Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)} + \|W_0\|_{\dot{H}^1(\mathbb{H}^1)} + \|Q_\beta - Q\|_{\dot{H}^1(\mathbb{H}^1)}) \\ &\leq \frac{1-\gamma}{1-\beta}4\sqrt{1-\beta}r \\ &\leq 4r^2,\end{aligned}$$

for $\beta \geq \beta_*(r)$ large enough.

To conclude, there exist $C_0 > 0$, $r_0 > 0$ and $N \in \mathbb{N}$, such that for all $n \geq N$, $r \in (0, r_0)$ and $\gamma \in [\beta, 1)$,

$$|\mathcal{E}_\gamma(U_0^{\gamma,n}) - \mathcal{E}(Q)| + |\mathcal{P}(U_0^{\gamma,n}) - \mathcal{P}(Q)| < C_0r^2. \qquad \square$$

From now on, we assume that $\beta \geq \beta_*(r)$ and $n \geq N(r)$.

As in Proposition 3.5 for the limiting equation, equation (22) admits a unique global solution in $H^3_n(\mathbb{H}^1) := \Pi^{(n)}(H^3(\mathbb{H}^1))$.

PROPOSITION 4.3. — *Let $U_0^{\gamma,n} \in H^3_n(\mathbb{H}^1)$ and $\gamma \in [0,1)$. Then there exists a unique $U_{\gamma,n} \in \mathcal{C}^\infty(\mathbb{R}, H^3_n(\mathbb{H}^1))$, such that (22) is satisfied in the distributional sense. Moreover, the solution map is continuous from $H^3_n(\mathbb{H}^1)$ to $\mathcal{C}^\infty(\mathbb{R}, H^3_n(\mathbb{H}^1))$.*

Proof. — Local well-posedness comes from the Cauchy–Lipschitz theory from ODEs. Indeed, $H^3(\mathbb{H}^1)$ is an algebra, and, moreover, the Hermite modes k are restricted to $k \geq n$, and the frequencies σ are restricted to the set $\{\frac{1}{n} \leq |\sigma| \leq n\}$. Therefore, the map $V \mapsto \frac{\Delta_{\mathbb{H}^1}+\gamma D_s}{1-\gamma}V + \Pi^{(n)}(|V|^2V)$ is well defined and locally Lipschitz from the Banach space $H^3_n(\mathbb{H}^1)$ to itself.

In order to show that the local maximal solutions are global, we prove that there is no blow-up in finite time due to the conservation of the momentum

$$\mathcal{P}(V) = (D_s V, V)_{L^2(\mathbb{H}^1)},$$

and the following inequality valid for $V \in H^3_n(\mathbb{H}^1)$:

$$(D_s V, V)_{L^2(\mathbb{H}^1)} \leq \|V\|^2_{H^3(\mathbb{H}^1)} \leq (n + (n+1)^3 n^2)(D_s V, V)_{L^2(\mathbb{H}^1)}. \qquad \square$$

4.2. Limit $\gamma \to 1$ for the n-th partial sum. — In this section, we use the conservation of energy and momentum to recover an upper bound on $d(U_{\gamma,n}(t), \mathcal{M})$ for $\gamma \geq \beta^*(n,t)$ close to 1. Then, we prove that the lower bound for γ can be chosen independently of n and t.

For $t \in \mathbb{R}$, we decompose $U_{\gamma,n}(t)$ as

$$U_{\gamma,n}(t) = U^+_{\gamma,n}(t) + W_{\gamma,n}(t),$$

where

$$U^+_{\gamma,n}(t) = \Pi^+_0(U_{\gamma,n}(t)) \in \dot{H}^1(\mathbb{H}^1) \cap V^+_0,$$

and, therefore,

$$W_{\gamma,n}(t) = (\mathrm{Id} - \Pi^+_0)(U_{\gamma,n}(t)) \in \bigoplus_{(k,\pm)\neq(0,+)} \dot{H}^1(\mathbb{H}^1) \cap V^\pm_k$$

(see Definition 2.1 for the definition of the spaces in the orthogonal sum $V^\pm_k$).

In what follows, we show that $U^+_{\gamma,n}(t)$ is the main part for which we control $\delta(U^+_{\gamma,n}(t))$, and $W_{\gamma,n}(t)$ is a remainder term that vanishes in the limit $\gamma \to 1$.

First, since $\mathcal{P}(U_{\gamma,n}(t)) = (D_s U_{\gamma,n}(t), U_{\gamma,n}(t))_{L^2(\mathbb{H}^1)}$ is conserved, bounded by $C_0 r^2 + \mathcal{P}(Q)$ for all $\gamma \in [\beta, 1)$ and equivalent to $\|U_{\gamma,n}(t)\|^2_{\dot{H}^1(\mathbb{H}^1)}$ in $H^2_n(\mathbb{H}^1)$, there exists some constant $C(n) > 0$, such that for all $t \in \mathbb{R}$ and $\gamma \in [\beta, 1)$,

$$\|U_{\gamma,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)} \leq C(n). \tag{23}$$

However, such a bound on $\|U_{\gamma,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)}$ and the conservation of energy imply that $W_{\gamma,n}(t)$ must vanish in $\dot{H}^1(\mathbb{H}^1)$ as γ tends to 1 due to the following lemma.

LEMMA 4.4. — *Let $r > 0$, $\beta \in (\beta_*(r), 1)$, $n \geq N(r)$ and assume that U_0 satisfies*

- *either $U_0 \in \dot{H}^1(\mathbb{H}^1) \cap V^+_0$ and*
$$\|U_0 - Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)} < r^2;$$
- *either*
$$\|U_0 - Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)} < \sqrt{1-\beta}\, r,$$

There exists $C_1 > 0$, such that if there exists $C > 0$ (possibly depending on n), $t \in \mathbb{R}$ and $\gamma \in [\beta, 1)$, such that $\|U_{\gamma,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)} \leq C$, then

$$\|W_{\gamma,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)} \leq C_1(1 + C^2)\sqrt{1-\gamma}.$$

Proof. — We use the conservation of energy

$$\mathcal{E}_\gamma(U_{\gamma,n}(t)) = \frac{1}{2}\left(-\frac{\Delta_{\mathbb{H}^1}+\gamma D_s}{1-\gamma}W_{\gamma,n}(t), W_{\gamma,n}(t)\right)_{L^2(\mathbb{H}^1)} + \frac{1}{2}(D_s U^+_{\gamma,n}(t), U^+_{\gamma,n}(t))_{L^2(\mathbb{H}^1)} - \frac{1}{4}\|U_{\gamma,n}(t)\|^4_{L^4(\mathbb{H}^1)},$$

then apply Lemma 4.2 to get

$$|\mathcal{E}_\gamma(U_0^{\gamma,n}) - \mathcal{E}(Q)| < C_0 r^2.$$

Due to the embedding $\dot{H}^1(\mathbb{H}^1) \hookrightarrow L^4(\mathbb{H}^1)$, we know that

$$\|U_{\gamma,n}(t)\|^4_{L^4(\mathbb{H}^1)} \leq K^4 \|U_{\gamma,n}(t)\|^4_{\dot{H}^1(\mathbb{H}^1)} = K^4 C^4.$$

Moreover, recall the equivalence of norms

$$\frac{1}{2}\|w\|^2_{\dot{H}^1(\mathbb{H}^1)} \leq (-(\Delta_{\mathbb{H}^1}+\gamma D_s)w, w)_{L^2(\mathbb{H}^1)} \leq 2\|w\|^2_{\dot{H}^1(\mathbb{H}^1)}, \qquad w \in \bigoplus_{(k,\pm)\neq(0,+)} \dot{H}^1(\mathbb{H}^1)\cap V_k^\pm.$$

We conclude that

$$\frac{1}{4(1-\gamma)}\|W_{\gamma,n}(t)\|^2_{\dot{H}^1(\mathbb{H}^1)} \leq \mathcal{E}(Q) + C_0 r^2 + \frac{1}{4}K^4 C^4,$$

which implies the lemma. □

Applying Lemma 4.4 and inequality (23), we know that for all $t \in \mathbb{R}$ and $\gamma \in [\beta, 1)$,

$$\|W_{\gamma,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)} \leq C_1(1 + C(n)^2)\sqrt{1-\gamma},$$

which vanishes as γ tends to 1.

Fix $t \in \mathbb{R}$. We establish in Lemma 4.5 below that $\gamma \in [\beta, 1) \mapsto W_{\gamma,n}(t)$ is continuous, so that we can define $\beta_0(n,t) \geq \beta$ as the minimal element in $[\beta, 1)$ satisfying:

$$\forall \gamma \in [\beta_0(n,t), 1), \quad \|W_{\gamma,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)} \leq r^2.$$

Lemma 4.5. — *For $t \in \mathbb{R}$, $\gamma \in [\beta, 1) \mapsto W_{\gamma,n}(t) \in \dot{H}^1(\mathbb{H}^1)$ is continuous.*

Proof. — Fix $t \in \mathbb{R}$. One knows that the orthogonal projection $(\mathrm{Id} - \Pi_0^+)$ onto the space $\bigoplus_{(k,\pm)\neq(0,+)} \dot{H}^1(\mathbb{H}^1) \cap V_k^\pm$ is continuous, since it satisfies the inequality $\|(\mathrm{Id} - \Pi_0^+)(V)\|_{\dot{H}^1(\mathbb{H}^1)} \leq \|V\|_{\dot{H}^1(\mathbb{H}^1)}$ for all $V \in \dot{H}^1(\mathbb{H}^1)$. However, by definition, $W_{\gamma,n}(t) = (\mathrm{Id} - \Pi_0^+)(U_{\gamma,n}(t))$, and, therefore, it is enough show

that the map $\gamma \in [\beta, 1) \mapsto U_{\gamma,n}(t) \in \dot{H}^1(\mathbb{H}^1)$ is continuous. Let $\gamma_1, \gamma_2 \in [\beta, 1)$ and set $R := U_{\gamma_1,n} - U_{\gamma_2,n}$. Then R is a solution to

$$i\partial_t R - \frac{\Delta_{\mathbb{H}^1} + \gamma_1 D_s}{1-\gamma_1} R - \left(\frac{\Delta_{\mathbb{H}^1} + \gamma_1 D_s}{1-\gamma_1} - \frac{\Delta_{\mathbb{H}^1} + \gamma_2 D_s}{1-\gamma_2} \right) U_{\gamma_2,n} \\ = \Pi^{(n)}(|U_{\gamma_1,n}|^2 U_{\gamma_1,n}) - \Pi^{(n)}(|U_{\gamma_2,n}|^2 U_{\gamma_2,n}).$$

We bound $\|\partial_t R(t)\|_{\dot{H}^1(\mathbb{H}^1)}$, which is equivalent to controlling $\|\partial_t R(t)\|_{\dot{H}^{-1}(\mathbb{H}^1)}$, since $\partial_t R(t) \in \Pi^{(n)}(\dot{H}^1(\mathbb{H}^1))$. We treat each term in the equation separately.

First,

$$\left\| -\frac{\Delta_{\mathbb{H}^1} + \gamma_1 D_s}{1-\gamma_1} R(t) \right\|_{\dot{H}^{-1}(\mathbb{H}^1)} \leq \frac{2}{1-\gamma_1} \|-\Delta_{\mathbb{H}^1} R(t)\|_{\dot{H}^{-1}(\mathbb{H}^1)} \\ \leq \frac{2}{1-\gamma_1} \|R(t)\|_{\dot{H}^1(\mathbb{H}^1)}.$$

Then,

$$\left\| \left(\frac{\Delta_{\mathbb{H}^1} + \gamma_1 D_s}{1-\gamma_1} - \frac{\Delta_{\mathbb{H}^1} + \gamma_2 D_s}{1-\gamma_2} \right) U_{\gamma_2,n}(t) \right\|_{\dot{H}^{-1}(\mathbb{H}^1)} \\ \leq \frac{|\gamma_2 - \gamma_1|}{(1-\gamma_1)(1-\gamma_2)} \big(\|-\Delta_{\mathbb{H}^1} U_{\gamma_2,n}(t)\|_{\dot{H}^{-1}(\mathbb{H}^1)} + \|D_s U_{\gamma_2,n}(t)\|_{\dot{H}^{-1}(\mathbb{H}^1)} \big) \\ \leq \frac{|\gamma_2 - \gamma_1|}{(1-\gamma_1)(1-\gamma_2)} 2C(n).$$

Finally, note that in the image $\dot{H}_n^{-1}(\mathbb{H}^1)$ of $\dot{H}^{-1}(\mathbb{H}^1)$ by $\Pi^{(n)}$, all the Sobolev norms are equivalent. Indeed, by definition of $\Pi^{(n)}$ (see equation (20)), the frequencies are restricted to the set $\{\frac{1}{n} \leq |\sigma| \leq n\}$, and the Hermite modes are bounded by n. Therefore, there exists $C_1'(n) > 0$, such that

$$\|\Pi^{(n)}(|U_{\gamma_1,n}|^2 U_{\gamma_1,n}(t)) - \Pi^{(n)}(|U_{\gamma_2,n}|^2 U_{\gamma_2,n}(t))\|_{\dot{H}^{-1}(\mathbb{H}^1)} \\ \leq C_1'(n) \|\Pi^{(n)}(|U_{\gamma_1,n}|^2 U_{\gamma_1,n}(t)) - \Pi^{(n)}(|U_{\gamma_2,n}|^2 U_{\gamma_2,n}(t))\|_{H^3(\mathbb{H}^1)}.$$

Using the algebra property of $H^3(\mathbb{H}^1)$, we get

$$\|\Pi^{(n)}(|U_{\gamma_1,n}|^2 U_{\gamma_1,n}(t)) - \Pi^{(n)}(|U_{\gamma_2,n}|^2 U_{\gamma_2,n}(t))\|_{\dot{H}^{-1}(\mathbb{H}^1)} \leq C_2'(n) \|R(t)\|_{H^3(\mathbb{H}^1)},$$

and again using the equivalence between $\dot{H}^1$ and H^3 norms, we deduce

$$\|\Pi^{(n)}(|U_{\gamma_1,n}|^2 U_{\gamma_1,n}(t)) - \Pi^{(n)}(|U_{\gamma_2,n}|^2 U_{\gamma_2,n}(t))\|_{\dot{H}^{-1}(\mathbb{H}^1)} \leq C_3'(n) \|R(t)\|_{\dot{H}^1(\mathbb{H}^1)}.$$

We now define $f(t) := \|R(t)\|^2_{\dot{H}^1(\mathbb{H}^1)}$ for $t \in \mathbb{R}$. Then there exists some constant $C''(n) > 0$, such that

$$\begin{aligned}
f'(t) &\leq 2\|\partial_t R(t)\|_{\dot{H}^1(\mathbb{H}^1)}\|R(t)\|_{\dot{H}^1(\mathbb{H}^1)} \\
&\leq C''(n)\|\partial_t R(t)\|_{\dot{H}^{-1}(\mathbb{H}^1)}\|R(t)\|_{\dot{H}^1(\mathbb{H}^1)} \\
&\leq C''(n)\bigg(\bigg(\frac{2}{1-\gamma_1} + C_3'(n)\bigg)\|R(t)\|^2_{\dot{H}^1(\mathbb{H}^1)} \\
&\qquad + \frac{|\gamma_2-\gamma_1|}{(1-\gamma_1)(1-\gamma_2)}2C(n)\|R(t)\|_{\dot{H}^1(\mathbb{H}^1)}\bigg) \\
&\leq C''(n)\bigg(\bigg(\frac{2}{1-\gamma_1} + C_3'(n) + \frac{|\gamma_2-\gamma_1|}{(1-\gamma_1)(1-\gamma_2)}C(n)\bigg)\|R(t)\|^2_{\dot{H}^1(\mathbb{H}^1)} \\
&\qquad + \frac{|\gamma_2-\gamma_1|}{(1-\gamma_1)(1-\gamma_2)}C(n)\bigg).
\end{aligned}$$

Therefore, $f(t)$ satisfies a Gronwall-type inequality

$$f(t)' \leq K_1(n)f(t) + K_2(n)\frac{|\gamma_2-\gamma_1|}{(1-\gamma_1)(1-\gamma_2)},$$

with

$$K_1(n) = C''(n)\left(\frac{2}{1-\gamma_1} + C_3'(n) + \frac{|\gamma_2-\gamma_1|}{(1-\gamma_1)(1-\gamma_2)}C(n)\right),$$

and

$$K_2(n) = C''(n)C(n).$$

This inequality implies that for all $t \in \mathbb{R}$,

$$f(t) \leq f(0)e^{K_1(n)|t|} + \frac{K_2(n)}{K_1(n)}\frac{|\gamma_2-\gamma_1|}{(1-\gamma_1)(1-\gamma_2)}(e^{K_1(n)|t|}-1),$$

with

$$f(0) = \|\Pi^{(n)}(U_0^{\gamma_1} - U_0^{\gamma_2})\|^2_{\dot{H}^1(\mathbb{H}^1)}.$$

Fix $t \in \mathbb{R}$ and $\gamma_1 \in [\beta, 1)$, we see that if γ_2 tends to γ_1, then $f(t)$ tends to 0. □

LEMMA 4.6. — *Let $r > 0$, $\beta \in (\beta_*(r), 1)$ and $n \geq N(r)$. There exists some constant $\beta^*(r) \in (0,1)$, such that if $\beta \geq \beta^*(r)$, then the solution $U_{\beta,n}$ to equation* (21)

$$\begin{cases} i\partial_t U_{\beta,n} - \frac{\Delta_{\mathbb{H}^1} + \beta D_s}{1-\beta}U_{\beta,n} = \Pi^{(n)}(|U_{\beta,n}|^2 U_{\beta,n}) \\ U_{\beta,n}(t=0) = U_0^n = \Pi^{(n)}(U_0) \end{cases},$$

satisfies for all $t \in \mathbb{R}$

$$\|W_{\beta,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)} \leq r^2,$$

and

$$\|U^+_{\beta,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)} \leq (\mathcal{P}(Q) + C_0 r^2)^{\frac{1}{2}}.$$

Proof. — Fix $\beta \in (0,1)$ and recall that $\beta_0(n,t) \geq \beta$ is the minimal element in $[\beta, 1)$ satisfying:

$$\forall \gamma \in [\beta_0(n,t), 1), \quad \|W_{\gamma,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)} \leq r^2,$$

and assume that $\beta < \beta_0(n,t) =: \beta_0$. We find an upper bound for β_0 in $[\beta, 1)$ independent of n and t. The continuity of $\gamma \mapsto W_{\gamma,n}(t)$ implies that

$$\|W_{\beta_0,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)} = r^2.$$

The projection of $U_{\beta_0,n}(t)$ on V_0^+ is bounded by

$$\begin{aligned}\|U^+_{\beta_0,n}(t)\|^2_{\dot{H}^1(\mathbb{H}^1)} &\leq \mathcal{P}(U_{\beta_0,n}(t)) \\ &\leq \mathcal{P}(Q) + C_0 r^2,\end{aligned}$$

and, therefore,

$$\|U_{\beta_0,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)} \leq C,$$

where $C := r^2 + (\mathcal{P}(Q) + C_0 r^2)^{\frac{1}{2}}$ no longer depends n or t. Lemma 4.4 now implies

$$\|W_{\beta_0,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)} \leq C_1(1 + C^2)\sqrt{1 - \beta_0}.$$

We conclude that

$$r^2 \leq C_1(1 + C^2)\sqrt{1 - \beta_0},$$

which means

$$\beta_0 \leq 1 - \left(\frac{r^2}{C_1(1 + C^2)}\right)^2 =: \beta^*(r),$$

and, therefore, $\beta < \beta^*(r)$. Taking the converse, we have proven that if $\beta \geq \beta^*(r)$, then $\beta_0 = \beta$. □

We now show that $U_{\beta,n}(t)$ is close to the orbit $\mathcal{M}$ of Q for $t \in \mathbb{R}$ and $\beta \geq \beta^*(r)$.

Proposition 4.7. — *There exist $r_0 > 0$ and $c_0 > 0$, such that if $r < r_0$, $\beta \in [\beta^*(r), 1)$ and $n \geq N(r)$, then for all $t \in \mathbb{R}$,*

$$d(U_{\beta,n}(t), \mathcal{M}) < c_0 r.$$

Proof. — Fix $t \in \mathbb{R}$. It suffices to estimate $\delta(U^+_{\beta,n}(t))$ and apply Proposition 1.5.

On the one hand, since $(D_s W_{\beta,n}(t), W_{\beta,n}(t))_{L^2(\mathbb{H}^1)} \leq \|W_{\beta,n}(t)\|^2_{\dot{H}^1(\mathbb{H}^1)} \leq r^2$, the conservation of momentum and Lemma 4.2 lead to

$$|(D_s U^+_{\beta,n}(t), U^+_{\beta,n}(t))_{L^2(\mathbb{H}^1)} - (D_s Q, Q)_{L^2(\mathbb{H}^1)}| \leq (C_0+1)r^2. \tag{24}$$

On the other hand, we estimate $|\|U^+_{\beta,n}(t)\|^4_{L^4(\mathbb{H}^1)} - \|Q\|^4_{L^4(\mathbb{H}^1)}|$ via the conservation of energy. We know that

$$\begin{aligned}|\|U_{\beta,n}(t)\|^4_{L^4(\mathbb{H}^1)} &- \|U^+_{\beta,n}(t)\|^4_{L^4(\mathbb{H}^1)}| \\ &\leq \|W_{\beta,n}(t)\|_{L^4(\mathbb{H}^1)}(\|U_{\beta,n}(t)\|_{L^4(\mathbb{H}^1)} + \|U^+_{\beta,n}(t)\|_{L^4(\mathbb{H}^1)})^3.\end{aligned}$$

Since $\|U_{\beta,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)}$ is bounded due to Lemma 4.6, there exists $C_1 > 0$, such that

$$|\|U_{\beta,n}(t)\|^4_{L^4(\mathbb{H}^1)} - \|U^+_{\beta,n}(t)\|^4_{L^4(\mathbb{H}^1)}| \leq C_1 \|W_{\beta,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)} \leq C_1 r^2. \tag{25}$$

Therefore, from (24) and (25), we get

$$\begin{aligned}\mathcal{E}_\beta(U_{\beta,n}(t)) &= \frac{1}{2}\left(-\frac{\Delta_{\mathbb{H}^1} + \beta D_s}{1-\beta} W_{\beta,n}(t), W_{\beta,n}(t)\right)_{L^2(\mathbb{H}^1)} \\ &\quad + \frac{1}{2}(D_s U^+_{\beta,n}(t), U^+_{\beta,n}(t))_{L^2(\mathbb{H}^1)} - \frac{1}{4}\|U_{\beta,n}(t)\|^4_{L^4(\mathbb{H}^1)} \\ &\geq \frac{1}{2}(D_s Q, Q)_{L^2(\mathbb{H}^1)} - \frac{1}{4}\|U^+_{\beta,n}(t)\|^4_{L^4(\mathbb{H}^1)} - \left(\frac{C_0+1}{2} + \frac{C_1}{4}\right) r^2.\end{aligned}$$

However, due to the conservation of energy and Lemma 4.2, we have

$$\mathcal{E}_\beta(U_{\beta,n}(t)) = \mathcal{E}_\beta(U^n_0) \leq \frac{1}{2}(D_s Q, Q)_{L^2(\mathbb{H}^1)} - \frac{1}{4}\|Q\|^4_{L^4(\mathbb{H}^1)} + C_0 r^2,$$

and, therefore,

$$\frac{1}{4}\|U^+_{\beta,n}(t)\|^4_{L^4(\mathbb{H}^1)} \geq \frac{1}{4}\|Q\|^4_{L^4(\mathbb{H}^1)} - \left(\frac{3C_0+1}{2} + \frac{C_1}{4}\right) r^2.$$

For the reverse inequality, recall the link between Q and the best constant in the embedding $\dot{H}^1(\mathbb{H}^1) \cap V_0^+ \hookrightarrow L^4(\mathbb{H}^1)$: if

$$\inf_{u \in \dot{H}^1(\mathbb{H}^1) \cap V_0^+} \frac{(D_s u, u)^2_{L^2(\mathbb{H}^1)}}{\|u\|^4_{L^4(\mathbb{H}^1)}} = I_+,$$

then

$$(D_s Q, Q)_{L^2(\mathbb{H}^1)} = \|Q\|^4_{L^4(\mathbb{H}^1)} = I_+ = \pi^2.$$

This leads to

$$\begin{aligned}\|U^+_{\beta,n}(t)\|^4_{L^4(\mathbb{H}^1)} &\leq \frac{1}{I_+}(D_sU^+_{\beta,n}(t),U^+_{\beta,n}(t))^2_{L^2(\mathbb{H}^1)}\\ &\leq \frac{1}{I_+}((D_sQ,Q)_{L^2(\mathbb{H}^1)}+(C_0+1)r^2)^2\\ &\leq \frac{1}{I_+}(I_++(C_0+1)r^2)^2\\ &\leq \|Q\|^4_{L^4(\mathbb{H}^1)}+\frac{1}{I_+}(2I_+(C_0+1)+(C_0+1)^2r^2)r^2.\end{aligned}$$

In the end, we have proven that if $r\leq 1$, then there exists $C_2>0$, such that

$$\delta(U^+_{\beta,n}(t))\leq C_2r^2,$$

and Proposition 1.5 immediately implies that for r small enough,

$$d(U^+_{\beta,n}(t),\mathcal{M})^2\leq CC_2r^2.$$

Since $\|W_{\beta,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)}\leq r^2$, we get the Proposition. □

4.3. Weak convergence. — We now know that if $\beta\geq\beta_*(r)$, then for all $n\geq N$ and $t\in\mathbb{R}$,

$$d(U_{\beta,n}(t),\mathcal{M})<c_0r. \tag{26}$$

The aim is now to pass to the limit $n\to+\infty$ in equation (21)

$$\begin{cases}i\partial_tU_{\beta,n}-\frac{\Delta_{\mathbb{H}^1}+\beta D_s}{1-\beta}U_{\beta,n}=\Pi^{(n)}(|U_{\beta,n}|^2U_{\beta,n})\\ U_{\beta,n}(t=0)=U_0^n=\Pi^{(n)}(U_0)\end{cases}$$

and in inequality (26) in order to get a weak solution U_β to equation (18)

$$\begin{cases}i\partial_tU_\beta-\frac{\Delta_{\mathbb{H}^1}+\beta D_s}{1-\beta}U_\beta=|U_\beta|^2U_\beta\\ U_\beta(t=0)=U_0\end{cases},$$

which satisfies

$$d(U_\beta(t),\mathcal{M})\leq c_0r,\quad t\in\mathbb{R}.$$

The method is identical to Sections 3.3 and 3.4 for the limiting equation: we use a uniform bound on $\|\partial_tU_{\beta,n}(t)\|_{\dot{H}^{-1}(\mathbb{H}^1)}$. Due to Ascoli's theorem, the sequence $(U_{\beta,n})_{n\in\mathbb{N}}$ admits a weak limit U_β, which is a weak solution to (18). Then, we construct bounded modulation parameters $X_{\beta,n}(t)$ in order to control the distance between U_β and $\mathcal{M}$.

Lemma 4.8. — *There exists $c_\beta>0$, such that for all $n\geq N$, $t\in\mathbb{R}$ and $X\in\mathbb{R}\times\mathbb{T}\times\mathbb{R}^*_+$,*

$$\|\partial_t(T_XU_{\beta,n})(t)\|_{\dot{H}^{-1}(\mathbb{H}^1)}\leq c_\beta.$$

Proof. — We know from Lemma 4.6 that there exists some constant $C_1 > 0$, such that for all $n \geq N$ and $t \in \mathbb{R}$,

$$\|U_{\beta,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)} \leq C_1.$$

Set $V_{\beta,n} := T_X U_{\beta,n}$. By symmetry invariance, $V_{\beta,n}$ satisfies that for all $n \geq N$ and $t \in \mathbb{R}$,

$$\|V_{\beta,n}(t)\|_{\dot{H}^1(\mathbb{H}^1)} \leq C_1.$$

Moreover, $V_{\beta,n}$ is a solution to some equation

$$\begin{cases} i\partial_t V_{\beta,n} - \frac{\Delta_{\mathbb{H}^1} + \beta D_s}{1-\beta} V_{\beta,n} = \widetilde{\Pi}^{(n)}(|V_{\beta,n}|^2 V_{\beta,n}) \\ V_{\beta,n}(t=0) = V_0^n = \widetilde{\Pi}^{(n)}(U_0) \end{cases}.$$

The projector $\widetilde{\Pi}^{(n)}$ is defined as follows. Write $X = (s, \theta, \alpha)$. For $u \in \dot{H}^{-1}(\mathbb{H}^1)$, we decompose

$$u = \sum_{k \in \mathbb{N}} \sum_{\pm} \Pi_{k,\pm}(u)$$

with $\Pi_{k,\pm}(u) \in \dot{H}^{-1}(\mathbb{H}^1) \cap V_k^{\pm}$ for $(k, \pm) \in \mathbb{N} \times \{\pm\}$. Then

$$\widehat{\widetilde{\Pi}^{(n)}(u)}(x, y, \sigma) = \sum_{k=0}^{n} \sum_{\pm} \widehat{\Pi_{k,\pm}(u)}(x, y, \sigma) \mathbb{1}_{\frac{\alpha^2}{n} \leq |\sigma| \leq \alpha^2 n}.$$

Due to the fact that $\widetilde{\Pi}^{(n)}$ is a projector and the embeddings $L^{\frac{4}{3}}(\mathbb{H}^1) \hookrightarrow \dot{H}^{-1}(\mathbb{H}^1)$ and $\dot{H}^1(\mathbb{H}^1) \hookrightarrow L^4(\mathbb{H}^1)$,

$$\begin{aligned} \|\partial_t V_{\beta,n}(t)\|_{\dot{H}^{-1}(\mathbb{H}^1)} &\leq \frac{1}{1-\beta} \|-(\Delta_{\mathbb{H}^1} + \beta D_s) V_{\beta,n}\|_{\dot{H}^{-1}(\mathbb{H}^1)} \\ &\quad + \|\widetilde{\Pi}^{(n)}(|V_{\beta,n}|^2 V_{\beta,n})\|_{\dot{H}^{-1}(\mathbb{H}^1)} \\ &\leq \frac{2}{1-\beta} \|-\Delta_{\mathbb{H}^1} V_{\beta,n}\|_{\dot{H}^{-1}(\mathbb{H}^1)} + \||V_{\beta,n}|^2 V_{\beta,n}\|_{\dot{H}^{-1}(\mathbb{H}^1)} \\ &\leq \frac{2}{1-\beta} \|V_{\beta,n}\|_{\dot{H}^1(\mathbb{H}^1)} + K_1 \||V_{\beta,n}|^2 V_{\beta,n}\|_{L^{\frac{4}{3}}(\mathbb{H}^1)} \\ &\leq \frac{2}{1-\beta} \|V_{\beta,n}\|_{\dot{H}^1(\mathbb{H}^1)} + K_2 \|V_{\beta,n}\|_{\dot{H}^1(\mathbb{H}^1)}^3 \\ &\leq \frac{2C_1}{1-\beta} + K_2 C_1^3. \end{aligned}$$

□

We deduce the weak convergence of $(U_{\beta,n})_{n \in \mathbb{N}}$, for which the proof is identical to that in Section 3.3 and is based on Ascoli's theorem.

LEMMA 4.9. — *Up to a subsequence,* $(U_{\beta,n})_n$ *converges weakly to a solution* $U_\beta \in \mathcal{C}(\mathbb{R}, \dot{H}^1(\mathbb{H}^1))$ *(with the weak topology) to* (18)

$$\begin{cases} i\partial_t U_\beta - \frac{\Delta_{\mathbb{H}^1} + \beta D_s}{1-\beta} U_\beta = |U_\beta|^2 U_\beta \\ U_\beta(t=0) = U_0 \end{cases}.$$

Moreover, one can see that for all $X \in \mathbb{R} \times \mathbb{T} \times \mathbb{R}_+^*$ and $t_0, t \in \mathbb{R}$, setting $V_{\beta,n} := T_{X^{-1}} U_{\beta,n}$,

$$\left| \frac{\mathrm{d}}{\mathrm{d}t} \|U_{\beta,n}(t) - T_X Q\|^2_{\dot{H}^1(\mathbb{H}^1)} \right| = \left| \frac{\mathrm{d}}{\mathrm{d}t} \|V_{\beta,n}(t) - Q\|^2_{\dot{H}^1(\mathbb{H}^1)} \right|.$$

Due to the conservation of the momentum for equation (18), we have

$$\begin{aligned} \left| \frac{\mathrm{d}}{\mathrm{d}t} \|U_{\beta,n}(t) - T_X Q\|^2_{\dot{H}^1(\mathbb{H}^1)} \right| &= \left| 2(\partial_t V_{\beta,n}(t), D_s Q)_{L^2(\mathbb{H}^1)} \right| \\ &\leq 2 \|\partial_t V_{\beta,n}(t)\|_{\dot{H}^{-1}(\mathbb{H}^1)} \|D_s Q\|_{\dot{H}^1(\mathbb{H}^1)}, \end{aligned}$$

which implies that there exists $c_\beta > 0$, such that for all $t_0, t \in \mathbb{R}$,

$$\|U_{\beta,n}(t) - T_X Q\|^2_{\dot{H}^1(\mathbb{H}^1)} \leq \|U_{\beta,n}(t_0) - T_X Q\|^2_{\dot{H}^1(\mathbb{H}^1)} + c_\beta |t - t_0|.$$

Set $\varepsilon \in (0,1)$ and define $t_1 := \frac{(1+\varepsilon)^2 - 1}{c_\beta} (c_0 r)^2$. Note that t_1 may depend on β, but this is not important because in this section, the varying parameter is n, whereas β is fixed. The construction of $X_{\beta,n}$ as a piecewise constant functional is now the same as for the limiting system. For $k \in \mathbb{Z}$, $X_{\beta,n}$ is constant on $[kt_1, (k+1)t_1[$, equal to some $X^k_{\beta,n} \in \mathbb{R} \times \mathbb{T} \times \mathbb{R}_+^*$ to be chosen. We first set $X^{-1}_{\beta,n} = X^0_{\beta,n} = (0,0,1)$. Then, at time $t_k = kt_1$, $k \geq 1$, we use the fact that $d(U_{\beta,n}(t_k), \mathcal{M}) < c_0 r$ and choose $X^k_{\beta,n}$, such that $\|U_{\beta,n}(t_k) - T_{X^k_{\beta,n}} Q\|_{\dot{H}^1(\mathbb{H}^1)} < c_0 r$. By definition of t_1, for all $k \geq 0$ and $t \in [t_k, t_k + t_1]$, $\|U_{\beta,n}(t) - T_{X^k_{\beta,n}} Q\|_{\dot{H}^1(\mathbb{H}^1)} \leq (1+\varepsilon) c_0 r$. We do a similar construction for negative times. The map $X_{\beta,n}$ satisfies

$$\|U_{\beta,n}(t) - T_{X_{\beta,n}(t)} Q\|_{\dot{H}^1(\mathbb{H}^1)} \leq (1+\varepsilon) c_0 r, \quad t \in \mathbb{R}. \tag{27}$$

It remains to show that $X_{\beta,n}$ is bounded independently of n on bounded intervals. In order to do so, it is enough to control the gap between $X^{k-1}_{\beta,n}$ and $X^k_{\beta,n}$. By construction, at time t_k,

$$\|U_{\beta,n}(t_k) - T_{X^{k-1}_{\beta,n}} Q\|_{\dot{H}^1(\mathbb{H}^1)} \leq (1+\varepsilon) c_0 r$$

and

$$\|U_{\beta,n}(t_k) - T_{X^k_{\beta,n}} Q\|_{\dot{H}^1(\mathbb{H}^1)} < c_0 r,$$

and, therefore,

$$\|T_{X^{k-1}_n} Q - T_{X^k_n} Q\|_{\dot{H}^1(\mathbb{H}^1)} \leq (2+\varepsilon) c_0 r.$$

Using Lemma 3.6 we conclude that if $r \leq r_0$ is small enough (for example, if $3c_0 r_0 \leq \sqrt{\pi} r_1$), then

$$|X_n^{k-1}(X_n^k)^{-1}| \leq c_1.$$

Now, for fixed $t \in \mathbb{R}$, the sequence $(X_{\beta,n}(t))_{n\in\mathbb{N}}$ is bounded, and, therefore, up to extraction, this sequence converges to some $X_\beta(t) \in \mathbb{R} \times \mathbb{T} \times \mathbb{R}_+^*$, and passing to the weak limit in (27),

$$\|U_\beta(t) - T_{X_\beta(t)}Q\|_{\dot{H}^1(\mathbb{H}^1)} \leq (1+\varepsilon)c_0 r.$$

Appendix A. Rate of convergence of Q_β to Q

In order to conclude the proof of Theorem 4.1 it only remains to make precise the convergence rate of $(Q_\beta)_\beta$ to Q as β tends to 1. Decompose

$$Q_\beta = Q_\beta^+ + R_\beta,$$

where $Q_\beta^+ \in \dot{H}^1(\mathbb{H}^1) \cap V_0^+$ and $R_\beta \in \bigoplus_{(k,\pm)\neq(0,+)} \dot{H}^1(\mathbb{H}^1) \cap V_k^\pm$. We improve the bound from [8]

$$\delta(Q_\beta^+) + \|R_\beta\|_{\dot{H}^1(\mathbb{H}^1)} = \mathcal{O}((1-\beta)^{\frac{1}{2}}).$$

Proposition A.1. — *Let $\varepsilon > 0$. Then, as β tends to 1,*

$$\delta(Q_\beta^+) + \|R_\beta\|_{\dot{H}^1(\mathbb{H}^1)} = \mathcal{O}((1-\beta)^{2-\varepsilon}),$$

which implies that

$$\|Q_\beta - Q\|_{\dot{H}^1(\mathbb{H}^1)} = \mathcal{O}((1-\beta)^{1-\frac{\varepsilon}{2}}).$$

Proof. — Assume that we have proven that

$$\delta(Q_\beta^+) + \|R_\beta\|_{\dot{H}^1(\mathbb{H}^1)} = \mathcal{O}((1-\beta)^\gamma)$$

for some exponent $\gamma > 0$ (for example, we already know that it is true for $\gamma = \frac{1}{2}$, see [8]), and, therefore,

$$\|Q_\beta - Q\|_{\dot{H}^1(\mathbb{H}^1)} = \mathcal{O}((1-\beta)^{\frac{\gamma}{2}}).$$

We increase the exponent γ by showing that actually

$$\delta(Q_\beta^+) + \|R_\beta\|_{\dot{H}^1(\mathbb{H}^1)} = \mathcal{O}((1-\beta)^{1+\frac{\gamma}{2}}).$$

Then, we conclude by iteration, since the sequence $\gamma_{n+1} = 1 + \frac{\gamma_n}{2}$ with $\gamma_0 = \frac{1}{2}$ is convergent to 2.

Since $R_\beta \in \bigoplus_{(k,\pm)\neq(0,+)} \dot{H}^1(\mathbb{H}^1) \cap V_k^\pm$, the norms $\|R_\beta\|_{\dot{H}^1(\mathbb{H}^1)}$ and $\|-(\Delta_{\mathbb{H}^1} + \beta D_s)R_\beta\|_{\dot{H}^{-1}(\mathbb{H}^1)}$ are equivalent

$$\|R_\beta\|_{\dot{H}^1(\mathbb{H}^1)} \leq 2\|-(\Delta_{\mathbb{H}^1} + \beta D_s)R_\beta\|_{\dot{H}^{-1}(\mathbb{H}^1)}.$$

Projecting the equation satisfied by Q_β

$$-\frac{\Delta_{\mathbb{H}^1}+\beta D_s}{1-\beta}Q_\beta = |Q_\beta|^2 Q_\beta$$

on $\bigoplus_{(k,\pm)\neq(0,+)} \dot{H}^1(\mathbb{H}^1)\cap V_k^\pm$, we deduce that

$$\|R_\beta\|_{\dot{H}^1(\mathbb{H}^1)} \leq 2(1-\beta)\|(\mathrm{Id}-\Pi_{0,+})(|Q_\beta|^2Q_\beta)\|_{\dot{H}^{-1}(\mathbb{H}^1)}.$$

Since $|Q|^2Q = D_sQ \in \dot{H}^{-1}(\mathbb{H}^1)\cap V_0^+$, one can make this term appear in the right-hand side term of the inequality:

$$\begin{aligned}\|R_\beta\|_{\dot{H}^1(\mathbb{H}^1)} &\leq 2(1-\beta)\|(\mathrm{Id}-\Pi_{0,+})(|Q_\beta|^2Q_\beta - |Q|^2Q)\|_{\dot{H}^{-1}(\mathbb{H}^1)}\\ &\leq 2K(1-\beta)\|(\mathrm{Id}-\Pi_{0,+})(|Q_\beta|^2Q_\beta - |Q|^2Q)\|_{L^{\frac{4}{3}}(\mathbb{H}^1)}.\end{aligned}$$

Now, since $(\mathrm{Id}-\Pi_{0,+})$ defines a bounded operator on $L^{\frac{4}{3}}(\mathbb{H}^1)$, there exist $C_1, C_2 > 0$, such that

$$\begin{aligned}\|R_\beta\|_{\dot{H}^1(\mathbb{H}^1)} &\leq C_1(1-\beta)\||Q_\beta|^2Q_\beta - |Q|^2Q\|_{L^{\frac{4}{3}}(\mathbb{H}^1)}\\ &\leq C_2(1-\beta)\|Q_\beta - Q\|_{\dot{H}^1(\mathbb{H}^1)}(\|Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)} + \|Q\|_{\dot{H}^1(\mathbb{H}^1)})^2.\end{aligned}$$

However, since $(Q_\beta)_\beta$ is bounded in $\dot{H}^1(\mathbb{H}^1)$, we get that there exists $C_3 > 0$, such that

$$\begin{aligned}\|R_\beta\|_{\dot{H}^1(\mathbb{H}^1)} &\leq C_3(1-\beta)\|Q_\beta - Q\|_{\dot{H}^1(\mathbb{H}^1)}\\ &= \mathcal{O}((1-\beta)^{1+\frac{\gamma}{2}}).\end{aligned}$$

Therefore,

$$\begin{aligned}|\|Q_\beta^+\|^2_{\dot{H}^1(\mathbb{H}^1)} - \|Q_\beta\|^2_{\dot{H}^1(\mathbb{H}^1)}| &\leq 2\|R_\beta\|_{\dot{H}^1(\mathbb{H}^1)}(\|R_\beta\|_{\dot{H}^1(\mathbb{H}^1)} + \|Q_\beta\|_{\dot{H}^1(\mathbb{H}^1)})\\ &= \mathcal{O}((1-\beta)^{1+\frac{\gamma}{2}})\end{aligned}$$

and

$$\begin{aligned}|\|Q_\beta^+\|^4_{L^4(\mathbb{H}^1)} - \|Q_\beta\|^4_{L^4(\mathbb{H}^1)}| &\lesssim \|R_\beta\|_{L^4(\mathbb{H}^1)}(\|R_\beta\|_{L^4(\mathbb{H}^1)} + \|Q_\beta\|_{L^4(\mathbb{H}^1)})^3\\ &= \mathcal{O}((1-\beta)^{1+\frac{\gamma}{2}}),\end{aligned}$$

which means that

$$\delta(Q_\beta^+) = \mathcal{O}((1-\beta)^{1+\frac{\gamma}{2}}).$$

It now remains to consider the sequence $\gamma_{n+1} = 1 + \frac{\gamma_n}{2}$, $\gamma_0 = \frac{1}{2}$, which is convergent to 2. □

Acknowledgements. — The author is grateful to her PhD advisor P. Gérard for his generous advice and encouragement. She would also like to thank the referee for carefully reading the manuscript and making constructive remarks.

BIBLIOGRAPHY

[1] T. Aubin – "Problèmes isopérimétriques et espaces de Sobolev", *J. Differential Geom.* **11** (1976), no. 4, p. 573–598.

[2] H. Bahouri, C. Fermanian-Kammerer & I. Gallagher – "Dispersive estimates for the Schrödinger operator on step-2 stratified Lie groups", *Analysis and PDE* **9** (2016), no. 3, p. 545–574.

[3] H. Bahouri, P. Gérard & C.-J. Xu – "Espaces de Besov et estimations de Strichartz généralisées sur le groupe de Heisenberg", *Journal d'Analyse Mathématique* **82** (2000), no. 1, p. 93–118.

[4] D. Békollé, A. Bonami, G. Garrigós, C. Nana, M. Peloso & F. Ricci – "Lecture notes on Bergman projectors in tube domains over cones: an analytic and geometric viewpoint", *IMHOTEP: African Journal of Pure and Applied Mathematics* **5** (2012).

[5] J. Bellazzini, V. Georgiev, E. Lenzmann & N. Visciglia – "On traveling solitary waves and absence of small data scattering for nonlinear half-wave equations", *Communications in Mathematical Physics* (2019).

[6] J. Bellazzini, V. Georgiev & N. Visciglia – "Long time dynamics for semi-relativistic nls and half wave in arbitrary dimension", *Mathematische Annalen* **371** (2018), no. 1, p. 707–740.

[7] N. Burq, P. Gérard & N. Tzvetkov – "Bilinear eigenfunction estimates and the nonlinear Schrödinger equation on surfaces", *Inventiones mathematicae* **159** (2005), no. 1, p. 187–223.

[8] L. Gassot – "On the radially symmetric traveling waves for the Schrödinger equation on the Heisenberg group", 2019, arXiv:1904.07010, to appear in Pure and Applied Analysis.

[9] P. Gérard & S. Grellier – "L'équation de Szegő cubique", *Séminaire X-EDP* (2008).

[10] P. Gérard, E. Lenzmann, O. Pocovnicu & P. Raphaël – "A two-soliton with transient turbulent regime for the cubic half-wave equation on the real line", *Annals of PDE* **4** (2018), no. 1, p. 7.

[11] P. Gérard & F. Rousset – "Propriétés qualitatives de solutions d'EDP non linéaires", 2016, Graduate course, Orsay.

[12] M. D. Hierro – "Dispersive and Strichartz estimates on H-type groups", *Studia Mathematica* **169** (2005), no. 1, p. 1–20.

[13] C. E. Kenig & F. Merle – "Global well-posedness, scattering and blow-up for the energy-critical, focusing, non-linear Schrödinger equation in the radial case", *Inventiones mathematicae* **166** (2006), no. 1, p. 645–675.

[14] J. Krieger, E. Lenzmann & P. Raphaël – "Nondispersive solutions to the L^2-critical half-wave equation", *Archive for Rational Mechanics and Analysis* **209** (2013), no. 1, p. 61–129.

[15] C. D. Lellis & L. Székelyhidi Jr. – "The Euler equations as a differential inclusion", *Annals of Mathematics* **170** (2009), p. 1417–1436.

[16] O. Pocovnicu – "Traveling waves for the cubic Szegő equation on the real line", *Analysis and PDE* **4** (2011), no. 3, p. 379–404.
[17] ———, "Soliton interaction with small Toeplitz potentials for the Szegő equation on $\mathbb{R}$", *Dynamics of Partial Differential Equations* **9** (2012), no. 1, p. 1–27.
[18] E. M. Stein & T. S. Murphy – *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton University Press, 1993.
[19] G. Talenti – "Best constant in Sobolev inequality", *Annali di Matematica Pura ed Applicata* **110** (1976), no. 1, p. 353–372.

Bull. Soc. Math. France
149 (1), 2021, p. 55-117

SPACES OF ALGEBRAIC MEASURE TREES AND TRIANGULATIONS OF THE CIRCLE

BY WOLFGANG LÖHR & ANITA WINTER

ABSTRACT. — In this paper, we present with *algebraic trees*, a novel notion of (continuum) trees that generalizes countable graph-theoretic trees to (potentially) uncountable structures. For this purpose, we focus on the tree structure given by the branch-point map, which assigns to each triple of points their branch point. We give an axiomatic definition of algebraic trees, define a natural topology, and equip them with a probability measure on the Borel-σ-field. Under an order-separability condition, algebraic (measure) trees can be considered as tree structure equivalence classes of metric (measure) trees (i.e., subtrees of $\mathbb{R}$-trees). Using Gromov-weak convergence (i.e., sample distance convergence) of the particular representatives given by the metric arising from the distribution of branch points, we define a metrizable topology on the space of equivalence classes of algebraic measure trees.

In many applications, binary trees are of particular interest. We introduce on that subspace with the sample shape and the sample subtree mass convergence two additional, natural topologies. Relying on the connection to triangulations of the circle, we show that all three topologies are actually the same, and the space of binary algebraic measure trees is compact. To this end, we provide a formal definition of triangulations of the circle and show that the coding map that sends a triangulation to an algebraic measure tree is a continuous surjection onto the subspace of binary algebraic nonatomic measure trees.

Texte reçu le 9 avril 2020, modifié le 10 juillet 2020, accepté le 21 juillet 2020.

WOLFGANG LÖHR, University of Duisburg-Essen, Mathematics, 45117 Essen, Germany •
E-mail : wolfgang.loehr@uni-due.de
ANITA WINTER, University of Duisburg-Essen, Mathematics, 45117 Essen, Germany •
E-mail : anita.winter@uni-due.de

Mathematical subject classification (2010). — 60B10, 05C05; 60D05, 54F50, 57R05, 60C05.

Key words and phrases. — Continuum tree, $\mathbb{R}$-tree, Metric tree, Branch-point map, Convergence of trees, Sample shape convergence, Gromov-weak convergence, Brownian CRT, β-splitting tree, Yule tree, State space.

Research supported by *DFG-RTG 2131* and by *DFG Priority Programme SPP 1590*. Wolfgang Löhr was supported by the *DFG project 415705084*.

0037-9484/2021/55/$ 5.00
doi:10.24033/bsmf.2825

Résumé (*Espaces d'arbres algébriques mesurés et triangulations du cercle*). — Nous présentons dans cet article une nouvelle notion d'arbres (continus), appelés arbres algébriques, qui généralise celle des arbres dénombrables (en théorie des graphes) à des structures (potentiellement) indénombrables. Pour cela, nous nous intéressons uniquement à la structure d'arbre donnée par la fonction de branchement, qui à chaque triplet de points associe leur point de branchement. Nous définissons les arbres algébriques de manière axiomatique et les munissons d'une topologie naturelle ainsi que d'une mesure de probabilité sur la tribu borélienne. Sous une condition de séparabilité de la structure d'ordre, les arbres algébriques mesurés peuvent être considérés comme des classes d'équivalence d'arbres métriques mesurés (i.e. des sous-arbres de $\mathbb{R}$-arbres). À chaque arbre algébrique mesuré on peut associer un arbre métrique en considérant la distance générée par la distribution des points de branchement. En utilisant la convergence Gromov-faible (i.e. la convergence des distances échantillonnées) de ces arbres métriques mesurés associés, nous définissons une topologie métrisable sur l'espace des classes d'équivalence d'arbres algébriques mesurés.

Le cas des arbres binaires est particulièrement intéressant en termes d'applications. Nous introduisons sur ce sous-espace deux autres topologies naturelles, la convergence des cladogrammes engendrés par un échantillon de points de l'arbre et la convergence des masses des sous-arbres associés à un échantillon. En utilisant le lien avec les triangulations du cercle, nous montrons que ces trois topologies sont identiques, et que l'espace des arbres algébriques mesurés binaires est compact. Nous donnons pour cela une définition formelle des triangulations du cercle, et nous montrons que la fonction de codage qui à une triangulation associe un arbre algébrique mesuré est une surjection continue sur le sous-espace des arbres algébriques binaires munis d'une mesure diffuse.

1. Introduction

Graph-theoretic trees are abundant in mathematics and its applications, from computer science to theoretical biology. A natural question is how to define limits and limit objects as the size of the trees tends to infinity. On the one hand, there are *local* approaches yielding countably infinite graphs or generalized so-called graphings with a Benjamini–Schramm-type approach (going back to [10], see [49, Part 4]). On the other hand, if one takes a more *global* point of view, as we are doing here, the predominant approach is to consider graph-theoretic trees as metric spaces equipped with the (rescaled) graph distance. Then the limit objects are certain "tree-like" metric spaces, most prominently the so-called $\mathbb{R}$-trees introduced in [56]. They are also of independent interest, e.g., to study isometry groups of hyperbolic space ([52]) or as generalized universal covering spaces in the study of the fundamental groups of one-dimensional spaces ([30]). The characterization of the topological structures induced by $\mathbb{R}$-trees has received considerable attention ([51, 50, 28]). Here, instead of the topological structures, we are more interested in the "tree structures" induced by $\mathbb{R}$-trees. We formalize the tree structure with a branch point map and call the resulting axiomatically defined objects *algebraic trees*.

While, unlike for metric spaces, we do not know any useful notion of convergence for topological spaces or topological measure spaces, it is essential for us that we can define a very useful convergence of algebraic measure trees.

Our main motivation lies in suitable state spaces for tree-valued stochastic processes. The construction and investigation of scaling limits of tree-valued Markov chains within a metric space setup started with the continuum analogues of the Aldous–Broder algorithm for sampling a uniform spanning tree from the complete graph ([26]) and of the tree-valued subtree-prune and regraft Markov chain used in the reconstruction of phylogenetic trees ([27]). It continued with the construction of evolving genealogies of infinite size populations in population genetics ([37, 20, 44, 54, 38]) and in population dynamics ([35, 45]). Moreover, continuum analogues of pruning procedures were constructed ([2, 1, 48, 42, 43]). All these constructions have in common that they encode trees as metric (measure) spaces or bimeasure $\mathbb{R}$-trees, and equip the respective space of trees with the Gromov–Hausdorff ([39]), Gromov-weak ([34, 36, 46]), Gromov–Hausdorff-weak ([58, 8]), or leaf-sampling weak-vague topology ([48]).

In the present paper, we shift the focus from the metric to the tree structure for several reasons. First, checking compactness or tightness criteria for (random) metric (measure) spaces is not always easy, and some natural sequences of trees do not converge as metric (measure) spaces with a uniform rescaling of edge lengths. At least for the subspace of binary algebraic measure trees that we introduce, the situation is much more favorable, because it turns out to be compact. Second, the metric is often less canonical than the tree structure in situations where it is not clear that every edge should have the same length, e.g., in a phylogenetic tree, where edges might correspond to very different evolutionary time spans. Third, one might want to preserve certain functionals of the tree structure in the limit. For instance, the limit of binary trees is not always binary in the metric space setup, while this will be the case for our algebraic measure trees. Also, the centroid function used in [7] is not continuous on spaces of metric measure trees, but it is continuous on our space.

The starting point of our construction is the notion of an $\mathbb{R}$-tree (see [56, 22, 13, 24]). There are many equivalent definitions, but the following one is the most convenient for us:

DEFINITION 1.1 ($\mathbb{R}$-trees). — A metric space (T, r) is an $\mathbb{R}$-*tree* iff it satisfies the following:

(RT1) (T, r) satisfies the so-called 4-*point condition*, i.e.,

$$r(x_1, x_2) + r(x_3, x_4) \leq \max\left\{r(x_1, x_3) + r(x_2, x_4),\, r(x_1, x_4) + r(x_2, x_3)\right\} \tag{1}$$

for all $x_1, x_2, x_3, x_4 \in T$.

(RT2) (T, r) is a connected metric space.

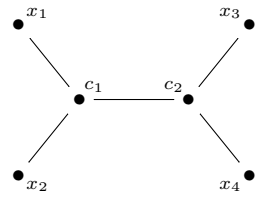

FIGURE 1.1. The only possible tree shape spanned by four points separates them into two pairs. Here, $r(x_1,x_2) + r(x_3,x_4) < \max\{r(x_1,x_3)+r(x_2,x_4),\, r(x_1,x_4)+r(x_2,x_3)\}$, while any other permutation yields equality. Furthermore, $c_1 = c(x_1,x_2,x_3) = c(x_1,x_2,x_4)$ and $c_2 = c(x_1,x_3,x_4) = c(x_2,x_3,x_4)$.

Note that any metric space (T,r) satisfying (RT1) and (RT2) admits a *branch-point map* $c\colon T^3 \to T$, i.e., for all $x_1,x_2,x_3 \in T$ there exists a unique point $c(x_1,x_2,x_3) \in T$, such that

$$\big\{c(x_1,x_2,x_3)\big\} = [x_1,x_2] \cap [x_1,x_3] \cap [x_2,x_3], \tag{2}$$

where for $x,y \in T$ the *interval* $[x,y]$ is defined as

$$[x,y] := \big\{z \in T :\, r(x,z) + r(z,y) = r(x,y)\big\}. \tag{3}$$

Given the branch-point map c, we can recover the intervals via the identity

$$[x,y] = \big\{z \in T :\, c(x,y,z) = z\big\}. \tag{4}$$

While condition (RT1) is crucial for trees, as it reflects the fact that there is only one possible shape for the subtree spanned by four points (as shown in Figure 1.1), the assumption of connectedness can be relaxed. In [9], the notion of a *metric tree* was introduced to allow for a unified setup in discrete and continuous situations. A metric tree (T,r) is defined as a metric space which can be embedded isometrically into an $\mathbb{R}$-tree, such that it contains all branch points $c(x_1,x_2,x_3)$, $x_1,x_2,x_3 \in T$, as defined by (2). To exclude nontree graphs satisfying the 4-point condition (see Figure 1.2) we have to require the property of containing the branch points explicitly.

DEFINITION 1.2 (metric trees). — A metric space (T,r) is a *metric tree* if the following holds:

(MT1) (T,r) satisfies the 4-point condition (RT1).

(MT2) (T,r) admits all branch points, i.e., for all $x_1,x_2,x_3 \in T$, there exists a (necessarily unique) $c(x_1,x_2,x_3) \in T$, such that

$$r\big(x_i,\, c(x_1,x_2,x_3)\big) + r\big(c(x_1,x_2,x_3),\, x_j\big) = r(x_i,x_j) \tag{5}$$

for all $i,j \in \{1,2,3\}$, $i \neq j$.

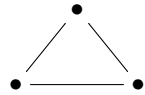

FIGURE 1.2. The graph shown here is not a tree, but the vertices satisfy the 4-point condition with respect to the graph-distance. Condition (MT2) fails.

Our main goal is to forget the metric while keeping the tree structure encoded by the branch-point map. To axiomatize the latter, note that for metric trees, the branch-point map satisfies the following obvious properties:

(BPM1) The map $c\colon T^3 \to T$ is symmetric.
(BPM2) The map $c\colon T^3 \to T$ satisfies the 2-*point condition* that

$$c(x,y,y) = y. \tag{6}$$

for all $x, y \in T$.
(BPM3) The map $c\colon T^3 \to T$ satisfies the 3-*point condition* that

$$c\big(x,y,\, c(x,y,z)\big) = c(x,y,z). \tag{7}$$

for all $x, y, z \in T$.
(BPM4) The map $c\colon T^3 \to T$ satisfies the 4-*point condition* that

$$c(x_1,x_2,x_3) \in \big\{c(x_1,x_2,x_4),\, c(x_1,x_3,x_4),\, c(x_2,x_3,x_4)\big\} \tag{8}$$

for all $x_1, x_2, x_3, x_4 \in T$.

DEFINITION 1.3 (algebraic tree). — An *algebraic tree* (T,c) consists of a set $T \neq \emptyset$ and a branch-point map $c\colon T^3 \to T$ satisfying (BPM1)–(BPM4).

We define a natural topology on an algebraic tree (T,c) as follows. For each $x \in T$, we define an equivalence relation $\sim_x$ on $T \setminus \{x\}$, such that for all $y, z \in T \setminus \{x\}$, $y \sim_x z$ iff $c(x,y,z) \neq x$. For $y \in T \setminus \{x\}$, we denote by

$$\mathcal{S}_x(y) \coloneqq \{z \in T :\, z \sim_x y\} \tag{9}$$

the equivalence class with respect to $\sim_x$, which contains y. $\mathcal{S}_x(y)$ should be thought of as a subtree rooted at (but not containing) x. We consider the topology generated by sets of the form (9) with $x \neq y$ and denote by $\mathcal{B}(T,c)$ the corresponding Borel σ-algebra.

Our first main result (Theorem 2.27) relates metric trees with algebraic trees. On the one hand, if (T,r) is a metric tree, then it is clear that T together with the map c from (MT2) yields an algebraic tree. On the other hand, we show that every order separable algebraic tree (Definition 2.21) is induced by a metric

tree in this way. More concretely, if ν is a measure on $\mathcal{B}(T,c)$, which is finite and nonzero on nondegenerate intervals, i.e., on sets of the form

$$[x,y] \coloneqq \big\{z \in T : c(x,y,z) = z\big\}, \tag{10}$$

for $x, y \in T$, $x \neq y$, then a metric representation of (T,c) is given by

$$r_\nu(x,y) \coloneqq \nu\big([x,y]\big) - \tfrac{1}{2}\nu\big(\{x\}\big) - \tfrac{1}{2}\nu\big(\{y\}\big). \tag{11}$$

Next, we equip an algebraic tree (T,c) with a sampling probability measure μ on $\mathcal{B}(T,c)$ and call the resulting triple a (T,c,μ) *algebraic measure tree*. Two algebraic measure trees (T,c,μ) and (T',c',μ') are equivalent (compare this with Definition 3.2), if there are $A \subseteq T$, $A' \subseteq T'$ and a bijection $\phi\colon A \to A'$, such that the following holds:

- $\mu(A) = \mu'(A') = 1$, $c(A^3) \subseteq A$ and $c'((A')^3) \subseteq A'$.
- ϕ is measure preserving, and $c'(\phi(x),\phi(y),\phi(z)) = \phi(c(x,y,z))$ for all $x,y,z \in T$.

Denote by $\mathbb{T}$ the space of all equivalence classes of order separable algebraic measure trees. We equip $\mathbb{T}$ with a topology based on Gromov-weak topology (introduced in [36] and shown in [46] to be equivalent to Gromov's $\square_1$-topology from [39]). For that purpose, we introduce a particular metric representation of an algebraic measure tree. As metric representations are far from being unique, we will consider the intrinsic metric r_ν, which comes from the branch-point distribution, i.e., the image measure $\nu \coloneqq c_*\mu^{\otimes 3}$ of $\mu^{\otimes 3}$ under the branch-point map c. We declare that

$$\begin{aligned} &(T_n,c_n,\mu_n) \xrightarrow[n\to\infty]{} (T,c,\mu) \quad \text{iff} \\ &(T, r_{(c_n)_*\mu_n^{\otimes 3}}, \mu_n) \xrightarrow[n\to\infty]{} (T, r_{c_*\mu^{\otimes 3}}, \mu) \text{ Gromov-weakly,} \end{aligned} \tag{12}$$

or equivalently, $\Phi((T_n,c_n,\mu_n)) \xrightarrow[n\to\infty]{} \Phi((T,c,\mu))$ for all test functions of the form

$$\Phi(T,c,\mu) = \Phi^{n,\phi}(T,c,\mu) \coloneqq \int_{T^n} \phi\big((r_{c_*\mu^{\otimes 3}}(x_i,x_j))_{1\le i,j\le n}\big)\, \mu^{\otimes n}(\mathrm{d}\underline{x}), \tag{13}$$

where $n \in \mathbb{N}$ and $\phi \in \mathcal{C}_b(\mathbb{R}^{n\times n})$. We refer to this convergence as the branch-point distribution distance Gromov-weak convergence, or shortly, *bpdd-Gromov-weak convergence*. It is important to keep in mind that—even though bpdd-Gromov-weak convergence is defined via the Gromov-weak convergence of particular metric representations—Gromov-weak convergence of a sequence $(T_n,r_n,\mu_n)_{n\in\mathbb{N}}$ of metric measure trees does not imply bpdd-Gromov-weak convergence of the corresponding sequence of algebraic measure trees. For instance, if the diameters $\sup_{x,y\in\mathbb{T}_n} r_n(x,y)$ converge to zero, the sequence of metric measure trees converges to the trivial (one-point) tree, while the corresponding sequence of algebraic measure trees might, or might not, converge to the same or a different

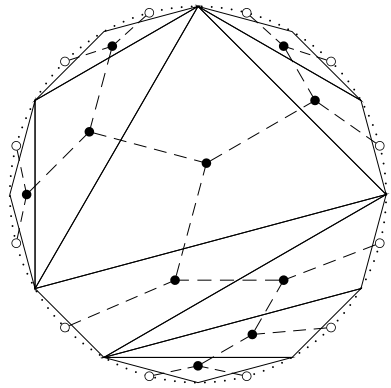

FIGURE 1.3. A triangulation of the 12-gon and the tree coded by it.

limit. The same reasoning also applies to the stronger Gromov-Hausdorff-weak topology.

A particular subclass of interest is the space of binary algebraic measure trees. Similarly to encoding compact $\mathbb{R}$-trees by a continuous excursion on the unit interval, binary algebraic trees can be encoded by *subtriangulations of the circle* (see Figure 1.3), where a subtriangulation of the circle $\mathbb{S}$ is a closed, nonempty subset C of $\mathbb{D}$ satisfying the following two conditions:

(Tri1) The complement of the convex hull of C consists of open interiors of triangles.

(Tri2) C is the union of noncrossing (nonintersecting except at endpoints), possibly degenerate closed straight line segments with endpoints in $\mathbb{S}$.

Such an encoding was introduced by David Aldous in [4, 5], and there has since then been an increasing amount of research in the random tree community using this approach (e.g., [18, 12, 17]). Also more general -angulations and dissections have been considered, which allow for encoding not necessarily binary trees ([14, 15]). Note, however, that the relation between triangulations and trees has never been made explicit, except for the finite case, where the tree is the dual graph.

Aldous originally defines a triangulation of the circle as a closed subset of the disc, the complement of which is a disjoint union of open triangles with vertices on the circle ([5, Definition 1]). We modify his definition in two respects. First, we add Condition (Tri2), which enforces the existence of branch points, and under which triangulations of the circle are precisely the Hausdorff-metric limits of triangulations of n-gons as $n \to \infty$. Second, we extend the definitions to subtriangulation of the circle (triangulations of a subset of the circle), which allow for encoding algebraic measure trees with point masses on leaves.

In fact, triangulations of the whole circle encode binary trees with nonatomic measures, which is relevant in the case of Aldous's CRT. We formally construct the coding map that associates to a subtriangulation of the circle the corresponding binary algebraic measure tree with point masses restricted to the leaves. Furthermore, we show that—similarly to the case of coding compact $\mathbb{R}$-trees by continuous excursions—the coding map is *surjective* and *continuous* when the set of subtriangulations is equipped with the Hausdorff metric topology and the set of binary algebraic measure trees with our bpdd-Gromov-weak topology (Theorem 4.8).

We also analyze the subspace of binary algebraic measure trees with point masses restricted to the leaves in more detail. Our third main result (Theorem 5.19) states that this space in the bpdd-Gromov-weak topology is topologically as nice as it gets, namely a compact, metrizable space. We also give two more notions of convergence, which turn out to be equivalent to bpdd-Gromov-weak convergence on this subspace. One is of combinatorial nature and is based on the weak convergence of test functions of the form

$$\begin{aligned}\Phi(T,c,\mu) &= \Phi^{n,\mathfrak{t}}(T,c,\mu)\\ &\coloneqq \mu\big(\big\{(u_1,\dots,u_n)\in T^n : \mathfrak{s}_{(T,c)}(u_1,\dots,u_n)=\mathfrak{t}\big\}\big),\end{aligned} \tag{14}$$

where $\mathfrak{t}$ is an n-cladogram (a binary graph-theoretic tree with n leaves), and $\mathfrak{s}_{(T,c)}$ denotes the shape spanned by a finite sample in (T,c) (Definitions 5.1 and 5.2). The other one is more in the spirit of stochastic analysis and is based on weak convergence of the *tensor of subtree masses* read off the algebraic measure subtree spanned by a finite sample (see Definition 5.12). This equivalence allows to switch between different perspectives and turns out to be very useful for the following reasons:

- Using convergence of sample bpd-distance matrices allows us to exploit well-known results about Gromov-weak convergence.
- Showing convergence of graph theoretic tree-valued Markov chains as the number of vertices tends to infinity is, due to the combinatorial nature of the Markov chains, often easiest by showing convergence of the sample shape distributions. This was recently successfully applied in the construction of the conjectured continuum limit of the Aldous chain ([7]) in [47] and of the continuum limit of the $\alpha = 1$-Ford chain ([31]) in [53].
- The convergence of sample subtree-mass tensor distributions allows us to analyze the limit process with stochastic analysis methods and gives more insight into the global structure of the evolving random trees.

Related work. — As an alternative with better compactness properties to Gromov–Hausdorff convergence of discrete trees, in [14] Curien suggested looking at convergence of coding triangulations (in Hausdorff metric topology). He

also proposed reading off a measured, ordered, *topological tree* from the limit triangulation and sketched the construction as the quotient with respect to some equivalence relation in the special case of the Brownian triangulation. Note, however, that the topological information cannot be completely encoded by the triangulation, because the latter only encodes the algebraic measure tree by Theorem 4.8, and the algebraic structure does not determine the topological structure uniquely (see Example 2.37). Therefore, Curien did not obtain a general map from the space of triangulations to a space of trees.

In order to turn the set of valuations on the ring $\mathbb{C}[\![x, y]\!]$ into the so-called *valuative tree*, in [29] Favre and Jonsson use partial orders to define the tree structure. Using partial orders is essentially equivalent to using branch-point maps, and under some additional assumptions (separability, order completeness and edge freeness), their *nonmetric trees* are equivalent to our algebraic trees. We want to stress, however, that for our theory, the branch-point map plays a much more crucial role than the partial order. The relation between partial orders and algebraic trees is discussed further in Section 2.

The random exchangeable *didendritic systems* recently introduced by Evans, Grübel, and Wakolbinger in [25] can be considered as rooted, ordered versions of binary algebraic measure trees with diffuse measure on the set of leaves. A didendritic system is an equivalence relation on $\mathbb{N} \times \mathbb{N}$ together with two partial orders on the set of equivalence classes. An exchangeable didendritic system is similar to our sequence of sample-shape distributions. The authors also introduce a particular metric representation as an $\mathbb{R}$-tree. Even though it is implicit in their work that they think of the set of exchangeable didendritic systems as being equipped with a kind of sample shape convergence, they do not define it explicitly and do not analyze the resulting topological space.

Close relatives of algebraic measure trees were recently studied independently by Forman in [32]. He uses ideas from [33] to represent rooted trees by so-called *hierarchies* (certain sets of subsets) on $\mathbb{N}$, which are similar to the didendritic systems in [25], but unordered. Thus, exchangeable random hierarchies can be thought of as rooted versions of algebraic measure trees. Forman shows that the resulting equivalence classes of rooted measure $\mathbb{R}$-trees coincide with the so-called *mass-structural* equivalence classes, which he defines by bijections preserving intervals, as well as masses of points, intervals, and certain subtrees. He also singles out a particular representative, which he calls an *interval partition tree*, with the essentially same metric as in [25] (not restricted to the binary case). This metric follows a similar idea to, but is different from, our r_ν. Note that [32] does not talk about convergence of trees or introduce the notion of a "continuum tree" without a measure.

Outline. — The rest of the paper is organized as follows. In Section 2, we introduce our concept of *algebraic trees* by formalizing the branch-point map as a tertiary operation on the tree. We also introduce an intrinsic Hausdorff

topology and characterize compactness (Proposition 2.19) and second countability (Proposition 2.20). We show that under a separability constraint, algebraic trees can be seen as metric trees (subtrees of $\mathbb{R}$-trees), where the metric structure has been "forgotten" (Theorem 2.27), and give an example that the separability condition cannot be dropped without replacement.

In Section 3, we introduce the space of (equivalence classes of) order-separable *algebraic measure trees* and equip it with the Gromov-weak topology with respect to the metric associated with the branch-point distribution. We show that the resulting space is separable and metrizable (Corollary 3.9). Furthermore, we prove a Carathéodory-type extension theorem, which is helpful for constructing algebraic measure trees (Propositions 3.12 and 3.13).

In Section 4, we give a definition of *triangulations of the circle* and show that they are precisely the limits of triangulations of n-gons (Proposition 4.3). We also formalize the notion of the algebraic measure tree associated with a given triangulation of the circle. This correspondence has been alluded to in the literature, but it has never been made precise (except for finite trees). We show that the resulting *coding map* (mapping triangulations to trees) is well defined and surjective onto the space of binary algebraic measure trees with nonatomic measure. Furthermore, the coding map is *continuous*, if the space of triangulation is equipped with the Hausdorff-metric topology and the space of trees with the bpdd-Gromov-weak topology (Theorem 4.8).

In Section 5, we consider the subspace of *binary* algebraic measure trees and introduce two other natural notions of convergence. We use the construction of the coding map from Section 4 to show that on this subspace, all three notions of convergence are actually equivalent and define the same topology (Theorem 5.19). This topology turns the subspace of binary algebraic measure trees into a *compact, metrizable* space, which in particular implies that it is a closed subset of the space of algebraic measure trees. In this section, we also finish the proof of Theorem 4.8 by showing continuity of the coding map.

In Section 6, we consider the example of the continuum limits of sampling consistent families of random trees and illustrate it with the example of so-called β-splitting trees introduced in [6]. This family includes the uniform binary tree (converging to the Brownian CRT) and the Yule tree (aka Kingman tree or random binary search tree).

2. Algebraic trees

In this section, we introduce algebraic trees. In Section 2.1, we formalize the "tree structure" common to both graph-theoretic trees and metric trees by a function that maps every triplet of points in the tree to the corresponding branch point. We show that the set of defining properties is rich enough to obtain known concepts, such as leaves, branch points, degree, edges, intervals,

subtrees spanned by a set, discrete and continuum trees, etc. In Section 2.2, we introduce the notion of structure-preserving morphisms. In Section 2.3, we equip algebraic trees with a canonical Hausdorff topology. We also characterize compactness and a concept that we call order separability, which is closely related to second countability of the topology. Finally, in Section 2.4, we show that any order-separable algebraic tree is induced by a metric tree (which is not true without order separability) and establish the condition under which this metric tree can be chosen to be a compact $\mathbb{R}$-tree.

2.1. The branch-point map. — In this section, we introduce algebraic trees. Recall from Definition 1.2 the definition of a metric tree and the properties (BPM1)–(BPM4) of the map that sends a triplet of three points in a metric tree to its branch point.

Definition 2.1 (algebraic trees). — An *algebraic tree* (T, c) consists of a set $T \neq \emptyset$ and a branch-point map $c\colon T^3 \to T$ satisfying (BPM1)–(BPM4).

The following useful property reflects the fact that any four points in an algebraic tree can be associated with a shape as illustrated in Figure 1.1.

Lemma 2.2 (a consequence of (BPM4)). — *Let (T, c) be an algebraic tree. Then, for all $x_1, x_2, x_3, x_4 \in T$, the following hold:*

(i) *If* $c(x_1, x_2, x_3) = c(x_1, x_2, x_4)$,
then $c(x_1, x_3, x_4) = c(x_2, x_3, x_4)$.

(ii) *If* $c(x_1, x_2, x_3) = c(x_1, x_2, x_4)$,
then $c(x_1, x_2, x_3) = c(x_1, x_2, c(x_1, x_3, x_4))$.

Proof. — Let $x_1, x_2, x_3, x_4 \in T$ with $c_1 := c(x_1, x_2, x_3) = c(x_1, x_2, x_4)$, and $c_2 := c(x_1, x_3, x_4)$.

(i) Condition (BPM4) implies that

$$c_2 \in \big\{c_1 = c(x_1, x_3, x_2),\, c(x_2, x_3, x_4),\, c_1 = c(x_1, x_2, x_4)\big\}. \tag{15}$$

Thus, $c_1 = c_2$ or $c_2 = c(x_2, x_3, x_4)$. The second case is the claim. In the first case, we apply Condition (BPM4) once more to find that

$$\begin{aligned} c(x_2, x_3, x_4) &\in \big\{c_1 = c(x_1, x_2, x_3),\, c_2 = c(x_1, x_3, x_4),\, c_1 = c(x_1, x_2, x_4)\big\} \\ &= \{c_1, c_2\} = \{c_2\}, \end{aligned} \tag{16}$$

so that the claim also holds in this case.

(ii) Condition (BPM3) implies that

$$c(x_1, x_3, c_2) = c\big(x_1,\, x_3,\, c(x_1, x_3, x_4)\big) = c(x_1, x_3, x_4) = c_2, \tag{17}$$

and similarly also $c(x_2, x_3, c_2) = c(x_2, x_3, x_4) = c_2$. Now part *(i)* with x_4 replaced by c_2 yields $c(x_1, x_2, x_3) = c(x_1, x_2, c_2)$, as claimed. □

We have seen that the four axiomatizing properties of the branch-point map are necessary. In many respects, they are also sufficient to capture the tree structure. For example, in analogy to (3) we can define for each $x, y \in T$ the *interval* $[x, y]$ by

$$[x, y] \coloneqq \big\{w \in T : c(x, y, w) = w\big\}. \tag{18}$$

We also use the notation $(x, y) \coloneqq [x, y] \setminus \{x, y\}$, and similarly $(x, y]$, $[x, y)$. The following properties of intervals are known to hold in $\mathbb{R}$-trees (see, e.g., [13, Chapter 2] or [24, Chapter 3]):

LEMMA 2.3 (properties of intervals). — *Let (T, c) be an algebraic tree. Then the following hold:*

(i) *If $x, v, w, z \in T$ are such that $w \in [x, z]$ and $v \in [x, w]$, then $v \in [x, z]$.*
(ii) *If $x, y, z \in T$, then*

$$[x, y] \cap [y, z] = \big[c(x, y, z),\, y\big]. \tag{19}$$

In particular,

$$\big[x,\, c(x, y, z)\big] \cap \big[c(x, y, z),\, z\big] = \big\{c(x, y, z)\big\}. \tag{20}$$

(iii) *If $x, y, z \in T$, then*

$$[x, y] \cup [y, z] = [x, z] \uplus \big(c(x, y, z),\, y\big]. \tag{21}$$

In particular,

$$[x, y] \cup [y, z] = [x, z] \quad \textit{iff} \quad y \in [x, z]. \tag{22}$$

(iv) *For all $x, y, z \in T$,*

$$[x, y] \cap [y, z] \cap [z, x] = \big\{c(x, y, z)\big\}. \tag{23}$$

Proof. — (i) Let $x, v, w, z \in T$ with $w = c(x, w, z)$ and $v = c(x, v, w)$. Then by Condition (BPM4),

$$c(x, v, z) \in \big\{c(x, v, w),\, w = c(x, w, z),\, c(v, w, z)\big\}. \tag{24}$$

We discuss the three cases separately.

If $c(x, v, z) = c(x, v, w)$, then $c(v, w, z) = c(x, w, z) = w$ by Lemma 2.2(i). It then follows that $c(x, v, z) = c(x, v, c(x, w, z)) = c(x, v, w) = v$ by Lemma 2.2(ii), which proves the claim in this case.

If $c(x, v, z) = w$, then $v = c(v, w, x) = c(v, w, z)$ by Lemma 2.2(i). It then follows that $c(x, v, z) = c(x, z, c(z, w, v)) = c(x, v, v) = v$ by Lemma 2.2(ii), which proves the claim in this case.

If $c(x, v, z) = c(v, w, z)$, then $v = c(x, w, v) = c(x, w, z) = w$ by Lemma 2.2(i). Thus, $v = w \in [x, z]$, and the claim holds also in this case.

(ii) Let $x, y, z \in T$, and $v \in [x, y] \cap [y, z]$. That is, $v = c(x, v, y) = c(y, v, z)$. It follows from Lemma 2.2(i) that $c(x, z, v) = c(x, z, y)$, and then from Lemma 2.2(ii) together with Condition (BPM2) that

$$v = c(x, v, y) = c\big(v, y, c(y, x, z)\big). \tag{25}$$

Equivalently, $v \in [c(x, y, z), y]$. This proves the inclusion $[x, y] \cap [y, z] \subseteq [c(x, y, z), y]$. The other inclusion follows from (i).

Note that (20) follows from (19) with the special choice $y = c(x, y, z)$.

(iii) Note first that it follows immediately from (i) that the union on the right-hand side is disjoint. We claim that

$$[x, z] \subseteq [x, y] \cup [y, z]. \tag{26}$$

Indeed, let $v \in [x, z]$, i.e., $c(x, z, v) = v$. Then, by (BPM4) applied to $\{v, x, y, z\}$,

$$v = c(x, z, v) \in \big\{c(x, y, v),\, c(x, y, z),\, c(y, z, v)\big\}, \tag{27}$$

which implies that $v \in [x, y]$ (if $v = c(x, y, v)$) or $v \in [x, z] \cap [x, y]$ (if $v = c(x, y, z)$) or $v \in [y, z]$ (if $v = c(y, z, v)$). Second, we claim that for all $x, y, z \in T$,

$$[x, z] \cup \big[c(x, y, z),\, y\big] \subseteq [x, y] \cup [y, z]. \tag{28}$$

To see this, recall from (ii) that $[c(x, y, z), y] = [x, y] \cap [z, y] \subseteq [x, y] \cap [z, y]$. As $[x, c(x, y, z)] \subseteq [x, y]$ by (i), we have $[x, y] \subseteq [x, c(x, y, z)] \cup [c(x, y, z), y] \subseteq [x, y] \uplus (c(x, y, z), y]$. The corresponding inclusion for $[y, z]$ is shown in the same way, and we have proven Equation (21).

(iv) This follows immediately from (ii). □

We say that $\{x, y\} \subseteq T$ with $x \neq y$ is an *edge* of (T, c), if and only if there is "nothing in between", i.e., $[x, y] = \{x, y\}$, and denote by

$$\mathrm{edge}(T, c) \coloneqq \big\{\{x, y\} \subseteq T : x \neq y,\ [x, y] = \{x, y\}\big\} \tag{29}$$

the *set of edges*. The following example explains that there is no need to distinguish between finite algebraic trees and graph-theoretical trees and that the definitions of edges are consistent.

Example 2.4 (finite algebraic trees correspond to graph-theoretic trees). — Finite algebraic trees are in one-to-one correspondence with finite (undirected) graph-theoretic trees. Let (T, E) be a graph-theoretic tree with vertex set T and edge set E. Then, (T, E) corresponds to the algebraic tree (T, c_E) with $c_E(u, v, w)$ defined as the unique vertex that is on the (graph-theoretic) path between any two of u, v, w. Conversely, if (T, c) is an algebraic tree with T finite, then (T, c) corresponds to the graph-theoretic tree (T, E_c) with $E_c \coloneqq \mathrm{edge}(T, c)$. Obviously, $c_{E_c} = c$.

For a graph-theoretic tree (T, E), we can allow the vertex set T to be countably infinite and still obtain a corresponding algebraic tree as in the previous example. Note, however, that countable algebraic trees do not necessarily correspond to graph-theoretic trees. Indeed, it is possible that T is countably infinite and $\mathrm{edge}(T, c) = \emptyset$. This can be seen by taking $T = \mathbb{Q}$ in the following example, which shows that every totally ordered space naturally corresponds to an algebraic tree.

EXAMPLE 2.5 (totally ordered spaces as algebraic trees). — For a totally ordered space $(T, \leq)$, define $c_{\leq}(x, y, z) \coloneqq y$ whenever $x \leq y \leq z$, $(x, y, z \in T)$. Then it is trivial to check that $(T, c_{\leq})$ is an algebraic tree, and the interval $[x, y]$ coincides with the order interval $\{z \in T : x \leq z \leq y\}$.

Conversely, given an algebraic tree (T, c) and a distinguished point ρ (often referred to as *root*), we can define a *partial order* $\leq_\rho$ by letting for $x, y \in T$,

$$x \leq_\rho y \quad \text{iff} \quad x \in [\rho, y]. \tag{30}$$

Partial orders provide an equivalent way of defining algebraic trees.

PROPOSITION 2.6 (algebraic trees and semilattices). — (i) *Let (T, c) be an algebraic tree and $\rho \in T$. Then, $(T, \leq_\rho)$ is a partially ordered set and a meet semi-lattice with infimum*

$$x \wedge y = c(\rho, x, y) \qquad \forall x, y \in T. \tag{31}$$

Furthermore, $\leq_\rho$ is a total order on $[\rho, x]$ for all $x \in T$.

(ii) *Let $(T, \leq)$ be a partially ordered set, such that all initial segments $\{y \in T : y \leq x\}$, $x \in T$, are totally ordered (in particular a meet semilattice). Then, for $x, y, z \in T$*

$$c(x, y, z) \coloneqq \max\{x \wedge y,\, y \wedge z,\, z \wedge x\} \tag{32}$$

exists, and (T, c) is an algebraic tree.

Proof. — *(i)* Let $x, y \in T$ with $x \leq_\rho y$ and $y \leq_\rho x$. That is, $x = c(\rho, x, y)$ and $y = c(\rho, y, x)$ which implies that $x = y$ and proves that $\leq_\rho$ is *antisymmetric*. As $x = c(\rho, x, x)$, $x \leq_\rho x$, which proves that $\leq_\rho$ is *reflexive*. Finally, to show *transitivity*, let $x, y, z \in T$ with $x \leq_\rho y$ and $y \leq_\rho z$. That is, $x \in [\rho, y]$ and $y \in [\rho, z]$, which implies that $x \in [\rho, z]$ by Lemma 2.3(i). Equivalently, $x \leq_\rho z$, which proves the transience and, thus, that $\leq_\rho$ is a partial order.

For the *infimum*, note that $v \leq_\rho x$ and $v \leq_\rho y$, if and only if $v \in [\rho, x] \cap [\rho, y]$, or equivalently by Lemma 2.3(ii), $v \in [\rho, c(\rho, x, y)]$. As for all $v \in [\rho, c(\rho, x, y)]$, we have $v \leq c(\rho, x, y)$, the claim (31) follows.

Fix $x \in T$. For *totality* on $[\rho, x]$, let $v, w \in [\rho, x]$, i.e., $v = c(\rho, v, x)$ and $w = c(\rho, w, x)$. Applying Condition (BPM4) to $\{\rho, v, w, x\}$ we find that one of the following three cases must occur: $c(\rho, v, w) = c(\rho, v, x)$ (which implies that $v = c(\rho, v, w)$, or equivalently, $v \leq_\rho w$), $c(\rho, w, v) = c(\rho, w, x)$ (which implies

that $w = c(\rho, w, v)$, or equivalently, $w \leq_\rho v$), or $c(\rho, x, v) = c(\rho, x, w)$ (which implies that $w = v$).

(ii) The maximum in (32) is over a totally ordered set (because initial segments are totally ordered), and thus exists. Furthermore, if $x \wedge y \leq x \wedge z \leq y \wedge z$, say, we also obtain $x \wedge y = x \wedge y \wedge z \geq x \wedge z$. This means that (at least) two of $x \wedge y$, $y \wedge z$, and $z \wedge x$ are identical, and the maximum $c(x, y, z)$ is the third one. That v satisfies (BPM1)–(BPM3) is obvious. To see the four-point condition (BPM4), let $x_1, \ldots, x_4 \in T$ and assume without loss of generality that $x_2 \wedge x_3 = c(x_1, x_2, x_3) =: v$, and, hence, $x_1 \wedge x_2 = x_1 \wedge x_3 \leq v$. We distinguish the cases: if $x_2 \wedge x_4 < v$, then $c(x_2, x_3, x_4) = \max\{x_2 \wedge x_4, v, x_3 \wedge x_4\} = v$, and (8) is satisfied. Otherwise, $x_2 \wedge x_4, x_3 \wedge x_4 \geq v$ and at most one of them can be strictly larger. If $x_2 \wedge x_4 > v \geq x_1 \wedge x_2$, then $x_1 \wedge x_2 = x_1 \wedge x_4 = x_1 \wedge x_3$, and $c(x_1, x_3, x_4) = x_3 \wedge x_4 = v$. The case $x_3 \wedge x_4 > v$ is analogous. In the last case, $x_2 \wedge x_4 = x_3 \wedge x_4 = v$, and $c(x_2, x_3, x_4) = v$. □

Remark 2.7 (Favre and Jonsson's nonmetric trees). — In [29], *rooted nonmetric trees* are introduced as partially ordered sets with a global minimum, totally ordered initial segments, and the additional property that all full, totally ordered subsets are order isomorphic to a real interval. Proposition 2.6 shows that they naturally induce algebraic trees.

Corollary 2.8. — *Let (T, c) be an algebraic tree and $\rho, x, y \in T$. If $v \in [x, y]$, then $v \geq_\rho c(x, y, \rho)$.*

Proof. — Let $\rho, x, y \in T$ and $v \in [x, y]$. That is, $v = c(x, v, y)$. We need to show that $c(\rho, v, c(\rho, x, y)) = c(\rho, x, y)$.

By Condition (BPM4) applied to $\{x, y, \rho, v\}$ we have one of the following three cases: $c(x, y, \rho) = c(x, y, v)$ (in which case $c(\rho, x, y) = v$) or $c(\rho, y, x) = c(\rho, y, v)$ (in which case $c(x, v, \rho) = c(x, v, y) = v$ by Lemma 2.2(i) and, thus, $v \in [\rho, x]$; the claim then follows, since this implies that $v \in [\rho, x] \cap [x, y] = [c(\rho, x, y), y]$ by Lemma 2.3(ii)) or $c(\rho, x, y) = c(\rho, x, v)$ (in which we conclude similarly to the second case that $v \in [c(\rho, x, y), x]$). □

The partial orders $\leq_\rho$ allow us to define a notion of completeness of algebraic trees.

Definition 2.9 (directed order completeness). — Let (T, c) be an algebraic tree. We call (T, r) *(directed) order complete*, if for all $\rho \in T$ the supremum of every totally ordered, nonempty subset exists in the partially ordered set $(T, \leq_\rho)$.

Obviously, in an order complete algebraic tree, infima of totally ordered sets exist, because they are either ρ if the set is empty or a nonempty supremum with respect to a different root. This notion of completeness allows us to define the analogs of complete $\mathbb{R}$-trees.

DEFINITION 2.10 (algebraic continuum tree). — We call an algebraic tree (T, c) *algebraic continuum tree*, if the following two conditions hold:

(ACT1) (T, c) is order complete.
(ACT2) $\mathrm{edge}(T, c) = \emptyset$.

2.2. Morphisms of algebraic trees. — Like any decent algebraic structure (or, in fact, mathematical structure), algebraic trees come with a notion of structure-preserving morphisms.

DEFINITION 2.11 (morphisms). — Let (T, c) and $(\widehat{T}, \hat{c})$ be algebraic trees. A map $f\colon T \to \widehat{T}$ is called a *tree homomorphism* (from T into $\widehat{T}$), if for all $x, y, z \in T$,

$$f\big(c(x, y, z)\big) = \hat{c}\big(f(x), f(y), f(z)\big). \tag{33}$$

We refer to a bijective tree homomorphism as *tree isomorphism*.

As we have seen, the tree structure can be expressed also in terms of intervals or partial orders rather than the branch-point map. This also works for the morphisms.

LEMMA 2.12 (equivalent definitions of morphisms). — *Let (T, c) and $(\widehat{T}, \hat{c})$ be algebraic trees and $f\colon T \to \widehat{T}$. Then the following are equivalent:*

1. *f is a tree homomorphism.*
2. *For all $\rho \in T$, f is an order-preserving map from $(T, \leq_\rho)$ to $(\widehat{T}, \leq_{f(\rho)})$.*
3. *For all $x, y \in T$, $f([x, y]) \subseteq [f(x), f(y)]$.*

Proof. — *"1⇒ 2".* Let $x, y, \rho \in T$ with $x \leq_\rho y$. Then $x = c(\rho, x, y)$, and, thus, $f(x) = \hat{c}(f(\rho), f(x), f(y))$. Therefore, $f(x) \leq_{f(\rho)} f(y)$.

"2⇒ 3". Let $x, y, z \in T$ with $z \in [x, y]$. Then $z \leq_x y$, and, thus, $f(z) \leq_{f(x)} f(y)$, i.e., $f(z) \in [f(x), f(y)]$. *"3⇒ 1".* Let $x, y, z \in T$. Then, $\{c(x, y, z)\} = [x, y] \cap [x, z] \cap [y, z]$. Hence,

$$\begin{aligned}\big\{f(c(x, y, z))\big\} &\subseteq \big[f(x), f(y)\big] \cap \big[f(y), f(z)\big] \cap \big[f(x), f(z)\big] \\ &= \big\{\hat{c}\big(f(x), f(y), f(z)\big)\big\}.\end{aligned} \tag{34}$$

Therefore, $f(c(x, y, z)) = \hat{c}(f(x), f(y), f(z))$. □

Lemma 2.12 shows that our notion of morphisms of algebraic trees is weaker than the morphisms of nonmetric trees used in [29], but the notion of isomorphism is the same. The image of an algebraic tree under a tree homomorphism is a subtree in the following sense.

DEFINITION 2.13 (subtree). — Let (T, c) be an algebraic tree and $\emptyset \neq A \subseteq T$; A is called a *subtree (of (T, c))* if

$$c(A^3) \subseteq A. \tag{35}$$

We refer to $c(A^3)$ as the *algebraic subtree generated by A*.

Obviously, a subtree A of (T,c), implicitly equipped with the restriction of c to A^3, is an algebraic tree in its own right. Furthermore, the following lemma is easy to check.

LEMMA 2.14 (tree homomorphisms). — *Let (T,c) and $(\widehat{T},\hat{c})$ be two algebraic trees and $f\colon T\to\widehat{T}$ a homomorphism. Then the image $f(T)$ is a subtree of $\widehat{T}$. If f is injective, f^{-1} is a tree homomorphism from $f(A)$ into T.*

In particular, if $(\tilde{T},\tilde{c})$ is another algebraic tree, and g is a homomorphism from $(\widehat{T},\hat{c})$ to $(\tilde{T},\tilde{c})$, then $g\circ f$ is a homomorphism from (T,c) to $(\tilde{T},c_{\tilde{T}})$.

2.3. Algebraic trees as topological spaces. — In contrast to metric trees, there is a priori no topology defined on a given algebraic tree. In this section, we therefore equip algebraic trees with a canonical topology.

For each $x\in T$, we introduce a (component) relation $\sim_x$ by letting $y\sim_x z$, if and only if $x\notin[y,z]$, where $y,z\in T$. Let for each $y\in T\setminus\{x\}$

$$\mathcal{S}_x(y)=\mathcal{S}_x^{(T,c)}(y)\coloneqq\big\{z\in T\setminus\{x\}:\ z\sim_x y\big\} \tag{36}$$

be the equivalence class of $T\setminus\{x\}$ containing y and note that $\mathcal{S}_x(y)$ is a subtree for all $x,y\in T$, and $\mathcal{S}_x(y)=\mathcal{S}_x(z)$ whenever $z\in\mathcal{S}_x(y)$. We refer to $\mathcal{S}_x(y)$ as the *component* of $T\setminus x$ containing y. Now and in the following, we equip (T,c) with the topology

$$\tau\coloneqq\tau\big(\big\{\mathcal{S}_x(y):\ x,y\in T,\ x\neq y\big\}\big) \tag{37}$$

generated by the set of components, i.e., with the coarsest topology, such that all components are open sets. We call τ the *component topology* of (T,c).

EXAMPLE 2.15 (on totally ordered trees, τ is the order topology). — If $(T,\leq)$ is a totally ordered space and $(T,c_{\leq})$ the corresponding algebraic tree as in Example 2.5, then τ coincides with the *order topology* (i.e., the one generated by sets of the form $\{y\in T:\ y>x\}$ and $\{y\in T:\ y<x\}$ for $x\in T$).

EXAMPLE 2.16 (intervals are closed sets). — Let (T,c) be an algebraic tree, and $x,y\in T$. Then

$$T\setminus[x,y]=\bigcup\big\{\mathcal{S}_u(v):\ u\in[x,y],\ v\in T,\ \mathcal{S}_u(v)\cap[x,y]=\emptyset\big\}\in\tau. \tag{38}$$

This means that $[x,y]$ is closed in the component topology τ.

LEMMA 2.17. — *Let (T,c) be an algebraic tree. Then c is continuous with respect to the component topology τ.*

Proof. — By definition of τ, it is sufficient to show that for any $x,y\in T$, $x\neq y$, the set $c^{-1}(\mathcal{S}_x(y))$ is open in T^3. By definition of $\mathcal{S}_x(y)$ and the property $c(u,v,w)\in[u,v]\cap[v,w]\cap[w,u]$ shown in Lemma 2.3, $c(u,v,w)\in\mathcal{S}_x(y)$, if and only if (at least) two of u,v,w are in $\mathcal{S}_x(y)$. Because $\mathcal{S}_x(y)$ is open, the

same is true for $\{(u,v,w) \in T^3 : u,v \in \mathcal{S}_x(y)\}$ in the product topology. Hence, $c^{-1}(\mathcal{S}_x(y))$ is a union of an open set and, thus, open. □

Next, we show that τ is a Hausdorff topology and characterize the compactness of algebraic trees in this topology.

LEMMA 2.18 (τ is Hausdorff). — *Let (T,c) be an algebraic tree. Then the component topology τ defined in* (37) *is a Hausdorff topology on T.*

Proof. — To show that (T,τ) is Hausdorff, let $x,y \in T$ be distinct. If $\mathcal{S}_y(x) \cap \mathcal{S}_x(y) = \emptyset$, then $\mathcal{S}_y(x)$ and $\mathcal{S}_x(y)$ are clearly disjoint neighborhoods of x and y, respectively. Assume that this is not the case and choose $z \in \mathcal{S}_x(y) \cap \mathcal{S}_y(x)$. Then $\rho \coloneqq c(x,y,z) \notin \{x,y\}$. Furthermore, $c(x,\rho,y) = c(x,y,z) = \rho$, and hence $x \nsim_\rho y$. Thus, $\mathcal{S}_\rho(x)$ and $\mathcal{S}_\rho(y)$ are disjoint neighborhoods of x and y, respectively. Hence, τ is Hausdorff. □

PROPOSITION 2.19 (characterizing compactness). — *Let (T,c) be an algebraic tree with component topology τ. Then (T,τ) is compact, if and only if (T,c) is directed order complete.*

Proof. — *"only if".* Assume first that (T,c) is not order complete. Then we can choose $\rho \in T$ and $\emptyset \neq A \subseteq T$, such that A is totally ordered with respect to $\leq_\rho$ but does not have a supremum in $(T,\leq_\rho)$. For $x,y \in T$, let $U_x \coloneqq \{z \in T : z \not\geq_\rho x\}$ and $V_y \coloneqq \{z \in T : z >_\rho y\}$. Then U_x and V_y are open sets. We claim that $\mathcal{U} \coloneqq \{U_x : x \in A\} \cup \{V_y : y \geq A\}$ is an open cover of T. Indeed, if $z \geq_\rho A$, then, because A has no supremum, there is $y \in T$ with $A \leq_\rho y \leq_\rho z$, and, hence $z \in V_y \in \mathcal{U}$. Otherwise, if $z \not\geq_\rho A$, there is $x \in A$ with $z \in U_x \in \mathcal{U}$. Thus, $\mathcal{U}$ is a cover of T.

$\mathcal{U}$ has no finite subcover, because if $\mathcal{U}' = \{U_{x_1},\dots,U_{x_n},V_{y_1},\dots,V_{y_m}\}$ were such a finite subcover, then $\{U_{x_1},\dots,U_{x_n}\}$ would cover A. This, however, would imply that $\max\{x_1,\dots,x_n\}$ would be a supremum of A, contradicting our assumption. Hence, (T,τ) is not compact.

"if". Assume that (T,c) is order complete. Consider a cover $\mathcal{U}$ of T with components, i.e., $\mathcal{U} \subseteq \{\mathcal{S}_y(x) : x,y \in T, x \neq y\}$. By the Alexander subbase theorem, for compactness of τ, it is sufficient to show that $\mathcal{U}$ has a finite subcover.

To this end, fix an element $\rho \in T$ and consider the set $\mathcal{U}_\rho \coloneqq \{U \in \mathcal{U} : \rho \in U\} \neq \emptyset$. By Hausdorff's maximal chain theorem (or Zorn's lemma), there is a maximal chain I in the partially ordered set $(\mathcal{U}_\rho,\subseteq)$. For every $U \in I$, we have $\rho \in U$, and, thus, there is $x_U \in T$, such that $U = \mathcal{S}_{x_U}(\rho)$. We claim that $U \subseteq V$ implies $x_U \leq_\rho x_V$. Indeed, $x_V \notin V$, and, hence, $x_V \notin U$, which is equivalent to $x_V \geq_\rho x_U$. Therefore, $z \coloneqq \sup\{x_U : U \in I\}$ exists in $(T,\leq_\rho)$ by directed order completeness of T. Because $\mathcal{U}$ is a cover, there is $V \in \mathcal{U}$ with $z \in V$, and, hence, $V = \mathcal{S}_y(z)$ for some $y \in T$. Because $V \notin I$ and I is a maximal chain, $y \not\geq_\rho z$. Hence, there is $U \in I$ with $y \not\geq_\rho x_U =: x$. We claim

that $T = \mathcal{S}_y(z) \cup \mathcal{S}_x(\rho)$. Indeed, let $w \in T \setminus \mathcal{S}_x(\rho)$. Then $w \geq_\rho x$. Using $z \geq_\rho x$ and $c(w,z,y) \in [w,z]$, we obtain $c(w,z,y) \geq_\rho x$, and, hence, $c(w,z,y) \neq y$. Thus, $w \in \mathcal{S}_y(z)$ as claimed, and $\{\mathcal{S}_y(z), \mathcal{S}_x(\rho)\}$ is the desired subcover. □

It turns out that the following separability condition, which we call order separability, is crucial for us.

PROPOSITION 2.20 (order separability). — *Let (T,c) be an algebraic tree with component topology τ. Then the following are equivalent:*

1. *There exists a countable set D, such that for all $x,y \in T$ with $x \neq y$,*

$$D \cap [x,y) \neq \emptyset. \tag{39}$$

2. *The topological space (T,τ) is second countable (i.e., τ has a countable base), and* $\mathrm{edge}(T,c)$ *is countable.*
3. *The topological space (T,τ) is separable, and* $\mathrm{edge}(T,c)$ *is countable.*

Proof. — *"3⇒ 1".* Assume that $\mathrm{edge}(T,c)$ is countable and that (T,τ) is separable. Then there exists a countable, dense subset $\tilde{D} \subseteq T$. We claim that

$$D \coloneqq c\big(\tilde{D}^3\big) \cup \big\{z \in T : \exists\, x \in T \text{ such that } \{x,z\} \in \mathrm{edge}(T,c)\big\} \tag{40}$$

satisfies (39). Indeed, D is countable by assumption. Moreover, let $x,y \in T$. Then two cases are possible: either $\mathcal{S}_x(y) \cap \mathcal{S}_y(x) = \emptyset$. In this case, $\{x,y\} \in \mathrm{edge}(T,c)$, which implies that $[x,y) \cap D \neq \emptyset$. Or $\mathcal{S}_x(y) \cap \mathcal{S}_y(x) \neq \emptyset$. In this case, as $\mathcal{S}_x(y) \cap \mathcal{S}_y(x)$ is open by definition of τ, there is $z \in \tilde{D} \cap \mathcal{S}_x(y) \cap \mathcal{S}_y(x)$. Let $v \coloneqq c(x,y,z)$. Then $v \in (x,y)$, and either $v = z \in D$, or the three components $\mathcal{S}_v(x)$, $\mathcal{S}_v(y)$, $\mathcal{S}_v(z)$ are distinct. In the second case, we can choose $x' \in \tilde{D} \cap \mathcal{S}_v(x)$ and $y' \in \tilde{D} \cap \mathcal{S}_v(y)$ to see that $v = c(x',y',z) \in D$. In any case, $v \in [x,y) \cap D$. *"1⇒ 2".* Let D be a countable set satisfying (39). Then, for all $\{x,y\} \in \mathrm{edge}(T,c)$, $D \cap [x,y) = \{x\}$. This implies that $\mathrm{edge}(T,c)$ is countable. We consider the countable set $\mathcal{U} = \big\{\mathcal{S}_v(u) : u,v \in D\big\} \subseteq \tau$ and claim that it is a subbase for τ (i.e., generates τ). To this end, let $x,y \in T$. We show that $U \coloneqq \mathcal{S}_x(y)$ is a union of sets from $\mathcal{U}$, i.e., for every $z \in U$, we construct $V \in \mathcal{U}$ with $z \in V \subseteq U$. By assumption on D, there is $v \in D \cap [x,z)$ and $u \in D \cap (v,z]$. Let $V \coloneqq \mathcal{S}_v(u) \in \mathcal{U}$. Because $c(u,v,z) = u \neq v$, we have $z \in V$. Let $w \in T \setminus U$. Because $u \in U$, we have $U = \mathcal{S}_x(u)$ and, therefore, $x \in [u,w]$. Similarly, $x \in [v,w]$. In particular, by Lemma 2.2, $c(u,v,w) = c(u,v,c(u,x,w)) = c(u,v,x) = v$, and, thus, $w \notin V$. Because $w \in T \setminus U$ is arbitrary, we obtain $V \subseteq U$. *"2⇒ 3".* Trivial, because every second countable topological space is separable. □

DEFINITION 2.21 (order separability). — We call an algebraic tree (T,c) *order separable* if the equivalent conditions of Proposition 2.20 are satisfied. We call a set $D \subseteq T$ *order dense* if it satisfies (39).

EXAMPLE 2.22 (uncountable star tree). — This example shows that in (39) we cannot replace $[x,y)$ by $[x,y]$. Let $T \coloneqq \{0\} \cup [1,2]$ with $c(x,y,z) \coloneqq 0$ whenever $x,y,z \in T$ are distinct. Then, if $D \subseteq T$ is such that $D \cap [x,0) \neq \emptyset$, for all $x \in [1,2]$, then $[1,2] \subseteq D$, and, thus, D is uncountable, and (T,c) not order separable. On the other hand, $D \coloneqq \{0\}$ has the property that $D \cap [x,y] \neq \emptyset$, for all $x,y \in T$ with $x \neq y$.

An order complete, order separable algebraic tree is, in its component topology τ, a compact, second countable Hausdorff space by Propositions 2.19 and 2.20. In particular, it is metrizable. In fact, order separability already implies metrizability, as we will see in Section 2.4. The following example shows that (topological) separability of (T,τ) alone, without requiring the number of edges to be countable, is not sufficient for either order separability or metrizability of (T,τ).

EXAMPLE 2.23 (a continuum ladder). — Let $T = [0,1] \times \{0,1\}$ with the lexicographic order $\leq$ on T and define the canonical branch point map $c_{\leq}$ as in Example 2.5. Then $\operatorname{edge}(T, c_{\leq}) = \big\{\{(x,0),\,(x,1)\} : x \in [0,1]\big\}$ is uncountable, and, hence, $(T, c_{\leq})$ is not order separable. Because $(\mathbb{Q} \cap [0,1]) \times \{0,1\}$ is a countable dense set, (T,τ) is (topologically) separable. The topological subspace $[0,1] \times \{1\}$ is the *Sorgenfrey line*, which is known to be nonmetrizable (see [55, Counterexample 51]). Thus, (T,τ) cannot be metrizable either.

DEFINITION 2.24 (Borel σ-algebra $\mathcal{B}(T,c)$). — Let (T,c) be an algebraic tree. We denote the Borel σ-algebra of the component topology τ by $\mathcal{B}(T,c)$ and call it the *Borel σ-algebra of* (T,c).

In general, $\mathcal{B}(T,c)$ is not generated by the set of components. Order separability, however, is sufficient to ensure this property because it implies second countability of the component topology.

COROLLARY 2.25 (Borel σ-algebra generated by components). — *Let (T,c) be an order separable algebraic tree, and $D \subseteq T$ an order dense set. Then its Borel σ-algebra is generated by the set of components indexed by D, i.e.,*

$$\mathcal{B}(T,c) = \sigma\big(\big\{\mathcal{S}_x(y) : x,y \in D,\ x \neq y\big\}\big). \tag{41}$$

Proof. — Define $\mathcal{U} \coloneqq \big\{\mathcal{S}_x(y) : x,y \in D,\ x \neq y\big\}$. By Proposition 2.20, (T,τ) is second countable. Hence, $\mathcal{B}(T,c)$ is generated by any subbase of τ. If D is order dense, $\mathcal{U}$ is such a subbase as shown in the proof of Proposition 2.20. □

2.4. Metric tree representations of algebraic trees. — In this section, we discuss the connection of metric trees with algebraic trees. Let (T,r) be a metric tree (recall Definition 1.2). Then by (MT2), there exists to any three points $x_1,x_2,x_3 \in T$ a unique branch point $c_{(T,r)}(x_1,x_2,x_3)$ satisfying (5). We refer

to $(T, c_{(T,r)})$ as the algebraic tree *induced by* (T, r) and to (T, r) as a *metric representation* of $(T, c_{(T,r)})$.

LEMMA 2.26 (the algebraic tree induced by a metric tree). — *Let (T, r) be a metric tree and $c_{(T,r)}$ the map that sends any three distinct points to their branch point. Then the following hold:*

(i) *$(T, c_{(T,r)})$ is an algebraic tree.*
(ii) *$(T, c_{(T,r)})$ is order separable, if and only if (T, r) is separable.*
(iii) *$(T, c_{(T,r)})$ is directed order complete, if and only if (T, r) is bounded and complete. In particular, it is an algebraic continuum tree, if and only if (T, r) is a bounded, complete $\mathbb{R}$-tree.*

Proof. — *(i)* It can be easily checked that $(T, c_{(T,r)})$ is an algebraic tree.

(ii) Let (T, r) be separable. Then $\operatorname{edge}(T, c_{(T,r)})$ is countable. The topology induced by r is obviously stronger than the topology τ introduced in (37), hence τ is separable and, therefore, the algebraic tree $(T, c_{(T,r)})$ is order separable. Conversely, if $(T, c_{(T,r)})$ is order separable, then any countable set D satisfying (39) is also dense in (T, r).

(iii) Clearly, $(T, c_{(T,r)})$ admits suprema along any linearly ordered set with respect to some root, if and only if (T, r) is bounded and complete. The "in particular" follows because a complete metric tree (T, r) is an $\mathbb{R}$-tree, if and only if $\operatorname{edge}(T, r) := \operatorname{edge}(T, c_{(T,r)}) = \emptyset$ ([9, Remark 1.2]). □

Our first main result states that under the assumption of order separability any algebraic tree can be embedded by an injective homomorphism into a compact $\mathbb{R}$-tree and, hence, is isomorphic to (the algebraic tree induced by) a totally bounded metric tree.

THEOREM 2.27 (characterization of order separable algebraic trees). — *Let T be a set, $c\colon T^3 \to T$.*

(i) *(T, c) is an order separable algebraic continuum tree, if and only if there exists a metric r on T, such that (T, r) is a compact $\mathbb{R}$-tree with*

$$c = c_{(T,r)}. \tag{42}$$

(ii) *(T, c) is an order separable algebraic tree, if and only if there is an order separable algebraic continuum tree $(\overline{T}, \bar{c})$, such that (T, c) is a subtree of $(\overline{T}, \bar{c})$. In particular, every order separable algebraic tree is induced by a totally bounded metric tree.*

The separability hypothesis in Theorem 2.27 is crucial and cannot be dropped. In Example 2.23, we already saw an algebraic tree where the component topology τ is not metrizable. Moreover, in this example, τ coincides with the order topology, which is also the case for the metric topology of any metric tree without branch points. Thus, the algebraic tree cannot be induced by a metric tree.

The following example shows that also algebraic continuum trees need not be induced by metric trees.

EXAMPLE 2.28 (algebraic continuum tree that is not induced by a metric tree). — Let $T = [0,1] \times [0,1]$ with lexicographic order and (T, c) the corresponding algebraic tree as in Example 2.5. It is easy to check that (T, c) is an algebraic continuum tree. It cannot be induced by a metric tree because in its order topology τ, it is connected but not path-wise connected. These two properties are equivalent for metric trees (see [24, Theorem 2.20]).

In order to prove Theorem 2.27, given an algebraic tree (T, c), we need to provide a metric r, such that (42) holds. For that purpose, we consider for any measure ν on $(T, \mathcal{B}(T, c))$, such that ν is finite on every interval, the following pseudometric,

$$r_\nu(x,y) \coloneqq \nu([x,y]) - \tfrac{1}{2}\nu(\{x\}) - \tfrac{1}{2}\nu(\{y\}), \quad x, y \in T. \tag{43}$$

LEMMA 2.29 (r_ν is a pseudometric). — *Let (T, c) be an algebraic tree and ν a measure on (T, c) with $\nu([x,y]) < \infty$, for all $x, y \in T$. Then r_ν is a pseudometric on T.*

Proof. — By Lemma 2.3 for all $x, y, z \in T$,

$$\begin{aligned} \nu\big([x,y]\big) + \nu\big([y,z]\big) &= \nu\big([x,y] \cup [y,z]\big) + \nu\big([x,y] \cap [y,z]\big) \\ &= \nu\big([x,z]\big) + \nu\big(\big(c(x,y,z),\, y\big]\big) + \nu\big(\big[c(x,y,z),\, y\big]\big). \end{aligned} \tag{44}$$

Hence,

$$\begin{aligned} &r_\nu(x,y) + \tfrac{1}{2}\nu\{x\} + \tfrac{1}{2}\nu\{y\} + r_\nu(y,z) + \tfrac{1}{2}\nu\{y\} + \tfrac{1}{2}\nu\{z\} \\ &\quad = r_\nu(x,z) + \tfrac{1}{2}\nu\{x\} + \tfrac{1}{2}\nu\{z\} + 2r_\nu\big(c(x,y,z),\, y\big) \\ &\qquad + \nu\{c(x,y,z)\} + \nu\{y\} - \nu\{c(x,y,z)\}, \end{aligned} \tag{45}$$

or equivalently,

$$r_\nu(x,y) + r_\nu(y,z) = r_\nu(x,z) + 2r_\nu\big(c(x,y,z),\, y\big). \tag{46}$$

This implies that r_ν satisfies the triangle inequality. □

We denote the quotient metric space by (T_ν, r_ν), i.e., T_ν is the set of equivalence classes of points in T with r_ν-distance zero, and the quotient metric on T_ν is again denoted by r_ν. Furthermore, let $\pi_\nu \colon T \to T_\nu$ be the canonical projection.

LEMMA 2.30 ((T_ν, r_ν) is a metric tree). — *Let (T, c) be an algebraic tree and ν a measure on (T, c) with $\nu([x,y]) < \infty$ for all $x, y \in T$. Then the quotient space (T_ν, r_ν) is a metric tree, and the canonical projection π_ν is a tree homomorphism.*

Proof. — Let $x_1, \dots, x_4 \in T$. By Condition (BPM4), we can assume without loss of generality that $c(x_1, x_2, x_3) = c(x_1, x_2, x_4)$. Then by Lemma 2.2(ii), $c(x_1, x_2, x_3) \in [x_1, x_2] \cap [x_1, x_3] \cap [x_2, x_3] \cap [x_1, x_4] \cap [x_2, x_4]$, and (46) yields that for $\{i, j\} \in \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}\}$,

$$r_\nu(x_i, x_j) = r_\nu\big(x_i, c(x_1, x_2, x_3)\big) + r_\nu\big(c(x_1, x_2, x_3), x_j\big). \tag{47}$$

Therefore,

$$\begin{aligned} r_\nu(x_1, x_3) &+ r_\nu(x_2, x_4) \\ &= r_\nu\big(x_1,\, c(x_1, x_2, x_3)\big) + r_\nu\big(c(x_1, x_2, x_3),\, x_3\big) \\ &\quad + r_\nu\big(x_2,\, c(x_1, x_2, x_3)\big) + r_\nu\big(c(x_1, x_2, x_3),\, x_4\big) \\ &= r_\nu(x_1, x_2) + r_\nu\big(c(x_1, x_2, x_3),\, x_3\big) + r_\nu\big(c(x_1, x_2, x_3),\, x_4\big) \\ &\geq r_\nu(x_1, x_2) + r_\nu(x_3, x_4), \end{aligned} \tag{48}$$

and analogously,

$$\begin{aligned} r_\nu\big(x_1, x_4\big) &+ r_\nu\big(x_2, x_3\big) \\ &= r_\nu\big(x_3,\, c(x_1, x_2, x_3)\big) + r_\nu\big(c(x_1, x_2, x_3),\, x_4\big) + r_\nu(x_1, x_2) \\ &\geq r_\nu(x_1, x_2) + r_\nu(x_3, x_4). \end{aligned} \tag{49}$$

This means that the four-point Condition (MT1) is satisfied. Moreover, (47) implies Condition (MT2) with branch point $\pi_\nu(c(x_1, x_2, x_3))$. In particular, π_ν is a tree homomorphism. □

Remark 2.31. — Lemma 2.30 also explains why we defined r_ν as in (43) and not just as $r'_\nu := \nu([x, y])$ for $x \neq y$. Namely, in the latter case, we would still have (MT1), but (MT2) might fail. Take, for example, $T := \{1, 2, 3\}$, $c(1, 2, 3) = 2$, and $\nu = \delta_2$. In this case, r'_ν is the discrete metric on T, and, thus, 2 does no longer lies on the interval $[1, 3]$.

Let (T, c) be an algebraic tree. For all $v \in T$, define the *degree* of v in (T, c) by

$$\deg(v) := \deg_{(T,c)}(v) := \#\big\{\mathcal{S}_v(y) : y \in T\big\}. \tag{50}$$

We say that $v \in T$ is a *leaf* if $\deg_{(T,c)}(v) = 1$, and a *branch point* if $\deg_{(T,c)}(v) \geq 3$. Note that

$$\mathrm{lf}(T, c) := \big\{u \in T : c(u, v, w) \neq u \ \forall v, w \in T \setminus \{u\}\big\} \tag{51}$$

equals the set of leaves of T, and

$$\mathrm{br}(T, c) := \big\{u \in T : c(x, v, w) = u \quad \text{for some } x, v, w \in T \setminus \{u\}\big\} \tag{52}$$

the set of branch points. Moreover, note that any ν-mass on $\mathrm{lf}(T, c)$ that is not atomic does not contribute to r_ν.

PROPOSITION 2.32 (metric representations of algebraic trees). — *Let (T,c) be an algebraic tree, ν a measure on $(T,\mathcal{B}(T,c))$ with $\nu([x,y])<\infty$, for all $x,y\in T$, and r_ν defined by* (11)*. Then the following hold:*

(i) *If (T,c) is order separable and ν has at most countably many atoms, then (T_ν,r_ν) is separable.*

(ii) *If $\#T>1$, (T,c) is order complete, and $[x,y]$ is order separable for every $x,y\in T$, then (T_ν,r_ν) is connected, if and only if ν is nonatomic. In this case, (T_ν,r_ν) is a complete $\mathbb{R}$-tree.*

Proof. — Throughout the proof denote by $\pi_\nu\colon T\to T_\nu$ the canonical projection.

(i) It is easy to see that if a set $A\subseteq T$ satisfies (39) and contains all atoms of ν, then $\pi_\nu(A)$ is dense in (T_ν,r_ν). Therefore, by Proposition 2.20 order separability of (T,c) implies separability of (T_ν,r_ν).

(ii) For all $x,y\in T$ with $x\neq y$, $r_\nu(x,y)\geq\frac{1}{2}\nu\{x\}$. Hence, (T_ν,r_ν) cannot be connected if ν has atoms. Conversely, assume that ν is nonatomic. For $x,z\in T$, consider $([x,z],\leq_x)$, which is a totally ordered space according to Proposition 2.6 and define $y:=\sup\{v\in[x,z]:2\nu([x,v])\leq\nu([x,z])\}$. The supremum exists due to order completeness of (T,c). Because of the order separability of $[x,z]$ and the nonatomicity of ν, we obtain $2\nu([x,y])=\nu([x,z])=2\nu([y,z])$, and, therefore, $2r_\nu(x,y)=r_\nu(x,z)=2r_\nu(y,z)$. From this equality, connectedness follows once we have shown completeness, and every connected metric tree is an $\mathbb{R}$-tree.

Recall from Lemma 2.30 that (T_ν,r_ν) is a metric tree. The same holds for its metric completion $\overline{T}_\nu$. Assume for a contradiction that there is a sequence $(x_n)_{n\in\mathbb{N}}$ in T_ν converging to some $x\in\overline{T}_\nu\setminus T_\nu$. Then x cannot be a branch point, and one of the at most two components of $\overline{T}_\nu\setminus\{x\}$ contains infinitely many x_n. Thus, we may assume without loss of generality that $x\in\mathrm{lf}(\overline{T}_\nu)$. Define $y_n:=c_{\overline{T}_\nu}(x_1,x_n,x)$. Then $y_n\to x$ and, for large enough m, we have $y_n=c_{\overline{T}_\nu}(x_1,x_n,x_m)$. Hence, $y_n\in T_\nu$, for all $n\in\mathbb{N}$, and we may choose representatives $x'_n\in\pi_\nu^{-1}(y_n)$, such that $x'_n=c(\rho,x'_n,x'_m)$ for $\rho:=x'_1$ and all sufficiently large m. By Proposition 2.6, $\{x'_n:n\in\mathbb{N}\}$ is totally ordered with respect to $\leq_\rho$, and, hence, $x':=\sup\{x'_n:n\in\mathbb{N}\}\in T$ exists by order completeness. Obviously, $\pi_\nu(x')=x$ and $x\in T_\nu$. □

In order to prove Theorem 2.27 using Proposition 2.32, we need a nonatomic probability measure ν (to ensure connectedness of (T_ν,r_ν)) charging all intervals (so that π_ν is injective). Such a measure always exists in the case of *order separable* algebraic continuum trees.

LEMMA 2.33. — *Let (T,c) be an order separable algebraic continuum tree with $\#T>1$. Then there exists a nonatomic probability measure ν on $(T,\mathcal{B}(T,c))$ with $\nu(\mathrm{lf}(T,c))=0$ and*

$$\nu\big([x,y]\big)>0\quad\forall\,x,y\in T,\ x\neq y. \tag{53}$$

Proof. — Fix $\rho \in T$. Then, for every $x \in T \setminus \{\rho\}$, the interval $([\rho, x], \leq_\rho)$ is a separable *linear continuum* in the sense of order theory, i.e., a totally ordered space (proven in Proposition 2.6) without *jumps* (which here we call edges) or *gaps* (which follows from directed order completeness). Due to Cantor's order characterization of $\mathbb{R}$ (e.g., [19, Theorem 560]), this means that $[\rho, x]$ is order isomorphic to the unit interval. Obviously, every order isomorphism is measurable and bijective, and the image of Lebesgue measure on the unit interval is a nonatomic probability measure ν_x on $[\rho, x]$. Then $\sum_{n\in\mathbb{N}} 2^{-n}\nu_{x_n}$, where $\{x_n : n \in \mathbb{N}\}$ satisfies (39), is a nonatomic probability satisfying (53) and $\nu(\mathrm{lf}(T, c)) = 0$. □

Any separable $\mathbb{R}$-tree (T, r) comes with an intrinsic measure, called length measure, that generalizes the Lebesgue measure on $\mathbb{R}$. More generally, if (T, r) is a complete, separable metric tree and $\rho \in T$ a fixed root, the *length measure* $\lambda = \lambda^{(T,r,\rho)}$ is uniquely defined by the two properties $\lambda([\rho, x]) = r(\rho, x)$, for all $x \in T$, and $\lambda(\mathrm{lf}_0(T, r)) = 0$, where lf_0 is the set of nonisolated leaves (see [9, Section 2.1]). Note that the total mass $\lambda(T)$ (the "total length" of the metric tree) does not depend on the choice of ρ.

Proposition 2.34 (total length of (T_ν, r_ν)). — *Let (T, c) be an order separable, order complete algebraic tree, ν a measure on $(T, \mathcal{B}(T, c))$ with $\nu([x, y]) < \infty$, for all $x, y \in T$, and such that $\nu\restriction_{\mathrm{lf}(T,c)}$ is purely atomic, and r_ν be defined by* (11)*. Then the following hold:*

(i) *The total length of the metric tree (T_ν, r_ν) is given by*

$$\lambda(T_\nu) = \tfrac{1}{2}\int_T \deg_{(T,c)}\, \mathrm{d}\nu. \tag{54}$$

(ii) $\int_T \deg_{(T,c)}\, \mathrm{d}\nu = \int_{T_\nu} \deg_{(T_\nu,r_\nu)} \circ \pi_\nu \,\mathrm{d}\nu$.

Proof. — *(i)* Let $D \coloneqq \{v_n : n \in \mathbb{N}\}$ be a subset of (T, c) which contains the atoms of ν and satisfies (39), and $\pi_\nu \colon T \to T_\nu$ be the canonical projection. We use $\rho \coloneqq \pi_\nu(v_1)$ as the root of (T_ν, r_ν). Then

$$T \setminus \mathrm{lf}(T, c) \subseteq [\![D]\!] = \bigcup_{n\in\mathbb{N}} [\![v_1, \dots, v_n]\!], \tag{55}$$

where $[\![A]\!] \coloneqq \bigcup_{x,y\in A}[x, y]$. Hence, $\nu(T \setminus [\![D]\!]) = 0$, and

$$\lambda^{(T_\nu,r_\nu,\rho)}(T_\nu) = \lim_{n\to\infty} \lambda^{(T_\nu,r_\nu,\rho)}\big(\pi_\nu([\![v_1, \dots, v_n]\!])\big). \tag{56}$$

Abbreviate $T_n \coloneqq [\![v_1, \dots, v_n]\!]$ and $\ell_n \coloneqq \lambda^{(T_\nu,r_\nu,\rho)}\big(\pi_\nu([\![v_1, \dots, v_n]\!])\big)$. If $v_{n+1} \in T_n$, then $T_{n+1} = T_n$ and $\lambda^{(T_\nu,r_\nu,\rho)}\big(\pi_\nu(T_{n+1})\big) = \lambda^{(T_\nu,r_\nu,\rho)}\big(\pi_\nu(T_n)\big)$. Otherwise, there exists a unique $u_n \in T$ with $T_{n+1} = T_n \uplus (u_n, v_{n+1}]$, and, thus,

$$\ell_{n+1} = \ell_n + r_\nu(u_n, v_{n+1}) = \ell_n + \nu\big((u_n, v_{n+1}]\big) - \tfrac{1}{2}\nu\{v_{n+1}\} + \tfrac{1}{2}\nu\{u_n\}. \tag{57}$$

For $v \in T_n$, let $\deg_n(v)$ be the degree of v in the tree $(T_n, c\restriction_{T_n})$. In the case $v_{n+1} \notin T_n$, we have $\deg_{n+1}(v) = \deg_n(v)$ for $v \in T_n \setminus \{u_n\}$, and $\deg_{n+1}(u_n) = \deg_n(u_n) + 1$. By induction over n, we obtain

$$\ell_n = \tfrac{1}{2} \int_{T_n} \deg_n \, \mathrm{d}\nu. \tag{58}$$

Note that $\deg_n(v)$ is monotonically increasing in n, and $\deg(v) = \lim_{n\to\infty} \deg_n(v)$ holds for all $v \in [\![D]\!]$. Thus, using the monotone convergence theorem, combining (56) and (58) yields (54).

(ii) If $\deg_{(T,c)}(v) \neq \deg_{(T_\nu,r_\nu)}(\pi_\nu(v))$, then either $\pi(\mathcal{S}_v(y)) = \{\pi(v)\}$ for some $y \in T$ (and, thus, $\deg_{(T,c)}(v) > \deg_{(T_\nu,r_\nu)}(\pi_\nu(v))$), or $\pi(v) = \pi(v')$ for some $v' \in \mathrm{Br}(T,c)$ (and, thus, $\deg_{(T,c)}(v) < \deg_{(T_\nu,r_\nu)}(\pi_\nu(v))$). In both cases, we have $\nu\{v\} = \nu\{\pi_\nu(v)\} = 0$, and, thus, the claim follows. □

Corollary 2.35 (compactness for bounded degree trees). — *Let (T,c) be an order separable algebraic tree, and ν a finite measure on $(T, \mathcal{B}(T,c))$ with $\nu\{v \in T : \deg(v) > d\} = 0$ for some $d \in \mathbb{N}$. Then the completion of (T_ν, r_ν) is compact.*

Proof. — Without loss of generality assume that $\nu\restriction_{\mathrm{lf}(T,c)}$ is nonatomic (if $\nu\restriction_{\mathrm{lf}(T,c)}$ has a nonatomic part, we can remove it without changing r_ν). Then by Proposition 2.34(i), (T_ν, r_ν) has finite total length. As complete metric trees with finite total length are necessarily compact, the statement follows. □

We are now in a position to prove Theorem 2.27.

Proof of Theorem 2.27.(i) "⟸" Since every compact metric space is bounded, complete, and separable, this step follows from Lemma 2.26.

"⟹" Let (T,c) be an order separable algebraic continuum tree. To avoid trivialities assume that T contains more than two points. By Lemma 2.33 we can choose a nonatomic probability measure ν on $(T, \mathcal{B}(T,c))$ satisfying (53). Define r_ν by (11). Then the equivalence classes in T_ν are singletons by (53), and we may identify T_ν with T.

By Proposition 2.32, (T, r_ν) is a complete $\mathbb{R}$-tree , and the identity is a tree homomorphism by Lemma 2.30. Thus, c is induced by r_ν. Moreover, $\nu(\mathrm{br}(T,c)) = 0$, because $\mathrm{br}(T,c)$ is countable, and ν is nonatomic. We can, therefore, conclude with Corollary 2.35 that (T, r_ν) is also compact.

(ii) "⟸" This is obvious because every order separable algebraic continuum tree is induced by a separable $\mathbb{R}$-tree according to part (i), and subspaces of separable metric spaces are separable.

"⟹" Let (T,c) be an order separable algebraic tree and $D \subseteq T$ a countable set satisfying (39). Let ν be any probability measure on D with $\nu\{x\} > 0$

for all $x \in D$, and r_ν defined by (11). The equivalence classes in T_ν are singletons, and we may again identify T_ν with T. By Proposition 2.34, (T, r_ν) is a metric tree with (42). As (T, c) is order separable, (T, r_ν) is separable by Proposition 2.32(i). Moreover, the diameter of (T, r_ν) is bounded by 1. Hence, by [24, Theorem 3.38] (known since [23]), there is a bounded, separable $\mathbb{R}$-tree $(\overline{T}, \bar{r})$ such that $T \subseteq \overline{T}$ and r_ν is the restriction of $\bar{r}$ to T. By Lemma 2.26, this $\mathbb{R}$-tree induces an algebraic continuum tree $(\overline{T}, \bar{c})$, and T is a subtree of $\overline{T}$.

"in particular". According to part *(i)*, there is a metric $\tilde{r}$ on $\overline{T}$, such that $(\overline{T}, \tilde{r})$ is a compact $\mathbb{R}$-tree inducing $(\overline{T}, \bar{c})$. Let r be the restriction of $\tilde{r}$ to T. Then, (T, r) is a totally bounded metric tree inducing (T, c). □

2.5. Tree homomorphisms versus homeomorphisms. — Since order separable algebraic continuum trees are $\mathbb{R}$-trees where we have "forgotten" the metric, the question arises as to how homeomorphisms of $\mathbb{R}$-trees relate to tree homomorphisms of the corresponding algebraic trees. A first observation is that homeomorphisms are necessarily tree homomorphisms. This statement relies on connectedness of the $\mathbb{R}$-trees , and we cannot replace "$\mathbb{R}$-tree" by "metric tree"; every bijection between finite metric trees is obviously a homeomorphism because the topologies are discrete but not necessarily a tree homomorphism.

LEMMA 2.36 (homeomorphisms are tree isomorphisms). — *Let (T, r), $(\widehat{T}, \hat{r})$ be $\mathbb{R}$-trees, and $f\colon T \to \widehat{T}$ a homeomorphism. Then f is a tree homomorphism.*

Proof. — The branch-point map can be expressed in terms of intervals by (2). In an $\mathbb{R}$-tree (T, r), the interval $[x, y]$, $x, y \in T$, is the unique simple path from x to y, which is a purely topological notion and, hence, preserved by homeomorphisms. □

EXAMPLE 2.37 (tree isomorphisms need not be homeomorphisms). — In Lemma 2.36, the converse is not true; bijective tree homomorphisms need not be homeomorphisms, even if the trees are order separable. To see this, let $r, \hat{r}$ the metrics on $\mathbb{N}$ defined by $r(n, m) = \frac{1}{n} + \frac{1}{m}$, $\hat{r}(n, m) = 2$ for distinct $n, m \in \mathbb{N}$. Let T and $\widehat{T}$ be the $\mathbb{R}$-trees generated by $(\mathbb{N}, r)$ and $(\mathbb{N}, \hat{r})$, respectively. Then both $\widehat{T}$ and T are the countable star with a set $\mathbb{N}$ of leaves. In T, the distance from the branch point to leaf n is $\frac{1}{n}$, while it is 1 in $\widehat{T}$. Hence, T is compact, while $\widehat{T}$ is not. The identity on $\mathbb{N}$ can be extended to a bijective tree homomorphism $f\colon T \to \widehat{T}$, which cannot be continuous.

Example 2.37 shows that it is possible for nonhomeomorphic (topologically nonequivalent) $\mathbb{R}$-trees to induce isomorphic (equivalent) algebraic continuum trees. This can only happen if at least one of the trees is noncompact.

PROPOSITION 2.38 (tree isomorphisms of compact $\mathbb{R}$-trees are homeomorphisms). — *Let $T, \widehat{T}$ be $\mathbb{R}$-treesand $f\colon T \to \widehat{T}$.*

(i) *If $\widehat{T}$ is compact, $f(T)$ is connected, and f a tree homomorphism, then f is continuous.*
(ii) *If both T and $\widehat{T}$ are compact and f is bijective, then f is a homeomorphism, if and only if it is a tree homomorphism.*

Proof. — (ii) is obvious from (i) and Lemma 2.36.

Assume f is a tree homomorphism, $f(T)$ is connected, and $\widehat{T}$ is compact. Choose a root $\rho \in T$. Let $v_n \to v$ be a convergent sequence in T and $w \in \widehat{T}$ an accumulation point of $f(v_n)$. Then there is a subsequence $(n_k)_{k\in\mathbb{N}}$ with $f(v_{n_k}) \to w$. We have

$$v = \sup_{k\in\mathbb{N}} \inf_{i>k} v_{n_i} \quad \text{and} \quad w = \sup_{k\in\mathbb{N}} \inf_{i>k} f(v_{n_i}), \tag{59}$$

where sup and inf are with respect to the partial orders $\leq_\rho$ and $\leq_{f(\rho)}$ in the first and second equality, respectively. In the following, we show that $w = f(v)$. Because f is order preserving for these partial orders due to Lemma 2.12, we obtain $w \leq_{f(\rho)} f(v)$. Assume for a contradiction that $w \neq f(v)$. Because $f(T)$ is connected, there is $y \in \widehat{T}$ with $w <_{f(\rho)} y <_{f(\rho)} f(v)$ and $x \in T$ with $y = f(x)$. For $u \coloneqq c(\rho, x, v)$, we have $f(u) = \hat{c}(f(\rho), y, f(v)) = y$, $u \leq_\rho v$, and $u \neq v$. Therefore, $u \leq_\rho v_{n_i}$ for all sufficiently large i, and, thus, $y = f(u) \leq_{f(\rho)} f(v_{n_i})$ for those i. Now (59) implies $y \leq_{f(\rho)} w$ in contradiction to the choice of y, finishing the proof of $w = f(v)$. Compactness of $\widehat{T}$ and uniqueness of accumulation points implies $f(v_n) \to f(v)$, and f is continuous. □

In view of Theorem 2.27, Proposition 2.38 implies that order separable algebraic continuum trees are in one-to-one correspondence with homeomorphism classes of compact $\mathbb{R}$-trees. Furthermore, the unique metric topology induced by the compact $\mathbb{R}$-tree coincides with the component topology τ introduced in Section 2.3. However, be aware that there may be other, nonhomeomorphic, noncompact $\mathbb{R}$-trees inducing the same order separable algebraic continuum tree as shown in Example 2.37.

COROLLARY 2.39 (uniqueness of inducing $\mathbb{R}$-tree). — *Every order separable algebraic continuum tree is induced by a compact $\mathbb{R}$-tree that is unique up to homeomorphism, and the unique induced topology coincides with the component topology τ defined in* (37).

Proof. — That an order separable algebraic continuum tree is induced by a compact $\mathbb{R}$-tree is Theorem 2.27(i). Any two such compact $\mathbb{R}$-trees are isomorphic as algebraic trees, and, hence, homeomorphic by Proposition 2.38. The component topology is a Hausdorff topology and is clearly weaker than the topology induced by the $\mathbb{R}$-tree, because components are open sets of $\mathbb{R}$-trees. Hence, by compactness of the $\mathbb{R}$-treethe two topologies coincide. □

3. The space of algebraic measure trees

In this section, we define algebraic measure trees and equip the space of (equivalence classes of) algebraic measure trees with a topology. In the following, the order separability of the underlying algebraic tree is crucial. Therefore, we include it already in the following definition of algebraic measure trees.

DEFINITION 3.1 (algebraic measure trees). — An *algebraic measure tree* (T, c, μ) is an order-separable algebraic tree (T, c) together with a probability measure μ on $\mathcal{B}(T, c)$.

DEFINITION 3.2 (equivalence of algebraic measure trees). —
(i) We call two algebraic measure trees (T_i, c_i, μ_i), $i = 1, 2$, *equivalent*, if there exist subtrees A_i of T_i with $\mu_i(A_i) = 1$ and a measure preserving tree isomorphism f from A_1 onto A_2. In this case, we call f *isomorphism* of the algebraic measure trees.
(ii) A metric measure tree (T, r, μ) is called a *metric representation* of the algebraic measure tree (T', c', μ'), if its induced algebraic measure tree $(T, c_{(T,r)}, \mu)$ is equivalent to (T', c', μ').

In the following, we denote for an algebraic measure tree $\boldsymbol{x} \coloneqq (T, c, \mu)$ by $\operatorname{supp}(\boldsymbol{x})$ the algebraic subtree generated by the support of μ, i.e.,

$$\operatorname{supp}(\boldsymbol{x}) \coloneqq c\big(\operatorname{supp}(\mu)^3\big), \tag{60}$$

and by

$$\operatorname{br}(\boldsymbol{x}) \coloneqq \operatorname{br}(T, c) \cap \operatorname{supp}(\boldsymbol{x}) \tag{61}$$

the set of *branch points* of $\boldsymbol{x}$. It is easy to check that an isomorphism f from $\boldsymbol{x} = (T, c, \mu)$ to $\boldsymbol{x}' = (T', c', \mu')$ induces a bijection between $\operatorname{br}(\boldsymbol{x})$ and $\operatorname{br}(\boldsymbol{x}')$ (although it need neither be defined nor injective on all of $\operatorname{supp}(\boldsymbol{x})$). Also note that $\boldsymbol{x}$ is equivalent to $\operatorname{supp}(\boldsymbol{x})$ equipped with the appropriate restrictions of c and μ.

REMARK 3.3 (a note on our definition of equivalence). — Every algebraic measure tree is equivalent to an algebraic continuum measure tree and has a metric representation with a compact $\mathbb{R}$-tree by Theorem 2.27. For the definition of the equivalence of algebraic measure trees, it is important that we do not require the whole trees to be isomorphic (see Example 3.11 below). On the other hand, it is also important that the isomorphism is injective on a subtree (as opposed to only a subset) of full measure, because otherwise it would not be an equivalence relation, and every tree with n leaves and uniform distribution on them would be equivalent to the n-star.

EXAMPLE 3.4 (the linear nonatomic measure tree). — There is only one equivalence class of linearly ordered algebraic measure trees with nonatomic measure. Indeed, let (T, c, μ) be an algebraic measure tree with $\operatorname{br}(T, c) = \emptyset = \operatorname{at}(\mu)$.

Then, by Theorem 2.27 there is a tree isomorphism from T into $[0,1]$, and we may assume $T \subseteq [0,1]$ to begin with. Let $F_\mu \colon [0,1] \to [0,1]$ be the distribution function of μ. Then F_μ is continuous and maps μ to Lebesgue measure $\lambda_{[0,1]}$. Let $A \coloneqq \{x \in \operatorname{supp}(\mu) : \text{there is no } y_n \in [0,1] \setminus \operatorname{supp}(\mu) : y_n < x,\ y_n \to x\}$ be the support of μ with left boundary points removed. Then F_μ restricted to A is bijective, and, hence, a measure preserving tree isomorphism onto $[0,1]$ (with Lebesgue measure and canonical branch-point map). Thus, (T, c, μ) is equivalent to $[0,1]$.

Let

$$\mathbb{T} \coloneqq \{\text{equivalence classes of algebraic measure trees}\}. \tag{62}$$

Next, we equip $\mathbb{T}$ with a topology. We shall base this notion of convergence on the fact that algebraic measure trees allow for metric representations (see Theorem 2.27) and require convergence in Gromov-weak topology of particular representations. To this end, let

$$\mathbb{H} \coloneqq \{\text{equivalence classes of (separable) metric measure trees}\}, \tag{63}$$

where we consider two metric measure trees (T, r, μ) and (T', r', μ') as *equivalent*, if there exists a measure-preserving isometry between the metric completions of $\operatorname{supp}(\mu)$ and $\operatorname{supp}(\mu')$.

In order to get a useful topology on $\mathbb{T}$, we cannot take arbitrary (optimal) metric representations. Instead, given an algebraic measure tree (T, c, μ), we use the metric r_ν defined in (43) for the *branch-point distribution* ν, namely the distribution of the random branch point obtained by sampling three points with the sampling measure μ.

Definition 3.5 (branch-point distribution). — The *branch-point distribution* of an algebraic measure tree (T, c, μ) is the push-forward of $\mu^{\otimes 3}$ under the branch-point map,

$$\nu \coloneqq c_* \mu^{\otimes 3}. \tag{64}$$

Note that the branch-point distribution is not necessarily supported by $\operatorname{br}(T, c)$. For instance, every atom of μ is also an atom of ν. If (T, c, μ) and (T', c', μ') are equivalent algebraic measure trees with branch-point distributions ν and ν', respectively, then the isomorphism is also an isometry with respect to r_ν and $r_{\nu'}$. Therefore, the following selection map, which associates a particular metric representation to every algebraic measure tree, is well defined.

Definition 3.6 (selection map ι). — Define the map $\iota \colon \mathbb{T} \to \mathbb{H}$ by

$$\iota(T, c, \mu) \coloneqq (T_\nu, r_\nu, \mu_\nu), \tag{65}$$

where $\nu = c_*\mu^{\otimes 3}$ is the branch-point distribution of (T, c, μ), (T_ν, r_ν) is the quotient metric space, and μ_ν is the image of μ under the canonical projection π_ν.

The topology we use on $\mathbb{T}$ is the Gromov-weak topology with respect to the branch-point distribution distance. That is, it is the topology induced by the selection map ι, i.e., the weakest (coarsest) topology on $\mathbb{T}$, such that ι is continuous.

DEFINITION 3.7 (bpdd-Gromov-weak topology). — Let $\mathbb{H}$ be equipped with the Gromov-weak topology. We call the topology induced on $\mathbb{T}$ by the selection map ι *branch-point distribution distance Gromov-weak topology* (*bpdd-Gromov-weak topology*).

The following reconstruction theorem is crucial for the usefulness of bpdd-Gromov-weak convergence. It shows that the selection map ι is an embedding and, indeed, selects metric representations.

PROPOSITION 3.8 (ι is injective). — *The selection map $\iota\colon \mathbb{T} \to \mathbb{H}$ is injective, and $\iota(\boldsymbol{x})$ is a metric representation of $\boldsymbol{x} \in \mathbb{T}$.*

Proof. — If we show that $\iota(\boldsymbol{x})$ is a metric representation of $\boldsymbol{x} = (T, c, \mu) \in \mathbb{T}$, it is obvious that ι is injective, because equivalence of metric measure spaces implies equivalence of the corresponding algebraic measure trees by Lemma 2.36.

Choosing an appropriate representative, we can assume that $\nu\{v\} > 0$ for all $v \in \mathrm{br}(T, c)$. The canonical projection $\pi_\nu\colon T \to T_\nu$ is a tree homomorphism by Lemma 2.30. To show equivalence of (T, c, μ) and $(T_\nu, c_{(T_\nu, r_\nu)}, \mu_\nu)$, we have to show that π_ν is injective on a subtree $A \subseteq T$ with $\mu(A) = 1$. Let $N \coloneqq \{v \in T : \pi_\nu(v) \neq \{v\}\}$. Then $\mu(\pi_\nu(v)) = 0$, for all $v \in N$, and $w \in \pi_\nu(v)$ implies $[v, w] \subseteq \pi_\nu(v)$, because π_ν is a tree homomorphism. Because there are at most countably many nondegenerate, disjoint closed intervals in T due to order separability, this implies that $\pi_\nu(N)$ is countable, and, thus, $\mu(N) = 0$. Define $A = T \setminus N$. Then $\mu(A) = 1$, and π_ν is injective on $T \setminus N$. To see that A is a subtree, pick $x, y, z \in A$. If $v \coloneqq c(x, y, z) \in \{x, y, z\}$, then $v \in A$. Otherwise, $v \in \mathrm{br}(T, c)$, and, hence, $\nu\{v\} > 0$. This implies $\pi_\nu(v) = \{v\}$, i.e., $v \in A$. □

COROLLARY 3.9 (metrizability). — $\mathbb{T}$ *equipped with bpdd-Gromov-weak topology is a separable, metrizable space.*

Proof. — The Gromov-weak topology on $\mathbb{H}$ is separable and metrizable, e.g., by the Gromov–Prohorov metric d_{GP} (see [36]). Because ι is injective by Proposition 3.8, $d_{\mathrm{BGP}}(\boldsymbol{x}, \boldsymbol{y}) \coloneqq d_{\mathrm{GP}}(\iota(\boldsymbol{x}), \iota(\boldsymbol{y}))$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{T}$ is a metric on $\mathbb{T}$ inducing bpdd-Gromov-weak topology. □

REMARK 3.10 (distance polynomials). — By definition, a sequence $(\mathscr{x}_n)_{n\in\mathbb{N}}$ in $\mathbb{T}$ converges to $\mathscr{x}\in\mathbb{T}$ bpdd-Gromov-weakly, if and only if $\iota(\mathscr{x}_n)\to\iota(\mathscr{x})$ Gromov-weakly. It has been shown that the Gromov-weak convergence is equivalent to the convergence of the distribution of the distance matrix ([36, Theorem 5]). Therefore, the bpdd-Gromov-weak convergence is equivalent to

$$\Phi(\mathscr{x}_n)\xrightarrow[n\to\infty]{}\Phi(\mathscr{x}), \tag{66}$$

for all so-called *polynomials* $\Phi\colon\mathbb{T}\to\mathbb{R}$, which are test functions of the form (13). Note that the set Π_ι of all polynomials is an algebra and, therefore, also convergence determining for $\mathbb{T}$-valued random variables (see [46, 11]).

As pointed out in Remark 3.3, the equivalence class of every algebraic measure tree contains an algebraic *continuum* measure tree. The following example shows that ι would not be injective if we had defined it on the set of algebraic continuum measure trees with the stricter notion of equivalence, where the whole algebraic continuum trees have to be measure-preserving isomorphic.

EXAMPLE 3.11. — For $x\geq 0$, let T_x be the $\mathbb{R}$-tree generated by the interval $I_x=[-x,1]$ together with additional leaves $\{v_n\}$, $n\in\mathbb{N}$, where $c(0,1,v_n)=\frac{1}{n}$ and $r(\frac{1}{n},v_n)=\frac{1}{n}$, i.e., at each point $\frac{1}{n}\in I_x$, there is a branch of length $\frac{1}{n}$ attached. Then T_x is a compact $\mathbb{R}$-tree for every $x\geq 0$ and, hence, induces an algebraic continuum tree by Theorem 2.27. Let $\mu_x\{-x\}=\frac{1}{2}$, and $\mu_x\{v_n\}=2^{-n-1}$ for $n\in\mathbb{N}$. Then $\mathscr{x}_x:=(T_x,\mu_x)\in\mathbb{T}_2$. Now $\iota(\mathscr{x}_x)=\iota(\mathscr{x}_y)$ for every $x,y\geq 0$, but T_x and T_0 are not homeomorphic and, hence, not isomorphic by Proposition 2.38.

Note that $A_x:=\{x\}\cup\{v_n:n\in\mathbb{N}\}\cup\{\frac{1}{n}:n\in\mathbb{N}\}$ is a subtree of T_x with $\mu_x(A_x)=1$, and A_x is isomorphic (although not homeomorphic) to A_0.

In order to construct algebraic measure trees, it is, of course, not necessary to specify the mass of every Borel subset. On the contrary, we can use the following Carathéodory-type extension result. To this end, recall for $x,y\in T$ with $x\neq y$ from (36) the component $\mathcal{S}_x(y)=\mathcal{S}_x^{(T,c)}(y)$ of $T\setminus\{x\}$, which contains y. In this section, it is convenient to define

$$\mathcal{S}_x(x):=\{x\}. \tag{67}$$

Then T is the disjoint union of the $\deg(x)+1$ sets in

$$\mathcal{C}_x:=\big\{\mathcal{S}_x(y):y\in T\big\}. \tag{68}$$

Note that $\mathcal{C}_x=\{\mathcal{S}_x(y):y\in V\}$ for order-dense $V\subseteq T$ with $x\in V$. In particular, $\mathcal{C}_x$ is countable if (T,c) is order separable. For $y\in T$, $V\subseteq T$, we call a function $f\colon V\to\mathbb{R}$ *order left-continuous* on V with respect to $\leq_y$, if the following holds. For all $x,x_n\in V$ with $x_1\leq_y x_2\leq_y\cdots$ and $x=\sup_{n\in\mathbb{N}}x_n$ with respect to $\leq_y$ (in short $x_n\uparrow x$), we have $\lim_{n\to\infty}f(x_n)=f(x)$. Recall the notion of the algebraic continuum tree from Definition 2.10.

PROPOSITION 3.12 (extension to a measure). — *Let (T,c) be an order separable algebraic continuum tree and $V \subseteq T$ order dense. Then a set function $\mu_0 \colon \mathcal{C}_V \coloneqq \bigcup_{x\in V} \mathcal{C}_x \to [0,1]$ has a unique extension to a probability measure on $\mathcal{B}(T,c)$ if it satisfies*

1. *For all $x \in V$, $\sum_{A\in\mathcal{C}_x} \mu_0(A) = 1$.*
2. *For all $x, y \in V$ with $x \neq y$,*

$$\mu_0\big(\mathcal{S}_x(y)\big) + \mu_0\big(\mathcal{S}_y(x)\big) \geq 1. \tag{69}$$

3. *For every $y \in V$, the function $\psi_y \colon x \mapsto \mu_0(\mathcal{S}_x(y))$ is order left-continuous on V with respect to $\leq_y$.*

Proof. — Note that $\psi_y(x) = \psi_z(x)$ for $z \in \mathcal{S}_x(y)$. We, therefore, may write $\psi_A(x) \coloneqq \psi_y(x)$ for any $A \subseteq \mathcal{S}_x(y)$. Define the $\cap$-stable set system

$$\mathcal{A} \coloneqq \Big\{ \bigcap_{k=1}^{n} A_k \,:\, n \in \mathbb{N},\, A_k \in \mathcal{C}_V \Big\}. \tag{70}$$

By Corollary 2.25, $\mathcal{A}$ generates the Borel σ-algebra $\mathcal{B}(T,c)$. Let $\emptyset \neq A \in \mathcal{A}$ and $y \in A$. Because (T,c) has no edges and is order complete, we have $A = \bigcap_{x\in\partial A} \mathcal{S}_x(y)$, where ∂ denotes the boundary with respect to the component topology τ, which is a finite set in the case of A. Using (69), we obtain for $v \in V$, $x_0, \dots, x_n \in V \setminus \{v\}$, such that $\mathcal{S}_v(x_0), \dots, \mathcal{S}_v(x_n)$ are distinct, that

$$\psi_{x_0}(v) \leq 1 - \sum_{k=1}^{n} \psi_{x_k}(v) \leq 1 - \sum_{k=1}^{n} \big(1 - \psi_v(x_k)\big). \tag{71}$$

This implies for $\emptyset \neq A \in \mathcal{A}$, by induction over $\#\partial A$, that

$$\mu(A) \coloneqq 1 - \sum_{x\in\partial A} \big(1 - \psi_A(x)\big) \geq 0, \tag{72}$$

hence μ is a nonnegative extension of μ_0 to $\mathcal{A}$. We claim that μ is superadditive, additive, and inner regular for compact sets. From this, it follows by standard arguments that it has a unique extension to a measure on the generated σ-algebra $\sigma(\mathcal{A}) = \mathcal{B}(T,c)$.

Additivity. Let $n \in \mathbb{N} \setminus \{1\}$, and $A_1, \dots, A_n \in \mathcal{A} \setminus \{\emptyset\}$ disjoint with $A \coloneqq \biguplus_{k=1}^{n} A_k \in \mathcal{A}$. Define $D \coloneqq \bigcup_{k=1}^{n} \partial A_k$. Then $\partial A \subseteq D$ and there is $x \in D \setminus \partial A \subseteq A$. Let $I_x \coloneqq \{k \in \{1,\dots,n\} : x \in \partial A_k\}$ and choose $y_k \in A_k$. Then, because the A_k are disjoint, the $\mathcal{S}_x(y_k)$, $k \in I$, are distinct, and because the A_k cover A, we have $\{\mathcal{S}_x(y_k) : k \in I_x\} = \mathcal{C}_x$. In particular, $\sum_{k\in I_x} \psi_{y_k}(x) = 1$, and $B_x \coloneqq \bigcup_{k\in I_x} A_k \in \mathcal{A}$ with $\partial B_x = \biguplus_{k\in I_x} \partial A_k \setminus \{x\}$. We obtain

$$\begin{aligned} \sum_{k\in I_x} \mu(A_k) &= \sum_{k\in I_x} \Big(1 - \big(1 - \psi_{y_k}(x)\big) - \sum_{z\in\partial A_k\setminus\{x\}} \big(1 - \psi_{y_k}(z)\big)\Big) \\ &= \sum_{k\in I_x} \psi_{y_k}(x) - \sum_{z\in\partial B_x} \big(1 - \psi_x(z)\big) = \mu(B_x). \end{aligned} \tag{73}$$

By induction over n, this implies additivity of μ.

Superadditivity. Let $A_1, \dots, A_n \in \mathcal{A} \setminus \{\emptyset\}$ be disjoint and $\biguplus_{k=1}^n A_k \subseteq A \in \mathcal{A}$. The case $n = 1$ is trivial, and we proceed by induction over n. Choose $y \in A_1$ and let $D \coloneqq \partial A_1 \setminus \partial A$. For $x \in D$, $C \in \mathcal{C}'_x \coloneqq \mathcal{C}_x \setminus \mathcal{S}_x(y)$ and $k \in \{2, \dots, n\}$, either $A_k \subseteq C$, or $A_k \cap C = \emptyset$. Therefore, we have the decomposition $\{2, \dots, n\} = \biguplus_{x \in D} \biguplus_{C \in \mathcal{C}'_x} I_C$ with $I_C \coloneqq \{k : A_k \subseteq C\}$. Because $C \cap A \in \mathcal{A}$, and $A_k \subseteq C \cap A$ for $k \in I_C$, we can use the induction hypothesis to obtain

$$\sum_{k \in I_C} \mu(A_k) \leq \mu(C \cap A) = \psi_C(x) - \sum_{z \in \partial A \cap C} \big(1 - \psi_A(x)\big). \tag{74}$$

Therefore,

$$\begin{aligned} \mu(A_1) &= 1 - \sum_{x \in \partial A_1 \cap \partial A} \big(1 - \psi_y(x)\big) - \sum_{x \in D} \big(1 - \psi_y(x)\big) \\ &= \mu(A) + \sum_{x \in \partial A \setminus \partial A_1} \big(1 - \psi_y(x)\big) - \sum_{x \in D} \sum_{C \in \mathcal{C}'_x} \psi_C(x) \\ &\leq \mu(A) - \sum_{x \in D} \sum_{C \in \mathcal{C}'_x} \sum_{k \in I_C} \mu(A_k) \\ &= \mu(A) - \sum_{k=2}^n \mu(A_k). \end{aligned} \tag{75}$$

Compact regularity. According to Proposition 2.19, all closed subsets of T are compact. Let $y \in A \in \mathcal{A}$. Because (T, c) is an order-separable algebraic continuum tree, and V is order dense, we find for $z \in \partial A$ a sequence $(x_n(z))_{n \in \mathbb{N}}$ in $A \cap V$ with $x_n(z) \uparrow z$ with respect to $\leq_y$ as $n \to \infty$. Define $A_n \coloneqq \bigcap_{z \in \partial A} \mathcal{S}_{x_n(z)}(y) \in \mathcal{A}$ and $K_n \coloneqq A_n \cup \partial A_n$. Then K_n is compact, $A_n \subseteq K_n \subseteq A$, and because ∂A is finite, we have $\partial A_n = \{x_n(z) : z \in \partial A\}$ for sufficiently large n. Thus, by order left continuity of ψ_y,

$$\begin{aligned} \lim_{n \to \infty} \mu(A_n) &= 1 - \lim_{n \to \infty} \sum_{z \in \partial A} \big(1 - \psi_y(x_n(z))\big) \\ &= 1 - \sum_{z \in \partial A} \big(1 - \psi_y(z)\big) = \mu(A), \end{aligned} \tag{76}$$

and μ is inner compact regular as claimed. □

We conclude this section with an extension result, which will be very useful for reading off algebraic measure trees from (sub)triangulations of the circle in Section 4. In Proposition 3.12, we assumed the whole tree to be known and considered the question of constructing a probability measure on it. Now, we assume that not the whole tree is given a priori but only the (countably many) branch points. The question is, whether there is an extension of the tree that is rich enough to carry a measure with the specified masses of components.

PROPOSITION 3.13 (construction of algebraic measure trees). — *Let (V, c_V) be a countable algebraic tree and for each $x \in V$, let $A \mapsto \psi_A(x)$ be a probability measure on $\mathcal{C}_x$. Define $\psi_y(x) \coloneqq \psi_{\mathcal{S}_x(y)}(x)$. Assume that for $x, y \in V$ with $x \neq y$,*

$$\psi_x(y) + \psi_y(x) \geq 1. \tag{77}$$

Then there is a unique (up to equivalence) algebraic measure tree $\mathscr{x} = (T, c, \mu)$, such that

(i) $V \subseteq T$, $\mathrm{br}(T, c) = \mathrm{br}(V, c_V)$.
(ii) $\mu\big(\mathcal{S}_x^{(T,c)}(y)\big) = \psi_y(x)$ *for all* $x, y \in V$.
(iii) $\mathrm{at}(\mu) \subseteq V$, *where* $\mathrm{at}(\mu)$ *denotes the set of atoms of* μ.

Note that, in general, we cannot obtain $\mathrm{lf}(T, c) \subseteq \mathrm{lf}(V, c_V)$. To the contrary, $\mathrm{lf}(T, c)$ can be uncountable (for every representative of $\mathscr{x}$).

Proof. — Existence. First note that for $y \in V$, ψ_y is monotonic with respect to $\leq_y$. Indeed, $z \leq_y x$ implies $\psi_y(z) \leq 1 - \psi_x(z) \leq \psi_z(x) = \psi_y(x)$.

We need to enlarge the tree to make ψ_y order left-continuous. Because V is countable, we may consider one y and one point x at a time. If $x, y \in V$ are such that there exists $x_n \in V$ with $x_n \uparrow x$, then by monotonicity $\phi_y(x) \coloneqq \lim_{n\to\infty} \psi_y(x_n) \leq \psi_y(x)$ exists and is independent of the choice of x_n. If $\phi_y(x) \neq \psi_y(x)$, we extend the tree by adding one extra point $z \notin V$, i.e., we consider $\tilde{V} \coloneqq V \uplus \{z\}$ with the unique extension $\tilde{c}$ of c_V, such that $(\tilde{V}, \tilde{c})$ is an algebraic tree with $x_n \leq_y z \leq_y x$ for all n. Furthermore, we extend ψ to $\tilde{\psi}$ on $\tilde{V}$ by defining $\tilde{\psi}_y(z) \coloneqq \phi_y(x)$, $\tilde{\psi}_z(z) = 0$ and $\tilde{\psi}_x(z) = 1 - \phi_y(x)$. It is easy to check that $(\tilde{V}, \tilde{c})$ together with $\tilde{\psi}$ satisfies the prerequisites of the proposition, $\mathrm{br}(\tilde{V}, \tilde{c}) = \mathrm{br}(V, c)$, and $\{x \in \tilde{V} : \tilde{\psi}_x(x) > 0\} = \{x \in V : \psi_x(x) > 0\} \subseteq V$.

Now assume without loss of generality that ψ_y is already order left-continuous for all $y \in V$. Because V is countable, it is, in particular, order separable and according to Theorem 2.27, there is an order separable algebraic continuum tree (T, c), such that (V, c_V) is a subtree. We can choose (T, c), such that $\mathrm{br}(T, c) = \mathrm{br}(V, c_V)$. Consider the closure $\overline{V}$ of V with respect to the component topology τ. For $x \in \overline{V} \setminus V$, we define

$$\psi_y(x) \coloneqq \sup\{\psi_y(z) : z \in V \cap \mathcal{S}_x(y)\}. \tag{78}$$

Then (77) holds for $x, y \in \overline{V}$, $x \neq y$, and ψ_y is order left-continuous. For every $\{x, y\} \in \mathrm{edge}(\overline{V}, \bar{c})$, where $\bar{c}$ is the restriction of c to $\overline{V}^3$, we fix an order isomorphism $\varphi_{x,y} \colon [x, y] \to [0, 1]$, which exists by Cantor's order characterization of $\mathbb{R}$ because $[x, y]$ is a linearly ordered, separable algebraic continuum tree. For every $z \in T \setminus \overline{V}$, there exists $\{x, y\} \in \mathrm{edge}(\overline{V}, \bar{c})$, with $z \in [x, y]$. We define

$$\psi_y(z) \coloneqq (1 - \varphi_{x,y}(z))\psi_y(x) + \varphi_{x,y}(z)\big(1 - \psi_x(y)\big), \tag{79}$$

$\psi_x(z) \coloneqq 1 - \psi_y(z)$ and $\psi_z(z) \coloneqq 0$. Now we can use Proposition 3.12 to see that

$$\mu_0\big(\mathcal{S}_x(y)\big) \coloneqq \psi_y(x) \tag{80}$$

has a unique extension to a probability measure μ on $\mathcal{B}(T,c)$.

The last step in the construction is to remove point masses outside V by expanding them to intervals. To this end, let $P \coloneqq \mathrm{at}(\mu) \setminus V$, and $\overline{T} \coloneqq (T \setminus P) \uplus (P \times [0,1])$. Because $P \subseteq T \setminus V$ contains no branch points, we can extend the restriction of c to $T \setminus P$ to a branch-point map $\tilde{c}$ on $\overline{T}$ in a canonical way, such that $[(x,0),(x,1)] = \{x\} \times [0,1]$ for $x \in P$. Define the Markov kernel κ from T to $\overline{T}$ by

$$\kappa(x) \coloneqq \begin{cases} \delta_x, & x \in T \setminus P, \\ \delta_x \otimes \lambda_{[0,1]}, & x \in P, \end{cases} \tag{81}$$

where δ_x is the Dirac measure in x, and $\lambda_{[0,1]}$ is a Lebesgue measure. Let $\bar{\mu} \coloneqq \kappa_*(\mu)$ be the push-forward of μ under κ. Then $(\overline{T}, \tilde{c}, \bar{\mu})$ is a separable algebraic measure tree, and by construction $\mathrm{br}(\overline{T}, \tilde{c}) = \mathrm{br}(V, c_V)$, as well as $\mathrm{at}(\bar{\mu}) = \mathrm{at}(\mu) \cap V \subseteq V$. Furthermore, for $x, y \in V$, we have $\bar{\mu}(\mathcal{S}_x^{(\overline{T},\tilde{c})}(y)) = \mu(\mathcal{S}_x^{(T,c)}(y)) = \psi_y(x)$, as claimed. *Uniqueness.* follows similarly, where we note that it does not matter how we distribute the mass on an edge of $(\overline{V}, \bar{c})$ in a nonatomic way, because all algebraic measure trees without branch points and nonatomic measure are equivalent by Example 3.4. □

4. Triangulations of the circle

In this section, we encode binary algebraic measure trees by triangulations of subsets of the circle. This is comparable with the encoding of compact (ordered, rooted) metric (probability) measure trees by excursions over the unit interval, where the height profile encodes the branch-point map, as well as the metric distances. Moreover, also the measure can be encoded by the excursion by identifying the lengths of subexcursions with the mass of the corresponding subtrees. Similarly, it turns out that we can encode binary algebraic measure trees by what we call subtriangulations of the circle. As in the case of coding metric measure trees with excursions, the resulting *coding map* associating the algebraic measure tree to a subtriangulation is continuous.

In Section 4.1, we introduce the space of subtriangulations of the circle. In Section 4.2, we construct the coding map.

4.1. The space of subtriangulations of the circle. — Let $\mathbb{D}$ be a (fixed) closed disc of circumference 1, and $\mathbb{S} \coloneqq \partial\mathbb{D}$ the circle. As usual, for a subset $A \subseteq \mathbb{D}$,

we denote by $\bar{A}$, $\mathring{A}$, ∂A, and $\operatorname{conv}(A)$ the closure, the interior, the boundary, and the convex hull of A, respectively. Furthermore, let

$$\Delta(A) \coloneqq \big\{\text{connected components of } \operatorname{conv}(A) \setminus A\big\}, \tag{82}$$

and

$$\nabla(A) \coloneqq \big\{\text{connected components of } \mathbb{D} \setminus \operatorname{conv}(A)\big\}. \tag{83}$$

Then we have the disjoint decomposition $\mathbb{D} = A \uplus \bigcup \Delta(A) \uplus \bigcup \nabla(A)$.

Definition 4.1 ((sub)triangulations of the circle). — A *subtriangulation of the circle* is a closed, nonempty subset C of $\mathbb{D}$ satisfying the following two conditions:

(Tri1) $\Delta(C)$ consists of open interiors of triangles.
(Tri2) C is the union of noncrossing (nonintersecting except at endpoints), possibly degenerate, closed straight line segments with endpoints in $\mathbb{S}$.

We denote the set of subtriangulations of the circle by $\mathcal{T}$, i.e.,

$$\mathcal{T} \coloneqq \big\{\text{subtriangulations of the circle}\big\} \tag{84}$$

and call $C \in \mathcal{T}$ the *triangulation of the circle*, if and only if $\mathbb{S} \subseteq C$.

In particular, (Tri1) implies that $\partial \operatorname{conv}(C) \subseteq C$, and we may call C triangulation of $\partial \operatorname{conv}(C)$. Given (Tri1), (Tri2) implies that $\nabla(A)$ consists of circular segments with the bounding straight line excluded and the rest of the bounding arc included. We want to point out that our definition of the triangulation of the circle differs from the one given by Aldous in [5, Definition 1]. Namely, Aldous required only Condition (Tri1). For the characterization of triangulations of the circle as limits of triangulations of n-gons given in Proposition 4.3, however, Condition (Tri2) is necessary. See Figure 4.1 for an example of a triangulation in the sense of Aldous that is excluded by Condition (Tri2), a subtriangulation of the circle that is no triangulation of the circle, and a triangulation of the circle.

For a metric space (X, d), let

$$\mathcal{F}(X) \coloneqq \{F \subseteq X : F \neq \emptyset,\ F \text{ closed}\} \tag{85}$$

and equip $\mathcal{F}(X)$ with the *Hausdorff metric topology*. That is, we say that a sequence $(F_n)_{n\in\mathbb{N}}$ converges to F in $\mathcal{F}(X)$, if and only if for all $\varepsilon > 0$ and all large enough $n \in \mathbb{N}$,

$$F_n^\epsilon \supseteq F \qquad \text{and} \qquad F^\epsilon \supseteq F_n, \tag{86}$$

where for all $A \in \mathcal{F}(X)$, as usual, $A^\epsilon \coloneqq \{x \in X : d(x, A) < \epsilon\}$. It is well known that if (X, d) is compact, then $\mathcal{F}(X)$ is a compact metrizable space as well. As subtriangulations of the circle are elements of $\mathcal{F}(\mathbb{D})$, we naturally equip $\mathcal{T}$ with the Hausdorff metric topology. A first observation is that $\mathcal{T}$ is actually a closed and, therefore, compact subspace of $\mathcal{F}(\mathbb{D})$.

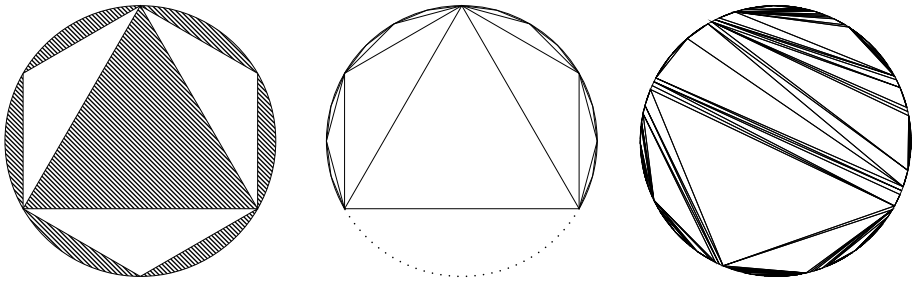

FIGURE 4.1. *Left*: an Aldous triangulation of the circle that is not a triangulation of the circle (Condition (Tri2) does not hold as the black triangle in the middle is not the union of noncrossing straight lines with endpoints on the circle). *middle*: A subtriangulation of the circle (compare with Example 4.12). *right*: A triangulation of the circle. It is a realization of the Brownian triangulation (compare with Example 4.5).

LEMMA 4.2 (compactness of $\mathcal{T}$). — *Both the space of triangulations of the circle and the space $\mathcal{T}$ of subtriangulations of the circle are compact metrizable spaces in the Hausdorff metric topology.*

Proof. — Because $\mathbb{D}$ is compact, $\mathcal{F}(\mathbb{D})$ is compact as well, and it is sufficient to show that $\mathcal{T}$ and the set of triangulations of the circle are closed subsets of $\mathcal{F}(\mathbb{D})$.

Let $C_n \in \mathcal{T}$ with $C_n \xrightarrow[n\to\infty]{} C \in \mathcal{F}(\mathbb{D})$ in the Hausdorff metric topology. (Tri1) is easily seen to be a closed property, and, thus, C satisfies (Tri1). Let L_n be a set of noncrossing line segments with endpoints in $\mathbb{S}$, such that $C_n = \bigcup L_n$. The closure of L_n in $\mathcal{F}(\mathbb{D})$ has the same property (it possibly differs from L_n by a set of degenerated one-point segments contained in nondegenerate segments of L_n), so we may assume L_n is closed to begin with, so that $L_n \in \mathcal{F}(\mathcal{F}(\mathbb{D}))$. Because $\mathcal{F}(\mathcal{F}(\mathbb{D}))$ is compact, we may assume, taking a subsequence if necessary, that $L_n \to L$ for some $L \in \mathcal{F}(\mathcal{F}(\mathbb{D}))$. Obviously, L_n consists of noncrossing line segments with endpoints in $\mathbb{S}$. Because the union operator $\bigcup\colon \mathcal{F}(\mathcal{F}(\mathbb{D})) \to \mathcal{F}(\mathbb{D})$ is continuous, we have $\bigcup L = C$. In particular, (Tri2) holds for C and $C \in \mathcal{T}$. Obviously, also the property that $\mathbb{S} \subseteq C$ is preserved by Hausdorff metric limits, and, thus, the set of triangulations of the circle is closed as well. □

We now show two characterizations of subtriangulations of the circle. Namely, condition (Tri2) can be replaced by existence of "triangles in the middle", which is the major technical ingredient for the construction of the branch-point map in the next section. Furthermore, they are precisely the limits of

finite subtriangulations, where we consider a subtriangulation C as *finite*, if $C \cap \mathbb{S}$ is a finite set, or equivalently, C consists of finitely many line segments.

PROPOSITION 4.3 (characterization of (sub)triangulations). — *Let $\emptyset \neq C \subseteq \mathbb{D}$ be closed. Then the following are equivalent.*

1. *C is a subtriangulation of the circle.*
2. *Condition (Tri1) holds, all extreme points of $\operatorname{conv}(C)$ are contained in $\mathbb{S}$, and*

(Tri2′) *For $x, y, z \in \Delta(C) \cup \nabla(C)$ pairwise distinct, there exists a unique $c_{xyz} \in \Delta(C)$, such that x, y, z are subsets of pairwise different connected components of $\mathbb{D} \setminus \partial c_{xyz}$.*

3. *There exists a sequence $(C_n)_{n\in\mathbb{N}}$ of finite subtriangulations of the circle with $C_n \xrightarrow[n\to\infty]{} C$ in the Hausdorff metric topology.*

Furthermore, C is a triangulation of the circle, if and only if C_n in 3. can be chosen as a triangulation of a regular n-gon inscribed in $\mathbb{S}$.

REMARK 4.4 (condition (Tri2)'). — That x, y, z are subsets of different connected components of $\mathbb{D} \setminus \partial c_{xyz}$ means that either $c_{xyz} \in \{x, y, z\}$, and the two elements of $\{x, y, z\} \setminus \{c_{xyz}\}$ are subsets of different connected components of $\mathbb{D} \setminus \overline{c_{xyz}}$, or $c_{xyz} \notin \{x, y, z\}$ and x, y, z are subsets of pairwise different connected components of $\mathbb{D} \setminus \overline{c_{xyz}}$.

Proof of Proposition 4.3. — "1⇒ 2". Because C is the union of line segments with endpoints on $\mathbb{S}$, it is obvious that the extreme points of $\operatorname{conv}(C)$ are contained in $\mathbb{S}$. We have to show (Tri2)', so let $x, y, z \in \Delta(C) \cup \nabla(C)$ be pairwise distinct and note that uniqueness is obvious. If one of the elements of $\{x, y, z\}$, say x, is such that the other two are subsets of two different connected components of $\mathbb{D} \setminus \bar{x}$, then necessarily $x \in \Delta(C)$, and $c_{xyz} \coloneqq x$ has the desired properties. So assume that this is not the case.

Fix a set L of noncrossing, closed lines with endpoints in $\mathbb{S}$, such that $C = \bigcup L$. Define

$$L_x \coloneqq \{\ell \in L : \ell \text{ separates } x \text{ from } y \cup z \text{ in } \mathbb{D}\}, \tag{87}$$

note that $L_x \neq \emptyset$ because y and z are in the same connected component of $\mathbb{D} \setminus \bar{x}$ by assumption, and order L_x by distance from x. Similarly, define L_y as set of lines separating y from $x \cup z$ ordered by the distance from y, and L_z as set of lines separating z from $x \cup y$, ordered by the distance from z. Define $\ell_x \coloneqq \sup L_x$, $\ell_y \coloneqq \sup L_y$, and $\ell_z \coloneqq \sup L_z$, which exist because C is closed. In particular, they are noncrossing, and because $\operatorname{conv}(C) \setminus C$ may only consist of triangles, they have to be the sides of some $c_{xyz} \in \Delta(C)$, which has the desired properties.

"2⇒ 3". Because the extreme points of $\operatorname{conv}(C)$ are on the circle, for every $x \in \nabla(C)$, the boundary $\partial_{\mathbb{D}} x$ in $\mathbb{D}$ is a single straight line with endpoints in $\mathbb{S}$. Let $(V_n)_{n\in\mathbb{N}}$ be an increasing sequence of finite subsets of $\Delta(C) \cup \nabla(C)$, such

that $c_{xyz} \in V_n$ for pairwise distinct $x, y, z \in V_n$, and $V_n \uparrow \Delta(C) \cup \nabla(C)$. Let $A_n := \mathbb{D} \setminus \bigcup V_n$. Then $A_n \to C$ in the Hausdorff metric topology. Because $c_{xyz} \in V_n$ for distinct $x, y, z \in V_n$, the boundary of each of the finitely many connected components of $A_n \setminus \mathbb{S}$ consists of one or two line segments and one or two connected subarcs of $\mathbb{S}$. Therefore, there is a finite subtriangulation $C_n \subseteq A_n$ of the circle with Hausdorff distance from A_n less than e^{-n}. Thus $C_n \to C$. *"3⇒ 1".* Obviously, because $\mathcal{T}$ is a closed subset of $\mathcal{F}(\mathbb{D})$ by Lemma 4.2. *"Furthermore".* If C_n is a triangulation of the n-gon, it contains the n-gon, and, hence, any Hausdorff metric limit as $n \to \infty$ contains the circle, and, hence, is a triangulation of the circle. That triangulations of the circle can be approximated by triangulations of regular n-gons is a slight modification of the arguments above. The details are left to the reader. □

The most prominent random tree is Aldous's Brownian CRT, which is the limit of uniform random trees. Similarly, one can define the Brownian triangulation of the circle.

Example 4.5 (Brownian triangulation). — The uniform random triangulation of the n-gon converges in law with respect to the Hausdorff metric topology to the so-called *Brownian triangulation* C_{CRT}, see [4, 5, 16]. A realization is shown on the right-hand side of Figure 4.1. It has a.s. Hausdorff dimension $\frac{3}{2}$ (see [4]).

4.2. Coding binary measure trees with (sub)triangulations of the circle. — Given an algebraic tree (T, c), recall the set of leaves $\mathrm{lf}(T, c)$ and the degree $\deg_{(T,c)}(v)$ of $v \in T$ from (51) and (50), respectively. In this section, we are interested in the following subspace of the space of all binary algebraic measure trees.

Definition 4.6 (our space $\mathbb{T}_2$). — Let $\mathbb{T}_2 \subseteq \mathbb{T}$ be the set of (equivalence classes of) algebraic measure trees (T, c, μ) with (T, c) binary (i.e., $\deg_{(T,c)}(v) \leq 3$ for all $v \in T$) and $\mathrm{at}(\mu) \subseteq \mathrm{lf}(T, c)$.

The space $\mathbb{T}_2$ is of particular interest to us, as it is invariant under the dynamics of the Aldous diffusion on cladograms, the construction of which was one of the motivations for studying algebraic measure trees, and because, as we will see, it is precisely the space of algebraic measure trees that can be coded by subtriangulations of the circle.

To illustrate the construction of the tree coded by a subtriangulation, we first consider a triangulation C of the regular n-gon into necessarily $n-2$ triangles (see Figure 1.3). Here, the coded tree is the dual graph. That is, every triangle corresponds to a branch point of the tree, and two branch points are connected by an edge if and only if the triangles share a common edge. We then add a leaf for every edge of the n-gon and obtain a graph-theoretic binary tree with n leaves and $n-2$ internal vertices. Recall from Example 2.4 that the finite

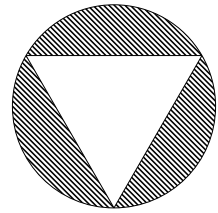

FIGURE 4.2. Triangulation C with $\#\Delta(C) = 1$, $\nabla(C) = \emptyset$, and $\#\square(C) = 3$. The coded tree consists of three line segments with nonatomic measure of $\frac{1}{3}$ each, glued together at one branch point.

graph-theoretic tree corresponds to a unique algebraic tree. We finally assign to each leaf mass n^{-1} (which corresponds to the length of the arcs of the circle connecting two endpoints of edges of the n-gon, if we inscribe it in a circle of unit length), and obtain an algebraic measure tree.

The main result of this section is that there is a natural, surjective coding map from $\mathcal{T}$ onto $\mathbb{T}_2$, which is also continuous. To state that formally we need further notation. Given a subtriangulation $C \subseteq \mathbb{D}$, recall $\Delta(C)$ and $\nabla(C)$ from (83) and (82), respectively. For $x \in \Delta(C) \cup \nabla(C)$ and $y \subseteq \mathbb{D}$ connected and disjoint from $\partial_{\mathbb{D}} x$, where $\partial_{\mathbb{D}}$ denotes the boundary in the space $\mathbb{D}$, let

$$\mathrm{comp}_x(y) \coloneqq \text{the connected component of } \mathbb{D} \setminus \partial_{\mathbb{D}} x \text{ which contains } y. \tag{88}$$

For $x \in \Delta(C)$, let $p_i(x)$, $i = 1, 2, 3$, be the mid-points of the three arcs of $\mathbb{S} \setminus \partial x$, and define

$$\square(C) \coloneqq \big\{\{p_i(x)\} : x \in \Delta(C),\, i \in \{1,2,3\},\, \mathrm{comp}_x(\{p_i(x)\}) \subseteq C\big\}, \tag{89}$$

as well as $\mathrm{comp}_p(p) \coloneqq p$ for $p \in \square(C)$ (see Figure 4.2). Recall the definition of components $\mathcal{S}_v(w)$ in an algebraic tree from (36).

LEMMA 4.7 (induced branch-point map). — *For $C \in \mathcal{T}$, let $V_C \coloneqq \Delta(C) \cup \nabla(C) \cup \square(C)$. If $V_C \neq \emptyset$, then there is a unique branch-point map $c_V : V_C^3 \to V_C$, such that (V_C, c_V) is an algebraic tree with $\mathcal{S}_x^{(V_c, c_V)}(y) = \{v \in V_C : \mathrm{comp}_x(y) = \mathrm{comp}_x(v)\}$ for $x, y \in V_C$. Furthermore, $\deg(x) = 3$ for all $x \in \Delta(C)$, and $\deg(x) = 1$, for $x \in \nabla(C) \cup \square(C)$.*

Proof. — Recall from Proposition 4.3 that for a subtriangulation C of the circle and pairwise distinct $x, y, z \in \Delta(C) \cup \nabla(C)$, there is a triangle $c_{xyz} \in \Delta(C)$ "in the middle". It is straightforward to see that this defines a branch-point map and can naturally be extended to V_C^3. □

The following theorem states that all subtriangulations C of the circle can be associated with an element in $\mathbb{T}_2$, for which $\Delta(C)$ corresponds to the set of

branch points, $\nabla(C)$ corresponds to the set

$$\mathrm{lf}_{\mathrm{atom}}(\boldsymbol{x}) \coloneqq \big\{x \in \mathrm{lf}(T,c) : \mu(\{x\}) > 0\big\} \tag{90}$$

of leaves that carry an atom, and $\mathrm{comp}_v(w)$ corresponds to the component $\mathcal{S}_v(w)$.

THEOREM 4.8 (algebraic measure tree associated to a subtriangulation). —

(i) *For every subtriangulation $C \subseteq \mathbb{D}$ of the circle, there is a unique (up to equivalence) algebraic measure tree $\boldsymbol{x}_C = (T_C, c_C, \mu_C) \in \mathbb{T}_2$ with the following properties:*

(CM1) *$V_C \subseteq T_C$, $\mathrm{br}(\boldsymbol{x}_C) = \Delta(C)$, and c_C is an extension of c_V, where (V_C, c_V) is defined in Lemma 4.7.*

(CM2) *$\mu_C\big(\mathcal{S}_x^{(T_C,c_C)}(y)\big) = \lambda_{\mathbb{S}}\big(\mathbb{S} \cap \mathrm{comp}_x(y)\big)$ for all $x, y \in V_C$, where $\lambda_{\mathbb{S}}$ denotes the Lebesgue measure on $\mathbb{S}$.*

(CM3) *$\mathrm{at}(\mu_C) = \nabla(C)$.*

(ii) *The* coding map *$\tau\colon \mathcal{T} \to \mathbb{T}_2$, $C \mapsto \boldsymbol{x}_C$ is surjective and continuous, where $\mathcal{T}$ is equipped with the Hausdorff metric topology and $\mathbb{T}_2$ with the bpdd-Gromov-weak topology.*

Proof. — *(i)* Let C be a subtriangulation of the circle. If $C = \mathbb{D}$, then $\Delta(C) = \nabla(C) = \emptyset$, which requires by (CM1) that $\mathrm{br}(\boldsymbol{x}_C) = \emptyset$, and by (CM3) that $\mathrm{at}(\mu) = \emptyset$. There is a unique algebraic measure tree without branch points and atoms, namely the line segment with no atoms (see Example 4.11). We may, therefore, assume without loss of generality that $C \neq \mathbb{D}$ and, consequently, that $T_C \neq \emptyset$.

We claim that (V_C, c_V) together with $\psi_y(x) \coloneqq \lambda_{\mathbb{S}}\big(\mathbb{S} \cap \mathrm{comp}_x(y)\big)$ satisfies the assumptions of Proposition 3.13. Indeed, V_C is obviously countable and an algebraic tree by Lemma 4.7, $\psi_y(x)$ depends on y only through its equivalence class with respect to $\sim_x$, and the lengths of all the arcs add up to the total length of $\lambda_{\mathbb{S}}(\mathbb{S}) = 1$. Furthermore, $\psi_x(y) + \psi_y(x) \geq \lambda_{\mathbb{S}}(\mathbb{S}) = 1$, and Proposition 3.13 yields the existence and uniqueness of the desired algebraic measure tree.

(ii) Let $\boldsymbol{x} = (T, c, \mu) \in \mathbb{T}_2$. We construct a subtriangulation C such that $\tau(C) = \boldsymbol{x}$. Fix $\rho \in \mathrm{lf}(T,c)$ and recall that ρ induces a partial order relation $\leq_\rho$. We can extend this partial order to a total (planar) order $\leq$ by picking for every $v \in \mathrm{br}(T,c)$ an order of the two components of $T \setminus \{v\}$ that do not contain ρ. That is, we define $S_0(v) \coloneqq \mathcal{S}_v(\rho)$, denote the two remaining components of $T \setminus \{v\}$ by $S_1(v)$, $S_2(v)$, and define

$$v \leq w \quad :\Leftrightarrow \quad v \leq_\rho w \text{ or } v \in S_1\big(c(x,y,\rho)\big),\, w \in S_2\big(c(x,y,\rho)\big). \tag{91}$$

Identify $\mathbb{S}$ with $[0,1]$, where the endpoints are glued. For $a \in [0,1]$ and $b, c > 0$ with $a + b + c \leq 1$, let $\Delta(a,b,c) \subseteq \mathbb{D}$ be the open triangle with vertices $a, a+b$, $a+b+c \in \mathbb{S}$, $\ell(a,b) \subseteq \mathbb{D}$ the straight line from a to $a+b$, and $L(a,b)$ the

connected component of $\mathbb{D}\setminus\ell(a,b)$ containing $a+\frac{b}{2}\in\mathbb{S}$. The first vertex of the triangle or circular segment corresponding to $v\in \mathrm{br}(T,c)\cup \mathrm{lf}_{\mathrm{atom}}(\boldsymbol{x})$ is given by the total mass of points smaller than v (with respect to $\leq$ defined in (91)), i.e., by

$$\alpha(v) \coloneqq \mu\big(\{u\in T:\ u<v\}\big). \tag{92}$$

Define

$$\mathbb{D}\setminus C \coloneqq \biguplus_{v\in\mathrm{br}(T,c)} \Delta\big(\alpha(v),\mu(S_1(v)),\mu(S_2(v))\big) \uplus \biguplus_{v\in\mathrm{lf}_{\mathrm{atom}}(\boldsymbol{x})} L\big(\alpha(v),\mu\{v\}\big). \tag{93}$$

By definition of C, $\mathrm{conv}(C)\setminus C$ consists of open triangles, i.e., condition (Tri1) is satisfied. Furthermore, the extreme points of $\mathrm{conv}(C)$ are contained in $\mathbb{S}$, and for $x,y,z\in\Delta(C)\cup\nabla(C)$ distinct, there are corresponding $u,v,w\in T$ and a triangle $c_{xyz}\in\Delta(C)$ corresponding to $c(u,v,w)$, which satisfies the requirements of (Tri2)'. Thus, by Proposition 4.3 C is a subtriangulation of the circle. It is straightforward to check that $\tau(C)=\boldsymbol{x}$.

We defer the proof of continuity of τ to the next section, where we prove it in Lemma 5.21. □

The following is now obvious.

Lemma 4.9 (nonatomicity). — *A subtriangulation C of the circle is a triangulation of the circle if and only if, for $(T_C,c_C,\mu_C)\coloneqq\tau(C)$, the measure μ_C is nonatomic.*

Corollary 4.10 (finite tree approximation). — *Let $\boldsymbol{x}=(T,c,\mu)\in\mathbb{T}_2$. Then there is a sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ of finite algebraic measure trees in $\mathbb{T}_2$ with $\boldsymbol{x}_n\to\boldsymbol{x}$ bpdd-Gromov-weakly. Furthermore, if μ is nonatomic, then $\boldsymbol{x}_n$ can be chosen as a tree with n leaves and uniform distribution on the leaves.*

Proof. — By Theorem 4.8, there is a subtriangulation $C\in\mathcal{T}$ with $\tau(C)=\boldsymbol{x}$, and by Proposition 4.3, there are finite subtriangulations C_n with $C_n\to C$. Obviously, $\boldsymbol{x}_n\coloneqq\tau(C_n)$ is a finite algebraic measure tree and by continuity of τ we have $\boldsymbol{x}_n\to\boldsymbol{x}$. If μ is nonatomic, then, by Lemma 4.9, C is a triangulation of the circle, and hence, by Proposition 4.3, C_n can be chosen as triangulation of the n-gon, which means that $\boldsymbol{x}_n$ has n leaves and uniform distribution on them. □

We conclude this section with a few illustrative examples.

Example 4.11 (coding algebraic measure trees without branch points). — Let $\boldsymbol{x}$ be an algebraic measure tree without branch points. If $\boldsymbol{x}=\boldsymbol{x}_C$ for some subtriangulation C, then $\Delta(C)=\mathrm{br}(\boldsymbol{x}_C)=\emptyset$, and the following five cases can occur (see Figure 4.3): a) $\boldsymbol{x}_C$ consists of one single point of mass 1. Then $C=\{x\}$, for some $x\in\mathbb{S}$. b) $\boldsymbol{x}_C$ consists of an interval with two leaves, where

FIGURE 4.3. Subtriangulations of the circle which correspond to the five cases of algebraic measure trees without branch points as explained in Example 4.11.

each carries positive mass adding up to 1, in which case, C is a single line segment dividing the circle into two arcs with lengths corresponding to the masses of the two leaves. c) $\varkappa_C$ consists of an interval with two leaves, where each has positive mass adding up to $a < 1$. In this case, C is the area of the disc bounded by two distinct line segments and two arcs (possibly one of them degenerated) of $\mathbb{S}$, and the lengths of the remaining two arcs are given by the masses of the leaves. d) $\varkappa_C$ consists of an interval with two leaves, where one has positive mass $a < 1$, and the other one has zero mass. Then C is a circular segment with arc length $1 - a$. e) $\varkappa_C$ consists of an interval with no atoms on the leaves, which implies $C = \mathbb{D}$.

EXAMPLE 4.12 (a complete binary tree). — Let C be the subtriangulation of the circle drawn in the middle of Figure 4.1. Then $\#\nabla(C) = \#\operatorname{lf}_{\mathrm{atom}}(\tau(C)) = 1$. We refer to this only leaf with positive mass as the root ρ and obtain $\mu(\{\rho\}) = \frac{1}{3}$, corresponding to the length of the dotted arc. Moreover, $\tau(C)$ consists of a complete rooted binary tree in the sense of graph theory (with the convention that the root has degree 1), together with an uncountable set of leaves given by the ends at infinity and carrying the remaining $\frac{2}{3}$ of the mass.

EXAMPLE 4.13 (coding the Brownian CRT). — Recall the Brownian triangulation C_{CRT} from Example 4.5, which is defined as the limit in distribution of uniform random triangulations C_n of the n-gon. A realization is shown on the right-hand side of Figure 4.1. It is easy to see that $\tau(C_n)$ is the uniform binary tree with n leaves and uniform distribution on the leaves. Thus, by Theorem 4.8, the uniform binary tree converges bpdd-Gromov-weakly to $\tau(C_{\mathrm{CRT}})$. At this point, it is not entirely clear that $\tau(C_{\mathrm{CRT}})$ is the algebraic measure tree induced by the metric measure Brownian CRT. We will see in Section 6 that this is, indeed, the case.

5. The subspace of binary algebraic measure trees

In Sections 5.1 and 5.2 with the *sample shape convergence* and the *sample subtree-mass convergence*, we introduce two more notions of convergence of

algebraic measure trees, which seem more natural when thinking of algebraic trees as combinatorial objects. We then show in Section 5.3 that on $\mathbb{T}_2$, both of these notions are equivalent to the bpdd-Gromov-weak convergence. The main tools are a uniform Glivenko–Cantelli argument and that the coding map sending a subtriangulation of the circle to an element in $\mathbb{T}_2$ is continuous.

5.1. Convergence in distribution of sampled tree shapes. — The basic idea behind Gromov-weak convergence for metric measure spaces is to sample finite metric subspaces with the sampling measure μ and then require these to converge in distribution. In this section, we propose a corresponding construction for binary algebraic measure trees, where we sample finite tree shapes with μ.

First, we have to make precise what we mean by "tree shape", which we understand to be a cladogram with the peculiarity that leaves may carry more than one label. The multilabel case is necessary to allow for sampling the same point several times due to a possible atom at that point.

Definition 5.1 (m-labelled cladogram). — For $m \in \mathbb{N}$, an *m-labelled cladogram* is a binary, finite algebraic tree $C = (C, c)$ together with a surjective labeling map $\ell\colon \{1, \dots, m\} \to \mathrm{lf}(C)$. Two m-labelled cladograms (C_1, ℓ_1) and (C_2, ℓ_2) are equivalent if they are label-preserving isomorphic, i.e., there exists a tree isomorphism $f\colon C_1 \to C_2$ with $f(\ell_1(i)) = \ell_2(i)$ for all $i = 1, \dots, m$.

Define

$$\mathfrak{C}_m \coloneqq \{\text{isomorphism classes of } m\text{-labelled cladograms}\}. \tag{94}$$

In the following, we will use cladograms to encode the shape of a subtree spanned by a finite sample of leaves.

Definition 5.2 (tree shape). — For a binary algebraic tree (T, c), $m \in \mathbb{N}$, and $u_1, \dots, u_m \in T \setminus \mathrm{br}(T, c)$, the *tree shape* $\mathfrak{s}_T(u_1, \dots, u_m)$ of the m-labeled cladogram spanned by $(u_1, \dots, u_m)$ in (T, c) is the unique (up to isomorphism) m-labelled cladogram $\mathfrak{s}_T(u_1, \dots, u_m) = (C, c_C, \ell)$ with $\mathrm{lf}(C) = \{u_1, \dots, u_m\}$ and $\ell(i) = u_i$, for all $i = 1, \dots, m$, and such that the identity on $\mathrm{lf}(C)$ extends to a tree homomorphism from C onto $c\big(\{u_1, \dots, u_m\}^3\big)$.

Remark 5.3 (spanned subtree and cladogram are not necessarily isomorphic). — The tree homomorphism from $\mathfrak{s}_T(u_1, \dots, u_m)$ onto $c(\{u_1, \dots, u_m\}^3)$ does not need to be injective. This is the case if (and only if) $u_i \in (u_j, u_k)$, for some $i, j, k \in \{1, \dots, m\}$. See Figure 5.1.

Example 5.4 (shape of a totally ordered algebraic tree). — Let (T, c) be a totally ordered algebraic tree, and $u_1, \dots, u_m \in T$. Then $\mathfrak{s}_T(u_1, \dots, u_m)$ is a so-called *comb tree*, which has a totally ordered spine of binary branch points with attached leaves (see Figure 5.2).

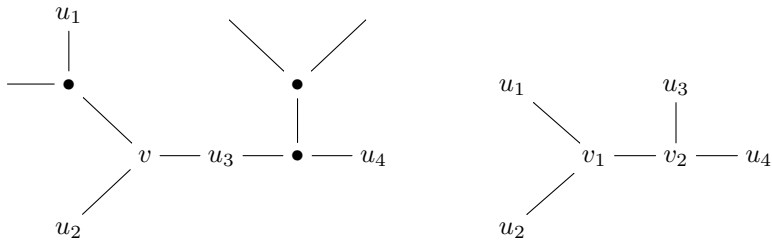

FIGURE 5.1. A tree T and the shape $\mathfrak{s}_T(u_1, u_2, u_3, u_4)$. Here, we are considering the homomorphism $f\colon C \to c^3(\{u_1, \dots, u_4\}^3)$ given by $f(u_i) := u_i$, $i = 1, \dots, 4$ and then necessarily $f(v_1) = v$, $f(v_2) = u_3$; f is clearly no isomorphism, and the cladogram is not isomorphic to the subtree $c(\{u_1, u_2, u_3, u_4\}^3)$, because $c(u_1, u_4, u_3) = u_3$.

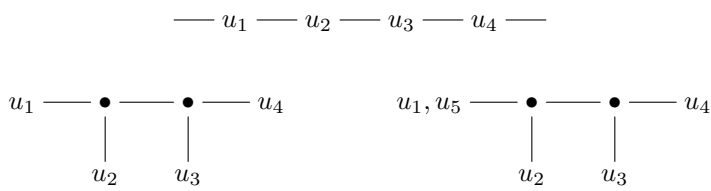

FIGURE 5.2. The upper graph shows a totally ordered binary algebraic tree and four distinct points $u_1, \dots, u_4$. The lower left shows the shape $\mathfrak{s}_T(u_1, \dots, u_4)$ of the cladogram which forms a comb tree. The lower right illustrates what happens if a fifth point is equal to u_1. Now, one of the leaves of $\mathfrak{s}_T(u_1, \dots, u_5)$ has two labels.

In the following, we build a topology on the convergence of tree shapes of m randomly sampled points. We, therefore, need the measurability of the shape map.

LEMMA 5.5 (measurability of the shape map). — *For every binary algebraic tree (T, c) and $m \in \mathbb{N}$, the tree shape map $\mathfrak{s}_T\colon (T \setminus \mathrm{br}(T, c))^m \to \mathfrak{C}_m$ is a measurable function.*

Proof. — Restricted to the open subset $\big\{v \in (T \setminus \mathrm{br}(T, c))^m : v_1, \dots, v_m \text{ distinct}\big\}$, $\mathfrak{s}_T$ is locally constant and, hence, continuous. The same is true on the set $\big\{v \in (T \setminus \mathrm{br}(T, c))^m : v_1 = v_2,\ v_2, \dots, v_m \text{ distinct}\big\}$, which is an intersection of a closed and an open set and, hence, measurable. We can continue this way to see that $\mathfrak{s}_T$ is measurable on $(T \setminus \mathrm{br}(T, c))^m$. □

DEFINITION 5.6 (tree shape distribution). — For $\boldsymbol{x} = (T, c, \mu) \in \mathbb{T}_2$ and $m \in \mathbb{N}$, the *m-tree shape distribution* of $\boldsymbol{x}$ is defined by

$$\mathfrak{S}_m(\boldsymbol{x}) \coloneqq \mu^{\otimes m} \circ \mathfrak{s}_T^{-1} \in \mathcal{M}_1(\mathfrak{C}_m). \tag{95}$$

EXAMPLE 5.7 (shape of the linear nonatomic measure tree). — Let $\boldsymbol{x} = (T, c, \mu)$ be the linear nonatomic algebraic measure tree (Example 3.4). Then any sample $(u_1, \dots, u_m)$ with μ consists of pairwise different points, and $\mathfrak{S}_m(\boldsymbol{x})$ is the mixture of Dirac measures on labeled comb trees where the mixture is over all (up to isometry) permutations of the labels.

We refer to the weakest topology on $\mathbb{T}_2$, such that for every $m \in \mathbb{N}$, the m-tree shape distribution is continuous as the sample shape topology.

DEFINITION 5.8 (sample shape topology). — The topology induced on $\mathbb{T}_2$ by the set $\{\mathfrak{S}_m : m \in \mathbb{N}\}$ of tree shape distributions is called the *sample shape topology.*

We say that a sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ is *sample shape convergent* to $\boldsymbol{x}$ in $\mathbb{T}_2$ if it converges with respect to the sample shape topology, i.e., if $\mathfrak{S}_m(\boldsymbol{x}_n)$ converges to $\mathfrak{S}_m(\boldsymbol{x})$ as $n \to \infty$ for every $m \in \mathbb{N}$.

In analogy to the set Π_ι of polynomials introduced in Remark 3.10, we also introduce a set of test functions that evaluate the tree shape distributions. We refer to $\Phi = \Phi^{m,\varphi} \colon \mathbb{T}_2 \to \mathbb{R}$,

$$\Phi(\boldsymbol{x}) = \int_{\mathfrak{C}_m} \varphi \, \mathrm{d}\mathfrak{S}_m(\boldsymbol{x}) = \int_{T^m} \varphi \circ \mathfrak{s}_T \, \mathrm{d}\mu^{\otimes m}, \tag{96}$$

where $m \in \mathbb{N}$ and $\varphi \colon \mathfrak{C}_m \to \mathbb{R}$, as *shape polynomial.* We also define

$$\Pi_{\mathfrak{s}} \coloneqq \{\, \text{shape polynomials on } \mathbb{T}_2 \,\}. \tag{97}$$

Obviously, the sample shape topology is induced by the set $\Pi_{\mathfrak{s}}$ of shape polynomials.

PROPOSITION 5.9 (sample shape implies bpdd-Gromov-weak convergence). — *On $\mathbb{T}_2$, the sample shape topology is stronger than the bpdd-Gromov-weak topology (i.e., any open set in the bpdd-Gromov-weak topology is open in the sample shape topology).*

Proof. — The bpdd-Gromov-weak topology is induced by the set Π_ι of polynomials (see Remark 3.10). Because the set of $\phi \in \mathcal{C}_b(\mathbb{R}^{m\times m})$ that are Lipschitz continuous is convergence determining for probability measures on $\mathbb{R}^{m\times m}$, the subset of those $\Psi \in \Pi_\iota$ with

$$\Psi(T, c, \mu) = \int_{T^m} \phi\big((\nu[u_i, u_j] - \tfrac{1}{2}\nu\{u_i\} - \tfrac{1}{2}\nu\{u_j\})_{i,j=1,\dots,m}\big) \, \mu^{\otimes m}(\mathrm{d}\underline{u}) \tag{98}$$

for some $m \in \mathbb{N}$ and Lipschitz continuous $\phi \in \mathcal{C}_b(\mathbb{R}^{m\times m})$ also induces the bpdd-Gromov-weak topology. Therefore, it is enough to show that such a Ψ

is continuous on $\mathbb{T}_2$ with respect to the sample shape topology. We do so by showing that the restriction to $\mathbb{T}_2$ of Ψ is in the uniform closure of $\Pi_{\mathfrak{s}}$. Let L be the Lipschitz constant of ϕ with respect to the ℓ_∞-norm on $\mathbb{R}^{m\times m}$.

For $n \in \mathbb{N}$ with $3n \geq m$, we define

(99)
$$\Phi_n(T,c,\mu) \coloneqq \int_{T^{3n}} \phi\big((\nu_{n,\underline{u}}[u_i,u_j] - \tfrac{1}{2}\nu_{n,\underline{u}}\{u_i\} - \tfrac{1}{2}\nu_{n,\underline{u}}\{u_j\})_{i,j=1,\dots,m}\big)\, \mu^{\otimes 3n}(\mathrm{d}\underline{u}),$$

with the empirical branch-point distribution

$$\nu_{n,\underline{u}} \coloneqq \tfrac{1}{n}\sum_{k=0}^{n-1} \delta_{c(u_{3k+1},u_{3k+2},u_{3k+3})}. \tag{100}$$

Note that the restriction of Φ_n to $\mathbb{T}_2$ belongs to $\Pi_{\mathfrak{s}}$ because whether or not $c(u_{k+1},u_{k+2},u_{k+3})$ lies on $[u_i,u_j]$, $k \in \{0,\dots,n-1\}$, $i,j \in \{1,\dots,m\}$ only depends on the shape $\mathfrak{s}_{3n}(\underline{u})$.

Finally, we observe

$$\begin{aligned} \|\Psi - \Phi_n\|_\infty &\leq \sup_{(T,c,\mu)\in\mathbb{T}_2} \int_{T^{3n}} L\cdot 3 \sup_{I\in\mathcal{I}_T} |\nu(I) - \nu_{n,\underline{u}}(I)|\, \mu^{\otimes 3n}(\mathrm{d}\underline{u}) \\ &\leq 3L\cdot\epsilon_n \xrightarrow[n\to\infty]{} 0, \end{aligned} \tag{101}$$

with $\mathcal{I}_T \coloneqq \{[x,y];\, x,y \in T\}$ and $(\epsilon_n)_{n\in\mathbb{N}} \xrightarrow[n\to\infty]{} 0$, where we have used a uniform Glivenko–Cantelli estimate that is an upper bound of the distance of the empirical branch-point distribution to the branch-point distribution. Such an estimate should be known, but as we could not come up with a reference, we show it in Lemma A.4 in the Appendix. We note that $\dim_{\mathrm{VC}}(\mathcal{I}_T) = 2$ (compare Example A.2). □

Corollary 5.10 (metrizability). — *The sample shape topology is metrizable.*

Proof. — Because the sample shape topology is induced by a countable family of functions $(\mathfrak{S}_m)_{m\in\mathbb{N}}$ with values in metrizable spaces, it is pseudo-metrizable. By Proposition 5.9, it is stronger than the bpdd-Gromov-weak topology and, hence, a Hausdorff topology. Therefore, it is metrizable. □

5.2. Convergence in distribution of sampled subtree masses. — In this section, we introduce yet another notion of convergence of algebraic measure trees, which, in contrast to sampling tree shapes, is based on sampling branch points and evaluating the masses of the subtrees that are joined at these branch points. This approach might be more similar to the case of metric measure spaces and distance matrix distributions, because we sample a tensor of real numbers (masses of subtrees) as opposed to a combinatorial object (tree shape). Thus, the typical tools of analysis are more readily applicable to the corresponding class of test functions.

Let $(T,c,\mu)\in\mathbb{T}_2$ and recall from (36) for $u,v,w\in T$ the subtree component $\mathcal{S}_{c(u,v,w)}(x)$ of $T\setminus\{c(u,v,w)\}$, which contains $x\neq c(u,v,w)$. Here, we always take the component containing $x=u$, and consider its mass

$$\eta(u,v,w)\coloneqq \mathbb{1}_{u\neq c(u,v,w)}\cdot\mu\big(\mathcal{S}_{c(u,v,w)}(u)\big). \tag{102}$$

Lemma 5.11 (measurability of the subtree masses). — *For every binary algebraic measure tree $\boldsymbol{x}=(T,c,\mu)\in\mathbb{T}_2$ and $m\in\mathbb{N}$, the function $\eta\colon T^3\to[0,1]$ is measurable.*

Proof. — First, we claim that the map $\psi\colon T^2\to[0,1]$,

$$\psi(u,v)\coloneqq \mathbb{1}_{u\neq v}\cdot\mu\big(\mathcal{S}_v(u)\big) \tag{103}$$

is lower semicontinuous. Indeed, let (u_n,v_n) be a sequence converging to (u,v). We may assume without loss of generality that $v\neq u$, $u_n\in\mathcal{S}_v(u)$, and either $v_n\notin\mathcal{S}_v(u)$ for all $n\in\mathbb{N}$, or $v_n\in\mathcal{S}_v(u)$ for all $n\in\mathbb{N}$. In the first case, $\mathcal{S}_v(u)\subseteq\mathcal{S}_{v_n}(u_n)$ and, hence, $\psi(u,v)\le\psi(u_n,v_n)$. In the second case, for every $x\in\mathcal{S}_v(u)$ and $n\ge n_x$ sufficiently large, we have $u\in\mathcal{S}_{v_n}(u_n)$ and $v_n\notin[x,u]$. This means $x\in\mathcal{S}_{v_n}(u)=\mathcal{S}_{v_n}(u_n)$ and, hence,

$$\psi(u,v)-\liminf_{n\to\infty}\psi(u_n,v_n)\le\lim_{n\to\infty}\mu\big(\mathcal{S}_v(u)\setminus\mathcal{S}_{v_n}(u_n)\big)=0. \tag{104}$$

Therefore, ψ is lower semicontinuous. Because the branch-point map c is continuous due to Lemma 2.17, the same applies to $\eta(u,v,w)=\psi((u,\,c(u,v,w))$, and η is measurable. □

Given a vector $\underline{u}=(u_1,\dots,u_m)\in T^m$, $m\in\mathbb{N}$, we consider the masses of all the subtrees we obtain as branch points of entries of $\underline{u}$. To this end, let

$$\underline{\eta}(u,v,w)\coloneqq\big(\eta(u,v,w),\,\eta(v,u,w),\,\eta(w,u,v)\big) \tag{105}$$

and define the function $\mathfrak{m}_x\colon T^m\to[0,1]^{3\cdot\binom{m}{3}}$, given by

$$\mathfrak{m}_x(\underline{u})\coloneqq\big(\underline{\eta}(u_i,u_j,u_k)\big)_{1\le i<j<k\le m}. \tag{106}$$

Definition 5.12 (subtree-mass tensor distribution). — For $\boldsymbol{x}=(T,c,\mu)\in\mathbb{T}_2$ and $m\in\mathbb{N}$, the *m-subtree-mass tensor distribution* of $\boldsymbol{x}$ is defined by

$$\vartheta_m(\boldsymbol{x})\coloneqq\mu^{\otimes m}\circ\mathfrak{m}_x^{-1}\in\mathcal{M}_1\big([0,1]^{3\cdot\binom{m}{3}}\big), \tag{107}$$

Example 5.13 (symmetric binary tree). — Let for each $n\in\mathbb{N}$, $\boldsymbol{x}_n=(T_n,c_n,\mu_n)$ be the symmetric binary tree with $N=2^n$ leaves and the uniform distribution

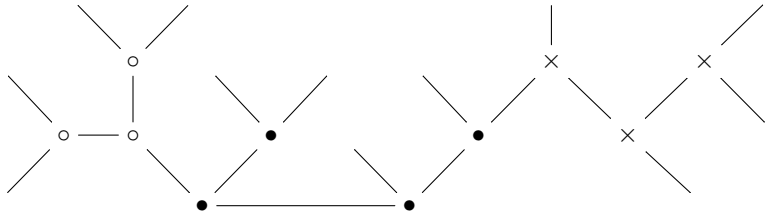

FIGURE 5.3. μ is the uniform distribution on the leaves. Swap the $\circ$-part with the $\times$-part to obtain a nonisomorphic tree giving the same value for ϑ_3.

on the set of leaves. Then the 3-*subtree-mass tensor distribution* of $\boldsymbol{x}_n$ is equal to

(108)

$$\begin{aligned}
&\vartheta_3(\boldsymbol{x}_n) = \mu_n^{\otimes 3} \circ \mathfrak{m}_x^{-1} \\
&= \sum_{k=1}^{n-1} \frac{1-2^{-k}}{2^{k+1}} \Big(\delta_{(\frac{1}{2^{k+1}}, \frac{1}{2^{k+1}}, 1-\frac{1}{2^k})} + \delta_{(\frac{1}{2^{k+1}}, 1-\frac{1}{2^k}, \frac{1}{2^{k+1}})} + \delta_{(1-\frac{1}{2^k}, \frac{1}{2^{k+1}}, \frac{1}{2^{k+1}})} \Big) \\
&\quad + \frac{1}{N}(1-\frac{1}{N}) \big(\delta_{(\frac{1}{N}, \frac{1}{N}, 1)} + \delta_{(\frac{1}{N}, 1, \frac{1}{N})} + \delta_{(1, \frac{1}{N}, \frac{1}{N})} \big) + \frac{1}{N^2} \delta_{(\frac{1}{N}, \frac{1}{N}, \frac{1}{N})} \\
&\xrightarrow[n\to\infty]{} \sum_{k=1}^{\infty} \frac{1-2^{-k}}{2^{k+1}} \Big(\delta_{(\frac{1}{2^{k+1}}, \frac{1}{2^{k+1}}, 1-\frac{1}{2^k})} + \delta_{(\frac{1}{2^{k+1}}, 1-\frac{1}{2^k}, \frac{1}{2^{k+1}})} + \delta_{(1-\frac{1}{2^k}, \frac{1}{2^{k+1}}, \frac{1}{2^{k+1}})} \Big).
\end{aligned}$$

REMARK 5.14 (3-subtree-mass tensor distribution is not enough). — It is not enough to consider only the 3-*subtree-mass tensor distribution*. Indeed, ϑ_3 cannot distinguish all nonisomorphic binary algebraic measure trees, i.e., it does not separate the points of $\mathbb{T}_2$. To see this, take the tree from Figure 5.3 with uniform distribution on its 12 leaves, and the same tree with the subtrees marked by $\times$ and $\circ$, respectively, exchanged. These two trees are clearly nonisomorphic, and because the two marked subtrees have the same number of leaves, every vertex in one tree corresponds to a vertex in the other with the same value for $\mathfrak{m}_x$.

We consider the weakest topology on $\mathbb{T}_2$ such that for every $m \in \mathbb{N}$ the m-subtree-mass tensor distribution is continuous. Here, as usual, we equip $\mathcal{M}_1([0,1]^{3\cdot\binom{m}{3}})$ with the weak topology.

DEFINITION 5.15 (sample subtree-mass topology). — The topology induced on $\mathbb{T}_2$ by the set $\{\vartheta_m : m \in \mathbb{N}\}$ of subtree-mass tensor distributions is called *sample subtree-mass topology*.

We say that a sequence $(\mathcal{x}_n)_{n\in\mathbb{N}}$ is *sample subtree-mass convergent* to $\mathcal{x}$ in $\mathbb{T}_2$ if it converges with respect to the sample subtree-mass topology, i.e., if $\vartheta_m(\mathcal{x}_n)$ converges to $\vartheta_m(\mathcal{x})$ as $n\to\infty$ for every $m\in\mathbb{N}$. To see that the sample subtree-mass topology is a Hausdorff topology on $\mathbb{T}_2$, we need the following reconstruction theorem.

PROPOSITION 5.16 (reconstruction theorem). — *The set of subtree-mass tensor distributions $\{\vartheta_m : m\in\mathbb{N}\}$ separates points of $\mathbb{T}_2$, i.e., if $\mathcal{x}_1,\mathcal{x}_2\in\mathbb{T}_2$ are such that $\vartheta_m(\mathcal{x}_1)=\vartheta_m(\mathcal{x}_2)$ for all $m\in\mathbb{N}$, then $\mathcal{x}_1=\mathcal{x}_2$.*

Proof. — We always assume that the representative (T,c,μ) of an algebraic measure tree is chosen such that $\mu(\mathcal{S}_v(u))>0$, whenever $u,v\in T$, $u\neq v$.

Because the set $\{\mathfrak{S}_m : m\in\mathbb{N}\}$ of tree shape distributions separates points by Corollary 5.10, it is enough to show that $\mathfrak{S}_m$ is determined by the m-subtree-mass tensor distribution ϑ_m for every $m\in\mathbb{N}$. We do so by showing that there exists a (noncontinuous) function $h\colon [0,1]^{3\cdot\binom{m}{3}}\to\mathfrak{C}_m$, such that for every $\mathcal{x}=(T,c,\mu)\in\mathbb{T}_2$, we have $\mathfrak{s}_T = h\circ\mathfrak{m}_{\mathcal{x}}$ on $\big(T\setminus \mathrm{br}(T,c)\big)^m$. This is enough, because $\mu(\mathrm{br}(T,c))=0$ by countability of $\mathrm{br}(T,c)$ and the assumption that $\mathrm{at}(\mu)\subseteq \mathrm{lf}(T,c)$.

Fix $\underline{u}=(u_1,\dots,u_m)\in(T\setminus\mathrm{br}(T,c))^m$ and set $C=(C,c_C,\ell)\coloneqq\mathfrak{s}_T(\underline{u})$. For $i\neq j$, we have $u_i=u_j$ if and only if $\eta(u_i,u_j,u_k)=\eta(u_j,u_i,u_k)=0$ for any and, hence, all $k\in\{1,\dots,m\}\setminus\{i,j\}$. Thus, we can determine multiple labels of C by $\mathfrak{m}_{\mathcal{x}}(\underline{u})$ and may assume in the following that $u_1,\dots,u_m$ are distinct. Then, the m-labelled cladogram C is uniquely determined by the set of pairs $(\underline{x}_1,\underline{x}_2)$ of triples $\underline{x}_i=(x_{i,1},x_{i,2},x_{i,3})\in\{u_1,\dots,u_m\}^3$, $x_{i,j}\neq x_{i,k}$ for $j\neq k$, $i=1,2$, such that

$$c_C\big(x_{1,1},\,x_{1,2},\,x_{1,3}\big)=c_C\big(x_{2,1},\,x_{2,2},\,x_{2,3}\big). \tag{109}$$

We claim that (109) holds if and only if we can reorder the three entries of $\underline{x}_2$, such that we can replace every entry of $\underline{x}_1$ by the corresponding entry of $\underline{x}_2$ and obtain the same masses of subtrees. More precisely,

$$\underline{\eta}(x_{1,1},x_{1,2},x_{1,3})=\underline{\eta}(x_{i,1},x_{j,2},x_{k,3})\quad\forall i,j,k\in\{1,2\}. \tag{110}$$

Indeed, if $c_C(\underline{x}_1)=c_C(\underline{x}_2)$, then $c(\underline{x}_1)=c(\underline{x}_2)$ by definition of $\mathfrak{s}_T$. Because none of the u_i is a branch point, every component of $T\setminus\{c(\underline{x}_1)\}$ contains precisely one of the $x_{1,i}$, as well as one of the $x_{2,i}$. We can reorder the entries of x_2, such that $x_{1,i}$ is in the same component as $x_{2,i}$, $i=1,\dots,3$. Then it is easy to check that (110) holds.

Conversely, assume that $c_C(\underline{x}_1)\neq c_C(\underline{x}_2)$. Because the restriction of the tree homomorphism $C\to c(\{u_1,\dots,u_m\}^3)$ to the branch points of C is injective, this implies $v_1\coloneqq c(\underline{x}_1)\neq c(\underline{x}_2)=:v_2$. There must be an i with $x_{1,i}\in\mathcal{S}_{v_1}(v_2)$, say $i=3$. Also, $x_{2,j}\in\mathcal{S}_{v_1}(v_2)$ for at least two different j, so at least one which is different from i, say $j=2$ (see Figure 5.4). Then $v_3\coloneqq$

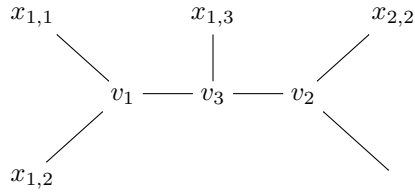

FIGURE 5.4. The situation in the proof of Proposition 5.16.

$c(x_{1,1}, x_{2,2}, x_{1,3}) \in \mathcal{S}_{v_1}(v_2)$, and, in particular, $x_{1,1}, x_{1,2} \in \mathcal{S}_{v_3}(x_{1,1})$. Thus $\eta(\underline{x}_1) < \eta(x_{1,1}, x_{2,2}, x_{1,3})$, and (110) does not hold. □

COROLLARY 5.17 (metrizability). — *The sample subtree-mass topology is metrizable.*

Proof. — Because the sample subtree-mass topology is induced by a countable family of functions $(\vartheta_m)_{m\in\mathbb{N}}$ with values in metrizable spaces, it is pseudo-metrizable. By Proposition 5.16, it is a Hausdorff topology and, hence, it is metrizable. □

In analogy to the sets Π_ι and $\Pi_\mathfrak{s}$ of polynomials and shape polynomials, respectively, the sample subtree-mass topology also comes with a canonical set of test functions. We call $\Psi\colon \mathbb{T}_2 \to \mathbb{R}$ *subtree-mass polynomial* if there is $m \in \mathbb{N}$, and $\psi \in \mathcal{C}_b([0,1]^{3\cdot\binom{m}{3}})$ with

$$\Psi(\boldsymbol{x}) = \int_{[0,1]^{3\cdot\binom{m}{3}}} \psi \, \mathrm{d}\vartheta_m(\boldsymbol{x}) = \int_{T^m} \psi \circ \mathfrak{m}_x \, \mathrm{d}\mu^{\otimes m}. \tag{111}$$

We also define

$$\Pi_\mathfrak{m} \coloneqq \{\, \text{subtree-mass polynomials on } \mathbb{T}_2 \,\}. \tag{112}$$

Obviously, the sample subtree-mass topology is induced by the set $\Pi_\mathfrak{m}$ of subtree-mass polynomials.

PROPOSITION 5.18 (sample shape convergence implies sample subtree-mass convergence). — *The sample shape topology is stronger than the sample subtree-mass topology.*

Proof. — The proof is similar to that of Proposition 5.9. We will show that each subtree-mass polynomial in $\Psi \in \Pi_\mathfrak{m}$,

$$\Psi(T, c, \mu) = \int_{T^m} \psi\big(\big(\underline{\eta}(u_i, u_j, u_k)\big)_{1\le i<j<k\le m}\big) \, \mu^{\otimes m}(\mathrm{d}\underline{u}), \tag{113}$$

with $m \in \mathbb{N}$ and $\psi \in \mathcal{C}([0,1]^{3\cdot\binom{m}{3}})$ Lipschitz continuous with respect to the ℓ_∞-norm on $[0,1]^{3\cdot\binom{m}{3}}$ is in the uniform closure of $\Pi_\mathfrak{s}$. Let L be the Lipschitz

constant of Ψ. For $n \in \mathbb{N}$ with $n \geq m$, we define

$$\Phi_n(T,c,\mu) \coloneqq \int_{T^n} \psi\big(\big(\underline{\eta}^{\mu_{n,\underline{u}}}(u_i,u_j,u_k)\big)_{1\leq i<j<k\leq m}\big)\,\mu^{\otimes n}(\mathrm{d}\underline{u}), \tag{114}$$

where $\underline{\eta}^{\mu_{n,\underline{u}}}$ is defined in the same way as $\underline{\eta}$ but with μ replaced by the empirical sample distribution

$$\mu_{n,\underline{u}} \coloneqq \tfrac{1}{n}\sum_{\ell=1}^{n} \delta_{u_\ell}. \tag{115}$$

Note that $\Phi_n \in \Pi_{\mathfrak{s}}$, because whether or not $u_\ell \in \mathcal{S}_{c(u_i,u_j,u_k)}(u_i)$ for some $\ell \in \{1,\dots,n\}$, $i,j,k \in \{1,\dots,m\}$ depends only on the shape $\mathfrak{s}_T(\underline{u})$.

Finally, applying the uniform Glivenko–Cantelli estimate Lemma A.4, we have

$$\begin{aligned}\|\Psi - \Phi_n\|_\infty &\leq \sup_{(T,c,\mu)\in\mathbb{T}_2} \int_{T^n} L \cdot \sup_{S\in\mathcal{S}_T} \big|\mu(S) - \mu_{n,\underline{u}}(S)\big|\;\mu^{\otimes n}(\mathrm{d}\underline{u}) \\ &\leq L\epsilon_n \xrightarrow[n\to\infty]{} 0,\end{aligned} \tag{116}$$

where $\mathcal{S}_T \coloneqq \big\{\mathcal{S}_v(u) : u,v \in T\big\}$ and $(\epsilon_n)_{n\in\mathbb{N}} \xrightarrow[n\to\infty]{} 0$. We note that $\dim_{\mathrm{VC}}(\mathcal{S}_T) \leq 3$ (compare Example A.3). □

5.3. Equivalence and compactness of topologies. — In this section, we show that sample shape convergence (Definition 5.8), sample subtree-mass convergence (Definition 5.15), and branch-point distribution distance Gromov-weak convergence (Definition 3.7) on $\mathbb{T}_2$ are equivalent. While spaces of metric measure spaces are usually far from being locally compact, $\mathbb{T}_2$ is in this topology even a compact metrizable space.

THEOREM 5.19 (equivalence of topologies and compactness). — *The sample shape topology, the sample subtree-mass topology, and the bpdd-Gromov-weak topology coincide on $\mathbb{T}_2$. Furthermore, $\mathbb{T}_2$ is compact and metrizable in this topology.*

Because compact subsets of a Hausdorff space are closed, a direct corollary is that unlike the situation in the space of metric measure trees (with Gromov-weak or Gromov-Hausdorff-weak topology), the set of binary trees is closed with respect to the bpdd-Gromov-weak topology. In particular, Gromov(-Hausdorff)-weak convergence does not imply bpdd-Gromov-weak convergence of the induced trees.

COROLLARY 5.20. — *The subspace $\mathbb{T}_2$ of binary algebraic measure trees with atoms restricted to leaves is closed in $\mathbb{T}$ (with bpdd-Gromov-weak topology).*

As a preparation of the proof for the theorem, we show that binary algebraic measure trees continuously depend on their encoding as subtriangulations of

the circle. Together with Proposition 5.9, this also finishes the proof of Theorem 4.8. Recall the space $\mathcal{T}$ of subtriangulations of the circle equipped with the Hausdorff metric topology from (84) and the coding map $\tau\colon \mathcal{T} \to \mathbb{T}_2$ from Theorem 4.8.

LEMMA 5.21 (continuity of the coding map). — *Let $\mathbb{T}_2$ be equipped with the sample shape topology and $\mathcal{T}$ with the Hausdorff metric topology. Then the coding map $\tau\colon \mathcal{T} \to \mathbb{T}_2$ is continuous.*

Proof. — Fix $C \in \mathcal{T}$ and $m \in \mathbb{N}$. By definition of the sample shape topology, it is enough to show that $\mathfrak{S}_m \circ \tau\colon \mathcal{T} \to \mathcal{M}_1(\mathfrak{C}_m)$ is continuous at C. Let $U_1, \dots, U_m$ be independent, identically distributed points on the circle $\mathbb{S}$ chosen with the Lebesgue measure.

Recall from (83) the set $\nabla(C)$ of connected components of $\mathbb{D} \setminus \mathrm{conv}(C)$ and from (88) the connected component $\mathrm{comp}_x(y)$ of $\mathbb{D} \setminus \partial_{\mathbb{D}} x$ that contains y, where $x \in \Delta(C) \cup \nabla(C)$, and $y \subseteq \mathbb{D}$ connected and disjoint from $\partial_{\mathbb{D}} x$. Furthermore, recall the set $\square(C)$ from (89) and the subtree components $\mathcal{S}_x(y)$ from (9).

For $\epsilon > 0$, there exists $N = N_{C,m,\epsilon} \in \mathbb{N}$ and $v_1, \dots, v_N \in \Delta(C) \cup \nabla(C)$ distinct, such that with probability at least $1 - \epsilon$ the following holds:

- if $\{U_1, \dots, U_m\} \cap v \neq \emptyset$ for $v \in \nabla(C)$, then $v \in \{v_1, \dots, v_N\}$, and
- if $\{U_1, \dots, U_m\} \cap \mathrm{comp}_v(w) \neq \emptyset$ for some $v \in \Delta(C)$ and all $w \in \Delta(C) \cup \nabla(C) \cup \square(C)$ with $w \neq v$, then $v \in \{v_1, \dots, v_N\}$.

Put $\epsilon' \coloneqq \epsilon \cdot (12mN)^{-1}$. Then

$$\mathbb{P}\big(\big\{d(U_i, \partial v_j) \geq \epsilon',\, \forall\, i = 1, \dots, m; j = 1, \dots, N\big\}\big) \geq 1 - \epsilon. \tag{117}$$

There is a $\delta = \delta(\epsilon) > 0$ sufficiently small, such that for any $C' \in \mathcal{T}$ with Hausdorff metric $d_{\mathrm{H}}(C, C') < \delta$, there are distinct $v_1', \dots, v_N' \in \Delta(C') \cup \nabla(C')$, such that $d_{\mathrm{H}}(v_i, v_i') \leq \epsilon'$ for $i = 1, \dots, N$. Let $\boldsymbol{x} = (T, c, \mu) \coloneqq \tau(C)$, and $V_1, \dots, V_m$ be independent, identically distributed, μ-distributed, coupled to $U_1, \dots, U_m$, such that $V_k \in \mathcal{S}_v(w)$ if and only if $U_k \in \mathrm{comp}_v(w)$, which is possible due to the properties of τ established in Theorem 4.8. Define $\boldsymbol{x}'$ and $V_1', \dots, V_m'$ similarly with C' instead of C. Then

$$\mathbb{P}\big(\big\{\mathfrak{s}_T(V_1, \dots, V_m) = \mathfrak{s}_{T'}(V_1', \dots, V_m')\big\}\big) \geq 1 - 2\epsilon, \tag{118}$$

which implies that $d_{\mathrm{Pr}}\big(\mathfrak{S}_m(\tau(C)),\, \mathfrak{S}_m(\tau(C'))\big) \leq 2\epsilon$ (with d_{Pr} denoting the Prokhorov distance). This shows that $\mathfrak{S}_m \circ \tau$ is continuous at C and, since m and C are arbitrary, that τ is continuous. □

Now we are in a position to combine our results to a proof of the main theorem of Section 5.

Proof of Theorem 5.19. — The space $\mathcal{T}$ of subtriangulations of the circle with Hausdorff metric topology is compact according to Lemma 4.2. The coding map $\tau\colon \mathcal{T} \to \mathbb{T}_2$ is surjective by Theorem 4.8, and continuous when $\mathbb{T}_2$ is equipped

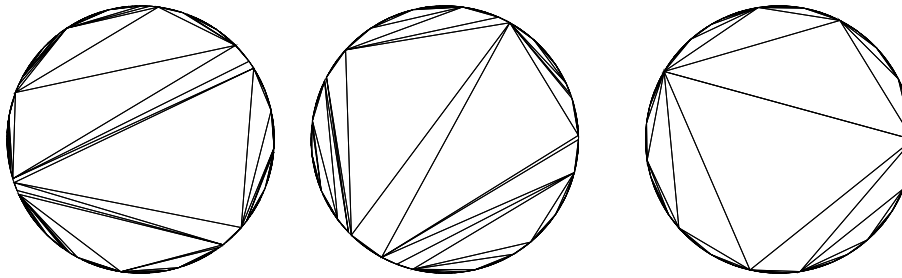

FIGURE 6.1. Realizations of β-splitting trees for (from left to right) $\beta = -1$, $\beta = 0$ (Yule tree), $\beta = 10$

with the sample shape topology by Lemma 5.21. Therefore, the sample shape topology is a compact topology on $\mathbb{T}_2$. Moreover, the sample shape topology is Hausdorff by Corollary 5.10. As the sample subtree-mass topology is a weaker Hausdorff topology by Proposition 5.18 and Corollary 5.17, it coincides with the sample shape topology. The same is true for the bpdd-Gromov-weak topology by Proposition 5.9. □

Recall from Remark 3.10 that the set of distance polynomials is convergence determining for measures on $\mathbb{T}_2$. It directly follows from the construction that the same is true for the sets of shape polynomials and subtree-mass polynomials. This property is very useful for proving convergence in law of random variables.

COROLLARY 5.22 (convergence determining classes of functions). — *The sets* $\Pi_{\mathfrak{s}} \subseteq \mathcal{C}_b(\mathbb{T}_2)$ *(defined in* (96)*) and* $\Pi_{\mathfrak{m}}$ *(defined in* (111)*) are convergence determining for measures on* $\mathbb{T}_2$ *with bpdd-Gromov-weak topology.*

Proof. — $\mathbb{T}_2$ is a compact metrizable space, and both $\Pi_{\mathfrak{s}}$ and $\Pi_{\mathfrak{m}}$ induce the bpdd-Gromov-weak topology on $\mathbb{T}_2$ by Theorem 5.19. Furthermore, each of $\Pi_{\mathfrak{s}}$ and $\Pi_{\mathfrak{m}}$ is closed under multiplication. Thus, the claim follows by the Stone–Weierstrass theorem. □

6. Example: sampling consistent families

Consider a family $(T_n, c_n)_{n\in\mathbb{N}}$ of random, finite binary (algebraic) trees, where (T_n, c_n) has n leaves. Let K_n be the Markov kernel that takes such a tree and removes a leaf uniformly chosen at random, together with the branch point to which it is attached, thus obtaining a binary tree with $n-1$ leaves. We say that the family is *sampling consistent* if $K_n(T_n, \cdot\,) = \mathcal{L}(T_{n-1})$, where $\mathcal{L}$ denotes the law of a random variable.

EXAMPLE 6.1 (β-splitting trees). — For every $\beta \in [-2, \infty]$, let T_n^β be the β-splitting tree on n leaves from [6] (with forgotten labels). For $-2 < \beta < \infty$, the β-splitting tree T_n^β can be constructed recursively as follows: T_2^β consists of two leaves connected by a distinguished root edge. If $n > 2$, choose $i \in \{1, \dots, n-1\}$ with probability

$$q_n^\beta(i) = \frac{1}{a_n(\beta)} \binom{n}{i} \int_0^1 x^{i+\beta}(1-x)^{n-i+\beta} \, \mathrm{d}x, \tag{119}$$

where $a_n(\beta)$ is a normalization constant. Then construct two independent β-splitting trees T_i^β and T_{n-i}^β, introduce a new branch point in the middle of each of the two root edges, and connect these new branch points with the new root edge to obtain T_n^β.

It is easy to see (and was observed in [6]) that $(T_n^\beta)_{n\in\mathbb{N}}$ is sampling consistent. Note the special cases $\beta = -2$, which is the *comb tree*, $\beta = -\frac{3}{2}$, which is the *uniform cladogram*, $\beta = 0$, which is the *Yule tree*, and $\beta = \infty$, which is the *symmetric binary tree*. See Figure 6.1 for triangulations of a realization of β-splitting trees for different values of β and large n. The Aldous Brownian CRT, which is the limit for $\beta = -\frac{3}{2}$, is shown in Figure 4.1.

LEMMA 6.2 (convergence of sampling consistent families). — *Let $((T_n, c_n))_{n\in\mathbb{N}}$ be a sampling consistent family of random binary trees and μ_n the uniform distribution on $\mathrm{lf}(T_n, c_n)$. Then we have the convergence in law*

$$(T_n, c_n, \mu_n) \xrightarrow[n\to\infty]{\mathcal{L}} (T, c, \mu) \quad \text{on } \mathbb{T}_2 \text{ with bpdd-Gromov-weak topology}, \tag{120}$$

for some random algebraic measure tree $(T, c, \mu) \in \mathbb{T}_2$ with the nonatomic measure μ.

Proof. — Recall the m-tree shape distribution $\mathfrak{S}_m$ from Definition 5.8. Let $n, m \in \mathbb{N}$ with $m < n$ and define

$$\epsilon_{n,m} \coloneqq \mu_n^{\otimes m}\big\{x \in T^m : x_1, \dots, x_m \text{ not distinct}\big\} \le \tfrac{m^2}{n}. \tag{121}$$

Because (T_n) is sampling consistent, we obtain for the annealed shape distribution

$$\mathbb{E}\big(\mathfrak{S}_m(T_n, c_n, \mu_n)\big) = (1 - \epsilon_{n,m})\mathcal{L}(T_m^*) + \epsilon_{n,m}\mu_{n,m}, \tag{122}$$

where T_m^* is obtained from T_m by randomly labeling the leaves, and $\mu_{n,m} \in \mathcal{M}_1(\mathfrak{C}_m)$ is some law of m-labelled cladograms supported by cladograms where at least one leaf has more than one label. This shows that, for every fixed m, the expected m-tree shape distribution converges as $n \to \infty$. Because the m-tree shape distribution is convergence determining for the bpdd-Gromov-weak topology by Corollary 5.22, all limit points of $\mathcal{L}(T_n, c_n, \mu_n)$ in $\mathcal{M}_1(\mathbb{T}_2)$ coincide. According to Theorem 5.19, $\mathbb{T}_2$, and, hence, $\mathcal{M}_1(\mathbb{T}_2)$, is compact, and, thus, a unique limit exists. That the limiting measure is nonatomic is

obvious, because the probability that a sampled shape is single-labeled tends to 1 by (122). □

In the parameter range $\beta \in [-2, -1)$, the height (in graph distance) of the β-splitting tree with n leaves is asymptotically of power-law order $\Theta(n^{-\beta-1})$. In this case, after rescaling edge lengths with the factor $n^{\beta+1}$, Gromov-Hausdorff convergence in law to a fragmentation tree is shown in [40, Corollary 16]. In the case $\beta > -1$, the height of the tree is only of logarithmic order $\Theta(\log(n))$, and it is easy to see that no nontrivial Gromov-Hausdorff scaling limit (with uniform edge rescalings) exists. Seen as algebraic measure trees, however, it easily follows from sampling consistency that the bpdd-Gromov-weak limit exists in the full parameter range $\beta \in [-2, \infty]$.

EXAMPLE 6.3 (β-splitting trees continued). — By Lemma 6.2, for every $\beta \in [-2, \infty]$, the sequence $(T_n^\beta, c_n^\beta, \mu_n^\beta)_{n\in\mathbb{N}}$ of increasing β-splitting trees converges in distribution to some limiting random algebraic measure tree $(T^\beta, c^\beta, \mu^\beta)$. In the case of the uniform cladogram ($\beta = -\frac{3}{2}$), the limit is the Brownian algebraic continuum random tree, which can be obtained as tree $\tau(C_{\mathrm{CRT}})$ coded by the Brownian triangulation (see Example 4.5), or as the algebraic measure tree induced by the metric measure Brownian CRT, which is known to have uniform shape distribution ([3]). In the case of the comb tree ($\beta = -2$), the limit is the unit interval with Lebesgue measure (a coding triangulation is shown in the very right-hand part of Figure 4.3).

Appendix A. A uniform Glivenko–Cantelli theorem

In Sections 5.1 and 5.2, we made use of uniform estimates of the speed of convergence in the approximation of the branch-point distribution and the measure of a algebraic measure tree by empirical distribution. Such uniform Glivenko–Cantelli estimates under a bound on the Vapnik–Chervonenkis dimension (VC-dimension) of the type presented below should be well known. As we did not find it explicitly in sufficient generality in the literature, we will present it here.

We recall the definition of VC-dimension, going back to the seminal work of Vapnik and Chervonenkis, [57]. Let E be a nonempty set and $\mathcal{I}$ a nonempty collection of subsets of E. For $n \in \mathbb{N}$ and $x \in E^n$, put

$$\mathcal{I}(x) \coloneqq \big\{(\mathbb{1}_I(x_1), \dots, \mathbb{1}_I(x_n)) : I \in \mathcal{I}\big\} \subseteq \{0,1\}^n. \tag{123}$$

Then, obviously, $1 \le \#\mathcal{I}(x) \le 2^n$.

DEFINITION A.1 (Vapnik–Chervonenkis dimension). — The *Vapnik–Chervonenkis dimension* of $\mathcal{I}$ is defined as

$$\dim_{\mathrm{VC}}(\mathcal{I}) \coloneqq \sup\big\{n \in \mathbb{N} : \max_{x\in E^n} \#\mathcal{I}(x) = 2^n\big\}. \tag{124}$$

EXAMPLE A.2 (collection of intervals of an algebraic tree). — Let (T, c) be a separable algebraic tree with $\#T > 2$ and

$$\mathcal{I} \coloneqq \mathcal{I}_T \coloneqq \big\{[u, v] : u, v \in T\big\}. \tag{125}$$

For $x_1, x_2, u \in T$ distinct, we have

$$\#\mathcal{I}(x) \geq \#\{[u, u],\, [x_1, x_1],\, [x_2, x_2],\, [x_1, x_2]\} = 2^2,$$

and, hence, $\dim_{\mathrm{VC}}(\mathcal{I}_T) \geq 2$. Conversely, for $x \in T^3$, either there is $u, v \in T$ with $x_1, x_2, x_3 \in [u, v]$. Then without loss of generality $x_2 \in [x_1, x_3]$ and $(1, 0, 1) \notin \mathcal{I}_T(x)$. Or there is no such $u, v \in T$, which means $(1, 1, 1) \notin \mathcal{I}_T(x)$. Therefore,

$$\dim_{\mathrm{VC}}(\mathcal{I}_T) = 2. \tag{126}$$

Recall the notion $\mathcal{S}_x(y)$ of the equivalence class of $T \setminus \{x\}$ containing y.

EXAMPLE A.3 (collection of subtrees branching of a branch point). — Let (T, c) be a separable algebraic tree and

$$\mathcal{I} \coloneqq \mathcal{S}_T \coloneqq \big\{\mathcal{S}_v(u) : u, v \in T\big\}. \tag{127}$$

We claim that

$$\dim_{\mathrm{VC}}(\mathcal{S}_T) \leq 3. \tag{128}$$

For this upper bound, let $x = (x_1, x_2, x_3, x_4) \in T^4$. By the four-point condition of the branch-point map, we can assume without loss of generality that

$$c(x_1, x_2, x_3) = c(x_1, x_2, x_4). \tag{129}$$

In this case, it is not possible to cover $\{x_1, x_3\}$, but not x_2 or x_4 either, with a single subtree in $\mathcal{S}_T$, which proves the claim.

The leading constant in the following Glivenko–Cantelli lemma is clearly not optimal. For us, it is only important that it is universal and does not depend on the measure space (E, μ).

LEMMA A.4 (rate of convergence in Glivenko–Cantelli). — *Let E be a Polish space, μ a probability measure on E, $(X_n)_{n\in\mathbb{N}}$ independent and identically distributed, μ-distributed, and $\mu_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k}$ the empirical measure. Then, for every $\mathcal{I} \subseteq \mathcal{B}(E)$ with $\dim_{\mathrm{VC}}(\mathcal{I}) < \infty$ and $n > 1$,*

$$\mathbb{E}\Big(\sup_{I\in\mathcal{I}} \big|\mu(I) - \mu_n(I)\big|\Big) \leq 96\sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{I})}{n}}. \tag{130}$$

Proof. — By the Kuratowski isomorphism theorem, all uncountable Polish spaces are Borel-isomorphic. Therefore, we may assume without loss of generality that $E = \mathbb{R}$. Theorem 3.2 in [21] yields

$$\Delta := \mathbb{E}\big(\sup_{I\in\mathcal{I}}\big|\mu(I) - \mu_n(I)\big|\big) \le \frac{24}{\sqrt{n}} \sup_{x\in\mathbb{R}^n} \int_0^1 \sqrt{\log\big(2N(r,\mathcal{I}(x))\big)}\,\mathrm{d}r, \tag{131}$$

where $N(r,\mathcal{I}(x))$ is the covering number of $\mathcal{I}(x)$ with respect to the metric $\frac{1}{\sqrt{n}}\cdot d_{\ell^2}$, where d_{ℓ^2} is the Euclidean metric on $\{0,1\}^n$. This covering number can have an upper bound in terms of the separation number $M(r,\mathcal{I})$ with respect to the metric $\frac{1}{n}\cdot d_{\ell^1}$ used by Haussler in [41], and Theorem 1 there yields

$$N\big(r,\mathcal{I}(x)\big) \le M\big(r^2,\mathcal{I}(x)\big) \le e(\dim_{\mathrm{VC}}(\mathcal{I})+1)\Big(\frac{2e}{r^2}\Big)^{\dim_{\mathrm{VC}}(\mathcal{I})}, \tag{132}$$

provided that $nr^2\in\mathbb{N}$. For $r^2\le\frac{1}{n}$, we use the trivial estimate $M(r^2,\mathcal{I}(x))\le 2^n$. For general $r^2\ge\frac{1}{n}$, we estimate $M(r^2,\mathcal{I}(x)) \le M(\frac{1}{n}\lfloor nr^2\rfloor,\mathcal{I}(x))$, and inserting (132) into (131) yields

(133)

$$\begin{aligned}\Delta &\le \frac{24}{\sqrt{n}}\Big(\sqrt{\tfrac{n+1}{n}} \\ &\qquad + \int_{\frac{1}{\sqrt{n}}}^1 \sqrt{\log(2e(\dim_{\mathrm{VC}}(\mathcal{I})+1)) + \dim_{\mathrm{VC}}(\mathcal{I})\log\big(2e(r^2-\tfrac{1}{n})^{-1}\big)}\,\mathrm{d}r\Big) \\ &\le \frac{24}{\sqrt{n}}\sqrt{\dim_{\mathrm{VC}}(\mathcal{I})}\Big(\sqrt{\tfrac{n+1}{n}} + \int_0^1 \sqrt{3+\log(2e)-2\log(r)}\,\mathrm{d}r\Big),\end{aligned}$$

where we used that $\log(2e(d+1))\le 3d$ for $d\ge 1$, and $r^2-\frac{1}{n}\ge(r-\frac{1}{\sqrt{n}})^2$. The last bracket is less than 4 for $n>1$, and the claim follows. □

BIBLIOGRAPHY

[1] R. Abraham & J.-F. Delmas – "A continuum-tree-valued Markov process", *Ann. Probab.* **40** (2012), no. 3, p. 1167–1211.

[2] R. Abraham, J.-F. Delmas & G. Voisin – "Pruning a Lévy continuum random tree", *Electron. J. Probab.* **15** (2010), no. 46, p. 1429–1473.

[3] D. Aldous – "The continuum random tree III", *Ann. Probab.* **21** (1993), p. 248–289.

[4] ———, "Recursive self-similarity for random trees, random triangulations and Brownian excursion", *Ann. Probab.* **22** (1994), no. 2, p. 527–545.

[5] ———, "Triangulating the circle, at random", *Amer. Math. Monthly* **101** (1994), no. 3, p. 223–233.

[6] D. Aldous – "Probability distributions on cladograms", in *Random discrete structures (Minneapolis, MN, 1993)*, IMA Vol. Math. Appl., vol. 76, Springer, New York, 1996, p. 1–18.
[7] D. Aldous – "Mixing time for a Markov chain on cladograms", *Combin. Probab. Comput.* **9** (2000), no. 3, p. 191–204.
[8] S. Athreya, W. Löhr & A. Winter – "The gap between Gromov-vague and Gromov-Hausdorff-vague topology", *Stochastic Process. Appl.* **126** (2016), no. 9, p. 2527–2553.
[9] ———, "Invariance principle for variable speed random walks on trees", *Ann. Probab.* **45** (2017), no. 2, p. 625–667.
[10] I. Benjamini & O. Schramm – "Recurrence of distributional limits of finite planar graphs", *Electron. J. Probab.* **6** (2001), p. no. 23, 13.
[11] D. Blount & M. A. Kouritzin – "On convergence determining and separating classes of functions", *Stochastic Process. Appl.* **120** (2010), no. 10, p. 1898–1907.
[12] N. Broutin & H. Sulzbach – "The dual tree of a recursive triangulation of the disk", *Ann. Probab.* **43** (2015), no. 2, p. 738–781.
[13] I. Chiswell – *Introduction to* Λ*-trees*, World Scientific Publishing Co. Inc., River Edge, NJ, 2001.
[14] N. Curien – "Dissecting the circle, at random", in *Journées MAS 2012*, ESAIM Proc., vol. 44, EDP Sci., Les Ulis, 2014, p. 129–139.
[15] N. Curien, B. Haas & I. Kortchemski – "The CRT is the scaling limit of random dissections", *Random Structures Algorithms* **47** (2015), no. 2, p. 304–327.
[16] N. Curien & I. Kortchemski – "Random non-crossing plane configurations: a conditioned Galton-Watson tree approach", *Random Structures Algorithms* **45** (2014), no. 2, p. 236–260.
[17] ———, "Percolation on random triangulations and stable looptrees", *Probab. Theory Relalted Fields* **163** (2015), no. 1-2, p. 303–337.
[18] N. Curien & J.-F. Le Gall – "Random recursive triangulations of the disk via fragmentation theory", *Ann. Probab.* **39** (2011), no. 6, p. 2224–2270.
[19] A. Dasgupta – *Set theory*, Birkhäuser/Springer, New York, 2014, With an introduction to real point sets.
[20] A. Depperschmidt, A. Greven & P. Pfaffelhuber – "Tree-valued Fleming-Viot dynamics with selection", *Ann. Appl. Probab.* **22** (2012), no. 6, p. 2560–2615.
[21] L. Devroye & G. Lugosi – *Combinatorial methods in density estimation*, Springer Series in Statistics, Springer-Verlag, New York, 2001.
[22] A. Dress, V. Moulton & W. Terhalle – "T-theory: an overview", *European J. Combin.* **17** (1996), no. 2-3, p. 161–175, Discrete metric spaces (Bielefeld, 1994).

[23] A. W. M. Dress – "Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: a note on combinatorial properties of metric spaces", *Adv. in Math.* **53** (1984), no. 3, p. 321–402.
[24] S. N. Evans – *Probability and real trees*, Lecture Notes in Mathematics, vol. 1920, Springer, Berlin, 2008, Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005.
[25] S. N. Evans, R. Grübel & A. Wakolbinger – "Doob-Martin boundary of Rémy's tree growth chain", *Ann. Probab.* **45** (2017), no. 1, p. 225–277.
[26] S. N. Evans, J. Pitman & A. Winter – "Rayleigh processes, real trees, and root growth with re-grafting", *Probab. Theory Relalted Fields* **134** (2006), no. 1, p. 81–126.
[27] S. N. Evans & A. Winter – "Subtree prune and re-graft: A reversible real-tree valued Markov chain", *Ann. Probab.* **34** (2006), no. 3, p. 918–961.
[28] P. Fabel – "A topological characterization of the underlying spaces of complete R-trees", *Michigan Math. J.* **64** (2015), no. 4, p. 881–887.
[29] C. Favre & M. Jonsson – *The valuative tree*, Lecture Notes in Mathematics, vol. 1853, Springer-Verlag, Berlin, 2004.
[30] H. Fischer & A. Zastrow – "Combinatorial $\mathbb{R}$-trees as generalized Cayley graphs for fundamental groups of one-dimensional spaces", *Geom. Dedicata* **163** (2013), p. 19–43.
[31] D. J. Ford – "Probabilities on cladograms: introduction to the alpha model", 2005, arxiv:0511246.
[32] N. Forman – "Exchangeable hierarchies and mass-structure of weighted real trees", *Electron. J. Probab.* **25** (2020), paper no. 131.
[33] N. Forman, C. Haulk & J. Pitman – "A representation of exchangeable hierarchies by sampling from random real trees", *Probab. Theory Related Fields* **172** (2018), no. 1-2, p. 1–29.
[34] K. Fukaya – "Collapsing of Riemannian manifolds and eigenvalues of Laplace operators", *Invent. Math.* **87** (1987), p. 517–547.
[35] P. K. Glöde – "Dynamics of Genealogical Trees for Autocatalytic Branching Processes", Dissertation, FAU Erlangen-Nürnberg, 2012, `http://d-nb.info/1033029912/34`.
[36] A. Greven, P. Pfaffelhuber & A. Winter – "Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees)", *Probab. Theory Relalted Fields* **145** (2009), p. 285–322.
[37] ———, "Tree-valued resampling dynamics: Martingale Problems and applications", *Probab. Theory Relalted Fields* **155** (2013), no. 3–4, p. 789–838.
[38] A. Greven, R. Sun & A. Winter – "Continuum space limit of the genealogies of interacting Fleming-Viot processes on $\mathbb{Z}$", *Electron. J. Probab.* **21** (2016), no. 58, p. 1–64.

[39] M. Gromov – *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser Boston Inc., Boston, MA, 1999.
[40] B. Haas, G. Miermont, J. Pitman & M. Winkel – "Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models", *Ann. Probab.* **36** (2008), no. 5, p. 1790–1837.
[41] D. Haussler – "Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension", *J. Combin. Theory Ser. A* **69** (1995), no. 2, p. 217–232.
[42] H. He & M. Winkel – "Invariance principles for pruning processes of Galton-Watson trees", 2014, arxiv:1409.1014v1.
[43] ———, "Gromov-Hausdorff-Prokhorov convergence of vertex cut-trees of n-leaf Galton-Watson trees", *Bernoulli* **25** (2019), no. 3, p. 2301–2329.
[44] S. Kliem & W. Löhr – "Existence of mark functions in marked metric measure spaces", *Electron. J. Probab.* **20** (2015), no. 73, p. 1–24.
[45] S. Kliem & A. Winter – "Evolving phylogenies of trait-dependent branching with mutation and competition, Part I: Existence", *Stochastic Process. Appl.* **129** (2019), no. 12, p. 4837–4877.
[46] W. Löhr – "Equivalence of Gromov-Prohorov- and Gromov's $\square_\lambda$-metric on the space of metric measure spaces", *Electron. Commun. Probab.* **18** (2013), no. 17, p. 1–10.
[47] W. Löhr, L. Mytnik & A. Winter – "The Aldous chain on cladograms", *Ann. Probab.* **48** (2020), no. 5, p. 2565–2590.
[48] W. Löhr, G. Voisin & A. Winter – "Convergence of bi-measure $\mathbb{R}$-trees and the pruning process", *Ann. Inst. H. Poincaré Probab. Statist.* **51** (2015), no. 4, p. 1342–1368.
[49] L. Lovász – *Large networks and graph limits*, American Mathematical Society Colloquium Publications, vol. 60, American Mathematical Society, Providence, RI, 2012.
[50] J. C. Mayer, J. Nikiel & L. G. Oversteegen – "Universal spaces for $\mathbf{R}$-trees", *Trans. Amer. Math. Soc.* **334** (1992), no. 1, p. 411–432.
[51] J. C. Mayer & L. G. Oversteegen – "A topological characterisation of $\mathbf{R}$-trees", *Trans. Amer. Math. Soc.* **320** (1990), no. 1, p. 395–415.
[52] J. W. Morgan & P. B. Shalen – "Valuations, trees, and degenerations of hyperbolic structures. I", *Ann. of Math. (2)* **120** (1984), no. 3, p. 401–476.
[53] J. Nussbaumer – "The Kingman coalescent: old and new aspects", `http://www-stud.uni-due.de/~snjonuss/master_thesis_kingman_coalescent.pdf`, Master thesis.
[54] S. Piotrowiak – "Dynamics of Genealogical Trees for Type- and State-dependent Resampling Models", Dissertation, FAU Erlangen-Nürnberg, 2010, `http://d-nb.info/1009840509/34`.

[55] L. A. Steen & J. A. Seebach, Jr. – *Counterexamples in topology*, second ed., Springer-Verlag, New York-Heidelberg, 1978.
[56] J. Tits – "A "theorem of Lie-Kolchin" for trees", in *Contributions to algebra (collection of papers dedicated to Ellis Kolchin)*, Academic Press, New York, 1977, p. 377–388.
[57] V. N. Vapnik & A. Y. Chervonenkis – "On the uniform convergence of relative frequencies of events to their probabilities", *Theory Probab. Appl.* **16** (1971), no. 2, p. 264–280, English translation of Russian original.
[58] C. Villani – *Optimal transport*, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer, Berlin-Heidelberg, 2009.

Bull. Soc. Math. France
149 (1), 2021, p. 119-153

CHARACTERIZATION OF THE TWO-DIMENSIONAL FIVEFOLD TRANSLATIVE TILES

BY QI YANG & CHUANMING ZONG

ABSTRACT. — In 1885, Fedorov discovered that a convex domain can form a lattice tiling of the Euclidean plane, if and only if it is a parallelogram or a centrally symmetric hexagon. This paper proves the following results. Besides parallelograms and centrally symmetric hexagons, there is no other convex domain that can form any two, three or fourfold translative tiling in the Euclidean plane. In particular, it characterizes all two-dimensional fivefold translative tiles, which are parallelograms, centrally symmetric hexagons, two classes of octagons and one class of decagons.

RÉSUMÉ (*Caractérisation des pavages translatifs quintuples à deux dimensions*). — En 1885, Fedorov découvrait qu'un domaine convexe peut former un réseau-pavage de la plane euclidienne si et seulement s'il est un parallélogramme ou un hexagone symmétrique centralement. Cet article démontre les résultats suivants: outre les parallélogrammes et les hexagones symmétriques centralement, il n'y aucun autre domaine convexe qui peut former dans la plane eucldienne un pavage translatif double ou triple ou quadruple. En particulier, il caractérise tous les pavages translatifs quintuples en deux dimensions, qui sont parallélogrammes, hexagones symmétriques centralement, deux classes d'octogones, et une classe de décagons.

Texte reçu le 12 avril 2019, modifié le 5 octobre 2020, accepté le 20 octobre 2020.

QI YANG, School of Mathematical Sciences, Peking University, Beijing 100871, China
CHUANMING ZONG, Center for Applied Mathematics, Tianjin University, Tianjin 300072, China • *E-mail :* cmzong@math.pku.edu.cn

Mathematical subject classification (2010). — 52C20, 05B45, 51M20, 52C15.

Key words and phrases. — Multiple tiling, Translative tiling, Lattice tiling.

This work is supported by the National Natural Science Foundation of China (NSFC11921001), the National Key Research and Development Program of China (2018YFA0704701) and 973 Program 2013CB834201.

BULLETIN DE LA SOCIÉTÉ MATHÉMATIQUE DE FRANCE 0037-9484/2021/119/$ 5.00

doi:10.24033/bsmf.2826

1. Introduction

In 1885, Fedorov [6] proved that *a convex domain can form a lattice tiling in the plane if and only if it is a parallelogram or a centrally symmetric hexagon; a convex body can form a lattice tiling in the space if and only if it is a parallelotope, a hexagonal prism, a rhombic dodecahedron, an elongated dodecahedron, or a truncated octahedron.* As a generalized inverse problem of Fedorov's discovery, in 1900 Hilbert [13] listed the following question in the second part of his 18th problem: *Whether polyhedra also exist which do not appear as fundamental regions of groups of motions, by means of which nevertheless by a suitable juxtaposition of congruent copies a complete filling up of all space is possible.* To verify Hilbert's problem in the plane, in 1917 Bieberbach suggested to Reinhardt (see [19]) to determine all the two-dimensional convex tiles. However, to complete the list turns out to be challenging and dramatic. Over the years, the list has been successively extended by Reinhardt, Kershner, James, Rice, Stein, Mann, McLoud-Mann and Von Derau (see [15, 27]); its completeness has been mistakenly announced several times! In 2017, M. Rao [18] announced a completeness proof based on computer checks.

Let K be a convex body with (relative) interior $\mathrm{int}(K)$ and (relative) boundary $\partial(K)$, and let X be a discrete set, both in $\mathbb{E}^n$. We call $K+X$ a *translative tiling* of $\mathbb{E}^n$ and call K a *translative tile*, if $K+X=\mathbb{E}^n$ and the translates $\mathrm{int}(K)+\mathbf{x}_i$ are pairwise disjoint. In other words, if $K+X$ is both a packing and a covering in $\mathbb{E}^n$. In particular, we call $K+\Lambda$ a *lattice tiling* of $\mathbb{E}^n$ and call K a *lattice tile*, if Λ is an n-dimensional lattice. It is apparent that a translative tile must be a convex polytope. Usually, a lattice tile is called a *parallelohedron*.

As one can predict, to determine the parallelohedra in higher dimensions is complicated. According to Fedorov [6], there are exact five types of parallelohedra in $\mathbb{E}^3$. Through the works of Delone [3], Štogrin [23] and Engel [5], we know that there are exact 52 combinatorially different types of parallelohedra in $\mathbb{E}^4$. A computer classification for the five-dimensional parallelohedra was announced by Dutour Sikirić, Garber, Schürmann and Waldmann [4] only in 2015.

Let Λ be an n-dimensional lattice. The *Dirichlet–Voronoi cell* of Λ is defined by

$$C=\{\mathbf{x}:\mathbf{x}\in\mathbb{E}^n,\ |\mathbf{x},\mathbf{o}|\le|\mathbf{x},\Lambda|\},$$

where $|X,Y|$ denotes the Euclidean distance between X and Y. Clearly, $C+\Lambda$ is a lattice tiling, and the Dirichlet–Voronoi cell C is a parallelohedron. In 1908, Voronoi [22] made a conjecture that *every parallelohedron is a linear transformation image of the Dirichlet–Voronoi cell of a suitable lattice.* In $\mathbb{E}^2$, $\mathbb{E}^3$ and $\mathbb{E}^4$, this conjecture was confirmed by Delone [3] in 1929. In higher dimensions, it is still open.

To characterize the translative tiles is another fascinating problem. At the first glance, translative tilings should be more complicated than lattice tilings. However, the dramatic story had a happy ending! It was shown by Minkowski [17] in 1897 that *every translative tile must be centrally symmetric.* In 1954, Venkov [21] proved that *every translative tile must be a lattice tile* (*parallelohedron*) (see [1] for generalizations). Later, a new proof for this beautiful result was independently discovered by McMullen [16].

Let X be a discrete multiset in $\mathbb{E}^n$ and let k be a positive integer. We call $K+X$ a *k-fold translative tiling* of $\mathbb{E}^n$ and call K a *translative k-tile*, if every point $\mathbf{x}\in\mathbb{E}^n$ belongs to at least k translates of K in $K+X$, and every point $\mathbf{x}\in\mathbb{E}^n$ belongs to at most k translates of $\mathrm{int}(K)$ in $\mathrm{int}(K)+X$. In other words, $K+X$ is both a k-fold packing and a k-fold covering in $\mathbb{E}^n$ (see [7, 27]). In particular, we call $K+\Lambda$ a k-fold lattice tiling of $\mathbb{E}^n$ and call K a *lattice k-tile*, if Λ is an n-dimensional lattice. Apparently, a translative k-tile must be a convex polytope. In fact, similarly to Minkowski's characterization, it was shown by Gravin, Robins and Shiryaev [10] that *a translative k-tile must be a centrally symmetric polytope with centrally symmetric facets.*

Multiple tilings were first investigated by Furtwängler [8] in 1936 as a generalization of Minkowski's conjecture on cube tilings. Let C denote the n-dimensional unit cube. Furtwängler made a conjecture that *every k-fold lattice tiling $C+\Lambda$ has twin cubes. In other words, every multiple lattice tiling $C+\Lambda$ has two cubes sharing a whole facet.* In the same paper, he proved the two and three-dimensional cases. Unfortunately, when $n\geq 4$, this beautiful conjecture was disproved by Hajós [12] in 1941. In 1979, Robinson [20] determined all the integer pairs $\{n,k\}$ for which Furtwängler's conjecture is false. We refer to Zong [25, 26] for detailed accounts on this fascinating problem and to pages 82–84 of Gruber and Lekkerkerker [11] for some generalizations.

Let P denote an n-dimensional centrally symmetric convex polytope, let $\tau(P)$ be the smallest integer k, such that P can form a k-fold translative tiling in $\mathbb{E}^n$, and let $\tau^*(P)$ be the smallest integer k, such that P can form a k-fold lattice tiling in $\mathbb{E}^n$. For convenience, we define $\tau(P)=\infty$, if P cannot form translative tiling of any multiplicity. Clearly, for every centrally symmetric convex polytope, we have

$$\tau(P)\leq\tau^*(P).$$

In 1994, Bolle [2] proved that *every centrally symmetric lattice polygon is a lattice multiple tile.* However, little is known about the multiplicity. Let Λ denote the two-dimensional integer lattice and let P_8 denote the octagon with vertices $(\frac{1}{2},\frac{3}{2})$, $(\frac{3}{2},\frac{1}{2})$, $(\frac{3}{2},-\frac{1}{2})$, $(\frac{1}{2},-\frac{3}{2})$, $(-\frac{1}{2},-\frac{3}{2})$, $(-\frac{3}{2},-\frac{1}{2})$, $(-\frac{3}{2},\frac{1}{2})$ and $(-\frac{1}{2},\frac{3}{2})$, as shown in Figure 1. As a particular example of Bolle's theorem, it was discovered by Gravin, Robins and Shiryaev [10] that $P_8+\Lambda$ *is a sevenfold lattice tiling of* $\mathbb{E}^2$.

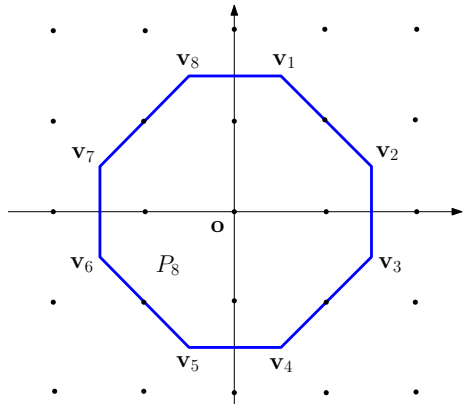

Figure 1

In 2000, Kolountzakis [14] proved that, if D is a two-dimensional convex domain, which is not a parallelogram, and $D + X$ is a multiple tiling in $\mathbb{E}^2$, then X must be a finite union of translated two-dimensional lattices. In 2013, a similar result in $\mathbb{E}^3$ was discovered by Gravin, Kolountzakis, Robins and Shiryaev [9].

In 2017, Yang and Zong [24] studied multiple lattice tilings by proving the following results. *Besides parallelograms and centrally symmetric hexagons, there is no other convex domain that can form any two, three or fourfold lattice tiling in the Euclidean plane. However, there are particular octagons and decagons that can form fivefold lattice tilings.* Afterwards, Zong [29] characterized all the two-dimensional fivefold lattice tiles. *A convex domain can form a fivefold lattice tiling of the Euclidean plane, if and only if it is a parallelogram, a centrally symmetric hexagon, under a suitable affine linear transformation, a centrally symmetric octagon with vertices* $\mathbf{v}_1 = (-\alpha, -\frac{3}{2})$, $\mathbf{v}_2 = (1-\alpha, -\frac{3}{2})$, $\mathbf{v}_3 = (1+\alpha, -\frac{1}{2})$, $\mathbf{v}_4 = (1-\alpha, \frac{1}{2})$, $\mathbf{v}_5 = -\mathbf{v}_1$, $\mathbf{v}_6 = -\mathbf{v}_2$, $\mathbf{v}_7 = -\mathbf{v}_3$ *and* $\mathbf{v}_8 = -\mathbf{v}_4$, *where* $0 < \alpha < \frac{1}{4}$, *or with vertices* $\mathbf{v}_1 = (\beta, -2)$, $\mathbf{v}_2 = (1+\beta, -2)$, $\mathbf{v}_3 = (1-\beta, 0)$, $\mathbf{v}_4 = (\beta, 1)$, $\mathbf{v}_5 = -\mathbf{v}_1$, $\mathbf{v}_6 = -\mathbf{v}_2$, $\mathbf{v}_7 = -\mathbf{v}_3$, $\mathbf{v}_8 = -\mathbf{v}_4$, *where* $\frac{1}{4} < \beta < \frac{1}{3}$, *or a centrally symmetric decagon with* $\mathbf{u}_1 = (0, 1)$, $\mathbf{u}_2 = (1, 1)$, $\mathbf{u}_3 = (\frac{3}{2}, \frac{1}{2})$, $\mathbf{u}_4 = (\frac{3}{2}, 0)$, $\mathbf{u}_5 = (1, -\frac{1}{2})$, $\mathbf{u}_6 = -\mathbf{u}_1$, $\mathbf{u}_7 = -\mathbf{u}_2$, $\mathbf{u}_8 = -\mathbf{u}_3$, $\mathbf{u}_9 = -\mathbf{u}_4$ *and* $\mathbf{u}_{10} = -\mathbf{u}_5$ *as the middle points of its edges.*

This paper proves the following theorems.

THEOREM 1.1. — *Besides parallelograms and centrally symmetric convex hexagons, there is no other convex domain that can form a two, three, or fourfold translative tiling of the Euclidean plane.*

THEOREM 1.2. — *A convex domain can form a fivefold translative tiling of the Euclidean plane, if and only if it is a parallelogram, a centrally symmetric hexagon, under a suitable affine linear transformation, a centrally symmetric octagon with vertices* $\mathbf{v}_1 = \left(\frac{3}{2} - \frac{5\alpha}{4}, -2\right)$, $\mathbf{v}_2 = \left(-\frac{1}{2} - \frac{5\alpha}{4}, -2\right)$, $\mathbf{v}_3 = \left(\frac{\alpha}{4} - \frac{3}{2}, 0\right)$, $\mathbf{v}_4 = \left(\frac{\alpha}{4} - \frac{3}{2}, 1\right)$, $\mathbf{v}_5 = -\mathbf{v}_1$, $\mathbf{v}_6 = -\mathbf{v}_2$, $\mathbf{v}_7 = -\mathbf{v}_3$ *and* $\mathbf{v}_8 = -\mathbf{v}_4$, *where* $0 < \alpha < \frac{2}{3}$, *or with vertices* $\mathbf{v}_1 = (2 - \beta, -3)$, $\mathbf{v}_2 = (-\beta, -3)$, $\mathbf{v}_3 = (-2, -1)$, $\mathbf{v}_4 = (-2, 1)$, $\mathbf{v}_5 = -\mathbf{v}_1$, $\mathbf{v}_6 = -\mathbf{v}_2$, $\mathbf{v}_7 = -\mathbf{v}_3$ *and* $\mathbf{v}_8 = -\mathbf{v}_4$, *where* $0 < \beta \leq 1$, *or a centrally symmetric decagon with* $\mathbf{u}_1 = (0, 1)$, $\mathbf{u}_2 = (1, 1)$, $\mathbf{u}_3 = (\frac{3}{2}, \frac{1}{2})$, $\mathbf{u}_4 = (\frac{3}{2}, 0)$, $\mathbf{u}_5 = (1, -\frac{1}{2})$, $\mathbf{u}_6 = -\mathbf{u}_1$, $\mathbf{u}_7 = -\mathbf{u}_2$, $\mathbf{u}_8 = -\mathbf{u}_3$, $\mathbf{u}_9 = -\mathbf{u}_4$ *and* $\mathbf{u}_{10} = -\mathbf{u}_5$ *as the middle points of its edges.*

REMARK 1.3. — Comparing this with Zong's work [29], it is easy to show that all fivefold translative tiles are fivefold lattice tiles.

2. Basic preparation

Let P_{2m} denote a centrally symmetric convex $2m$-gon centered at the origin, let $\mathbf{v}_1$, $\mathbf{v}_2$, …, $\mathbf{v}_{2m}$ be the $2m$ vertices of P_{2m} enumerated clock-wise, and let G_1, G_2, …, G_{2m} be the $2m$ edges, where G_i is ended by $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$. For convenience, we write

$$V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{2m}\}$$

and

$$\Gamma = \{G_1, G_2, \ldots, G_{2m}\}.$$

Assume that $P_{2m} + X$ is a $\tau(P_{2m})$-fold translative tiling in $\mathbb{E}^2$, where $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots\}$ is a discrete multiset with $\mathbf{x}_1 = \mathbf{o}$. Now, let us observe the local structures of $P_{2m} + X$ at the vertices $\mathbf{v} \in V + X$.

Let $X^{\mathbf{v}}$ denote the subset of X consisting of all points $\mathbf{x}_i$, such that

$$\mathbf{v} \in \partial(P_{2m}) + \mathbf{x}_i.$$

Since $P_{2m} + X$ is a multiple tiling, the set $X^{\mathbf{v}}$ can be divided into disjoint subsets $X_1^{\mathbf{v}}$, $X_2^{\mathbf{v}}$, …, $X_t^{\mathbf{v}}$, such that the translates in $P_{2m} + X_j^{\mathbf{v}}$ can be re-enumerated as $P_{2m} + \mathbf{x}_1^j$, $P_{2m} + \mathbf{x}_2^j$, …, $P_{2m} + \mathbf{x}_{s_j}^j$ satisfying the following conditions (as shown by Figure 2 in two cases):

1. $\mathbf{v} \in \partial(P_{2m}) + \mathbf{x}_i^j$ *holds for all* $i = 1, 2, \ldots, s_j$.
2. *Let* $\angle_i^j$ *denote the inner angle of* $P_{2m} + \mathbf{x}_i^j$ *at* $\mathbf{v}$ *with two half-line edges* $L_{i,1}^j$ *and* $L_{i,2}^j$, *where* $L_{i,1}^j$, $\mathbf{x}_i^j - \mathbf{v}$ *and* $L_{i,2}^j$ *are in clock order. Then, the inner angles join properly as*

$$L_{i,2}^j = L_{i+1,1}^j$$

holds for all $i = 1, 2, \ldots, s_j$, *where* $L_{s_j+1,1}^j = L_{1,1}^j$.

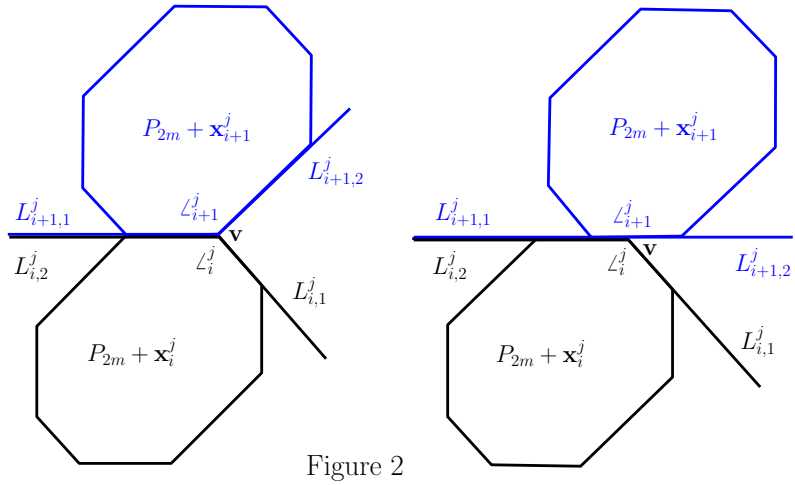

Figure 2

For convenience, we call such a sequence $P_{2m}+\mathbf{x}_1^j$, $P_{2m}+\mathbf{x}_2^j$, …, $P_{2m}+\mathbf{x}_{s_j}^j$ an *adjacent wheel* at $\mathbf{v}$. In other words, if $\mathbf{v}$ belongs to the boundary of a tile, then we follow this tile around, moving from tile to tile, until it closes up again. It is easy to see that

$$\sum_{i=1}^{s_j} \angle_i^j = 2w_j \cdot \pi$$

hold for positive integers w_j. Then we define

$$\varpi(\mathbf{v}) = \sum_{j=1}^{t} w_j = \frac{1}{2\pi} \sum_{j=1}^{t} \sum_{i=1}^{s_j} \angle_i^j$$

and

$$\varphi(\mathbf{v}) = \sharp\left\{\mathbf{x}_i : \ \mathbf{x}_i \in X, \ \mathbf{v} \in \mathrm{int}(P_{2m}) + \mathbf{x}_i\right\}.$$

In other words, $\varpi(\mathbf{v})$ is the tiling multiplicity produced by the boundary, and $\varphi(\mathbf{v})$ is the tiling multiplicity produced by the interior.

Clearly, if $P_{2m}+X$ is a $\tau(P_{2m})$-fold translative tiling of $\mathbb{E}^2$, then

$$\tau(P_{2m}) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \tag{1}$$

holds for all $\mathbf{v} \in V + X$.

Now we introduce some basic results which will be useful in this paper.

LEMMA 2.1. — *Assume that P_{2m} is a centrally symmetric convex $2m$-gon centered at the origin and $P_{2m}+X$ is a $\tau(P_{2m})$-fold translative tiling of the plane, where $m \geq 4$. If $\mathbf{v} \in V+X$ is a vertex and $G \in \Gamma+X$ is an edge with $\mathbf{v}$ as one*

of its two ends, then there are at least $\lceil (m-3)/2 \rceil$ *different translates* $P_{2m}+\mathbf{x}_i$ *satisfying both*

$$\mathbf{v} \in \partial(P_{2m}) + \mathbf{x}_i$$

and

$$G \setminus \{\mathbf{v}\} \subset \operatorname{int}(P_{2m}) + \mathbf{x}_i.$$

Proof. — Since adjacent wheels are circular, without loss of generality, let $P_{2m}+\mathbf{x}_1$, $P_{2m}+\mathbf{x}_2$, …, $P_{2m}+\mathbf{x}_s$ be an adjacent wheel at $\mathbf{v}$, such that G is the first edge appearing in the wheel and let $\angle_i$ denote the inner angle of $P_{2m}+\mathbf{x}_i$ at the vertex $\mathbf{v}$.

Let n denote the smallest index, such that

$$\sum_{i=1}^{n} \angle_i = \omega \cdot \pi \tag{2}$$

holds with some positive integer ω. Then the angle sequence $\angle_1$, $\angle_2$, …, $\angle_n$ has no pair $\angle_i$ and $\angle_j$ satisfying $\angle_i = \angle_j$. Otherwise, one can make the index n smaller. If $\angle_j$ and $\angle_{j+k}$ are two opposite angles of P_{2m} appearing in the angle sequence with $1 \le j < j+k \le n$, it is easy to see that

$$\sum_{i=0}^{k-1} \angle_{j+i} = \omega' \cdot \pi$$

holds with a positive integer ω' and $\omega \ge \omega'$. Therefore, to estimate ω we may assume that the angle sequence $\angle_1$, $\angle_2$, …, $\angle_n$ has no opposite angle pair of P_{2m}.

Clearly, $\angle_i = \pi$, if and only if $\mathbf{v}$ is a relative interior point of an edge of $P_{2m}+\mathbf{x}_i$ (such as $\angle_5$ in Figure 3) and, therefore,

$$\sum_{i=1}^{n} \angle_i < n \cdot \pi. \tag{3}$$

On the other hand, if ℓ of the n angles are π and $n-\ell < m$, then $m-n+\ell$ pairs of the opposite angles of P_{2m} do not appear in the angle sequence. Thus, we have

$$\sum_{i=1}^{n} \angle_i > \ell \cdot \pi + (m-1)\cdot\pi - (m-n+\ell)\cdot\pi = (n-1)\cdot\pi, \tag{4}$$

which together with (3) contradicts (2). Therefore, to avoid the contradiction, we must have

$$n - \ell = m,$$

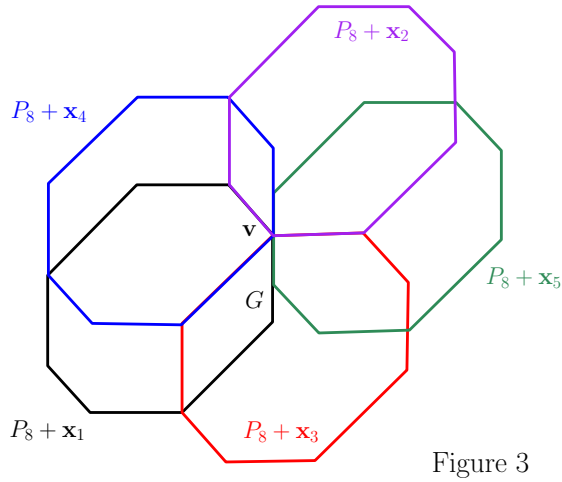

Figure 3

and each pair of the opposite angles of P_{2m} has a representative in the sequence $\angle_1, \angle_2, \ldots, \angle_n$. Consequently, we have

$$\sum_{i=1}^{n} \angle_i \geq \frac{(2m-2)\cdot\pi}{2} = (m-1)\cdot\pi. \tag{5}$$

If $\mathbf{v} \in \partial(P_{2m}) + \mathbf{x}_i$, $G \subset P_{2m} + \mathbf{x}_i$, and G is not an edge of $P_{2m} + \mathbf{x}_i$, then by the convexity and symmetry of P_{2m} it follows that $G \setminus \{\mathbf{v}\} \subset \mathrm{int}(P_{2m}) + \mathbf{x}_i$. Therefore, it follows by (5) that $G \setminus \{\mathbf{v}\}$ is covered by at least

$$\left\lceil \frac{m-1}{2} \right\rceil - 1 = \left\lceil \frac{m-3}{2} \right\rceil$$

of the s translates $\mathrm{int}(P_{2m}) + \mathbf{x}_i$. Lemma 2.1 is proved. □

LEMMA 2.2. — *Assume that P_{2m} is a centrally symmetric convex $2m$-gon centered at the origin, $P_{2m} + X$ is a translative multiple tiling of the plane, and $\mathbf{v} \in V + X$. Then we have*

$$\varpi(\mathbf{v}) = \kappa \cdot \frac{m-1}{2} + \ell \cdot \frac{1}{2},$$

where κ is a positive integer, and ℓ is the number of the edges in $\Gamma + X$, which take $\mathbf{v}$ as an interior point.

Proof. — Assume that $P_{2m} + \mathbf{x}_1$, $P_{2m} + \mathbf{x}_2$, …, $P_{2m} + \mathbf{x}_s$ is an adjacent wheel at $\mathbf{v}$ and let $\angle_i$ denote the inner angle of $P_{2m} + \mathbf{x}_i$ at $\mathbf{v}$. Of course, we have $\angle_i = \pi$, if $\mathbf{v}$ is not a vertex of $P_{2m} + \mathbf{x}_i$.

Assume that $\angle_1 < \pi$ and let n to be the smallest index, such that

$$\sum_{i=1}^{n} \angle_i = \omega\pi \tag{6}$$

holds with a positive integer ω. We proceed to show that each pair of the opposite angles of P_{2m} has one and only one representative in $\angle_1, \angle_2, \ldots, \angle_n$.

If, on the contrary, $\angle_j$ and $\angle_{j+k}$ are two of these n angles, $\angle_j < \pi$, which are either identical or opposite. Then, it is easy to see that

$$\sum_{i=0}^{k-1} \angle_{j+i} = \omega'\pi \tag{7}$$

holds with a positive integer ω'. For convenience, we assume that $\angle_j, \angle_{j+1}, \ldots, \angle_{j+k-1}$ have neither a identical nor an opposite pair. Then, by repeating the argument between (2) and (5) in the proof of Lemma 2.1, one can deduce that each pair of the opposite angles of P_{2m} has one and only one representative in $\angle_j, \angle_{j+1}, \ldots, \angle_{j+k-1}$. Consequently, one of these k angles is either identical or opposite to $\angle_1$, which contradicts the minimum assumption on n and ω.

Then, applying the argument between (2) and (5) to $\angle_1, \angle_2, \ldots, \angle_n$, it can be deduced that

$$\sum_{i=1}^{n} \angle_i = (m-1)\pi + \ell_1\pi, \tag{8}$$

where ℓ_1 is the number of the π angles in $\angle_1, \angle_2, \ldots, \angle_n$. In fact, it is $n - m$.

By repeating this process to $\angle_{n+1}, \angle_{n+2}, \ldots, \angle_s$ if necessary, it follows that

$$\sum_{i=1}^{s} \angle_i = \kappa'(m-1)\pi + \ell'\pi, \tag{9}$$

and, therefore,

$$\varpi(\mathbf{v}) = \frac{1}{2\pi} \sum \sum_{i=1}^{s} \angle_i = \kappa \cdot \frac{m-1}{2} + \ell \cdot \frac{1}{2}, \tag{10}$$

where the first sum is over all adjacent wheels at $\mathbf{v}$, κ' and κ are suitable positive integers, and ℓ' and ℓ are suitable nonnegative integers. In fact, ℓ is the number of the edges that take $\mathbf{v}$ as an interior point.

Lemma 2.2 is proved. □

LEMMA 2.3. — *If m is an odd positive integer, P_{2m} is a centrally symmetric convex $2m$-gon centered at the origin $\mathbf{o}$, and $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{2m}$ are the middle points of its edges enumerated clockwise, then we have*

$$\sum_{i=1}^{m} (-1)^i \mathbf{u}_i = \mathbf{o}.$$

Proof. — Since $\mathbf{u}_i$ is the middle point of G_i, we have

$$\begin{cases} \mathbf{v}_2 = 2\mathbf{u}_1 - \mathbf{v}_1, \\ \mathbf{v}_3 = 2\mathbf{u}_2 - \mathbf{v}_2, \\ \quad \cdots \\ \mathbf{v}_{m+1} = 2\mathbf{u}_m - \mathbf{v}_m, \end{cases}$$

which implies

$$-\mathbf{v}_1 = \mathbf{v}_{m+1} = -\mathbf{v}_1 - 2\sum_{i=1}^{m}(-1)^i\mathbf{u}_i \tag{11}$$

and, therefore,

$$\sum_{i=1}^{m}(-1)^i\mathbf{u}_i = \mathbf{o}.$$

The lemma is proved. □

The following lemma will be useful in the proofs of Lemma 3.5 and Lemma 3.8.

LEMMA 2.4 (Bolle [2]). — *A convex polygon is a k-fold lattice tile for a lattice Λ and some positive integer k, if and only if the following conditions are satisfied:*

1. *It is centrally symmetric.*
2. *When it is centred at the origin, in the relative interior of each edge G there is a point of $\frac{1}{2}\Lambda$.*
3. *If the midpoint of G is not in $\frac{1}{2}\Lambda$, then G is a lattice vector of Λ.*

3. Proofs of the theorems

LEMMA 3.1. — *Let P_{2m} be a centrally symmetric convex $2m$-gon, then*

$$\tau(P_{2m}) \geq \begin{cases} m-1, & \text{if } m \text{ is even,} \\ m-2, & \text{if } m \text{ is odd.} \end{cases}$$

Proof. — Assume that $P_{2m} + X$ is a $\tau(P_{2m})$-fold translative tiling in the Euclidean plane and assume that $\mathbf{v} \in V + X$. Then it follows by Lemma 2.1 that

$$\varphi(\mathbf{v}) \geq \left\lceil \frac{m-3}{2} \right\rceil. \tag{12}$$

Let $P_{2m} + \mathbf{x}_1$, $P_{2m} + \mathbf{x}_2$, …, $P_{2m} + \mathbf{x}_s$ be an adjacent wheel at $\mathbf{v}$ and let $\angle_1$, $\angle_2$, …, $\angle_s$ be the corresponding angle sequence. By (5) we have

$$\varpi(\mathbf{v}) \geq \frac{1}{2\pi}\sum_{i=1}^{s}\angle_i \geq \left\lceil \frac{m-1}{2} \right\rceil. \tag{13}$$

Then, it follows by (1), (12) and (13) that

$$\tau(P_{2m}) \geq \left\lceil \frac{m-3}{2} \right\rceil + \left\lceil \frac{m-1}{2} \right\rceil = \begin{cases} m-1, & \text{if } m \text{ is even,} \\ m-2, & \text{if } m \text{ is odd.} \end{cases}$$

Lemma 3.1 is proved. □

LEMMA 3.2. — *Let P_{14} be a centrally symmetric convex tetradecagon, then*

$$\tau(P_{14}) \geq 6.$$

Proof. — Assume that $P_{14} + X$ is a $\tau(P_{14})$-fold translative tiling in $\mathbb{E}^2$ and $\mathbf{v} \in V + X$. By Lemma 2.1 and Lemma 2.2, we have

$$\varphi(\mathbf{v}) \geq \left\lceil \frac{7-3}{2} \right\rceil = 2 \tag{14}$$

and

$$\varpi(\mathbf{v}) = \kappa \cdot 3 + \ell \cdot \frac{1}{2} \geq 3, \tag{15}$$

where κ is a positive integer and ℓ is a nonnegative integer.

Now, to show the lemma it is sufficient to deal with the following two cases.

Case 1. — *$\varpi(\mathbf{v}) \geq 4$ holds for a vertex $\mathbf{v} \in V + X$.* Then, by (1) and (14) we get

$$\tau(P_{14}) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{16}$$

Case 2. — *$\varpi(\mathbf{v}) = 3$ holds for every vertex $\mathbf{v} \in V + X$.* First, let us observe a simple fact. If $\varpi(\mathbf{v}) = 3$ holds at $\mathbf{v} \in V + X$ and $P_{14} + \mathbf{x}_1$, $P_{14} + \mathbf{x}_2$, …, $P_{14} + \mathbf{x}_s$ is an adjacent wheel at $\mathbf{v}$, then it follows from (15) that s must be seven and $\mathbf{v}$ is a common vertex of all these translates, as shown by Figure 4. Then, by Lemma 2.1, every vertex $\mathbf{v}_i^*$ connecting with $\mathbf{v}$ by an edge is an interior point of two of the seven translates in the wheel.

Then, we have

$$\varpi(\mathbf{v}_1^*) = \varpi(\mathbf{v}_2^*) = \varpi(\mathbf{v}_3^*) = \varpi(\mathbf{v}_4^*) = \varpi(\mathbf{v}_5^*) = \varpi(\mathbf{v}_6^*) = \varpi(\mathbf{v}_7^*) = 3. \tag{17}$$

Therefore, for each vertex $\mathbf{v}_i^*$, there are two different points $\mathbf{y}_{i,1}$, $\mathbf{y}_{i,2} \in X$, such that

$$\mathbf{v}_i^* \in \partial(P_{14}) + \mathbf{y}_{i,j}, \qquad j = 1,\ 2$$

and

$$\mathbf{v} \in \operatorname{int}(P_{14}) + \mathbf{y}_{i,j}, \qquad j = 1,\ 2.$$

If $\mathbf{y}_{i,j} \notin \{\mathbf{y}_{1,1}, \mathbf{y}_{1,2}\}$ holds for one of these points, and then we have

$$\varphi(\mathbf{v}) \geq 3$$

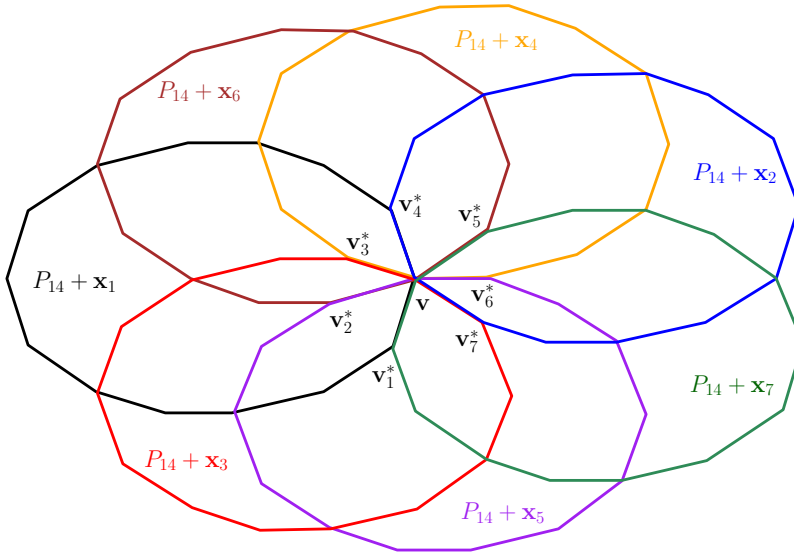

Figure 4

and, therefore,

$$\tau(P_{14}) \geq \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{18}$$

If $\mathbf{y}_{i,j} \in \{\mathbf{y}_{1,1}, \mathbf{y}_{1,2}\}$ holds for all of these points, then we must have

$$\{\mathbf{v}_1^*, \mathbf{v}_2^*, \mathbf{v}_3^*, \mathbf{v}_4^*, \mathbf{v}_5^*, \mathbf{v}_6^*, \mathbf{v}_7^*\} \subset \partial(P_{14}) + \mathbf{y}_{1,1}.$$

It is known that $(D + \mathbf{x}) \cap (D + \mathbf{y})$ is centrally symmetric for all $\mathbf{x}$ and $\mathbf{y}$ whenever D is centrally symmetric. Then, by Figure 4 it is easy to see that $(P_{14} + \mathbf{y}_{1,1}) \cap (P_{14} + \mathbf{x}_1)$ is a parallelogram with vertices $\mathbf{v}_1^*$, $\mathbf{v}$, $\mathbf{v}_4^*$ and $\mathbf{v}_1^* + (\mathbf{v}_4^* - \mathbf{v})$, and $(P_{14} + \mathbf{y}_{1,1}) \cap (P_{14} + \mathbf{x}_7)$ is a parallelogram with vertices $\mathbf{v}_1^*$, $\mathbf{v}$, $\mathbf{v}_5^*$ and $\mathbf{v}_1^* + (\mathbf{v}_5^* - \mathbf{v})$. Consequently, by symmetry one can deduce that $P_{14} + \mathbf{y}_{1,1}$ is an hexagon with vertices $\mathbf{v}_1^*$, $\mathbf{v}_1^* + (\mathbf{v}_4^* - \mathbf{v})$, $\mathbf{v}_4^*$, $\mathbf{v} + (\mathbf{v} - \mathbf{v}_1^*)$, $\mathbf{v}_5^*$ and $\mathbf{v}_1^* + (\mathbf{v}_5^* - \mathbf{v})$, which contradicts the assumption that P_{14} is a tetradecagon.

As a conclusion, for every centrally symmetric convex tetradecagon, we have

$$\tau(P_{14}) \geq 6. \tag{19}$$

The lemma is proved. □

LEMMA 3.3. — *Let P_{12} be a centrally symmetric convex dodecagon, then we have*

$$\tau(P_{12}) \geq 6.$$

Proof. — First of all, it follows by Lemma 2.1 that

$$\varphi(\mathbf{v}) \geq \left\lceil \frac{6-3}{2} \right\rceil = 2 \tag{20}$$

holds for all $\mathbf{v} \in V + X$. On the other hand, by Lemma 2.2 we have

$$\varpi(\mathbf{v}) = \kappa \cdot \frac{6-1}{2} + \ell \cdot \frac{1}{2} \geq 3. \tag{21}$$

Thus, to show the lemma it is sufficient to deal with the following two cases.

Case 1. — $\varpi(\mathbf{v}) \geq 4$ *holds for a vertex* $\mathbf{v} \in V + X$. Then it follows by (1) and (20) that

$$\tau(P_{12}) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{22}$$

Case 2. — $\varpi(\mathbf{v}) = 3$ *holds for a vertex* $\mathbf{v} \in V + X$. Assume that $P_{12} + \mathbf{x}_1$, $P_{12} + \mathbf{x}_2$, …, $P_{12} + \mathbf{x}_s$ is an adjacent wheel at $\mathbf{v}$. By (21) it can be deduced that there is a $G \in \Gamma + X$, such that

$$\mathbf{v} \in \operatorname{int}(G).$$

Let $\mathbf{v}'$ and $\mathbf{v}^*$ denote the two ends of G. By Lemma 2.1 and the convexity of P_{12} it follows that X has four different points $\mathbf{y}_1'$, $\mathbf{y}_2'$, $\mathbf{y}_1^*$ and $\mathbf{y}_2^*$ satisfying

$$\begin{aligned} \mathbf{v}' &\in \partial(P_{12}) + \mathbf{y}_i', \quad i = 1,\ 2, \\ \mathbf{v}^* &\in \partial(P_{12}) + \mathbf{y}_i^*, \quad i = 1,\ 2, \\ \mathbf{v} &\in \operatorname{int}(P_{12}) + \mathbf{y}_i', \quad i = 1,\ 2, \qquad \text{and} \\ \mathbf{v} &\in \operatorname{int}(P_{12}) + \mathbf{y}_i^*, \quad i = 1,\ 2. \end{aligned}$$

Consequently, we have

$$\varphi(\mathbf{v}) \geq 4,$$

and, therefore,

$$\tau(P_{12}) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 7. \tag{23}$$

The conclusion of these two cases that

$$\tau(P_{12}) \geq 6 \tag{24}$$

holds for every centrally symmetric dodecagon. Lemma 3.3 is proved. □

LEMMA 3.4 (Yang and Zong [24]). — *Let P_{10} be a centrally symmetric decagon centred at the origin, then we have*

$$\tau^*(P_{10}) \geq 5.$$

LEMMA 3.5. — *Let P_{10} be a centrally symmetric decagon centred at the origin, then we have*

$$\tau(P_{10}) \geq 5,$$

where the equality holds, if and only if P_{10} is a fivefold lattice tile.

Proof. — Let $\mathbf{v}_1$, $\mathbf{v}_2$, …, $\mathbf{v}_{10}$ denote the ten vertices of P_{10} enumerated clockwise, let G_i denote the edge with ends $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$ and let $\mathbf{u}_i$ denote the middle point of G_i. Suppose that X is a discrete subset of $\mathbb{E}^2$, and $P_{10}+X$ is a $\tau(P_{10})$-fold translative tiling of the plane. First of all, it follows from Lemma 2.1 that

$$\varphi(\mathbf{v}) \geq \left\lceil \frac{5-3}{2} \right\rceil = 1 \tag{25}$$

holds for every $\mathbf{v} \in V + X$. On the other hand, by Lemma 2.2 we have

$$\varpi(\mathbf{v}) = \kappa \cdot 2 + \ell \cdot \frac{1}{2}, \tag{26}$$

where κ is a positive integer, and ℓ is the number of the edges that contain $\mathbf{v}$ as a relative interior point.

Now we prove the lemma by dealing with two cases.

Case 1. — *$\ell \neq 0$ holds at a vertex $\mathbf{v} \in V + X$.* In other words, there is an edge $G \in \Gamma + X$, such that $\mathbf{v} \in \mathrm{int}(G)$. Clearly, by (26) we have $\varpi(\mathbf{v}) \geq 3$.

Suppose that $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$ are the two ends of G. By Lemma 2.1, there are two different points $\mathbf{y}_1 \in X^{\mathbf{v}_1^*}$ and $\mathbf{y}_2 \in X^{\mathbf{v}_2^*}$, such that

$$\mathbf{v} \in (\mathrm{int}(P_{10}) + \mathbf{y}_1) \cap (\mathrm{int}(P_{10}) + \mathbf{y}_2)\,.$$

Then we have $\varphi(\mathbf{v}) \geq 2$. If $\varpi(\mathbf{v}) \geq 4$, one can deduce that

$$\tau(P_{10}) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{27}$$

If $\varpi(\mathbf{v}) = 3$, by (26) one can deduce that $P_{10} + X^{\mathbf{v}}$ consists of seven translates $P_{10} + \mathbf{x}_1$, $P_{10} + \mathbf{x}_2$, …, $P_{10} + \mathbf{x}_7$, and there is another $G' \in \Gamma + X$, which contains $\mathbf{v}$ as an interior point. Suppose that G is an edge of $P_{10} + \mathbf{x}_6$, and G' has two ends $\mathbf{v}_5^*$ and $\mathbf{v}_6^*$. We deal with three subcases.

Subcase 1.1. — *$G' || G$ and $G' \neq G$.* Without loss of generality, we assume that $\mathbf{v}_5^*$ is between $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$. Then, by Lemma 2.1 we have $\mathbf{y}_i \in X^{\mathbf{v}_i^*}$, such that

$$\mathbf{v} \in \mathrm{int}(P_{10}) + \mathbf{y}_i, \quad i = 1,\ 2,\ 5.$$

It is obvious that $\mathbf{y}_1$, $\mathbf{y}_2$ and $\mathbf{y}_5$ are pairwise distinct. Thus, we have $\varphi(\mathbf{v}) \geq 3$ and, therefore,

$$\tau(P_{10}) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{28}$$

Subcase 1.2. — *$G' = G$.* Then $P_{10} + X^{\mathbf{v}}$ can be divided into two adjacent wheels, as shown by Figure 5.

Let $P_{10}+\mathbf{x}_6$ and $P_{10}+\mathbf{x}_7$ be the two translates that contain G as a common edge. Without loss of generality, suppose that $G = G_6 + \mathbf{x}_7$ and $\mathbf{v} = \mathbf{v}_7 + \mathbf{x}_1$, as shown in Figure 5. Let L be the straight line determined by $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$, let G_1^* be the edge of $P_{10} + \mathbf{x}_1$ lying on L with ends $\mathbf{v}$ and $\mathbf{v}_3^*$ and let G_2^* be the edge of $P_{10} + \mathbf{x}_1$ with ends $\mathbf{v}$ and $\mathbf{v}_4^*$.

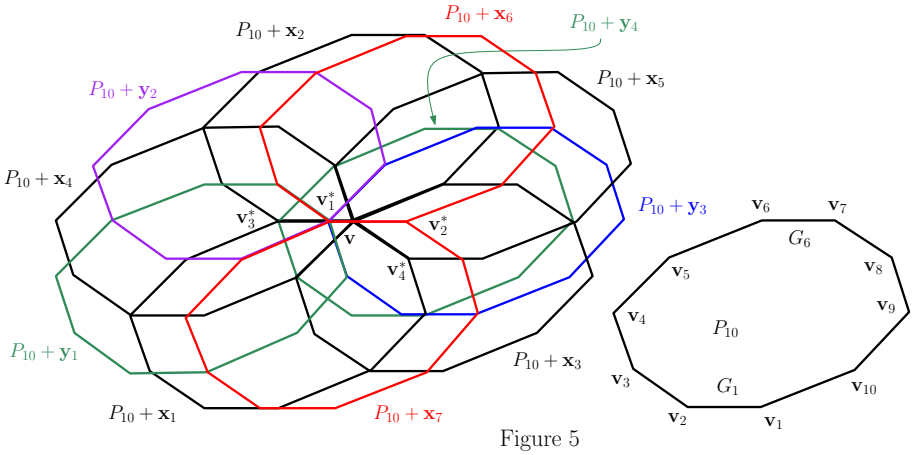

Figure 5

It is easy to see that $\varpi(\mathbf{v}_1^*) \geq 3$ and $\varphi(\mathbf{v}_1^*) \geq 2$, since $\mathbf{v}_1^*$ is an interior point of G_1^*. If $\varpi(\mathbf{v}_1^*) \geq 4$, then we have $\tau(P_{10}) \geq 6$. If $\varpi(\mathbf{v}_1^*) = 3$, the adjacent wheel at $\mathbf{v}_1^*$ can be divided into two adjacent wheels. Since $\mathbf{v}_1^* = \mathbf{v}_6 + \mathbf{x}_7$, by Lemma 2.1 and the structure of the adjacent wheel that consists of five translates, we have three points $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3 \in X^{\mathbf{v}_1^*}$, such that

$$\mathbf{v}_1^* = \mathbf{v}_8 + \mathbf{y}_1, \quad \mathbf{v}_3^* \in \operatorname{int}(P_{10}) + \mathbf{y}_1, \tag{29}$$

$$\mathbf{v}_1^* = \mathbf{v}_{10} + \mathbf{y}_2, \quad \mathbf{v}_3^* \in \operatorname{int}(P_{10}) + \mathbf{y}_2, \tag{30}$$

and

$$\mathbf{v}_1^* = \mathbf{v}_4 + \mathbf{y}_3, \quad \mathbf{v} \in \operatorname{int}(P_{10}) + \mathbf{y}_3. \tag{31}$$

Clearly, we also have $\mathbf{v}_3^* \in \operatorname{int}(P_{10}) + \mathbf{x}_4$. Since $\mathbf{v}_1^* \in \operatorname{int}(P_{10}) + \mathbf{x}_4$, we thus have $\mathbf{x}_4 \notin \{\mathbf{y}_1, \mathbf{y}_2\}$, $\varphi(\mathbf{v}_3^*) \geq 3$ and

$$\tau(P_{10}) = \varphi(\mathbf{v}_3^*) + \varpi(\mathbf{v}_3^*) \geq 5, \tag{32}$$

where the equality may hold only if $\varpi(\mathbf{v}_3^*) = 2$. When $\varpi(\mathbf{v}_3^*) = 2$, by Lemma 2.1 and the structure of the adjacent wheel with five translates, there is a point $\mathbf{y}_4 \in X^{\mathbf{v}_3^*}$, such that

$$\mathbf{v}_3^* = \mathbf{v}_4 + \mathbf{y}_4, \quad \mathbf{v} \in \operatorname{int}(P_{10}) + \mathbf{y}_4. \tag{33}$$

Furthermore, by Lemma 2.1 we have a point $\mathbf{y}_5 \in X^{\mathbf{v}_4^*}$, such that $\mathbf{v} \in \operatorname{int}(P_{10}) + \mathbf{y}_5$. By (31), (33) and convexity we have

$$\mathbf{v}_4^* \in (\operatorname{int}(P_{10}) + \mathbf{y}_3) \cap (\operatorname{int}(P_{10}) + \mathbf{y}_4),$$

$\mathbf{y}_5 \notin \{\mathbf{y}_3, \mathbf{y}_4\}$, $\varphi(\mathbf{v}) \geq 3$ and, thus,

$$\tau(P_{10}) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{34}$$

Subcase 1.3. — $G' \nparallel G$. Suppose that G is an edge of $P_{10}+\mathbf{x}_6$ with ends $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$, which contains $\mathbf{v}$ as an interior point. Since $G' \nparallel G$, there is a translate $P_{10}+\mathbf{x}'$ in $X^{\mathbf{v}}$ that meets $P_{10}+\mathbf{x}_6$ at a non-singleton part of G. Let L be the line determined by $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$. Let G_1^* be the edge of $P_{10}+\mathbf{x}'$ lying on L with ends $\mathbf{v}_3^*$ and $\mathbf{v}$, where $\mathbf{v}_1^* \in \mathrm{int}(G_1^*)$.

First, since $\ell \neq 0$ at $\mathbf{v}_1^*$, by (26) we have

$$\varpi(\mathbf{v}_1^*) \geq 3. \tag{35}$$

On the other hand, since $G' \nparallel G$, the local arrangement $P_{10}+X^{\mathbf{v}}$ cannot be divided into smaller adjacent wheels. Then, two of the seven translates in $P_{10}+X^{\mathbf{v}}$ contain both $\mathbf{v}_3^*$ and $\mathbf{v}_1^*$ as interior points. Furthermore, by Lemma 2.1, there is a translate $P_{10}+\mathbf{y}$ in $P_{10}+X^{\mathbf{v}_3^*}$ that contains $\mathbf{v}_1^*$ as an interior point and, therefore, $\varphi(\mathbf{v}_1^*) \geq 3$. Then, by (35) we get

$$\tau(P_{10}) = \varphi(\mathbf{v}_1^*) + \varpi(\mathbf{v}_1^*) \geq 6. \tag{36}$$

Case 2. — $\ell = 0$ *holds for all vertices* $\mathbf{v} \in V+X$. Then by (26) it is sufficient to assume that $\varpi(\mathbf{v})$ can take only two values, 2 or 4.

Subcase 2.1. — $\varpi(\mathbf{v}) = 4$ *holds at a vertex* $\mathbf{v} \in V+X$. Then the local arrangements $P_{10}+X^{\mathbf{v}}$ can be divided into two adjacent wheels, each containing five translates. Suppose that $P_{10}+\mathbf{x}_1$, $P_{10}+\mathbf{x}_2$, …, $P_{10}+\mathbf{x}_5$ is such a wheel at $\mathbf{v}$ and $\mathbf{v} = \mathbf{v}_k + \mathbf{x}_1$. Then, as shown in Figure 6, the wheel can be determined by $P_{10}+\mathbf{x}_1$ explicitly as follows:

$$\begin{aligned}
\mathbf{v} &= \mathbf{v}_{k+4} + \mathbf{x}_2, \quad G_{k+4} + \mathbf{x}_2 = G_{k-1} + \mathbf{x}_1,\\
\mathbf{v} &= \mathbf{v}_{k+8} + \mathbf{x}_3, \quad G_{k+8} + \mathbf{x}_3 = G_{k+3} + \mathbf{x}_2,\\
\mathbf{v} &= \mathbf{v}_{k+2} + \mathbf{x}_4, \quad G_{k+2} + \mathbf{x}_4 = G_{k+7} + \mathbf{x}_3,\\
\mathbf{v} &= \mathbf{v}_{k+6} + \mathbf{x}_5, \quad G_{k+6} + \mathbf{x}_5 = G_{k+1} + \mathbf{x}_4,
\end{aligned}$$

where $\mathbf{v}_{10+i} = \mathbf{v}_i$ and $G_{10+i} = G_i$.

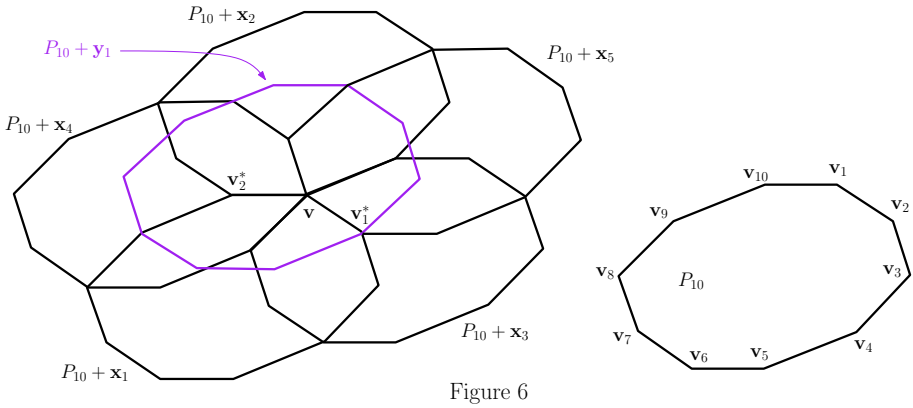

Figure 6

Without loss of generality, as shown by Figure 6, we take $\mathbf{v} = \mathbf{v}_1 + \mathbf{x}_1$, $\mathbf{v}_1^* = \mathbf{v}_2 + \mathbf{x}_1$ and $\mathbf{v}_2^* = \mathbf{v}_{10} + \mathbf{x}_1$. By Lemma 2.1, for each $\mathbf{v}_i^*$, there is a point $\mathbf{y}_i \in X^{\mathbf{v}_i^*}$, such that

$$\mathbf{v} \in \operatorname{int}(P_{10}) + \mathbf{y}_i.$$

In fact, by the previous analysis, we have $\mathbf{y}_1 = \mathbf{v}_1^* - \mathbf{v}_4$. Therefore, by convexity and symmetry,

$$\mathbf{v}_2^* \in \operatorname{int}(P_{10}) + \mathbf{y}_1.$$

Thus, the two points $\mathbf{y}_1$ and $\mathbf{y}_2$ are different. Then we have

$$\varphi(\mathbf{v}) \geq 2,$$

and

$$\tau(P_{10}) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{37}$$

Subcase 2.2. — *$\varpi(\mathbf{v}) = 2$ hold for all vertices $\mathbf{v} \in V + X$.* Let P_{10} be a centrally symmetric convex decagon centred at the origin with vertices $\mathbf{v}_1$, $\mathbf{v}_2$, …, $\mathbf{v}_{10}$ enumerated in anti-clock order. Let G_i denote the edge with ends $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$ and let $\mathbf{u}_i$ denote the middle point of G_i. Then, we define

$$\begin{cases} \mathbf{a}_1 = \mathbf{u}_1 - \mathbf{u}_6, \\ \mathbf{a}_2 = \mathbf{u}_2 - \mathbf{u}_7, \\ \mathbf{a}_3 = \mathbf{u}_3 - \mathbf{u}_8, \\ \mathbf{a}_4 = \mathbf{u}_4 - \mathbf{u}_9, \\ \mathbf{a}_5 = \mathbf{u}_5 - \mathbf{u}_{10}. \end{cases}$$

By Lemma 2.3 we have

$$\mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_3 - \mathbf{a}_4 + \mathbf{a}_5 = \mathbf{o}. \tag{38}$$

Assume that $\mathbf{x}_1 = \mathbf{o} \in X$. Since $\varpi(\mathbf{v}) = 2$ holds for every vertex $\mathbf{v} \in V + X$, by studying the structure of the adjacent wheel at $\mathbf{v}$ we have

$$\sum z_i \mathbf{a}_i \in X, \quad z_i \in \mathbb{Z}.$$

For convenience, we define

$$\Lambda = \left\{ \sum z_i \mathbf{a}_i : \ z_i \in \mathbb{Z} \right\}. \tag{39}$$

Suppose that the adjacent wheel at $\mathbf{v}_1$ is $P_{10} + \mathbf{x}_i$, $i = 1, 2, \ldots, 5$. Let $\mathbf{v}_i^*$ be the common vertex of $P_{10} + \mathbf{x}_i$ and $P_{10} + \mathbf{x}_{i+1}$ other than $\mathbf{v}_1$, as shown in Figure 7, where $\mathbf{x}_6 = \mathbf{x}_1$ and $\mathbf{x}_1 = \mathbf{o}$. By Lemma 2.1, we have $\mathbf{y}_i \in X^{\mathbf{v}_i^*}$, such

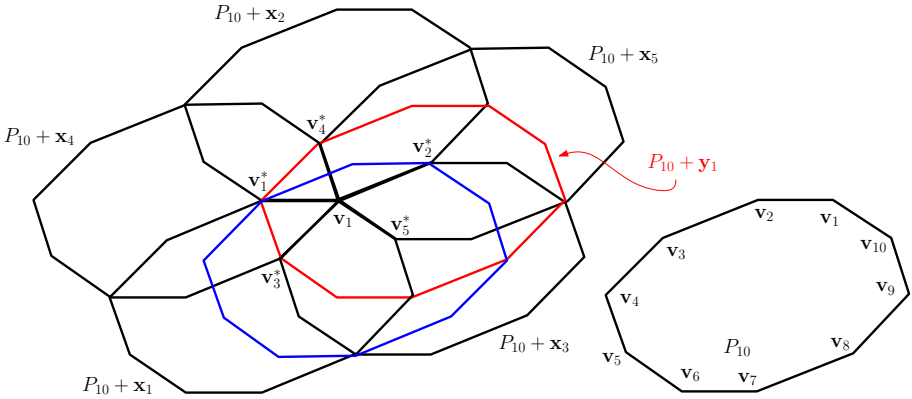

Figure 7

that $\mathbf{v}_1 \in \operatorname{int}(P_{10}) + \mathbf{y}_i$. In fact, it can be explicitly deduced by the adjacent wheels at $\mathbf{v}_1$, $\mathbf{v}_1^*$, $\mathbf{v}_2^*$, $\mathbf{v}_3^*$, $\mathbf{v}_4^*$ and $\mathbf{v}_5^*$ that

$$\begin{cases} \mathbf{y}_1 = \mathbf{v}_1^* - \mathbf{v}_4 = \mathbf{a}_2 - \mathbf{a}_3, \\ \mathbf{y}_2 = \mathbf{v}_2^* - \mathbf{v}_{10} = \mathbf{a}_1 - \mathbf{a}_3 + \mathbf{a}_4, \\ \mathbf{y}_3 = \mathbf{v}_3^* - \mathbf{v}_6 = \mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_4 - \mathbf{a}_5, \\ \mathbf{y}_4 = \mathbf{v}_4^* - \mathbf{v}_2 = -\mathbf{a}_5 + \mathbf{a}_3 - \mathbf{a}_2, \\ \mathbf{y}_5 = \mathbf{v}_5^* - \mathbf{v}_8 = -\mathbf{a}_5 - \mathbf{a}_1 + \mathbf{a}_2. \end{cases} \tag{40}$$

For example, if $P_{10} + \mathbf{y}_2$ satisfying $\mathbf{v}_{10} + \mathbf{y}_2 = \mathbf{v}_2^*$, one can obtain $P_{10} + \mathbf{y}_2$ by moving P_{10} successively to $P_{10}+\mathbf{a}_1$, $P_{10}+\mathbf{a}_1-\mathbf{a}_3$, and then to $P_{10}+\mathbf{a}_1-\mathbf{a}_3+\mathbf{a}_4$.

By (40) and symmetry it can be shown that $\mathbf{y}_i \neq \mathbf{y}_{i+1}$, where $\mathbf{y}_6 = \mathbf{y}_1$. For example, if $\mathbf{y}_1 = \mathbf{y}_2$ (as shown in Figure 7), then by symmetry we will get that $(P_{10}+\mathbf{x}_2)\cap(P_{10}+\mathbf{y}_1)$ is a parallelogram and $\mathbf{y}_1 = \mathbf{v}_1^*-\mathbf{v}_3$, which contradicts the first equation of (40). Thus, any triple of $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_5\}$ cannot be identical and, therefore, $\varphi(\mathbf{v}) \geq 3$. Consequently, we get

$$\tau(P_{10}) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 5, \tag{41}$$

where the equality may hold only if $\varphi(\mathbf{v}) = 3$.

When $\varphi(\mathbf{v}) = 3$, the five points $\mathbf{y}_1$, $\mathbf{y}_2$, …, $\mathbf{y}_5$ have to satisfy one of the following five groups of conditions:

(i) $\mathbf{y}_1 = \mathbf{y}_3$ and $\mathbf{y}_2 = \mathbf{y}_4$;
(ii) $\mathbf{y}_1 = \mathbf{y}_3$ and $\mathbf{y}_2 = \mathbf{y}_5$;
(iii) $\mathbf{y}_1 = \mathbf{y}_4$ and $\mathbf{y}_2 = \mathbf{y}_5$;
(iv) $\mathbf{y}_1 = \mathbf{y}_4$ and $\mathbf{y}_3 = \mathbf{y}_5$ and
(v) $\mathbf{y}_2 = \mathbf{y}_4$ and $\mathbf{y}_3 = \mathbf{y}_5$.

Case (i). — $\mathbf{y}_1 = \mathbf{y}_3$ *and* $\mathbf{y}_2 = \mathbf{y}_4$. Then, by (40) and (38) we get

$$\begin{cases} \mathbf{a}_2 - \mathbf{a}_3 = \mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_4 - \mathbf{a}_5, \\ \mathbf{a}_1 - \mathbf{a}_3 + \mathbf{a}_4 = -\mathbf{a}_2 + \mathbf{a}_3 - \mathbf{a}_5, \\ \mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_3 - \mathbf{a}_4 + \mathbf{a}_5 = \mathbf{o}, \end{cases}$$

$$\begin{cases} 2\mathbf{a}_2 - 2\mathbf{a}_4 + \mathbf{a}_5 = \mathbf{a}_1 + (\mathbf{a}_3 - \mathbf{a}_4), \\ \mathbf{a}_4 - 2\mathbf{a}_5 = 2\mathbf{a}_1 - (\mathbf{a}_3 - \mathbf{a}_4), \\ \mathbf{a}_2 - \mathbf{a}_5 = \mathbf{a}_1 + (\mathbf{a}_3 - \mathbf{a}_4) \end{cases}$$

and, therefore,

$$\begin{cases} \mathbf{a}_1 = \mathbf{a}_1, \\ \mathbf{a}_2 = -2\mathbf{a}_1 + 4(\mathbf{a}_3 - \mathbf{a}_4), \\ \mathbf{a}_3 = -4\mathbf{a}_1 + 6(\mathbf{a}_3 - \mathbf{a}_4), \\ \mathbf{a}_4 = -4\mathbf{a}_1 + 5(\mathbf{a}_3 - \mathbf{a}_4), \\ \mathbf{a}_5 = -3\mathbf{a}_1 + 3(\mathbf{a}_3 - \mathbf{a}_4), \end{cases}$$

which means that Λ is a lattice with a basis $\{\mathbf{a}_1,\ \mathbf{a}_3 - \mathbf{a}_4\}$. Furthermore, since $\mathbf{u}_i = \frac{1}{2}\mathbf{a}_i \in \frac{1}{2}\Lambda$, it follows by Lemma 2.4 that $P_{10} + \Lambda$ is, indeed, a multiple lattice tiling. Thus, for this particular P_{10} by (39) and Lemma 3.4 we have

$$\tau(P_{10}) \geq \tau^*(P_{10}) \geq 5, \tag{42}$$

where the equalities hold only if $P_{10} + X$ is a fivefold lattice tiling.

Case (ii). — $\mathbf{y}_1 = \mathbf{y}_3$ *and* $\mathbf{y}_2 = \mathbf{y}_5$. Then, by (40) and (38) we have

$$\begin{cases} \mathbf{a}_2 - \mathbf{a}_3 = \mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_4 - \mathbf{a}_5, \\ \mathbf{a}_1 - \mathbf{a}_3 + \mathbf{a}_4 = -\mathbf{a}_1 + \mathbf{a}_2 - \mathbf{a}_5, \\ \mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_3 - \mathbf{a}_4 + \mathbf{a}_5 = \mathbf{o}, \end{cases}$$

$$\begin{cases} \mathbf{a}_1 + \mathbf{a}_4 + \mathbf{a}_5 = -\mathbf{a}_3 + 2(\mathbf{a}_2 + \mathbf{a}_5), \\ 3\mathbf{a}_1 + 4\mathbf{a}_5 = 2(\mathbf{a}_2 + \mathbf{a}_5), \\ \mathbf{a}_1 - \mathbf{a}_4 + 2\mathbf{a}_5 = -\mathbf{a}_3 + (\mathbf{a}_2 + \mathbf{a}_5) \end{cases}$$

and, therefore,

$$\begin{cases} \mathbf{a}_1 = 8\mathbf{a}_3 - 6(\mathbf{a}_2 + \mathbf{a}_5), \\ \mathbf{a}_2 = 6\mathbf{a}_3 - 4(\mathbf{a}_2 + \mathbf{a}_5), \\ \mathbf{a}_3 = \mathbf{a}_3, \\ \mathbf{a}_4 = -3\mathbf{a}_3 + 3(\mathbf{a}_2 + \mathbf{a}_5), \\ \mathbf{a}_5 = -6\mathbf{a}_3 + 5(\mathbf{a}_2 + \mathbf{a}_5), \end{cases}$$

which means that Λ is a lattice with a basis $\{\mathbf{a}_3,\ \mathbf{a}_2 + \mathbf{a}_5\}$. Furthermore, since $\mathbf{u}_i = \frac{1}{2}\mathbf{a}_i \in \frac{1}{2}\Lambda$, it follows by Lemma 2.4 that $P_{10} + \Lambda$ is, indeed, a multiple

lattice tiling. Thus, for this particular P_{10} by (39) and Lemma 3.4 we have

$$\tau(P_{10}) \geq \tau^*(P_{10}) \geq 5, \tag{43}$$

where the equalities hold only if $P_{10} + X$ is a fivefold lattice tiling.

Case (iii). — $\mathbf{y}_1 = \mathbf{y}_4$ *and* $\mathbf{y}_2 = \mathbf{y}_5$. Then, by (40) and (38) we get

$$\begin{cases} \mathbf{a}_2 - \mathbf{a}_3 = -\mathbf{a}_2 + \mathbf{a}_3 - \mathbf{a}_5, \\ \mathbf{a}_1 - \mathbf{a}_3 + \mathbf{a}_4 = -\mathbf{a}_1 + \mathbf{a}_2 - \mathbf{a}_5, \\ \mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_3 - \mathbf{a}_4 + \mathbf{a}_5 = \mathbf{o}, \end{cases}$$
$$\begin{cases} 2\mathbf{a}_2 - \mathbf{a}_3 + 2\mathbf{a}_5 = -(\mathbf{a}_1 - \mathbf{a}_2) + \mathbf{a}_4, \\ \mathbf{a}_2 + 2\mathbf{a}_5 = -3(\mathbf{a}_1 - \mathbf{a}_2), \\ \mathbf{a}_3 + \mathbf{a}_5 = -(\mathbf{a}_1 - \mathbf{a}_2) + \mathbf{a}_4 \end{cases}$$

and, therefore,

$$\begin{cases} \mathbf{a}_1 = 4\mathbf{a}_4 + 6(\mathbf{a}_1 - \mathbf{a}_2), \\ \mathbf{a}_2 = 4\mathbf{a}_4 + 5(\mathbf{a}_1 - \mathbf{a}_2), \\ \mathbf{a}_3 = 3\mathbf{a}_4 + 3(\mathbf{a}_1 - \mathbf{a}_2), \\ \mathbf{a}_4 = \mathbf{a}_4, \\ \mathbf{a}_5 = -2\mathbf{a}_4 - 4(\mathbf{a}_1 - \mathbf{a}_2), \end{cases}$$

which means that Λ is a lattice with a basis $\{\mathbf{a}_4,\ \mathbf{a}_1 - \mathbf{a}_2\}$. Furthermore, since $\mathbf{u}_i = \frac{1}{2}\mathbf{a}_i \in \frac{1}{2}\Lambda$, it follows by Lemma 2.4 that $P_{10} + \Lambda$ is, indeed, a multiple lattice tiling. Thus, for this particular P_{10} by (39) and Lemma 3.4 we have

$$\tau(P_{10}) \geq \tau^*(P_{10}) \geq 5, \tag{44}$$

where the equalities hold only if $P_{10} + X$ is a fivefold lattice tiling.

Case (iv). — $\mathbf{y}_1 = \mathbf{y}_4$ *and* $\mathbf{y}_3 = \mathbf{y}_5$. Then, by (40) and (38) we have

$$\begin{cases} \mathbf{a}_2 - \mathbf{a}_3 = -\mathbf{a}_2 + \mathbf{a}_3 - \mathbf{a}_5, \\ \mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_4 - \mathbf{a}_5 = -\mathbf{a}_1 + \mathbf{a}_2 - \mathbf{a}_5, \\ \mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_3 - \mathbf{a}_4 + \mathbf{a}_5 = \mathbf{o}, \end{cases}$$
$$\begin{cases} \mathbf{a}_2 - \mathbf{a}_4 + \mathbf{a}_5 = \mathbf{a}_3 - (\mathbf{a}_1 + \mathbf{a}_5), \\ 3\mathbf{a}_2 + 2\mathbf{a}_5 = \mathbf{a}_3 + 3(\mathbf{a}_1 + \mathbf{a}_5), \\ \mathbf{a}_2 + \mathbf{a}_4 = \mathbf{a}_3 + (\mathbf{a}_1 + \mathbf{a}_5) \end{cases}$$

and therefore

$$\begin{cases} \mathbf{a}_1 = 4\mathbf{a}_3 - 5(\mathbf{a}_1 + \mathbf{a}_5), \\ \mathbf{a}_2 = 3\mathbf{a}_3 - 3(\mathbf{a}_1 + \mathbf{a}_5), \\ \mathbf{a}_3 = \mathbf{a}_3, \\ \mathbf{a}_4 = -2\mathbf{a}_3 + 4(\mathbf{a}_1 + \mathbf{a}_5), \\ \mathbf{a}_5 = -4\mathbf{a}_3 + 6(\mathbf{a}_1 + \mathbf{a}_5), \end{cases}$$

which means that Λ is a lattice with a basis $\{\mathbf{a}_3,\ \mathbf{a}_1 + \mathbf{a}_5\}$. Furthermore, since $\mathbf{u}_i = \frac{1}{2}\mathbf{a}_i \in \frac{1}{2}\Lambda$, it follows by Lemma 2.4 that $P_{10} + \Lambda$ is indeed a multiple lattice tiling. Thus, for this particular P_{10} by (39) and Lemma 3.4 we have

$$\tau(P_{10}) \geq \tau^*(P_{10}) \geq 5, \tag{45}$$

where the equalities hold only if $P_{10} + X$ is a fivefold lattice tiling.

Case (v). — $\mathbf{y}_2 = \mathbf{y}_4$ *and* $\mathbf{y}_3 = \mathbf{y}_5$. Then, by (40) and (38) we have

$$\begin{cases} \mathbf{a}_1 - \mathbf{a}_3 + \mathbf{a}_4 = -\mathbf{a}_2 + \mathbf{a}_3 - \mathbf{a}_5, \\ \mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_4 - \mathbf{a}_5 = -\mathbf{a}_1 + \mathbf{a}_2 - \mathbf{a}_5, \\ \mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_3 - \mathbf{a}_4 + \mathbf{a}_5 = \mathbf{o}, \end{cases}$$

$$\begin{cases} 2\mathbf{a}_1 - \mathbf{a}_3 = -2\mathbf{a}_5, \\ 3\mathbf{a}_1 + \mathbf{a}_3 - 3\mathbf{a}_4 = 3(\mathbf{a}_2 - \mathbf{a}_4) - \mathbf{a}_5, \\ \mathbf{a}_1 + \mathbf{a}_3 - 2\mathbf{a}_4 = (\mathbf{a}_2 - \mathbf{a}_4) - \mathbf{a}_5 \end{cases}$$

and, therefore,

$$\begin{cases} \mathbf{a}_1 = 3\mathbf{a}_5 + 3(\mathbf{a}_2 - \mathbf{a}_4), \\ \mathbf{a}_2 = 6\mathbf{a}_5 + 5(\mathbf{a}_2 - \mathbf{a}_4), \\ \mathbf{a}_3 = 8\mathbf{a}_5 + 6(\mathbf{a}_2 - \mathbf{a}_4), \\ \mathbf{a}_4 = 6\mathbf{a}_5 + 4(\mathbf{a}_2 - \mathbf{a}_4), \\ \mathbf{a}_5 = \mathbf{a}_5, \end{cases}$$

which means that Λ is a lattice with a basis $\{\mathbf{a}_5,\ \mathbf{a}_2 - \mathbf{a}_4\}$. Furthermore, since $\mathbf{u}_i = \frac{1}{2}\mathbf{a}_i \in \frac{1}{2}\Lambda$, it follows by Lemma 2.4 that $P_{10} + \Lambda$ is indeed a multiple lattice tiling. Thus, for this particular P_{10} by (39) and Lemma 3.4, we have

$$\tau(P_{10}) \geq \tau^*(P_{10}) \geq 5, \tag{46}$$

where the equalities hold only if $P_{10} + X$ is a fivefold lattice tiling.

As a conclusion of these cases, Lemma 3.5 is proved. □

LEMMA 3.6 (Zong [28, 29]). — *A centrally symmetric convex decagon can form a fivefold lattice tiling in* $\mathbb{E}^2$*, if and only if, under a suitable affine linear transformation, it takes* $\mathbf{u}_1 = (0,1)$, $\mathbf{u}_2 = (1,1)$, $\mathbf{u}_3 = (\frac{3}{2}, \frac{1}{2})$, $\mathbf{u}_4 = (\frac{3}{2}, 0)$, $\mathbf{u}_5 = (1, -\frac{1}{2})$, $\mathbf{u}_6 = -\mathbf{u}_1$, $\mathbf{u}_7 = -\mathbf{u}_2$, $\mathbf{u}_8 = -\mathbf{u}_3$, $\mathbf{u}_9 = -\mathbf{u}_4$ *and* $\mathbf{u}_{10} = -\mathbf{u}_5$ *as the middle points of its edges.*

REMARK 3.7 (Zong [28, 29]). — Let W denote the quadrilateral with vertices $\mathbf{w}_1 = (-\frac{1}{2}, 1)$, $\mathbf{w}_2 = (-\frac{1}{2}, \frac{3}{4})$, $\mathbf{w}_3 = (-\frac{2}{3}, \frac{2}{3})$ and $\mathbf{w}_4 = (-\frac{3}{4}, \frac{3}{4})$. A centrally symmetric convex decagon can take $\mathbf{u}_1 = (0, 1)$, $\mathbf{u}_2 = (1, 1)$, $\mathbf{u}_3 = (\frac{3}{2}, \frac{1}{2})$, $\mathbf{u}_4 = (\frac{3}{2}, 0)$, $\mathbf{u}_5 = (1, -\frac{1}{2})$, $\mathbf{u}_6 = -\mathbf{u}_1$, $\mathbf{u}_7 = -\mathbf{u}_2$, $\mathbf{u}_8 = -\mathbf{u}_3$, $\mathbf{u}_9 = -\mathbf{u}_4$ and $\mathbf{u}_{10} = -\mathbf{u}_5$ as the middle points of its edges, if and only if one of its vertices is an interior point of W.

LEMMA 3.8. — *For every centrally symmetric convex octagon P_8 we have*

$$\tau(P_8) \geq 5,$$

where the equality holds, if and only if, under a suitable affine linear transformation, it is one with vertices $\mathbf{v}_1 = \left(\frac{3}{2} - \frac{5\alpha}{4}, -2\right)$, $\mathbf{v}_2 = \left(-\frac{1}{2} - \frac{5\alpha}{4}, -2\right)$, $\mathbf{v}_3 = \left(\frac{\alpha}{4} - \frac{3}{2}, 0\right)$, $\mathbf{v}_4 = \left(\frac{\alpha}{4} - \frac{3}{2}, 1\right)$, $\mathbf{v}_5 = -\mathbf{v}_1$, $\mathbf{v}_6 = -\mathbf{v}_2$, $\mathbf{v}_7 = -\mathbf{v}_3$ *and* $\mathbf{v}_8 = -\mathbf{v}_4$, *where* $0 < \alpha < \frac{2}{3}$, *or with vertices* $\mathbf{v}_1 = (2 - \beta, -3)$, $\mathbf{v}_2 = (-\beta, -3)$, $\mathbf{v}_3 = (-2, -1)$, $\mathbf{v}_4 = (-2, 1)$, $\mathbf{v}_5 = -\mathbf{v}_1$, $\mathbf{v}_6 = -\mathbf{v}_2$, $\mathbf{v}_7 = -\mathbf{v}_3$ *and* $\mathbf{v}_8 = -\mathbf{v}_4$, *where* $0 < \beta \leq 1$.

Proof. — Suppose that X is a discrete subset of $\mathbb{E}^2$, and $P_8 + X$ is a $\tau(P_8)$-fold translative tiling of the plane. First of all, it follows from Lemma 2.1 that

$$\varphi(\mathbf{v}) \geq \left\lceil \frac{4-3}{2} \right\rceil = 1 \tag{47}$$

holds for all $\mathbf{v} \in V + X$. On the other hand, by Lemma 2.2 we have

$$\varpi(\mathbf{v}) = \kappa \cdot \frac{3}{2} + \ell \cdot \frac{1}{2}, \tag{48}$$

where κ is a positive integer and ℓ is a nonnegative integer. In fact, ℓ is the number of the edges that take $\mathbf{v}$ as an interior point. Thus, to prove the lemma, it is sufficient to deal with the following four cases:

Case 1. — $\varpi(\mathbf{v}) \geq 5$ *holds for a vertex* $\mathbf{v} \in V + X$. It follows by (1) and (47) that

$$\tau(P_8) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{49}$$

Case 2. — $\varpi(\mathbf{v}) = 4$ *holds for a vertex* $\mathbf{v} \in V + X$. It follows by (48) that $\ell \neq 0$ and therefore $\mathbf{v} \in \mathrm{int}(G)$ holds for some $G \in \Gamma + X$. Assume that $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$ are the two ends of G. Applying Lemma 2.1 to $\{\mathbf{v}_1^*, G\}$ and $\{\mathbf{v}_2^*, G\}$, respectively, one can deduce that

$$\varphi(\mathbf{v}) \geq 2$$

and, therefore,

$$\tau(P_8) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{50}$$

Case 3. — $\varpi(\mathbf{v}) = 3$ *holds for a vertex* $\mathbf{v} \in V + X$. Then (48) has and only has two groups of solutions $\{\kappa, \ell\} = \{1, 3\}$ or $\{2, 0\}$.

Subcase 3.1. — $\{\kappa,\ell\}=\{1,3\}$. Then, there are three edges G_1', G_2' and G_3' in $\Gamma+X$ satisfying

$$\mathbf{v}\in\operatorname{int}(G_i'),\quad i=1,\ 2,\ 3.$$

Next, we study the multiplicity by considering the relative positions of these edges.

Subcase 3.1.1. — $G_1'=G_2'=G_3'$. Assume that $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$ are the two ends of G_1'. Then $X^{\mathbf{v}_1^*}$ has two identical points. By computing the angle sum of all the adjacent wheels at $\mathbf{v}_1^*$ it can be deduced that

$$\varpi(\mathbf{v}_1^*)\geq 4.$$

Then, by Case 1 and Case 2 we get

$$\tau(P_8)=\varpi(\mathbf{v}_1^*)+\varphi(\mathbf{v}_1^*)\geq 6. \tag{51}$$

Subcase 3.1.2. — $G_2'=G_3'$ *and* $G_1'\nparallel G_2'$. Then there are two adjacent wheels at $\mathbf{v}$, one has five translates $P_8+\mathbf{x}_1$, $P_8+\mathbf{x}_2$, …, $P_8+\mathbf{x}_5$, and the other has two translates $P_8+\mathbf{x}_1'$ and $P_8+\mathbf{x}_2'$, as shown by Figure 8.

By re-enumeration we may assume that $\angle_1$, $\angle_2$, $\angle_3$ and $\angle_4$ are inner angles of P_8 and $\angle_5=\pi$, as shown by Figure 8. Guaranteed by linear transformation, we assume that the two edges G_1 and G_3 of P_8 are horizontal and vertical, respectively. Suppose that $G_1'=G_1+\mathbf{x}_5$. Let $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$ be the two ends of G_1', let L denote the straight line determined by $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$, let G_3^* denote the

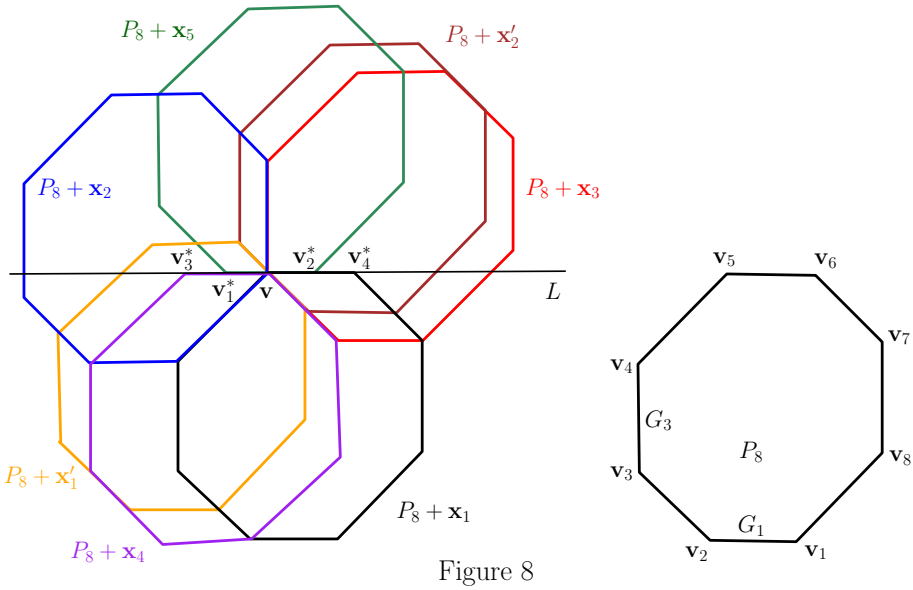

Figure 8

edge of $P_8+\mathbf{x}_4$ lying on L with two ends $\mathbf{v}$ and $\mathbf{v}_3^*$, and let G_4^* denote the edge of $P_8+\mathbf{x}_1$ lying on L with two ends $\mathbf{v}$ and $\mathbf{v}_4^*$.

By Lemma 2.1, there is a point $\mathbf{y}_1 \in X^{\mathbf{v}_1^*}$, such that $\mathbf{v}_3^* \in \mathrm{int}(P_8)+\mathbf{y}_1$. Clearly, by the convexity of P_8, both $\mathbf{v}_3^*$ and $\mathbf{v}_1^*$ belong to $\mathrm{int}(P_8)+\mathbf{x}_1'$. Thus, we have $\mathbf{y}_1 \neq \mathbf{x}_1'$. Meanwhile, since both $\mathbf{v}_3^*$ and $\mathbf{v}_1^*$ belong to $\mathrm{int}(P_8)+\mathbf{x}_2$, we have $\mathbf{x}_2 \neq \mathbf{y}_1$ and, therefore,

$$\varphi(\mathbf{v}_3^*) \geq 3.$$

Similarly, we have $\varphi(\mathbf{v}_1^*) \geq 3$, $\varphi(\mathbf{v}_2^*) \geq 3$ and $\varphi(\mathbf{v}_4^*) \geq 3$. Then, by (48) we get

$$\tau(P_8) = \varphi(\mathbf{v}_i^*) + \varpi(\mathbf{v}_i^*) \geq 5, \tag{52}$$

where the equality may hold only if

$$\varpi(\mathbf{v}_1^*) = \varpi(\mathbf{v}_2^*) = \varpi(\mathbf{v}_3^*) = \varpi(\mathbf{v}_4^*) = 2. \tag{53}$$

By (48) it is easy to see that the local configuration of $P_8+X^{\mathbf{v}}$ is essentially unique when $\varpi(\mathbf{v})=2$. In other words, it is determined by the one that $\mathbf{v}$ is not its vertex. Consequently, the set X has four points $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{y}_3$ and $\mathbf{y}_4$ satisfying

$$\mathbf{v}_1^* = \mathbf{v}_4+\mathbf{y}_1, \mathbf{v} \in \mathrm{int}(P_8)+\mathbf{y}_1, \tag{54}$$

$$\mathbf{v}_2^* = \mathbf{v}_7+\mathbf{y}_2, \mathbf{v} \in \mathrm{int}(P_8)+\mathbf{y}_2, \tag{55}$$

$$\mathbf{v}_3^* = \mathbf{v}_3+\mathbf{y}_3, \mathbf{v} \in \mathrm{int}(P_8)+\mathbf{y}_3, \quad \text{and} \tag{56}$$

$$\mathbf{v}_4^* = \mathbf{v}_8+\mathbf{y}_4, \mathbf{v} \in \mathrm{int}(P_8)+\mathbf{y}_4. \tag{57}$$

Clearly, by the convexity of P_8 we have $\mathbf{y}_1 \neq \mathbf{y}_2$, $\mathbf{y}_1 \neq \mathbf{y}_3$ and $\mathbf{y}_2 \neq \mathbf{y}_4$. For convenience, we write $\mathbf{v}_i = (x_i, y_i)$. If $\mathbf{y}_2 = \mathbf{y}_3$, then by (55) and (56) we have

$$y_3 = y_7. \tag{58}$$

If $\mathbf{y}_1 = \mathbf{y}_4$, then by (54) and (57) we get

$$y_4 = y_8. \tag{59}$$

However, it is obvious that (58) and (59) cannot hold simultaneously. Therefore, we still get

$$\varphi(\mathbf{v}) \geq 3$$

and, therefore,

$$\tau(P_8) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{60}$$

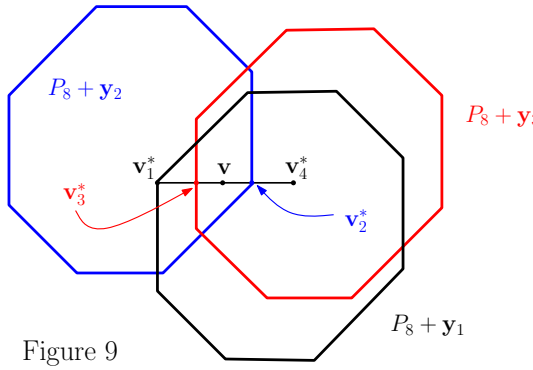

Figure 9

Subcase 3.1.3. — *$G_1' \neq G_2'$ and $G_1' \parallel G_2'$.* Let $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$ be the two ends of G_1', and let $\mathbf{v}_3^*$ and $\mathbf{v}_4^*$ be the two ends of G_2'. Without loss of generality, we suppose that $\mathbf{v}_3^*$ is between $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$, as shown by Figure 9. By Lemma 2.1, X has three points $\mathbf{y}_1$, $\mathbf{y}_2$ and $\mathbf{y}_3$ satisfying both

$$\mathbf{v}_i^* \in \partial(P_8)+\mathbf{y}_i, \quad i=1,\ 2,\ 3$$

and

$$\mathbf{v} \in \operatorname{int}(P_8)+\mathbf{y}_i, \quad i=1,\ 2,\ 3.$$

By the convexity of P_8 it is easy to see that these three points are pairwise distinct. Then, we get

$$\varphi(\mathbf{v}) \geq 3$$

and, therefore,

$$\tau(P_8) = \varpi(\mathbf{v}) + \varphi(\mathbf{v}) \geq 6. \tag{61}$$

Subcase 3.1.4. — *$G_1' \nparallel G_2'$, $G_1' \nparallel G_3'$ and $G_2' \nparallel G_3'$.* By studying the angle sum at $\mathbf{v}$ it can be deduced that $P_8+X^{\mathbf{v}}$ is an adjacent wheel of seven translates. Suppose that $\mathbf{x}_2 \in X^{\mathbf{v}}$ and G_1' is an edge of $P_8+\mathbf{x}_2$. Since G_1', G_2' and G_3' are mutually non-collinear, $X^{\mathbf{v}}$ has two points $\mathbf{x}_1$ and $\mathbf{x}_3$, such that $\mathbf{v}$ is a common vertex of both $P_8+\mathbf{x}_1$ and $P_8+\mathbf{x}_3$, and $P_8+\mathbf{x}_2$ joins both $P_8+\mathbf{x}_1$ and $P_8+\mathbf{x}_3$ at non-singleton parts of G_1', respectively. Let $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$ be the two ends of G_1', let L denote the straight line determined by $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$, let G_1^* denote the edge of $P_8+\mathbf{x}_1$ lying on L with ends $\mathbf{v}$ and $\mathbf{v}_3^*$, and let G_2^* denote the edge of $P_8+\mathbf{x}_3$ lying on L with ends $\mathbf{v}$ and $\mathbf{v}_4^*$, as shown in Figure 10.

By studying the corresponding angles of the adjacent wheel at $\mathbf{v}$, it is easy to see that $P_8+X^{\mathbf{v}}$ has exact two translates which contain both $\mathbf{v}_1^*$ and $\mathbf{v}_3^*$ as interior points. On the other hand, by Lemma 2.1, $P_8+X^{\mathbf{v}_3^*}$ has at least one more translate that contains $\mathbf{v}_1^*$ as an interior point. Thus, we have

$$\varphi(\mathbf{v}_1^*) \geq 3.$$

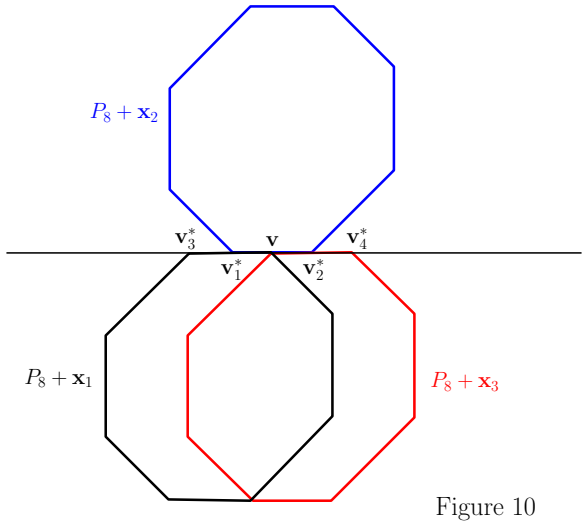

Figure 10

Similarly, we have $\varphi(\mathbf{v}_2^*) \geq 3$, $\varphi(\mathbf{v}_3^*) \geq 3$ and $\varphi(\mathbf{v}_4^*) \geq 3$. Then, by (48) we get

$$\tau(P_8) = \varphi(\mathbf{v}_i^*) + \varpi(\mathbf{v}_i^*) \geq 5, \tag{62}$$

where the equality may hold only if

$$\varpi(\mathbf{v}_1^*) = \varpi(\mathbf{v}_2^*) = \varpi(\mathbf{v}_3^*) = \varpi(\mathbf{v}_4^*) = 2. \tag{63}$$

By repeating the argument between (53) and (60), it can be deduced that

$$\tau(P_8) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{64}$$

Subcase 3.2. — $\{\kappa, \ell\} = \{2, 0\}$ *holds at every vertex* $\mathbf{v} \in V + X$. Then $P_8 + X^{\mathbf{v}}$ is an adjacent wheel of eight translates $P_8 + \mathbf{x}_1$, $P_8 + \mathbf{x}_2$, …, $P_8 + \mathbf{x}_8$, as shown in Figure 11. Let $\mathbf{v}_i^*$ be the second vertex of $P_8 + \mathbf{x}_i$ connecting to $\mathbf{v}$ by an edge. Since $\varpi(\mathbf{v}) = 3$, every $\mathbf{v}_i^*$ is an interior point of exactly two of these eight translates. Consequently, for every $\mathbf{v}_i^*$, there are two different translates $P_8 + \mathbf{y}_i$ and $P_8 + \mathbf{y}_i'$ in $P_8 + X^{\mathbf{v}_i^*}$ that both contain $\mathbf{v}$ as an interior point.

On the other hand, it can be easily deduced that there is only one point $\mathbf{x} \in X$, such that both $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$ belong to $\partial(P_8) + \mathbf{x}$ and $\mathbf{v} \in \operatorname{int}(P_8) + \mathbf{x}$. It is $\mathbf{v}_2^* - \mathbf{v} + \mathbf{x}_1$. Therefore, at least one of the two points $\mathbf{y}_2$ and $\mathbf{y}_2'$ is different from both $\mathbf{y}_1$ and $\mathbf{y}_1'$. Then, we get

$$\varphi(\mathbf{v}) \geq 3$$

and

$$\tau(P_8) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{65}$$

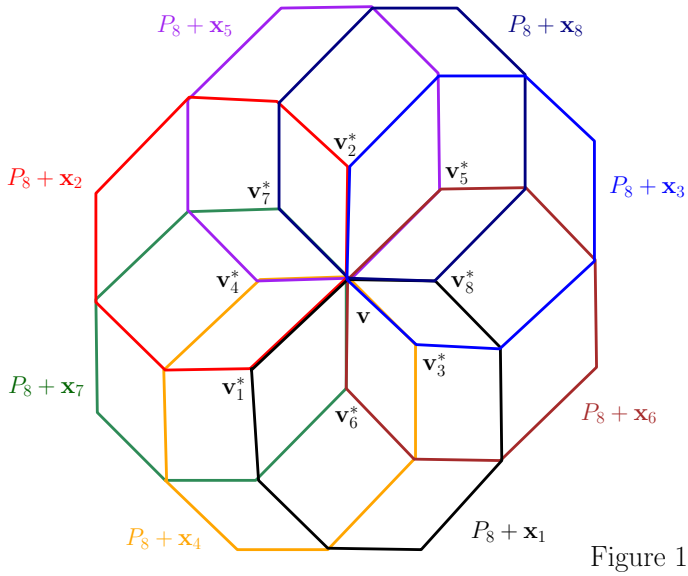

Figure 11

Case 4. — $\varpi(\mathbf{v}) = 2$ *holds for a vertex* $\mathbf{v} \in V + X$. It follows by (48) that $\varpi(\mathbf{v}) = 2$ holds, if and only if $\kappa = 1$ and $\ell = 1$. In other words, $P_8 + X^{\mathbf{v}}$ is an adjacent wheel of five translates. By re-enumeration we may assume that $\angle_1$, $\angle_2$, $\angle_3$ and $\angle_4$ are inner angles of P_8 and $\angle_5 = \pi$, as shown by Figure 12. Guaranteed by linear transformation, we assume that the edges G_1 and G_3 of P_8 are horizontal and vertical, respectively.

Let G_1^* denote the edge of $P_8 + \mathbf{x}_5$, such that $\mathbf{v} \in \mathrm{int}(G_1^*)$ with two ends $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$, let L denote the straight line determined by $\mathbf{v}_1^*$ and $\mathbf{v}_2^*$, let G_3^* denote the edge of $P_8 + \mathbf{x}_4$ lying on L with ends $\mathbf{v}$ and $\mathbf{v}_3^*$, and let G_4^* denote the edge of $P_8 + \mathbf{x}_1$ lying on L with ends $\mathbf{v}$ and $\mathbf{v}_4^*$. If $\varpi(\mathbf{v}_1^*) \geq 3$ or $\varpi(\mathbf{v}_2^*) \geq 3$, by Case 1, Case 2 and Subcase 3.1 we have $\tau(P_8) \geq 6$. Therefore, by considering the adjacent wheels at $\mathbf{v}_1^*$, $\mathbf{v}_2^*$, $\mathbf{v}_3^*$ and $\mathbf{v}_4^*$ successively, $\tau(P_8) \leq 5$ may happen only if

$$\varpi(\mathbf{v}_1^*) = \varpi(\mathbf{v}_2^*) = \varpi(\mathbf{v}_3^*) = \varpi(\mathbf{v}_4^*) = 2. \tag{66}$$

Since the configuration of $P_8 + X^{\mathbf{v}}$ is essentially unique if $\varpi(\mathbf{v}) = 2$, by considering the wheel structures at $\mathbf{v}$, $\mathbf{v}_1^*$, $\mathbf{v}_2^*$, $\mathbf{v}_3^*$ and $\mathbf{v}_4^*$, there are four points $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{y}_3$ and $\mathbf{y}_4$ in X satisfying

$$\mathbf{v}_1^* = \mathbf{v}_4 + \mathbf{y}_1, \mathbf{v} \in \mathrm{int}(P_8) + \mathbf{y}_1, \tag{67}$$

$$\mathbf{v}_2^* = \mathbf{v}_7 + \mathbf{y}_2, \mathbf{v} \in \mathrm{int}(P_8) + \mathbf{y}_2, \tag{68}$$

$$\mathbf{v}_3^* = \mathbf{v}_3 + \mathbf{y}_3, \mathbf{v} \in \mathrm{int}(P_8) + \mathbf{y}_3, \quad \text{and} \tag{69}$$

$$\mathbf{v}_4^* = \mathbf{v}_8 + \mathbf{y}_4, \mathbf{v} \in \mathrm{int}(P_8) + \mathbf{y}_4. \tag{70}$$

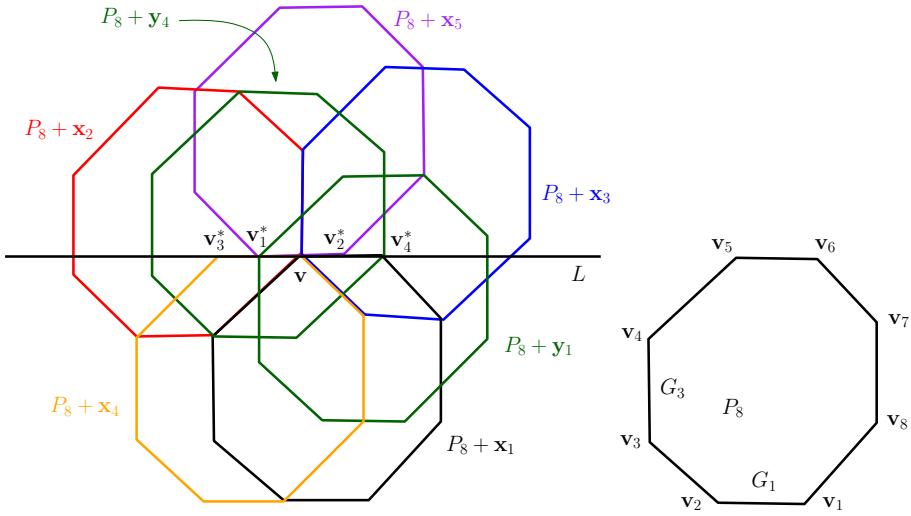

Figure 12

By the convexity of P_8 it follows that $\mathbf{y}_1 \neq \mathbf{y}_2$, $\mathbf{y}_1 \neq \mathbf{y}_3$ and $\mathbf{y}_2 \neq \mathbf{y_4}$. For convenience, we write $\mathbf{v}_i = (x_i, y_i)$. If $\mathbf{y}_1 = \mathbf{y}_4$, and then by (67) and (70) we have

$$y_4 = y_8. \tag{71}$$

If $\mathbf{y}_2 = \mathbf{y}_3$, then by (68) and (69) we have

$$y_3 = y_7. \tag{72}$$

It is obvious that (71) and (72) cannot hold simultaneously. Therefore, we have either $\mathbf{y}_1 \neq \mathbf{y}_4$ or $\mathbf{y}_2 \neq \mathbf{y}_3$.

On the other hand, since $\varpi(\mathbf{v}) = 2$, the three inequalities $\mathbf{y}_3 \neq \mathbf{y}_4$, $\mathbf{y}_2 \neq \mathbf{y}_3$ and $\mathbf{y}_1 \neq \mathbf{y_4}$ cannot hold simultaneously. Otherwise, it can be deduced that

$$\varphi(\mathbf{v}) \geq 4$$

and, therefore,

$$\tau(P_8) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{73}$$

Since $\mathbf{y}_1 = \mathbf{y}_4$ and $\mathbf{y}_2 = \mathbf{y}_3$ are symmetric, with respect to $\mathbf{v}$ in Figure 12, it is sufficient to deal with two subcases.

Subcase 4.1. — $\mathbf{y}_2 = \mathbf{y}_3$. Let $\mathbf{v}_1'$ and $\mathbf{v}_2'$ be the two vertices of $P_8 + \mathbf{x}_2$ that are adjacent to $\mathbf{v}$, as shown in Figure 13. First, we have $\mathbf{v}_1' \in \operatorname{int}(P_8) + \mathbf{x}_5$. Second, by convexity it is easy to see that $\mathbf{v}_1' \in \operatorname{int}(P_8) + \mathbf{y_4}$. Since $\mathbf{y}_2 = \mathbf{y}_3$, we have $y_3 = y_7$. Then $\mathbf{v}_1'$ is an interior point of $P_8 + \mathbf{y_2}$ as well. Thus we get

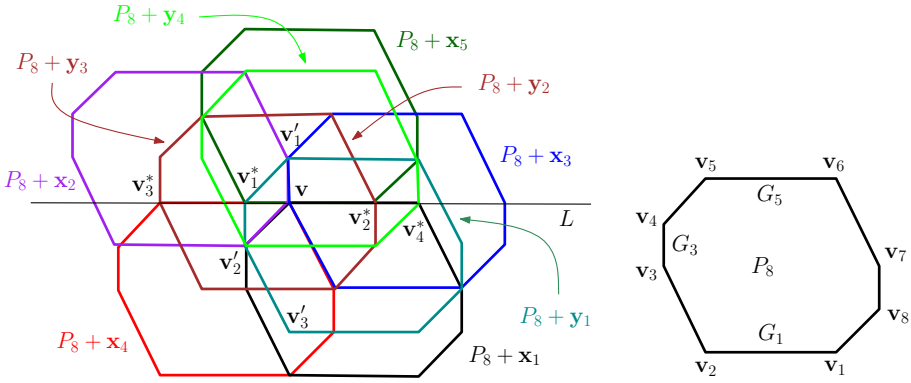

Figure 13

$\varphi(\mathbf{v}_1') \geq 3$ and

$$\tau(P_8) = \varpi(\mathbf{v}_1') + \varphi(\mathbf{v}_1') \geq 5, \tag{74}$$

where the equality may hold only if

$$\varpi(\mathbf{v}_1') = 2. \tag{75}$$

By Lemma 2.1, there is a point $\mathbf{y_5} \in X^{\mathbf{v}_1'}$, such that $\mathbf{v} \in \operatorname{int}(P_8) + \mathbf{y_5}$.

Subcase 4.1.1. — *$\mathbf{v}_1'$ is a vertex of $P_8 + \mathbf{y_5}$.* If $\mathbf{v}_1'$ is a vertex of $P_8 + \mathbf{y_1}$, as shown by Figure 13, then by symmetry one can deduce that $\mathbf{v}_2'$ is a vertex of $P_8 + \mathbf{y_1}$. Similarly, it follows by (70) that $\mathbf{v}_2'$ is a vertex of $P_8 + \mathbf{y}_4$ as well. Then some points near to $\mathbf{v}_2'$ belong to all $\operatorname{int}(P_8) + \mathbf{y}_1$, $\operatorname{int}(P_8) + \mathbf{y}_4$ and $\operatorname{int}(P_8) + \mathbf{x}_1$. Thus, we have

$$\varpi(\mathbf{v}_2') \geq 3. \tag{76}$$

Let $\mathbf{v}_3'$ denote the vertex $\mathbf{v}_2 + \mathbf{y}_1$ of $P_8 + \mathbf{y}_1$, as shown in Figure 13. By Lemma 2.1, there is a point $\mathbf{z} \in X^{\mathbf{v}_3'}$, such that $\mathbf{v}_2' \in \operatorname{int}(P_8) + \mathbf{z}$. Then it can be deduced that $\mathbf{v}_3' \in \operatorname{int}(P_8) + \mathbf{x_4}$ and, thus, $\mathbf{z} \neq \mathbf{x}_4$. Since $y_3 = y_7$, it can be shown that $\mathbf{v}_3' \notin P_8 + \mathbf{y_2}$ and, therefore, $\mathbf{z} \neq \mathbf{y}_2$. In addition, we have

$$\mathbf{v}_2' \in (\operatorname{int}(P_8) + \mathbf{x_4}) \cap (\operatorname{int}(P_8) + \mathbf{y_2}).$$

Thus, we have

$$\varphi(\mathbf{v}_2') \geq 3$$

and, consequently,

$$\tau(P_8) = \varphi(\mathbf{v}_2') + \varpi(\mathbf{v}_2') \geq 6. \tag{77}$$

If $\mathbf{v}_1'$ is not a vertex of $P_8 + \mathbf{y_1}$, remembering the subcase assumption, then we have $\mathbf{y_1} \neq \mathbf{y_5}$. In fact, in this case all $\mathbf{y}_1$, $\mathbf{y}_3$, $\mathbf{y}_4$ and $\mathbf{y}_5$ are pairwise distinct.

Thus, we have

$$\varphi(\mathbf{v}) \geq 4$$

and

$$\tau(P_8) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{78}$$

Subcase 4.1.2. — $\mathbf{v}_1'$ *is an interior point of an edge of* $P_8 + \mathbf{y_5}$. It follows from the convexity of P_8 that $\mathbf{v}_1'$ is an interior point of both $P_8 + \mathbf{y}_4$ and $P_8 + \mathbf{y}_3$. Therefore, we have $\mathbf{y_5} \notin \{\mathbf{y}_3, \mathbf{y}_4\}$. If $\mathbf{y_5} \neq \mathbf{y}_1$, then all $\mathbf{y}_1$, $\mathbf{y}_3$, $\mathbf{y}_4$ and $\mathbf{y}_5$ are pairwise distinct. Thus, we have

$$\varphi(\mathbf{v}) \geq 4$$

and

$$\tau(P_8) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{79}$$

If $\mathbf{y_5} = \mathbf{y}_1$, then all $\mathbf{y}_1$, $\mathbf{y}_3$ and $\mathbf{y}_4$ are pairwise distinct and, therefore,

$$\tau(P_8) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 5. \tag{80}$$

Now, we try to figure out the equality cases in (80).

Notice that $\mathbf{v}_1'$ is an interior point of $P_8 + \mathbf{x}_5$, and $P_8 + \mathbf{y_1}$ has only two edges $G_4 + \mathbf{y}_1$ and $G_5 + \mathbf{y}_1$, which contain interior points of $P_8 + \mathbf{x}_5$. Since $\varpi(\mathbf{v}_1') = 2$ (see (75)), by studying the structure of the adjacent wheel at $\mathbf{v}_1'$, one can deduce that $\mathbf{v}_1'$ must be an interior point of $G_5 + \mathbf{y}_1$. Then we have

$$y_5 - y_4 = y_4 - y_3 \tag{81}$$

and

$$y_3 - y_2 = 2(y_4 - y_3). \tag{82}$$

Let $\mathbf{v}_5^*$ and $\mathbf{v}_6^*$ be the two ends of $G_5 + \mathbf{y}_1$. Suppose that $\mathbf{v}_5^*$ is on the left-hand side of $\mathbf{v}_1'$. By Lemma 2.1, there is a point $\mathbf{y}_6 \in X^{\mathbf{v}_5^*}$, such that $\mathbf{v}_1' \in \operatorname{int}(P_8) + \mathbf{y}_6$.

It is obvious that $\mathbf{v}_5^*$ is an interior point of both $P_8 + \mathbf{x_5}$ and $P_8 + \mathbf{y_2}$. Thus, we have $\mathbf{y}_6 \notin \{\mathbf{y_2}, \mathbf{x_5}\}$. If $\mathbf{v}_5^*$ is not lying on the boundary of $P_8 + \mathbf{y_4}$, then we have $\mathbf{y_4} \neq \mathbf{y_6}$. Consequently, all $\mathbf{y}_2$, $\mathbf{y}_4$, $\mathbf{y}_6$ and $\mathbf{x}_5$ are pairwise distinct. Then we have

$$\varphi(\mathbf{v}_1') \geq 4$$

and

$$\tau(P_8) = \varphi(\mathbf{v}_1') + \varpi(\mathbf{v}_1') \geq 6. \tag{83}$$

Thus, to save $\tau(P_8) = 5$, the point $\mathbf{v}_5^*$ must belong to the boundary of $P_8 + \mathbf{y_4}$. Furthermore, since the y-coordinate of $\mathbf{v}_5^*$ is equal to the y-coordinates of both $\mathbf{v}_1'$ and $\mathbf{v}_3 + \mathbf{y}_4$, the point $\mathbf{v}_5^*$ must be the vertex $\mathbf{v}_3 + \mathbf{y}_4$ of $P_8 + \mathbf{y_4}$.

Let v denote the x-coordinate of $\mathbf{v}$ and let w_1, w_2 and w_3 denote the x-coordinates of $\mathbf{v}_3+\mathbf{y}_4$, $\mathbf{v}_1^*$ and $\mathbf{v}_5^*$, respectively. First, by computing the x-coordinate of $\mathbf{v}_4^*$ in two ways we get

$$w_1+(x_7-x_6)+(x_6-x_5)+(x_5-x_4)=v+(x_6-x_5)$$

and, thus,

$$w_1=v-(x_7-x_6)-(x_5-x_4). \tag{84}$$

On the other hand, since $\mathbf{y}_2=\mathbf{y}_3$, by computing the distance between $\mathbf{v}_3^*$ and $\mathbf{v}_4^*$ in two ways we get

$$(x_7-x_6)+(x_6-x_5)+(x_5-x_4)+v-w_2=2(x_6-x_5)$$

and, thus,

$$w_2=v+(x_7-x_6)-(x_6-x_5)+(x_5-x_4).$$

Since $\mathbf{v}_5^*$ is the left vertex of $G_5+\mathbf{y}_1$, we get

$$w_3=w_2+(x_5-x_4)=v+(x_7-x_6)-(x_6-x_5)+2(x_5-x_4). \tag{85}$$

Then, $\mathbf{v}_5^*=\mathbf{v}_3+\mathbf{y}_4$ implies $w_1=w_3$ and

$$2(x_7-x_6)+3(x_5-x_4)=x_6-x_5. \tag{86}$$

In conclusion, recalling (81) and (82), a centrally symmetric octagon P_8 with G_1 horizontal, G_3 vertical and $\mathbf{y}_2=\mathbf{y}_3$ is a fivefold translative tile only if

$$\begin{cases} y_5-y_4=y_4-y_3,\\ y_3-y_2=2(y_4-y_3),\\ x_6-x_5=2(x_7-x_6)+3(x_5-x_4). \end{cases} \tag{87}$$

Guaranteed by linear transformations, by choosing $y_4-y_3=1$, $x_6-x_5=2$ and $x_5-x_4=\alpha$ and keeping the symmetry in mind, one can deduce that the candidates are the octagons $D_8(\alpha)$ with vertices $\mathbf{v}_1=\left(\frac{3}{2}-\frac{5\alpha}{4},-2\right)$, $\mathbf{v}_2=\left(-\frac{1}{2}-\frac{5\alpha}{4},-2\right)$, $\mathbf{v}_3=\left(\frac{\alpha}{4}-\frac{3}{2},0\right)$, $\mathbf{v}_4=\left(\frac{\alpha}{4}-\frac{3}{2},1\right)$, $\mathbf{v}_5=-\mathbf{v}_1$, $\mathbf{v}_6=-\mathbf{v}_2$, $\mathbf{v}_7=-\mathbf{v}_3$ and $\mathbf{v}_8=-\mathbf{v}_4$, where $0<\alpha<\frac{2}{3}$.

Let $\Lambda(\alpha)$ denote the lattice generated by $\mathbf{u}_1=(2,0)$ and $\mathbf{u}_2=(1+\frac{\alpha}{2},1)$. By Lemma 2.4 it can be shown that $D_8(\alpha)+\Lambda(\alpha)$ is, indeed, a fivefold tiling of $\mathbb{E}^2$.

Subcase 4.2. — $\mathbf{y}_3=\mathbf{y}_4$. First of all, we have $\varphi(\mathbf{v})\geq 3$ and, therefore,

$$\tau(P_8)=\varphi(\mathbf{v})+\varpi(\mathbf{v})\geq 5. \tag{88}$$

Next, we try to figure out the equality cases in (88).

As shown by Figure 14, by (69) and (70) we get

$$y_3=y_8 \tag{89}$$

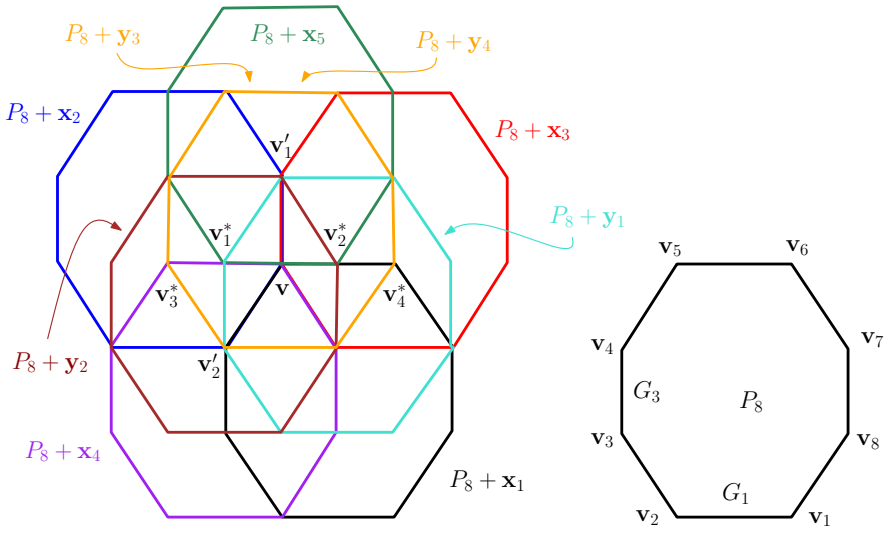

Figure 14

and

$$x_8 - x_3 = 2(x_1 - x_2). \tag{90}$$

By Lemma 2.1, there is $\mathbf{y}_5 \in X^{\mathbf{v}_1'}$, such that $\mathbf{v} \in \operatorname{int}(P_8) + \mathbf{y}_5$. Since $y_3 = y_8$, by convexity we have $\mathbf{v}_1' \in \operatorname{int}(P_8) + \mathbf{y}_3$, and then $\mathbf{y}_5 \neq \mathbf{y}_3$. If both $\mathbf{y}_5 \neq \mathbf{y}_1$ and $\mathbf{y}_5 \neq \mathbf{y}_2$ hold simultaneously, then we have

$$\varphi(\mathbf{v}) \geq 4,$$

and, therefore,

$$\tau(P_8) = \varphi(\mathbf{v}) + \varpi(\mathbf{v}) \geq 6. \tag{91}$$

Thus, the equality in (88) holds only if $\mathbf{y}_5 = \mathbf{y}_1$ or $\mathbf{y}_5 = \mathbf{y}_2$.

Suppose that $\mathbf{y}_5 = \mathbf{y}_1$. If $\mathbf{v}_1'$ is a vertex of $P_8 + \mathbf{y}_1$, then we have

$$y_5 - y_4 = y_4 - y_3. \tag{92}$$

If $\mathbf{v}_1'$ is an interior point of an edge of $P_8 + \mathbf{y}_1$, eliminated by Case 1, Case 2 and Subcase 3.1 at $\mathbf{v}_1'$, it may be assumed that $\varpi(\mathbf{v}_1') = 2$. Then, by studying the structure of the adjacent wheel at $\mathbf{v}_1'$, one can deduce that $\mathbf{v}_1'$ must be an interior point of $G_5 + \mathbf{y}_1$. Since $G_5 + \mathbf{y}_1$ is horizontal, we also obtain (92).

In conclusion, recalling (89) and (90), a centrally symmetric octagon P_8 with G_1 horizontal, G_3 vertical and $\mathbf{y}_3 = \mathbf{y}_4$ is a fivefold translative tile only if

$$\begin{cases} y_3 = y_8, \\ y_5 - y_4 = y_4 - y_3, \\ x_8 - x_3 = 2(x_1 - x_2). \end{cases} \tag{93}$$

Guaranteed by linear transformation, by choosing $y_4 - y_3 = 2$, $x_1 - x_2 = 2$ and $x_6 = \beta$ and keeping symmetry in mind, one can deduce that the candidates are the octagons $D_8(\beta)$ with vertices $\mathbf{v}_1 = (2-\beta, -3)$, $\mathbf{v}_2 = (-\beta, -3)$, $\mathbf{v}_3 = (-2, -1)$, $\mathbf{v}_4 = (-2, 1)$, $\mathbf{v}_5 = -\mathbf{v}_1$, $\mathbf{v}_6 = -\mathbf{v}_2$, $\mathbf{v}_7 = -\mathbf{v}_3$ and $\mathbf{v}_8 = -\mathbf{v}_4$, where $0 < \beta \leq 1$.

Let $\Lambda(\beta)$ denote the lattice generated by $\mathbf{u}_1 = (2, 0)$ and $\mathbf{u}_2 = (1 + \frac{\beta}{2}, 2)$. By Lemma 2.4, it can be proved that $D_8(\beta) + \Lambda(\beta)$ is, indeed, a fivefold tiling of $\mathbb{E}^2$.

Lemma 3.8 is proved. □

Proof of Theorem 1.1. — It has been shown by Gravin, Robins and Shiryaev [10] that a convex body can form a multiple translative tiling in $\mathbb{E}^n$ only if it is a centrally symmetric polytope with centrally symmetric facets. Then, Theorem 1.1 follows from Lemma 3.1, Lemma 3.5 and Lemma 3.8. □

Proof of Theorem 1.2. — Assume that P_{2m} is a centrally symmetric $2m$-gon satisfying $\tau(P_{2m}) = 5$. First, by Fedorov's theorem and Lemma 3.1 we have $4 \leq m \leq 7$. Second, by Lemma 3.3 and Lemma 3.2 we get $m \neq 6$ and 7, respectively. When $m = 5$, the theorem follows by Lemma 3.5 and Lemma 3.6. Finally, when $m = 4$, the theorem follows from Lemma 3.8. □

Acknowledgements. — The authors are grateful to Professor S. Robins, Professor J.Y. Yao, Professor G.M. Ziegler and the referee for their helpful comments and suggestions. We do acknowledge that the introduction of this paper is more or less the same as that in [24] and [29].

BIBLIOGRAPHY

[1] A. D. Aleksandrov – "On tiling space by polytopes", *Vestnik Leningrad Univ Ser. Mat. Fiz. Him.* **9** (1954), p. 33–43.

[2] U. Bolle – "On multiple tiles in R^2", in *Intuitive Geometry*, Colloq. Math. Soc. J. Bolyai, vol. 63, North-Holland, Amsterdam, 1994.

[3] B. N. Delone – "Sur la partition regulière de l'espace à 4 dimensions I, II", *Izv. Akad. Nauk SSSR, Ser. VII* (1929), p. 79–110, 147–164.

[4] M. Dutour Sikirić, A. Garber, A. Schürmann & C. Waldmann – "The complete classification of five-dimensional Dirichlet–Voronoi polyhedra of translational lattices", *Acta Crystallogr. Sect. A* **72** (2016), p. 673–683.
[5] P. Engel – "On the symmetry classification of the four-dimensional parallelohedra", *Z. Kristallographie* **200** (1992), p. 199–213.
[6] E. S. Fedorov – "Elements of the study of figures", *Zap. Mineral. Imper. S. Petersburgskogo Obšč* **21** (1885), no. 2, p. 1–279.
[7] G. Fejes Tóth & W. Kuperberg – "Packing and covering with convex sets", in *Handbook of Convex Geometry* (P. M. Gruber & J. M. Wills, eds.), North-Holland, Amsterdam, 1993, p. 799–860.
[8] P. Furtwängler – "Über Gitter konstanter Dichte", *Monatsh. Math. Phys.* **43** (1936), p. 281–288.
[9] N. Gravin, M. N. Kolountzakis, S. Robins & D. Shiryaev – "Structure results for multiple tilings in 3D", *Discrete Comput. Geom.* **50** (2013), p. 1033–1050.
[10] N. Gravin, S. Robins & D. Shiryaev – "Translational tilings by a polytope, with multiplicity", *Combinatorica* **32** (2012), p. 629–649.
[11] P. M. Gruber & C. G. Lekkerkerker – *Geometry of numbers*, 2 ed., North-Holland, 1987.
[12] G. Hajós – "Über einfache und mehrfache Bedeckung des n-dimensionalen Raumes mit einem Würfelgitter", *Math. Z.* **47** (1941), p. 427–467.
[13] D. Hilbert – "Mathematical problems", *Göttinger Nachr.* (1900), p. 253–297, *Bull. Amer. Math. Soc.* **8** (1902), p. 437–479, **37** (2000), p. 407–436.
[14] M. N. Kolountzakis – "On the structure of multiple translational tilings by polygonal regions", *Discrete Comput. Geom.* **23** (2000), p. 537–553.
[15] C. Mann, J. McLoud-Mann & D. V. Derau – "Convex pentagons that admit i-block transitive tilings", *Geom. Dedicata* **194** (2018), p. 141–167.
[16] P. McMullen – "Convex bodies which tiles space by translation", *Mathematika* **27** (1980), p. 113–121.
[17] H. Minkowski – "Allgemeine Lehrsätze über konvexen Polyeder", *Nachr. K. Ges. Wiss. Göttingen, Math.-Phys. Kl.* (1897), p. 198–219.
[18] M. Rao – "Exhaustive search of convex pentagons which tile the plane", arXiv:1708.00274.
[19] K. Reinhardt – "Über die Zerlegung der Ebene in Polygone", *Dissertation*, Universität Frankfurt am Main.
[20] R. M. Robinson – "Multiple tilings of n-dimensional space by unit cubes", *Math. Z.* **166** (1979), p. 225–275.
[21] B. A. Venkov – "On a class of Euclidean polytopes", *Vestnik Leningrad Univ. Ser. Mat. Fiz. Him.* **9** (1954), p. 11–31.
[22] G. F. Voronoi – "Nouvelles applications des parammètres continus à la théorie des formes quadratiques. Deuxième Mémoire. Recherches sur les

paralléloèdres primitifs", *J. reine angew. Math* **134** (1908), p. 198–287, **135** (1909), 67–181.

[23] M. I. Štogrin – "Regular Dirichlet–Voronoi partitions for the second triclinic group", *Proc. Steklov. Inst. Math.* **123** (1975), in Russian.

[24] Q. Yang & C. Zong – "Multiple lattice tiling in Euclidean spaces", *Canadian Math. Bull.* **62** (2019), p. 923–929.

[25] C. Zong – "What is known about unit cubes", *Bull. Amer. Math. Soc.* **42** (2005), p. 181–211.

[26] ———, *The cube: A window to convex and discrete geometry*, Cambridge University Press, Cambridge, 2006.

[27] ———, "Packing, covering and tiling in two-dimensional spaces", *Expo. Math.* **32** (2014), p. 297–364.

[28] ———, "Can you pave the plane with identical tiles?", *Notices Amer. Math. Soc.* **67** (2020), no. 5, p. 635–646.

[29] ———, "Characterization of the two-dimensional five-fold lattice tiles", arXiv:1712.01122.

Bull. Soc. Math. France
149 (1), 2021, p. 155-177

ON SETS WITH SMALL SUMSET AND m-SUM-FREE SETS IN $\mathbb{Z}/p\mathbb{Z}$

BY PABLO CANDELA, DIEGO GONZÁLEZ-SÁNCHEZ & DAVID J. GRYNKIEWICZ

ABSTRACT. — The $3k-4$ conjecture in groups $\mathbb{Z}/p\mathbb{Z}$ for p prime states that if A is a nonempty subset of $\mathbb{Z}/p\mathbb{Z}$ satisfying $2A \neq \mathbb{Z}/p\mathbb{Z}$ and $|2A| = 2|A| + r \leq \min\{3|A| - 4,\ p - r - 4\}$, then A is covered by an arithmetic progression of size at most $|A| + r + 1$. Previously, the best result toward this conjecture, without any additional constraint on $|A|$, was a theorem of Serra and Zémor proving the conjecture provided $r \leq 0.0001|A|$. Subject to the mild additional constraint $|2A| \leq 3p/4$, which is optimal in the sense explained in the paper, our first main result improves the bound on r, allowing $r \leq 0.1368|A|$. We also prove a variant that further improves this bound on r provided that A is sufficiently dense. We then give several applications. First, we apply the above variant to give a new upper bound for the maximal density of m-sum-free sets in $\mathbb{Z}/p\mathbb{Z}$, i.e., sets A having no solution $(x, y, z) \in A^3$ to the equation $x + y = mz$, where $m \geq 3$ is a fixed integer. The previous best upper bound for this maximal density was $1/3.0001$ (using the Serra-Zémor theorem). We improve this to $1/3.1955$. We also present a construction following an idea of Schoen, which yields a lower bound for this maximal density of the form $1/8 + o(1)_{p\to\infty}$. Another application of our main results concerns sets of the form $\frac{A+A}{A}$ in $\mathbb{F}_p$, and we also improve the structural description of large sum-free sets in $\mathbb{Z}/p\mathbb{Z}$.

Texte reçu le 11 septembre 2019, modifié le 12 mars 2020, accepté le 20 octobre 2020.

PABLO CANDELA, Universidad Autónoma de Madrid, and ICMAT, Madrid 28049, Spain • *E-mail :* pablo.candela@uam.es

DIEGO GONZÁLEZ-SÁNCHEZ, Universidad Autónoma de Madrid, and ICMAT, Madrid 28049, Spain • *E-mail :* diego.gonzalezs@predoc.uam.es

DAVID J. GRYNKIEWICZ, University of Memphis, Department of Mathematical Sciences, Memphis, TN 38152 • *E-mail :* diambri@hotmail.com

Mathematical subject classification (2010). — 11P70, 11B13, 05B10.

Key words and phrases. — Additive combinatorics, Small sumset, m-sum-free set, Freiman's $3k-4$ theorem, $3k-4$ conjecture.

This work has benefited from support from the Spanish Ministerio de Ciencia e Innovación project MTM2017-83496-P and from the La Caixa Foundation (ID 100010434) under agreement LCF/BQ/SO16/52270027.

0037-9484/2021/155/$ 5.00
doi:10.24033/bsmf.2827

Résumé (*Sur les ensembles de petite somme et les ensembles sans m-somme dans* $\mathbb{Z}/p\mathbb{Z}$). — La conjecture $3k-4$ dans les groupes $\mathbb{Z}/p\mathbb{Z}$, pour p premier, affirme que si A est un sous-ensemble non vide de $\mathbb{Z}/p\mathbb{Z}$ vérifiant $2A \neq \mathbb{Z}/p\mathbb{Z}$ et $|2A| = 2|A| + r \leq \min\{3|A| - 4,\ p - r - 4\}$, alors A est inclus dans une suite arithmétique de cardinalité au plus $|A|+r+1$. Le meilleur résultat précédent vers cette conjecture, sans contraintes supplémentaires sur $|A|$, est un théorème de Serra et Zémor qui confirme la conjecture pour $r \leq 0.0001|A|$. Sous la faible contrainte additionnelle $|2A| \leq 3p/4$, qui est optimale en un sens détaillé dans l'article, notre premier résultat principal améliore la borne supérieure sur r, permettant de prendre $r \leq 0.1368|A|$. Nous démontrons aussi une variante qui améliore davantage la borne sur r pour tout ensemble A suffisamment dense. Nous présentons ensuite plusieurs applications. Premièrement, la variante en question est employée pour obtenir une nouvelle borne supérieure pour la densité maximale des ensembles sans m-somme dans $\mathbb{Z}/p\mathbb{Z}$, i.e., les ensembles A tels qu'il n'existe aucune solution $(x, y, z) \in A^3$ de l'équation $x + y = mz$, où $m \geq 3$ est un entier fixé. Précédemment, la meilleure borne supérieure pour cette densité maximale était 1/3.0001 (comme conséquence du théorème de Serra–Zémor). Nous obtenons ici la borne améliorée 1/3.1955. Nous présentons aussi une construction suivant une idée de Schoen, qui fournit une borne inférieure $1/8 + o(1)_{p\to\infty}$ pour la densité maximale en question. Une autre application de nos résultats concerne les ensembles de la forme $\frac{A+A}{A}$ dans $\mathbb{F}_p$. Nous donnons aussi une description améliorée de la structure des grands ensembles sans somme dans $\mathbb{Z}/p\mathbb{Z}$.

1. Introduction

Given a subset A of an abelian group G, we often denote the sumset $A+A = \{x + y : x, y \in A\}$ by $2A$ and we denote the complement $G \setminus A$ by $\overline{A}$.

One of the central topics in additive number theory is the study of the structure of a finite subset A of an abelian group under the assumption that the sumset $2A$ is small. In this paper, we focus on groups $\mathbb{Z}/p\mathbb{Z}$ of integers modulo a prime p and on the regime in which the *doubling constant* $|2A|/|A|$ is within a small additive constant of the minimum possible value.

To put this into context, let us recall the basic fact that a finite set A of integers always satisfies $|2A| \geq 2|A| - 1$ and that this minimum is attained only if A is an arithmetic progression (see [12, Theorem 3.1]). This description of extremal sets is extended by a result of Freiman, known as the $3k-4$ theorem, which tells us that A is still efficiently covered by an arithmetic progression even when $|2A|$ is as large as $3|A| - 4$.

Theorem 1.1 (Freiman's $3k-4$ theorem). — *Let $A \subseteq \mathbb{Z}$ be a finite set satisfying $|2A| \leq 3|A| - 4$. Then there is an arithmetic progression $P \subseteq \mathbb{Z}$, such that $A \subseteq P$ and $|P| \leq |2A| - |A| + 1$.*

For sets A in $\mathbb{Z}/p\mathbb{Z}$ with $2A \neq \mathbb{Z}/p\mathbb{Z}$, the Cauchy–Davenport theorem [12, Theorem 6.2] gives the lower bound analogous to the one for $\mathbb{Z}$ mentioned above, namely $|2A| \geq 2|A| - 1$, and the description of extremal sets as arithmetic

progressions (when $|2A| < p-1$) is given by Vosper's theorem [12, Theorem 8.1].

It is widely believed that an analogue of Freiman's $3k-4$ theorem holds for subsets of $\mathbb{Z}/p\mathbb{Z}$ under some mild additional upper bound on $|2A|$ (or on $|A|$). More precisely, the following conjecture is believed to be true (see [12, Conjecture 19.2]), describing efficiently not just A but also $2A$, in terms of progressions.

CONJECTURE 1.2. — *Let p be a prime and let $A \subset \mathbb{Z}/p\mathbb{Z}$ be a nonempty subset satisfying $2A \neq \mathbb{Z}/p\mathbb{Z}$ and $|2A| = 2|A| + r \leq \min\{3|A| - 4,\ p - r - 4\}$. Then there exist arithmetic progressions $P_A, P_{2A} \subseteq \mathbb{Z}/p\mathbb{Z}$ with the same difference, such that $A \subseteq P_A$, $|P_A| \leq |A| + r + 1$, $P_{2A} \subseteq 2A$, and $|P_{2A}| \geq 2|A| - 1$.*

Progress toward this conjecture was initiated by Freiman himself, who proved in [10] that the conclusion concerning P_A holds provided that $|2A| \leq 2.4|A| - 3$ and $|A| < p/35$. Since then, there has been much work improving Freiman's result in various ways. For instance, Rødseth showed in [17] that the constraint $|A| < p/35$ can be weakened to $|A| < p/10.7$ while maintaining the doubling constant 2.4. In [11], Green and Ruzsa pushed the doubling constant up to 3, at the cost of a stronger constraint $|A| < p/10^{215}$. In [20], Serra and Zémor obtained a result with no constraint on $|A|$ other than the bounds on $|2A|$ in the conjecture, with the same conclusion concerning P_A but at the cost of reducing the doubling constant, namely, assuming that $|2A| \leq (2+\alpha)|A|$ with $\alpha < 0.0001$. See also [5, 14] for recent improvements on the doubling constant 2.4 in Freiman's result. The book [12] presents various other results towards Conjecture 1.2, in a treatment covering many of the methods from the works mentioned above.

In this paper, we establish the following new result regarding Conjecture 1.2, which noticeably improves the doubling constant obtained by Serra and Zémor in [20] at the cost of only adding the constraint $|2A| \leq \frac{3}{4}p$.

THEOREM 1.3. — *Let p be prime, let $A \subseteq \mathbb{Z}/p\mathbb{Z}$ be a nonempty subset with $|2A| = 2|A| + r$, and let $\alpha \approx 0.136861$ be the unique real root of the cubic $4x^3 + 9x^2 + 6x - 1$. Suppose*

$$|2A| \leq (2+\alpha)|A| - 3 \quad \text{and} \quad |2A| \leq \frac{3}{4}p.$$

Then there exist arithmetic progressions $P_A, P_{2A} \subseteq \mathbb{Z}/p\mathbb{Z}$ with the same difference, such that $A \subseteq P_A$, $|P_A| \leq |A| + r + 1$, $P_{2A} \subseteq 2A$, and $|P_{2A}| \geq 2|A| - 1$.

Unlike in [20], here we do have a constraint on $|A|$ in the form of the upper bound $|2A| \leq \frac{3}{4}p$. However, this upper bound is still optimal in the following weak sense. The conjectured upper bound on $|2A|$ (given by Conjecture 1.2) is $p - r - 4$. However, in the extremal case where $r = |A| - 4$ (the largest value of r allowed in Conjecture 1.2), the conjectured bound implies $3|A| - 4 = |2A| \leq$

$p - |A|$, whence $|A| \le \frac{p+4}{4}$ and $|2A| = 3|A| - 4 \le \frac{3p}{4} - 1$. Thus, the bound $p - r - 4$ becomes as small as $\frac{3p}{4} - 1$, as we *range* over all allowed values for α and $|A|$, making $\frac{3}{4}p$ the optimal bound independent of α and r.

Let us emphasize that our improvement upon the Serra–Zémor result (i.e., our weakening of the constraint on α) is valid for $|A| \le \frac{0.75p+3}{2+\alpha}$, whereas the natural upper bound on $|A|$ given by Conjecture 1.2 is larger, namely $|A| \le \frac{p+2}{2+2\alpha}$. Therefore, in the regime $\frac{0.75p+3}{2+\alpha} < |A| \le \frac{p+2}{2+2\alpha}$, our result does not improve on that of Serra and Zémor.

We also prove the following variant of Theorem 1.3, which is optimized for sets A whose density is large but at most $1/3$. This optimization is designed for an application concerning m-sum-free sets, which we discuss below.

THEOREM 1.4. — *Let p be prime, let $\eta \in (0,1)$, let $A \subseteq \mathbb{Z}/p\mathbb{Z}$ be a set with $|A| \ge \eta\, p > 0$ and $|2A| = 2|A| + r < p$, and let*

$$\alpha = -\frac{5}{4} + \frac{1}{4}\sqrt{9 + 8\,\eta\, p \sin(\pi/p)/\sin(\pi\eta/3)}.$$

Suppose

$$|2A| \le (2+\alpha)|A| - 3 \quad \textit{and} \quad |A| \le \frac{p-r}{3}.$$

Then there exist arithmetic progressions $P_A, P_{2A} \subseteq \mathbb{Z}/p\mathbb{Z}$ with the same difference such that $A \subseteq P_A$, $|P_A| \le |A| + r + 1$, $P_{2A} \subseteq 2A$, and $|P_{2A}| \ge 2|A| - 1$.

We apply this result to obtain new upper bounds for the size of m-sum-free sets in $\mathbb{Z}/p\mathbb{Z}$. For a positive integer m, a subset A of an abelian group is said to be *m-sum-free* if there is no triple $(x, y, z) \in A^3$ satisfying $x + y = mz$. These sets have been studied in numerous works on arithmetic combinatorics, including various types of abelian group settings [1, 8, 7, 16, 15] (see also [4, Section 3] for an overview of this topic). In $\mathbb{Z}/p\mathbb{Z}$, a central goal concerning these sets is to estimate the quantity

$$d_m(\mathbb{Z}/p\mathbb{Z}) = \max\big\{\tfrac{|A|}{p} : A \subseteq \mathbb{Z}/p\mathbb{Z} \text{ } m\text{-sum-free}\big\}. \tag{1}$$

This goal splits naturally into two problems of different nature. On the one hand, we have the case $m = 2$, which is the only one in which the solutions of the linear equation in question (i.e., three-term arithmetic progressions) form a translation invariant set. Roth's theorem [18] tells us that $d_2(\mathbb{Z}/p\mathbb{Z}) \to 0$ as $p \to \infty$, and the problem in this case is then the well-known one of determining the optimal bounds for Roth's theorem, i.e., how fast $d_2(\mathbb{Z}/p\mathbb{Z})$ vanishes as p increases (recent developments in this direction include [3, 19]). On the other hand, we have the cases $m \ge 3$. For each of these, the above-mentioned translation invariance fails, and it is known that $d_m(\mathbb{Z}/p\mathbb{Z})$ converges, as $p \to \infty$ through primes, to a positive constant d_m that can be modeled on the circle group (see [6]), the problem then being to determine this constant. Our application of Theorem 1.4 makes progress on the latter problem.

Note that, if A is m-sum-free, then the *dilate* $m\cdot A = \{mx : x \in A\} \subseteq \mathbb{Z}/p\mathbb{Z}$ satisfies $2A \cap m\cdot A = \emptyset$, whence, if m and p are coprime, we have $|2A|+|m\cdot A| = |2A| + |A| \leq p$. Combining this with the bound $|2A| \geq 2|A| - 1$ given by the Cauchy–Davenport Theorem, we deduce the simple bound $|A| \leq \frac{p+1}{3}$, which implies in particular that $d_m \leq 1/3$. It was noted in [4] that partial versions of Conjecture 1.2 can be used to improve on this bound, provided that these versions are applicable to sets of density up to $1/3$. The best version available for that purpose in [4] was given by the theorem of Serra and Zémor mentioned above, and this resulted in the first upper bound for d_m below $1/3$, namely $1/3.0001$ (see [4, Theorem 3.1]). In this paper, using Theorem 1.4 we obtain the following improvement.

THEOREM 1.5. — *Let $p \geq 80$ be a prime, let m be an integer in $[2, p-2]$, and let $c = c(p)$ be the solution to the equation $\left(7 + \sqrt{8\,c\,p \sin(\pi/p)/\sin(\pi c/3) + 9}\right) c = 4 + \frac{12}{p}$. Then $d_m(\mathbb{Z}/p\mathbb{Z}) < c$. In particular, $d_m \leq \frac{1}{3.1955}$.*

The following observation, relating this theorem to the study of sum-products in the field $\mathbb{F}_p$, was made by the anonymous referee: if $(A+A) \cap m\cdot A$ contains a nonzero element and $0 \notin A$, then m is in the set $\frac{A+A}{A} := \{(a_1 + a_2)a_3^{-1} : a_1, a_2, a_3 \in A\} \subset \mathbb{F}_p$, and, therefore, Theorem 1.5 has the following consequence.

COROLLARY 1.6. — *If $A \subset \mathbb{F}_p \setminus \{0\}$ satisfies $|A| \geq 0.313\,p$, then for p sufficiently large we have $\mathbb{F}_p \setminus \{-1, 0, 1\} \subseteq \frac{A+A}{A}$.*

This result is an analogue, for sets $\frac{A+A}{A}$, of Theorem 1.1 in [2], which says that if $A \subset \mathbb{F}_p$ has $|A| \geq 0.3051\,p$, then for p sufficiently large, we have $\mathbb{F}_p \setminus \{0\} \subseteq (A + A)A := \{(a_1 + a_2)a_3 : a_i \in A\}$.

Regarding lower bounds for $d_m(\mathbb{Z}/p\mathbb{Z})$, note that, identifying $\mathbb{Z}/p\mathbb{Z}$ with the integers $[0, p-1]$, the interval $(\frac{2}{m^2-4}p, \frac{m}{m^2-4}p)$ is an m-sum-free set. This set has asymptotic density $\frac{1}{m+2}$ and is still the greatest known example for $m \leq 6$. However, for larger values of m, a construction of Tomasz Schoen (personal communication), presented in this paper in Lemma 3.1 in an optimized form thanks to indications of the anonymous referee, yields the improved lower bound $d_m \geq \frac{1}{8}$. The following theorem summarizes these results.

THEOREM 1.7. — *For $m \leq 6$, we have $d_m \geq \frac{1}{m+2}$. For $m \geq 7$, we have $d_m \geq \frac{1}{8}$.*

Our final application concerns the study of large sum-free sets in $\mathbb{Z}/p\mathbb{Z}$ (i.e. the case $m = 1$ of m-sum-free sets as defined above). It is well-known, by the argument using the Cauchy–Davenport theorem mentioned above, that a sum-free set in $\mathbb{Z}/p\mathbb{Z}$ has size at most $\lfloor (p+1)/3 \rfloor$ and that this bound is attained by the interval $I = (p/3, 2p/3) \subset \mathbb{Z}/p\mathbb{Z}$ and by any nonzero dilate

of I. Several works have studied the question of the robustness of this structural description, namely, whether every sum-free set in $\mathbb{Z}/p\mathbb{Z}$ of density close to $1/3$ must resemble a dilate of I. In this direction, the following theorem was proved by Deshouillers and Lev in [9].

THEOREM 1.8. — *Let p be a sufficiently large prime and suppose that $A \subset \mathbb{Z}/p\mathbb{Z}$ is sum-free. If $|A| > 0.318\,p$, then there exists $d \in \mathbb{Z}$, such that $A \subset d \cdot [\,|A|, p - |A|\,]$.*

Applying Theorem 1.4, we improve the constant 0.318 to 0.313.

The paper is laid out as follows. In Section 2, we prove Theorems 1.3 and 1.4. Our results on m-sum-free sets are proved in Section 3. In Section 3.1, we present the above construction and deduce Theorem 1.7. In Section 3.2, we apply Theorem 1.4 to obtain Theorem 1.5. Finally, in Section 3.3, we obtain the above-mentioned improvement of Theorem 1.8.

2. New bounds toward the $3k-4$ conjecture in $\mathbb{Z}/p\mathbb{Z}$

Our first task in this section is to prove Theorem 1.3. We shall obtain this result as the special case $\varepsilon = 3/4$ of the following theorem.

THEOREM 2.1. — *Let p be prime, let $0 < \varepsilon \leq \frac{3}{4}$ be a real number, let α be the unique positive root of the cubic $4x^3 + (12 - 4\varepsilon)x^2 + (9 - 4\varepsilon)x + (8\varepsilon - 7)$, and let $A \subseteq \mathbb{Z}/p\mathbb{Z}$ be a nonempty subset with $|2A| = 2|A| + r$. Suppose*

$$|2A| \leq (2 + \alpha)|A| - 3 \quad \textit{and} \quad |2A| \leq \varepsilon\, p.$$

Then there exist arithmetic progressions $P_A, P_{2A} \subseteq \mathbb{Z}/p\mathbb{Z}$ with the same difference, such that $A \subseteq P_A$, $|P_A| \leq |A| + r + 1$, $P_{2A} \subseteq 2A$, and $|P_{2A}| \geq 2|A| - 1$.

The proof is a modification of the argument used to prove [12, Theorem 19.3], itself based on the original work of Freiman [10] and incorporating improvements to the calculations noted by Rødseth [17]. The main new contribution is an argument to allow the restriction $|2A| \leq \frac{1}{2}(p+3)$ from [12, Theorem 19.3] to be replaced by the above condition $|2A| \leq \varepsilon p$. For $\varepsilon = 3/4$, this is optimal in the sense explained in the Introduction.

In the proof of Theorem 2.1, we use the following version of the $3k-4$ theorem for $\mathbb{Z}$. Here, for $X \subseteq \mathbb{Z}$, we denote the greatest common divisor $\gcd(X - X)$ by $\gcd^*(X)$. Note, for $|X| \geq 2$, that $d = \gcd^*(X)$ is the minimal $d \geq 1$, such that X is contained in an arithmetic progression with difference d. We remark that, when $B = -A$, we have $P_{A-A} \subseteq A - A$ and $-P_{A-A} \subseteq -(A-A) = A - A$. Since $2|P_{A-A}| \geq 4|A| - 2 > |A - A|$, the progressions P_{A-A} and $-P_{A-A}$ intersect, ensuring that $P = P_{A-A} \cup -P_{A-A} \subseteq A - A$ is a progression contained in $A - A$ with $|P| \geq 2|A| - 1$ and $-P = P$. Thus, the progression P_{A-A} in Theorem 2.2 can be assumed to be symmetric (i.e., centered at the origin) when $B = -A$.

THEOREM 2.2. — *Let A, $B \subseteq \mathbb{Z}$ be finite, nonempty subsets with* $\gcd^*(A+B) = 1$ *and*

$$|A+B| = |A| + |B| + r \leq |A| + |B| + \min\{|A|, |B|\} - 3 - \delta,$$

where $\delta = 1$ if $x + A = B$ for some $x \in \mathbb{Z}$, and otherwise $\delta = 0$. Then there are arithmetic progressions P_A, P_B, $P_{A+B} \subseteq \mathbb{Z}$ with common difference 1*, such that $A \subseteq P_A$, $B \subseteq P_B$, $P_{A+B} \subseteq A + B$, $|P_A| \leq |A| + r + 1$, $|P_B| \leq |B| + r + 1$ and $|P_{A+B}| \geq |A| + |B| - 1$.*

Let G and G' be abelian groups and let A, $B \subseteq G$. A *Freiman isomorphism* is a well-defined map $\psi : A + B \to G'$ defined by two coordinate maps $\psi_A : A \to G'$ and $\psi_B : B \to G'$, such that $\psi(x + y) = \psi_A(x) + \psi_B(y)$ for all $x \in A$ and $y \in B$. That ψ is well-defined is equivalent to the statement that $\psi_A(x_1) + \psi_B(y_1) = \psi_A(x_2) + \psi_B(y_2)$ whenever $x_1 + y_1 = x_2 + y_2$, for x_1, $x_2 \in A$ and y_1, $y_2 \in B$, and $\psi_A(A)+\psi_B(B)$ is then the homomorphic image of $A + B$. It is an isomorphism if ψ is injective on $A + B$, which is equivalent to $\psi_A(x_1)+\psi_B(y_1) = \psi_A(x_2)+\psi_B(y_2)$ holding if and only if $x_1+y_1 = x_2+y_2$, for x_1, $x_2 \in A$ and y_1, $y_2 \in B$. We denote this by $A + B \cong \psi_A(A) + \psi_B(B)$. A Freiman homomorphism $\psi : A + B \to G'$ on the sumset defines a Freiman homomorphism $\psi' : A - B \to G'$ on the difference set given by $\psi'(x - y) = \psi_A(x) - \psi_B(y)$, for $x \in A$ and $y \in B$, which is an isomorphism when ψ is. In the special case when $A = B$, we find that $\psi_A(x) + \psi_B(y) = \psi_A(y) + \psi_B(x)$ for all x, $y \in A = B$, implying $\psi_B(x) = \psi_A(x) + (\psi_B(y) - \psi_A(y))$ for any x, $y \in A = B$. Fixing $y \in A$ and letting x range over all possible $x \in A$ shows that the map ψ_B is simply a translate of the map ψ_A. This means it can (and generally will) be assumed that $\psi_A = \psi_B$ for a Freiman homomorphism ψ when $A = B$. See [12, Chapter 20] for a fuller discussion regarding Freiman homomorphisms.

For a prime p, nonzero $g \in \mathbb{Z}/p\mathbb{Z}$ (which is then a generator of $\mathbb{Z}/p\mathbb{Z}$), and integers $m \leq n$, let

$$[m, n]_g = \{mg, (m + 1)g, \ldots, ng\}$$

denote the corresponding interval in $\mathbb{Z}/p\mathbb{Z}$. If $m > n$, then $[m, n]_g = \emptyset$. We define (for each $g \in \mathbb{Z}/p\mathbb{Z} \setminus \{0\}$) a function ℓ_g from the set of subsets $X \subset \mathbb{Z}/p\mathbb{Z}$ to $\mathbb{Z}_{\geq 0}$, by

$$\ell_g(X) := \min\{|P| : P \text{ is an arithmetic progression of difference } g \text{ with } X \subset P\}.$$

We let $\overline{X} = (\mathbb{Z}/p\mathbb{Z}) \setminus X$ denote the complement of X in $\mathbb{Z}/p\mathbb{Z}$. We say that a sumset $A + B \subseteq \mathbb{Z}/p\mathbb{Z}$ *is rectifiable*, if $\ell_g(A) + \ell_g(B) \leq p + 1$ for some nonzero $g \in \mathbb{Z}/p\mathbb{Z}$. In such a case, $A \subseteq a_0 + [0, m]_g$ and $B \subseteq b_0 + [0, n]_g$ with $m + n = \ell_g(A) + \ell_g(B) - 2 \leq p - 1$, for some a_0, $b_0 \in \mathbb{Z}/p\mathbb{Z}$, in which case the maps $a_0 + sg \mapsto s$ and $b_0 + tg \mapsto t$, for s, $t \in \mathbb{Z}$, when restricted to A and B, respectively, show that the sumset $A + B$ is Freiman isomorphic (see

[12, Section 2.8]) to an integer sumset. This allows us to canonically apply results from $\mathbb{Z}$ to the sumset $A+B$.

If G is an abelian group and $A, B \subseteq G$ are subsets, then we say that A is *saturated* with respect to B, if $(A \cup \{x\}) + B \neq A + B$ for all $x \in \overline{A}$. In the proof of Theorem 2.1, we shall also use the following basic result regarding saturation [12, Lemma 7.2], whose earlier form dates back to Vosper [21]. We include the short proof for completeness.

LEMMA 2.3. — *Let G be an abelian group and let $A, B \subseteq G$ be subsets. Then*

$$-B + \overline{A+B} \subseteq \overline{A}$$

with equality holding if and only if A is saturated with respect to B.

Proof. — First observe that $-B + \overline{A+B} \subseteq \overline{A}$, for if $b \in B$, $z \in \overline{A+B}$, and by contradiction $-b+z = a$ for some $a \in A$, then $z = a+b \in A+B$, contrary to its definition. If A is saturated with respect to B, then given any $x \in \overline{A}$, there exists some $b \in B$ and $z \in \overline{A+B}$ with $x+b = z$, whence $x = -b+z \in -B+\overline{A+B}$. This shows that $\overline{A} \subseteq -B+\overline{A+B}$, and as the reverse inclusion always holds (as was just shown), it follows that $\overline{A} = -B + \overline{A+B}$. Conversely, if $\overline{A} = -B+\overline{A+B}$, then given any $x \in \overline{A}$, there exists some $b \in B$ and $z \in \overline{A+B}$ with $x = -b+z$, implying $x+b = z \notin A+B$. Since $x \in \overline{A}$ is arbitrary, this shows that A is saturated with respect to B. □

Proof of Theorem 2.1. — Let $f(x) = 4x^3 + (12-4\varepsilon)x^2 + (9-4\varepsilon)x + (8\varepsilon - 7)$, so that $f'(x) = 12x^2 + (24-8\varepsilon)x + (9-4\varepsilon)$. Then $f'(x) > 0$ for $x \geq 0$ (in view of $\varepsilon \leq 3/4$), meaning that $f(x)$ is an increasing function for $x \geq 0$ with $f(0) = 8\varepsilon - 7 < 0$ and $f(1/2) = 1 + 5\varepsilon > 0$. Consequently, $f(x)$ has a unique positive root $0 < \alpha < \frac{1}{2}$.

Since $|2A| \leq \varepsilon p < p$, the Cauchy–Davenport theorem implies $r \geq -1$. Let

$$\beta = \frac{r+3}{|A|} > 0, \tag{2}$$

so that

$$r = \beta|A| - 3, \quad |2A| = 2|A| + r = (2+\beta)|A| - 3 \quad \text{and} \quad \beta \leq \alpha < \frac{1}{2}. \tag{3}$$

Since $2|A| + r = |2A| \leq \varepsilon p \leq \frac{3}{4}p$, it follows that $|A| \leq \frac{3}{8}p - \frac{1}{2}r$, and since $r \geq -1$, we deduce that

$$|A| \leq \frac{3p+4}{8}. \tag{4}$$

The proof naturally breaks into two parts: a first case where there is a large rectifiable subsumset and a second case where there is not. The latter case will lead to a contradiction.

Case 1. — Suppose there exist subsets $A' \subseteq A$ and $B' \subseteq A$ with $|B'| \le |A'|$ and

$$|A'| + 2|B'| - 4 \ge |2A| \tag{5}$$

such that $A' + B'$ is rectifiable. Furthermore, choose a pair of subsets $A' \subseteq A$ and $B' \subseteq A$ with these properties, such that $|A'|+|B'|$ is maximal, and for these subsets A' and B', let $g \in \mathbb{Z}/p\mathbb{Z}$ be a nonzero difference with $\ell_g(A') + \ell_g(B') \le p+1$ minimal. Note that $|A'| \ge |B'| \ge 2$; indeed, if $|B'| \le 1$, then combining this with the hypotheses $|B'| \le |A'| \le |A|$ and (5) yields the contradiction $|A| - 2 \ge |2A| \ge |A|$. Since $A' + B'$ is rectifiable, the Cauchy–Davenport theorem for $\mathbb{Z}$ [12, Theorem 3.1] ensures

$$|A' + B'| = |A'| + |B'| + r' \quad \text{with } r' \ge -1.$$

Moreover, we have

$$\begin{aligned} &A' \subseteq P_A := a_0 + [0,m]_g, \quad B' \subseteq P_B := b_0 + [0,n]_g, \quad \text{and} \\ &A' + B' \subseteq a_0 + b_0 + [0, m+n]_g, \end{aligned} \tag{6}$$

with $a_0,\, a_0 + mg \in A'$, $b_0,\, b_0 + ng \in B'$ and $m + n \le p - 1$, for some $a_0,\, b_0 \in \mathbb{Z}/p\mathbb{Z}$. Then, since $A' + B'$ is rectifiable, it follows that the map $\psi : \mathbb{Z}/p\mathbb{Z} \to [0, p-1] \subseteq \mathbb{Z}$ defined by $\psi(sg) = s$ for $s \in [0, p-1]$ gives a Freiman isomorphism of $A'+B'$ with the integer sumset $\psi(-a_0+A')+\psi(-b_0+B') \subseteq \mathbb{Z}$. Observe that

$$\gcd^*(\psi(-a_0 + A') + \psi(-b_0 + B')) = 1,$$

since if $\psi(-a_0+A')+\psi(-b_0+B')$ were contained in an arithmetic progression with difference $d \ge 2$, then this would also be the case for $\psi(-a_0 + A')$ and $\psi(-b_0+B')$, and then $\ell_{dg}(A')+\ell_{dg}(B') < \ell_g(A')+\ell_g(B')$ would follow in view of $|A'| \ge |B'| \ge 2$, contradicting the minimality of $\ell_g(A') + \ell_g(B')$ for g.

In view of (5) and $|B'| \le |A'|$, we have $|A' + B'| \le |2A| \le |A'| + |B'| + \min\{|A'|, |B'|\} - 4$. Thus, since $\gcd^*(\psi(-a_0 + A') + \psi(-b_0 + B')) = 1$, we can apply the $3k-4$ theorem (Theorem 2.2) to the isomorphic sumset $\psi(-a_0 + A') + \psi(-b_0 + B')$. Then, letting $P_A = a_0 + [0,m]_g$, $P_B = b_0 + [0,n]_g$ and letting $P_{A+B} \subseteq A' + B'$ be the resulting arithmetic progressions with common difference g, we conclude that

$$|P_A \setminus A'| \le r' + 1 \quad \text{und} \quad |P_B \setminus B'| \le r' + 1. \tag{7}$$

If $A' = A$ and $B' = A$, then the original sumset $2A$ is rectifiable, we have $r' = r$, and the theorem follows with $P_A = P_B$ and $P_{2A} = P_{A+B}$ as just defined. Therefore, we can assume otherwise, which in view of $|B'| \le |A'|$ means

$$A \setminus B' \ne \emptyset. \tag{8}$$

Let $\Delta = |2A| - |A' + B'| \geq 0$. Then

$$r' = |A \setminus A'| + |A \setminus B'| + r - \Delta. \tag{9}$$

Since $|A'| + |B'| + r' = |A' + B'| = |2A| - \Delta$, it follows from (5) and $|B'| \leq |A'|$ that

$$r' \leq |B'| - 4 - \Delta \quad \text{and} \quad r' \leq |A'| - 4 - \Delta. \tag{10}$$

Averaging both bounds in (10), using (9), and recalling that $|2A| = 2|A| + r$, we obtain

$$r' \leq \frac{1}{3}|2A| - \frac{8}{3} - \Delta. \tag{11}$$

Step A. — $| - A' + \overline{A' + A}| \leq |\overline{A' + A}| + 2|A'| - 4$.

Proof. — If Step A fails, then combining its failure with $p - |2A| = |\overline{2A}| \leq |\overline{A' + A}|$ and Lemma 2.3 yields

$$p - |2A| + 2|A'| - 3 \leq |\overline{A' + A}| + 2|A'| - 3 \leq | - A' + \overline{A' + A}| \leq |\overline{A}| = p - |A|,$$

which implies that $|A| + 2|A'| - 3 \leq |2A|$. This together with (5) and $|B'| \leq |A'| \leq |A|$ implies $|A| + 2|A'| - 3 \leq |A'| + 2|B'| - 4 \leq |A| + 2|A'| - 4$, which is not possible. □

Step B. — $| - A' + \overline{A' + A}| \leq |A'| + 2|\overline{A' + A}| - 3$.

Proof. — If Step B fails, then combining its failure with $2p - 4|A| - 2r = 2|\overline{2A}| \leq 2|\overline{A' + A}|$ and Lemma 2.3 yields

$$\begin{aligned} |A'| + 2p - 4|A| - 2r - 2 &\leq |A'| + 2|\overline{A' + A}| - 2 \\ &\leq | - A' + \overline{A' + A}| \leq |\overline{A}| = p - |A|. \end{aligned}$$

Collecting terms in the above inequality, multiplying by 2, and applying the estimates $|B'| \leq |A'|$ and (11) yields

$$\begin{aligned} 2p &\leq 6|A| + 4r - 2|A'| + 4 \leq 3|2A| + r - |A'| - |B'| + 4 \\ &= 3|2A| - |A' + B'| + r + r' + 4 = 2|2A| + \Delta + r + r' + 4 \\ &\leq \frac{7}{3}|2A| + r + \frac{4}{3}. \end{aligned}$$

Hence, $|2A| \geq \frac{6}{7}p - \frac{3}{7}r - \frac{4}{7}$. Combined with (3) and (4), we conclude that

$$\frac{6}{7}p - \frac{3}{7}\alpha\Big(\frac{3p+5}{8}\Big) + \frac{5}{7} < \frac{6}{7}p - \frac{3}{7}\beta|A| + \frac{5}{7} = \frac{6}{7}p - \frac{3}{7}r - \frac{4}{7} \leq |2A| \leq \varepsilon p \leq \frac{3}{4}p,$$

which yields the contradiction $0 < (\frac{6}{7} - \frac{3}{4} - \frac{9}{56}\alpha)p < \frac{15}{56}\alpha - \frac{5}{7} < 0$ (in view of $\alpha < \frac{1}{2}$), completing Step B. □

By our application of the $3k-4$ theorem (Theorem 2.2) to $\psi(-a_0+A')+\psi(-b_0+B')$, we know that $A'+B'$ contains an arithmetic progression P_{A+B} with difference g and length $|P_{A+B}| \geq |A'|+|B'|-1$, which implies

$$\ell_g(\overline{A'+B'}) \leq p-|A'|-|B'|+1.$$

By (7) and (10), we obtain

$$\ell_g(-A') = \ell_g(A') \leq |A'|+r'+1 \leq |A'|+|B'|-3, \tag{12}$$

whence $\ell_g(-A')+\ell_g(\overline{A'+B'}) \leq p-2$, ensuring that $-A'+\overline{A'+B'}$ is rectifiable via the difference g. Since $\overline{A'+A} \subseteq \overline{A'+B'}$, it follows that $-A'+\overline{A'+A}$ is also rectifiable via the difference g.

By our application of the $3k-4$ theorem (Theorem 2.2) to $\psi(-a_0+A')+\psi(-b_0+B')$ we know that $\psi(-a_0+A')$ is contained in the arithmetic progression $\psi(-a_0+P_A)=[0,m]$ with difference 1 and length $|P_A| \leq |A'|+r'+1$, with the latter inequality by (7). Moreover, $r'+1 \leq |B'|-3 \leq |A'|-3$ (by (10)), so that $|A'| > \lceil \frac{1}{2}|P_A| \rceil$, meaning that $\psi(-a_0+A')$ must contain at least two consecutive elements. Hence,

$$\gcd^*(\psi(-a_0+A')) = 1. \tag{13}$$

Since $-A'+\overline{A'+A}$ is rectifiable via the difference g, it is then isomorphic to the integer sumset $\psi(a_0+mg-A')+\psi(x+\overline{A'+A})$ for an appropriate $x \in \mathbb{Z}/p\mathbb{Z}$. Hence, in view of (13), Step A, and Step B, we can apply the $3k-4$ theorem (Theorem 2.2) to the isomorphic sumset $\psi(a_0+mg-A')+\psi(x+\overline{A'+A})$ and thereby conclude that there is an arithmetic progression $P \subseteq -A'+\overline{A'+A}$ with difference g and length $|P| \geq |A'|+|\overline{A'+A}|-1 \geq |A'|+|\overline{2A}|-1 = p-|2A|+|A'|-1$. Consequently, since Lemma 2.3 ensures that $P \subseteq -A'+\overline{A'+A} \subseteq \overline{A}$, it follows that $\ell_g(A) \leq |2A|-|A'|+1$. Combined with (12), we find that

$$\ell_g(A')+\ell_g(A) \leq |2A|+r'+2. \tag{14}$$

If $A'+A$ is not rectifiable, then $\ell_g(A')+\ell_g(A) \geq p+2$, and, hence, by (11) and (14) we have $p \leq |2A|+r' \leq \frac{4}{3}|2A|-\frac{8}{3}$, whence $|2A| \geq \frac{3}{4}p+2 > \varepsilon p$, contrary to the hypothesis. Therefore, $A'+A$ is rectifiable. This contradicts the maximality of $|A'|+|B'|$, since by (8) we have $|A| > |B'|$, which completes Case 1.

Case 2. — Every pair of subsets $A' \subseteq A$ and $B' \subseteq A$ with $|B'| \leq |A'|$, whose sumset $A'+B'$ is rectifiable, has

$$|A'|+2|B'| \leq |2A|+3. \tag{15}$$

Let $\ell := |2A| = 2|A|+r$. For the rest of this proof, let us identify $\mathbb{Z}/p\mathbb{Z}$ with the set of integers $[0,p-1]$ with addition mod p. Then, for every $X \subseteq \mathbb{Z}/p\mathbb{Z}$ and $d \in \mathbb{Z}/p\mathbb{Z}$, we define the exponential sum $S_X(d) = \sum_{x \in X} e^{\frac{2\pi i}{p}dx} \in \mathbb{C}$.

The idea is to use Freiman's estimate [13, Theorem 1] for such sums to show that the assumption (15) implies

$$|S_A(d)| \le \tfrac{1}{3}|A| + \tfrac{2}{3}r + 2 \quad \text{for all nonzero } \ d \in \mathbb{Z}/p\mathbb{Z}. \tag{16}$$

For any $u \in [0, 2\pi)$, consider the open arc $C_u = \{e^{ix} : \ x \in (u, u+\pi)\}$ of length π in the unit circle in $\mathbb{C}$. Let $A' = \{x \in A : e^{\frac{2\pi i}{p}dx} \in C_u\}$. Since the set of p-th roots of unity contained in C_u correspond to an arithmetic progression of difference 1 in $\mathbb{Z}/p\mathbb{Z}$, it is clear that for d^*, the multiplicative inverse of d modulo p, we have $\ell_{d^*}(A') \le \frac{p+1}{2}$. Hence, the sumset $A' + A'$ is rectifiable. Then the assumption (15) implies that $3|A'| \le |2A| + 3$. This shows that every open half-arc of the unit circle contains at most $n = \frac{1}{3}|2A| + 1$ of the $|A|$ terms involved in the sum $S_A(d)$. By [13, Theorem 1] applied with this n, $N = |A|$, and $\varphi = \pi$, we obtain $|S_A(d)| \le 2n - N = \frac{2}{3}|2A| + 2 - |A|$, and (16) follows.

To complete the proof, we now exploit (16) to obtain a contradiction, using in particular the following manipulations, which are standard in the additive combinatorial use of Fourier analysis (e.g. [12, pp. 290–291])

By Fourier inversion and the fact that $S_A(0) = |A|$ and $S_{2A}(0) = \ell$, we have

$$\begin{aligned}
|A|^2 p &= \sum_{x \in \mathbb{Z}/p\mathbb{Z}} S_A(x) S_A(x) \overline{S_{2A}(x)} \\
&= S_A(0) S_A(0) \overline{S_{2A}(0)} + \sum_{x \in (\mathbb{Z}/p\mathbb{Z}) \setminus \{0\}} S_A(x) S_A(x) \overline{S_{2A}(x)} \\
&= |A|^2 \ell + \sum_{x \in (\mathbb{Z}/p\mathbb{Z}) \setminus \{0\}} S_A(x) S_A(x) \overline{S_{2A}(x)} \\
&\le |A|^2 \ell + \sum_{x \in (\mathbb{Z}/p\mathbb{Z}) \setminus \{0\}} |S_A(x)||S_A(x)||S_{2A}(x)| \\
&\le |A|^2 \ell + (\tfrac{1}{3}|A| + \tfrac{2}{3}r + 2) \sum_{x \in (\mathbb{Z}/p\mathbb{Z}) \setminus \{0\}} |S_A(x)||S_{2A}(x)|.
\end{aligned}$$

This last sum is at most $\big(\sum_{x \in \mathbb{Z}/p\mathbb{Z} \setminus \{0\}} |S_A(x)|^2\big)^{1/2} \big(\sum_{x \in \mathbb{Z}/p\mathbb{Z} \setminus \{0\}} |S_{2A}(x)|^2\big)^{1/2}$ by the Cauchy–Schwarz inequality. We thus conclude that

$$|A|^2 p \le |A|^2 \ell + \frac{|A| + 2r + 6}{3} (|A|p - |A|^2)^{1/2} (\ell p - \ell^2)^{1/2}.$$

Rearranging this inequality, we obtain

$$\frac{|A| + 2r + 6}{3|A|} \ge \frac{|A|(p-\ell)}{|A|^{1/2}(p-|A|)^{1/2} \ell^{1/2} (p-\ell)^{1/2}} = \left(\frac{\frac{p}{\ell} - 1}{\frac{p}{|A|} - 1} \right)^{1/2}. \tag{17}$$

By hypothesis $r = \beta|A| - 3$ and $\ell = |2A| = (2+\beta)|A| - 3$, so $|A| = \frac{\ell+3}{2+\beta} > \frac{\ell}{2+\beta}$. Using these estimates in (17) yields

$$\frac{1+2\beta}{3} = \frac{|A| + 2(\beta|A| - 3) + 6}{3|A|} = \frac{|A| + 2r + 6}{3|A|}$$
$$\geq \left(\frac{\frac{p}{\ell} - 1}{\frac{p}{|A|} - 1}\right)^{1/2} > \left(\frac{\frac{p}{\ell} - 1}{(2+\beta)\frac{p}{\ell} - 1}\right)^{1/2}.$$

Rearranging the above inequality yields (in view of $0 < \beta \leq \alpha < 1$)

$$\varepsilon p \geq \ell > \frac{1 - (\frac{1+2\beta}{3})^2(2+\beta)}{1 - (\frac{1+2\beta}{3})^2} p. \tag{18}$$

Since $\beta \leq \alpha < 1$, rearranging the above inequality yields

$$4\beta^3 + (12 - 4\varepsilon)\beta^2 + (9 - 4\varepsilon)\beta + 8\varepsilon - 7 > 0. \tag{19}$$

Thus, $f(\beta) > 0$, with $f(x) = 4x^3 + (12 - 4\varepsilon)x^2 + (9 - 4\varepsilon)x + 8\varepsilon - 7$. As noted at the start of the proof, $f(x)$ is increasing for $x \geq 0$ with a unique positive root α. As a result, (19) ensures that $\beta > \alpha$, which is contrary to the hypothesis, completing the proof. □

REMARK 2.4. — Our restriction $|2A| \leq \frac{3}{4}p$ in Theorem 2.1 could be relaxed somewhat further, but at increasingly greater cost to the resulting constant α. One simply needs to strengthen the hypothesis of (5) and appropriately adjust the Fourier analytic calculation in Case 2 in the above proof, using the correspondingly weakened inequality for (15).

Proof of Theorem 1.3. — As mentioned earlier, Theorem 1.3 is just the special case of Theorem 2.1 with $\varepsilon = \frac{3}{4}$. □

We now proceed to prove the variant that we shall apply in the next section.

Proof of Theorem 1.4. — The proof is very close to that of Theorem 2.1, with the most significant difference occurring in Case 2. We only highlight the few differences in the argument.

First observe that, if $p = 2$, then $|2A| < p$ forces $|A| = 1$, in which case, the theorem holds trivially. Therefore, we can assume $p \geq 3$. Next, observe (via Taylor series expansion) that $p\sin(\pi/p)$ is an increasing function for $p > 1$ with limit π. The function $\eta/\sin(\pi\eta/3)$ is also an increasing function for $\eta \in (0, 1)$. Thus, $\alpha \leq -\frac{5}{4} + \frac{1}{4}\sqrt{9 + 8\pi/\sin(\pi/3)} < 0.3$. By hypothesis, $|A| \leq \frac{p-r}{3} = \frac{1}{3}p - \frac{1}{3}\beta|A| + 1$, implying

$$|A| \leq \frac{p+3}{\beta+3} < \frac{p+3}{3}, \tag{20}$$

which replaces (4) for the proof. Also, $|2A| = 2|A| + r \leq 2(\frac{p-r}{3}) + r = \frac{2p+r}{3}$.

At the end of Step B in Case 1, we instead obtain $\frac{6}{7}p-\frac{3}{7}r-\frac{4}{7} \le |2A| \le \frac{2p+r}{3}$, which implies

$$\frac{2}{3}p \ge \frac{6}{7}p-\frac{16}{21}r-\frac{4}{7} \ge \frac{6}{7}p-\frac{16}{21}\alpha|A|+\frac{16}{7}-\frac{4}{7} > \frac{6}{7}p-\frac{16}{21}\alpha\Big(\frac{p+3}{3}\Big)+\frac{16}{7}-\frac{4}{7},$$

with the final inequality above in view of (20). Thus, $0 < (\frac{6}{7}-\frac{2}{3}-\frac{16}{63}\alpha)p < \frac{16}{21}\alpha-\frac{12}{7} < 0$ (in view of $0<\alpha<0.3$), which is the contradiction that instead completes Step B.

At the end of Case 1, we instead likewise obtain

$$\frac{3}{4}p+2 \le |2A| \le \frac{2p+r}{3} \le \frac{2}{3}p+\frac{1}{3}\alpha|A|-1 < \frac{2}{3}p+\frac{1}{3}\alpha\Big(\frac{p+3}{3}\Big)-1.$$

This yields the contradiction $0 < (\frac{3}{4}-\frac{2}{3}-\frac{\alpha}{9})p < \frac{\alpha}{3}-3 < 0$ (in view of $0<\alpha<0.3$) in order to complete Case 1.

For Case 2, we begin by following the argument that proves (16), except that we use Lev's sharper estimate [13, Theorem 2] instead of [13, Theorem 1]. Thus, using that any two distinct terms in S_A have the shortest arc between them of length at least $\delta = 2\pi/p$, we obtain by [13, Theorem 2] applied with $n = \frac{1}{3}|2A|+1 \le p/2$ (so $\delta n \le \pi$) that for every such nonzero d, we have

$$|S_A(d)| \le \frac{\sin\Big(\big(\frac{1}{3}|2A|+1-\frac{1}{2}|A|\big)\frac{2\pi}{p}\Big)}{\sin(\frac{\pi}{p})} = \frac{\sin\big((\frac{1}{3}|A|+\frac{2}{3}r+2)\frac{\pi}{p}\big)}{\sin(\frac{\pi}{p})}. \tag{21}$$

Let $M = \frac{1}{3}|A|+\frac{2}{3}r+2$ and let $y = M/p$. Note $M \le (\frac{1}{3}+\frac{2}{3}\alpha)|A| < (\frac{1}{3}+\frac{2}{3}(0.3))\frac{p+3}{3} < \frac{p}{2}$ in view of $r \le \alpha|A|-3$ and (20), ensuring $y \in (\frac{\eta}{3},\frac{1}{2})$. Then the inequality in (21) becomes $|S_A(d)| \le \frac{\sin(y\pi)}{yp\sin(\frac{\pi}{p})}\,M$. The function $f(p,y) = \frac{\sin(y\pi)}{yp\sin(\frac{\pi}{p})}$ is decreasing in $y \in (0,1/2)$ for any fixed $p \ge 3$, as can be seen by considering the Taylor series expansion of its partial derivative. It is also decreasing in p for every fixed $y \in (0,1/2)$ by a similar analysis. Letting $\gamma = f(p,\frac{\eta}{3}) > 0$, we can, therefore, replace (16) by the bound

$$|S_A(d)| \le \gamma(\tfrac{1}{3}|A|+\tfrac{2}{3}r+2). \tag{22}$$

Since $M\frac{\pi}{p} < \frac{\pi}{2}$, $M>1$ and $p \ge 3$, it follows that $\sin(M\frac{\pi}{p}) - M\sin(\frac{\pi}{p}) \le 0$ (as can be seen by considering derivatives with respect to M and using the Taylor series expansion of $\tan(\frac{\pi}{p})$ to note $\tan(\frac{\pi}{p}) > \frac{\pi}{p}$). Consequently, we see that the bound in (21) is at most M, ensuring $\gamma \le 1$. We now obtain the following inequality instead of (17):

$$\begin{aligned}\gamma\frac{1+2\beta}{3} &= \frac{\gamma(\frac{1}{3}|A|+\frac{2}{3}r+2)}{|A|} \\ &\ge \frac{|A|(p-\ell)}{|A|^{1/2}(p-|A|)^{1/2}\ell^{1/2}(p-\ell)^{1/2}} = \left(\frac{\frac{p}{\ell}-1}{\frac{p}{|A|}-1}\right)^{1/2}.\end{aligned} \tag{23}$$

A similar rearrangement to the one that yielded (18) now leads to

$$\frac{2p+\frac{\beta}{3+\beta}(p+3)-3}{3} \geq \frac{2p+\beta|A|-3}{3} = \frac{2p+r}{3} \geq |2A| > \frac{1-\gamma^2(\frac{1+2\beta}{3})^2(2+\beta)}{1-\gamma^2(\frac{1+2\beta}{3})^2}p, \tag{24}$$

with the first inequality following from (20). Since $0 \leq \beta < 1$ and $0 < \gamma \leq 1$, we have $\frac{\beta}{3+\beta} < 1$ and also $1-\gamma^2(\frac{1+2\beta}{3})^2 > 0$, so (24) implies $\big(\frac{\beta+2}{\beta+3}\big)\big(1-\gamma^2\big(\frac{1+2\beta}{3}\big)^2\big) > 1-\gamma^2\big(\frac{1+2\beta}{3}\big)^2(2+\beta)$. Multiplying both sides by $\beta+3>0$ and grouping on the left-hand side the terms involving γ, we obtain $(\beta+2)^2\gamma^2\big(\frac{1+2\beta}{3}\big)^2 > 1$. Taking square roots and expanding, we deduce $2\beta^2+5\beta+2-3\gamma^{-1} > 0$. The quadratic formula thus implies that either $\beta < \frac{-5-\sqrt{9+24\gamma^{-1}}}{4} < 0$ or $\beta > \frac{-5+\sqrt{9+24\gamma^{-1}}}{4} = \alpha$. Since $\beta > 0$, this contradicts the hypothesis $\beta \leq \alpha$, completing the proof. □

3. Bounds for m-sum-free sets in $\mathbb{Z}/p\mathbb{Z}$

In this section, we give new bounds for the quantity $d_m(\mathbb{Z}/p\mathbb{Z})$ defined in formula (1) and for the associated limit

$$d_m = \lim_{\substack{p\to\infty \\ p \text{ prime}}} d_m(\mathbb{Z}/p\mathbb{Z}).$$

In Section 3.1, we present some examples of large m-sum-free sets and in Section 3.2, we apply Theorem 1.4 to give a new upper bound for $d_m(\mathbb{Z}/p\mathbb{Z})$. In Section 3.3 we obtain an improvement of Theorem 1.8.

3.1. Lower bounds for $d_m(\mathbb{Z}/p\mathbb{Z})$. — As mentioned in the Introduction, a simple example of a large m-sum-free set is the interval $(\frac{2}{m^2-4}p, \frac{m}{m^2-4}p)$, having the asymptotic density $\frac{1}{m+2}$ as $p\to\infty$. This gives the largest known size of m-sum-free sets for $m \leq 6$ but not for greater values of m. Indeed, there is the following construction, following an idea due to Tomasz Schoen. The version given below incorporates a suggestion of the anonymous referee that, with some additional modification, yielded the result as stated below. As noted in the proof, for fixed m, the constant $\frac{1}{8}$ can be improved by a small factor tending to 0 as $m\to\infty$.

LEMMA 3.1. — *For each integer $m \geq 7$, we have $d_m(\mathbb{Z}/p\mathbb{Z}) \geq \frac{1}{8}(1-\frac{1}{p})$ for every prime p of the form $p = 4m^2n+1$. In particular, $d_m \geq \frac{1}{8}$.*

Proof. — We identify $\mathbb{Z}/p\mathbb{Z}$ with the interval of integers $[0, p-1]$ with addition mod p. Let $\lambda \in [3, m]$ and $\mu \in [0, m-1]$ be integer parameters to be fixed later

and consider the interval

$$J = \{4mn + 1, 4mn + 2, \dots, 2\lambda mn\} \subset [0, p - 1].$$

We define an m-sum-free set A by picking elements from J in appropriate congruence classes mod m:

$$A := \{x \in J : x \bmod m \in [0, \mu]\}.$$

Note that the sumset $A + A$ taken in $\mathbb{Z}$ is a subset of $[0, p - 1]$ because $\lambda \leq m$ guarantees $2 \max A = 4mn\lambda \leq 4m^2 n = p - 1$. We therefore have $y \in 2A \Rightarrow y \bmod m \in [0, 2\mu]$.

Now,

$$\begin{aligned} J &= \bigcup_{i=1}^{\lceil \frac{\lambda}{2} \rceil - 2} [(4i)mn + 1, 4(i + 1)mn] \cup [(4(\lceil \tfrac{\lambda}{2} \rceil - 1)mn + 1, 2\lambda mn] \\ &\subseteq \bigcup_{i=1}^{\lceil \frac{\lambda}{2} \rceil - 1} [(4i)mn + 1, 4(i + 1)mn]. \end{aligned}$$

Since $p = 4m^2 n + 1$, we have $m \cdot [(4i)mn + 1, 4(i + 1)mn] = \{m - i, 2m - i, \dots, 4m^2 n - i\}$ for $i \in [1, \lceil \frac{\lambda}{2} \rceil - 1]$, meaning that $m \cdot J$ is covered by the progressions

$$U_i = \{m - i, 2m - i, \dots, 4m^2 n - i\}, \quad i \in [1, \lceil \tfrac{\lambda}{2} \rceil - 1].$$

For A to be m-sum-free it suffices to ensure that $2A \cap (m \cdot J) = \emptyset$, and for this, it suffices for λ and μ to satisfy $2\mu \leq m - (\lceil \frac{\lambda}{2} \rceil - 1) - 1$, that is,

$$2\mu + \lceil \tfrac{\lambda}{2} \rceil \leq m.$$

We also have $|A| = (\mu + 1) \frac{|J|}{m} = 2n(\mu + 1)(\lambda - 2)$, so

$$\frac{|A|}{p - 1} = \frac{(\mu + 1)(\lambda - 2)}{2m^2}.$$

Suppose $m \equiv 0 \bmod 4$, so $m \geq 8$. Considering $\lambda = m$ and $\mu = \frac{m}{4}$ yields $\frac{|A|}{p-1} = \frac{(\mu+1)(\lambda-2)}{2m^2} = \frac{m^2+2m-8}{8m^2}$, which is at least $\frac{1}{8}$ for $m \geq 4$.

Suppose $m \equiv 1 \bmod 4$, so $m \geq 9$. Considering $\lambda = m$ and $\mu = \frac{m-1}{4}$ yields $\frac{|A|}{p-1} = \frac{(\mu+1)(\lambda-2)}{2m^2} = \frac{m^2+m-6}{8m^2}$, which is at least $\frac{1}{8}$ for $m \geq 6$. We remark that taking $\lambda = m - 3$ and $\mu = \frac{m+3}{4}$ instead yields $\frac{|A|}{p-1} = \frac{(\mu+1)(\lambda-2)}{2m^2} = \frac{m^2+2m-35}{8m^2}$, which is slightly better for larger values of m.

Suppose $m \equiv 2 \bmod 4$, so $m \geq 10$. In this case, we will modify the above construction with $\lambda = m - 1$ and $\mu = \frac{m-2}{4}$. For these parameters, we have

$\lceil \frac{\lambda}{2} \rceil - 1 = \frac{m-2}{2}$, $2(\mu+1) = \frac{m+2}{2} = m - \frac{m-2}{2}$, and

$$m \cdot [(4(\lceil \tfrac{\lambda}{2} \rceil - 1)mn + 1, 2\lambda mn] = m \cdot [2(m-2)mn + 1, 2(m-1)mn]$$
$$= \{m - \tfrac{m-2}{2}, 2m - \tfrac{m-2}{2}, \dots, 2m^2 n - \tfrac{m-2}{2}\}, \tag{25}$$

which is the subset of $m \cdot J \subseteq [0, p-1]$ congruent to $m - \frac{m-2}{2}$ modulo m. Let $B = \{tm + \frac{m+2}{4} : \ mn \le t \le 2(m-1)n - 1\}$ and set $A' = A \cup B$. Since $mn \le t \le 2(m-1)n-1$, we have $B \subseteq J$, while $2B = \{2m^2 n + m - \frac{m-2}{2}, 2m^2 n + 2m - \frac{m-2}{2}, \dots, 4(m-1)mn - m - \frac{m-2}{2}\}$, which is disjoint from the set in (25). Since $A + B$ is also disjoint from $m \cdot J$, it follows that $2A' \cap (m \cdot J) = \emptyset$, so A' is m-sum-free. We have $\frac{|A'|}{p-1} = \frac{|A|+|B|}{p-1} = \frac{2n(\mu+1)(\lambda-2)+(m-2)n}{p-1} = \frac{m^2+m-10}{8m^2}$, which is at least $\frac{1}{8}$ for $m \ge 10$. We remark that taking $\lambda = m-2$ and $\mu = \frac{m+2}{4}$ in the original construction instead yields $\frac{|A|}{p-1} = \frac{(\mu+1)(\lambda-2)}{2m^2} = \frac{m^2+2m-24}{8m^2}$, which is slightly better for larger values of m.

Suppose $m \equiv 3 \bmod 4$, so $m \ge 7$. We modify the original construction with $\lambda = m$ and $\mu = \frac{m-3}{4}$. For these parameters, we have $\lceil \frac{\lambda}{2} \rceil - 1 = \frac{m-1}{2}$, $2(\mu+1) = \frac{m+1}{2} = m - \frac{m-1}{2}$, and

$$m \cdot [(4(\lceil \tfrac{\lambda}{2} \rceil - 1)mn + 1, 2\lambda mn] = m \cdot [2(m-1)mn + 1, 2m^2 n]$$
$$= \{m - \tfrac{m-1}{2}, 2m - \tfrac{m-1}{2}, \dots, 2m^2 n - \tfrac{m-1}{2}\},$$

which is the subset of $m \cdot J \subseteq [0, p-1]$ congruent to $m - \frac{m-1}{2}$ modulo m. Let $B = \{tm + \frac{m+1}{4} : \ mn \le t \le 2mn - 1\}$ and set $A' = A \cup B$. Since $mn \le t \le 2mn-1$, we have $B \subseteq J$, while $2B = \{2m^2 n + m - \frac{m-1}{2}, 2m^2 n + 2m - \frac{m-1}{2}, \dots, 4m^2 n - m - \frac{m-1}{2}\}$. Thus, similarly to the previous case, $2A' \cap (m \cdot J) = \emptyset$, so A' is m-sum-free. We have $\frac{|A'|}{p-1} = \frac{|A|+|B|}{p-1} = \frac{2n(\mu+1)(\lambda-2)+mn}{p-1} = \frac{m^2+m-2}{8m^2}$, which is at least $\frac{1}{8}$ for $m \ge 2$. We remark that taking $\lambda = m-1$ and $\mu = \frac{m+1}{4}$ in the original construction instead yields $\frac{|A|}{p-1} = \frac{(\mu+1)(\lambda-2)}{2m^2} = \frac{m^2+2m-15}{8m^2}$, which is slightly better for larger m.

In all four cases above, we obtain a set A, such that $d_m(\mathbb{Z}/p\mathbb{Z}) \ge \frac{|A|}{p} \ge \frac{|A|}{p-1}(1 - \frac{1}{p}) \ge \frac{1}{8}(1 - \frac{1}{p})$, and now the claim about the limit follows from the fact that by Dirichlet's theorem there exist infinitely many primes in the arithmetic progression $\{4m^2 n + 1 : n \ge 1\}$. □

3.2. Upper bound for $d_m(\mathbb{Z}/p\mathbb{Z})$. — In this section we prove Theorem 1.5, which we restate here for convenience.

THEOREM 3.2. — *Let $p \ge 80$ be a prime, m be an integer in $[2, p-2]$, and $c = c(p)$ be the solution to the equation $c = \frac{1+3/p}{3+\alpha(c,p)}$, where $\alpha = \alpha(c,p)$ is the parameter in Theorem 1.4 with $\eta = c$. Then, $d_m(\mathbb{Z}/p\mathbb{Z}) < c$. In particular, $d_m \le \frac{1}{3.1955}$.*

The idea of the proof is roughly the following: either an m-sum-free set A has a doubling constant at least $2+\alpha$, in which case, since $(m\cdot A)\cap 2A=\emptyset$, we have $(3+\alpha)|A|\le |(m\cdot A)|+|2A|\le p$ and we are done; or we can apply Theorem 1.4, and thus, working with the two arithmetic progressions provided by the theorem, we reduce the problem essentially to bounding the size that two progressions I and J of equal difference can have if the dilate $m\cdot J$ has small intersection with I. Let us begin by establishing this result about progressions.

LEMMA 3.3. — *Let $p\ge 80$ be prime, $0<\alpha\le 1/5$, $d\in[2,p-2]$, and $N\in\mathbb{N}$. Let I and J be progressions in $\mathbb{Z}/p\mathbb{Z}$ having the same difference and satisfying $|I|=2N-1$, $|J|>(1+\alpha)N-3$, and $|I\cap(d\cdot J)|\le \alpha N-2$. Then, $N<\frac{p+3}{3+\alpha}$.*

Proof. — First note that without loss of generality, we can assume $d\le\frac{p-1}{2}$, since if the lemma is proved with this assumption, then, given $d>\frac{p-1}{2}$, we can multiply by -1 and apply the lemma with the intervals $-I$ and J. Let us proceed by contradiction supposing that there exists some N (along with p, d, α, I, and J), such that the hypotheses of the lemma are satisfied, but $N\ge\frac{p+3}{3+\alpha}$. Note that the supposed properties of I and J are conserved, if we dilate by the inverse of their difference mod p and if we translate, replacing I by $I+dz$ and J by $J+z$. It follows that identifying $\mathbb{Z}/p\mathbb{Z}$ with the integers $[0,p-1]$ with addition mod p, we can assume that $I=[p-|I|,p-1]$ and $J=x+[0,|J|-1]\bmod p$, for some $x\in[0,p-1]$.

We claim that we can assume without loss of generality that

$$d\cdot x\in[0,d-1]\bmod p. \tag{26}$$

Indeed, if this does not hold, then either $d\cdot x\in[d,p-|I|+d-1]\bmod p$ or $d\cdot x\in[p-|I|,p-1]$. If the former holds, then $d\cdot(x-1)\notin I\bmod p$, so the interval $J'=(x-1)+[0,|J|-1]$ satisfies the hypotheses with $|I\cap(d\cdot J')|\le|I\cap(d\cdot J)|$. On the other hand, if $d\cdot x\in[p-|I|,p-1]$, then letting $J'=(x+1)+[0,|J|-1]$ we have $d\cdot x\in I\cap(d\cdot J)$ and $d\cdot x\notin I\cap(d\cdot J')$, so this interval J' satisfies the hypotheses with $|I\cap(d\cdot J')|\le|I\cap(d\cdot J)|$. In either case, by repeatedly shifting the interval J, we eventually obtain (26).

Given (26), we may partition $d\cdot J$ into successive progressions U_i (with difference d) for $i\in[1,s+1]$, such that $U_i=(\min U_i+d\mathbb{Z})\cap[0,p-1]$ with $\min U_i\in[0,d-1]$ for $i\in[1,s]$, and U_{s+1} is either empty or consists of an initial portion of $(\min U_{s+1}+d\mathbb{Z})\cap[0,p-1]$ with $\min U_{s+1}\in[0,d-1]$. Then, $|U_i\cap I|\ge\left\lfloor\frac{|I|}{d}\right\rfloor$ for $i\in[1,s]$. It follows that $|(d\cdot J)\cap I|\ge s\left\lfloor\frac{|I|}{d}\right\rfloor$, whence

$$\alpha N-2\ge s\left\lfloor\frac{|I|}{d}\right\rfloor. \tag{27}$$

Now, as $d\cdot x\in[0,d-1]\bmod p$, each U_i with $i\le s$ starts in $[0,d-1]$ and ends in $[p-d,p-1]$, so s is at least the number of consecutive intervals of length p

that fit inside $[0, |J|d-1]$:

$$s \geq \left\lfloor \frac{|J|d}{p} \right\rfloor > \frac{((1+\alpha)N-3)d}{p} - 1. \tag{28}$$

Substituting this lower bound for s in (27), as well as the bound $\lfloor \frac{|I|}{d} \rfloor \geq \frac{|I|}{d} - \frac{d-1}{d} = \frac{2N}{d} - 1$, and expanding the resulting product, we obtain $\alpha N - 2 > \frac{2(1+\alpha)}{p} N^2 - \big(\frac{(1+\alpha)d}{p} + \frac{6}{p} + \frac{2}{d}\big)N + 1 + \frac{3d}{p}$. We group all terms involving N on the right-hand side, note that the other terms grouped on the left-hand side amount to a negative number, and multiply through by $\frac{p}{2(1+\alpha)N}$ to deduce that

$$N < \frac{1}{2(1+\alpha)} \left(d(1+\alpha) + 6 + \frac{2p}{d} + \alpha p \right). \tag{29}$$

We want to obtain a contradiction from this, using that $N \geq \frac{p+3}{3+\alpha}$. To this end, using the bounds $2 \leq d \leq \frac{p-1}{2}$ on the right-hand side of (29) is not enough. However, we shall now show that we can assume $11 \leq d < p/6$, which will be enough.

First, we claim that $s \geq 1$. Indeed, otherwise $|J| \leq |(d \cdot J) \cap I| + |(d \cdot J) \cap [0, p - |I| - 1]| \leq \alpha N - 2 + \lceil \frac{p-|I|}{d} \rceil$. Using the assumptions on $|I|, |J|$, and $d \geq 2$, we deduce that $N < \frac{p+2d}{d+2} \leq \frac{p}{4} + 2$. This, combined with our assumptions $N \geq (p+3)/(3+\alpha)$ and $\alpha < 1/5$, contradicts $p \geq 80$.

Since $s \geq 1$, (27) yields $\alpha N - 2 \geq \lfloor |I|/d \rfloor \geq \frac{2N}{d} - 1$. It follows that $(\alpha N - 1)d \geq 2N > 0$. Hence, $\alpha N - 1 > 0$ and $d \geq \frac{2N}{\alpha N - 1} > \frac{2}{\alpha}$, whence $d \geq 11$ follows in view of $\alpha \leq \frac{1}{5}$.

Note that $\lfloor |I|/d \rfloor \geq 1$, for otherwise $2N = |I| + 1 < d + 1 \leq \frac{p+1}{2}$, contradicting our assumptions $N \geq \frac{p+3}{3+\alpha}$ and $\alpha \leq 1/5$. Combining this with (27) and (28), we obtain $\alpha N - 2 > \frac{((1+\alpha)N-3)d}{p} - 1$, which means $d \leq \big(\frac{\alpha N - 1}{(1+\alpha)N - 3}\big)p < \frac{\alpha}{1+\alpha} p$. As $\alpha \leq 1/5$, we conclude that $d < p/6$.

Now using the bounds $11 \leq d < p/6$ in (29) and the assumption that $N \geq \frac{p+3}{3+\alpha}$, we deduce that $\frac{p+3}{3+\alpha} < \frac{p}{12} + \frac{p}{11(1+\alpha)} + \frac{\alpha p}{2(1+\alpha)} + \frac{3}{1+\alpha}$, implying $\frac{1}{3.2} < \frac{1}{12} + \frac{1}{11} + \frac{1}{10} + \frac{3}{p}$, contradicting $p \geq 80$. □

Remark 3.4. — It is possible to extend the validity of Lemma 3.3 to all primes $p \geq 5$, at the cost of lengthening the proof with several technicalities. The lemma has potential generalizations that seem of independent interest, although we do not need to pursue them for our purposes in this paper. For instance, the anonymous referee raised the question of which values of coefficients α, β and which functions $f(\alpha, \beta), g(\alpha, \beta) > 0$ ensure that the following statement holds: if I, J are arithmetic progressions in $\mathbb{Z}/p\mathbb{Z}$ with common difference and respective sizes $\alpha N + a$, $\beta N + b$, then $N > f(\alpha, \beta)p$ implies $|I \cap (d \cdot J)| > g(\alpha, \beta)N$.

We can now prove the main result.

Proof of Theorem 3.2. — Let $A \subseteq \mathbb{Z}/p\mathbb{Z}$ be an m-sum-free subset of maximum size, with $|A| = \eta p$, and let $\alpha = \alpha(\eta, p) = -\frac{5}{4} + \frac{1}{4}\sqrt{9 + 8\,\eta\, p \sin(\pi/p)/\sin(\pi\eta/3)}$. Assume by contradiction that $\eta \geq c$. Then, since $x \mapsto \frac{1+3/p}{3+\alpha(x,p)}$ is decreasing in $x \in (0,1)$, and $c = \frac{1+3/p}{3+\alpha(c,p)}$, we deduce that $\eta \geq c \geq \frac{1+3/p}{3+\alpha}$, whence

$$|A| \geq \frac{p+3}{3+\alpha} > 1. \tag{30}$$

As noted at the start of the proof of Theorem 1.4, $\alpha(\eta, p)$ is increasing for $\eta \in (0,1)$ with $p\sin(\pi/p) \to \pi$ monotonically. Since $2A$ and $m \cdot A$ are disjoint, we have $|2A| \leq p - |A|$, while $|2A| \geq 2|A| - 1$ by the Cauchy–Davenport theorem. Thus, $2|A| - 1 \leq |2A| \leq p - |A|$, implying $|A| \leq \frac{p+1}{3}$ and $\eta \leq \frac{p+1}{3p}$. Since $p \geq 80$, we have $\eta \leq \frac{3}{8}$ and $\alpha \leq -\frac{5}{4} + \frac{1}{4}\sqrt{9 + 3\pi/\sin(\pi/8)} < 0.2$.

Let $|2A| = 2|A| + r$. Since A is m-sum-free, the sets $2A$ and $m \cdot A$ are disjoint, which implies that $|2A| < p$ (as A is nonempty) and that $p \geq |2A| + |m \cdot A| = 3|A| + r$. Thus,

$$|A| \leq \frac{p-r}{3} \quad \text{and} \quad |2A| = 2|A| + r \leq \frac{2p+r}{3}.$$

Since $|2A| < p$, the Cauchy–Davenport theorem implies $r \geq -1$.

If $|2A| = 2|A| + r > (2+\alpha)|A| - 3$, then $r > \alpha|A| - 3$, in which case $|A| \leq \frac{p-r}{3} < \frac{p-\alpha|A|+3}{3}$, which contradicts (30). Therefore, $|2A| \leq (2+\alpha)|A| - 3$ and $r \leq \lfloor \alpha|A| - 3 \rfloor$. We can now apply Theorem 1.4. As a result, there are arithmetic progressions P_A and P_{2A} with common difference g such that $A \subseteq P_A$, $P_{2A} \subseteq 2A$, $|P_A| = \lfloor (1+\alpha)|A| - 2 \rfloor \leq p$, and $|P_{2A}| = 2|A| - 1$. It follows that $P := m \cdot P_A$ is an arithmetic progression with difference $mg \neq \pm g$, such that

$$|P \cap P_{2A}| \leq |P \cap 2A| \leq |P_A \setminus A| \leq \alpha|A| - 2.$$

We can, therefore, apply Lemma 3.3 with $N = |A|$ (as $\alpha < 0.2$), deducing that $|A| < \frac{p+3}{3+\alpha}$, which is a contradiction. Therefore, we must have $\eta < c$, so $d_m(\mathbb{Z}/p\mathbb{Z}) < c$, which proves the first claim in the theorem. Taking the limit of c as $p \to \infty$, we deduce that $d_m \leq t$, where t is defined by the equation $t = F(t)$ for the function $F(t) = (\frac{7}{4} + \frac{1}{4}\sqrt{9 + 8\,t\,\pi/\sin(\pi t/3)})^{-1}$. Since F is monotonically decreasing and satisfies $F(3.1955^{-1}) < 3.1955^{-1}$, we must have $t < 3.1955^{-1}$, which proves the second claim in the theorem. □

3.3. The structure of large sum-free sets in $\mathbb{Z}/p\mathbb{Z}$. — In this final part of the paper, we apply Theorem 1.4 to obtain the following improvement of Theorem 1.8.

THEOREM 3.5. — *Let $p \geq 14\,000$ be prime and let $A \subseteq \mathbb{Z}/p\mathbb{Z}$ be sum-free with $|A| \geq (0.313)p$. Then, $m \cdot A \subseteq [\,|A|, p - |A|\,] \subseteq \mathbb{Z}/p\mathbb{Z}$, for some $m \in [1, p-1]$.*

Proof. — By hypothesis, $|A| = \eta p > 0$ with $\eta \geq 0.313$. Set $|2A| = 2|A| + r$. Since A is sum-free, we have $(2A) \cap A = \emptyset$, implying $2|A| + r = |2A| \leq p - |A| < p$, whence $|A| \leq \frac{p-r}{3}$. As in the proof of Theorem 3.2, let $\alpha = \alpha(\eta, p) = -\frac{5}{4} + \frac{1}{4}\sqrt{9 + 8\,\eta\, p \sin(\pi/p)/\sin(\pi\eta/3)}$. Observe that $\alpha(\eta, p)$ is increasing as a function of $p \geq 2$ and $\eta \in (0, 1)$, so $\alpha = \alpha(\eta, p) \geq \alpha(0.313, 14000) \geq \beta := 0.195579$. If $|2A| > (2+\beta)|A| - 3$, then we have $(2+\beta)|A| - 3 < |2A| \leq p - |A|$, implying $(0.313)p \leq |A| < \frac{p+3}{3+\beta}$, and, thus, $p \leq 13\,875$, which is contrary to the hypothesis. Therefore, we instead conclude that $|2A| \leq (2 + \beta)|A| - 3 \leq (2 + \alpha)|A| - 3$, allowing us to apply Theorem 1.4 to conclude that there is an arithmetic progression $P \subseteq \mathbb{Z}/p\mathbb{Z}$ with $A \subseteq P$ and $|P| \leq |A| + r + 1$. By dilating A by the inverse of the difference of the progression P, we can assume without loss of generality that P has difference 1. Since $2|A| + r = |2A| \leq (2+\beta)|A| - 3$, we have $r \leq \beta|A| - 3$, and, thus, $|P| \leq |A| + r + 1 \leq (1 + \beta)|A| - 2$. The bound $|A| \leq (p + 1)/3$ given by the Cauchy–Davenport theorem then implies $|P| \leq (1 + \beta)(p + 1)/3 < \frac{p+1}{2}$. It follows that the sumset $A + A$ is rectifiable.

Let $\psi : A + A \to \mathbb{Z}$ be the associated Freiman isomorphism, with coordinate map $\psi_A : A \to \mathbb{Z}$. Note that the map of the form $a_0 + sg \mapsto s$ involved in the definition of ψ_A (see the remarks before Lemma 2.3) can be assumed to be just a translation (since the element g here, being the difference of P, is assumed to be 1). By slight abuse of notation, we drop the subscript from ψ_A, denoting this map also by ψ. Let $\psi' : A - A \to \mathbb{Z}$ be the Freiman isomorphism defined by $\psi'(x - y) = \psi(x) - \psi(y)$, for $x, y \in A$ (see the remarks after Theorem 2.2). Since $|P| \leq |A| + r + 1 \leq 2|A| - 2$ implies $|A| > \frac{|P|+1}{2}$, we are assured that A contains two consecutive elements in P, whence $\gcd^*(\psi(A)) = 1$. Since A is sum-free, we have $(A - A) \cap A = \emptyset$, and, thus, $|A - A| \leq p - |A|$. Since $A - A \cong \psi(A) - \psi(A)$, we have $|\psi(A) - \psi(A)| = |A - A|$ and $|\psi(A)| = |A|$. As a result, if $|\psi(A) - \psi(A)| \geq 3|\psi(A)| - 3$, then $p - |A| \geq |A - A| = |\psi(A) - \psi(A)| \geq 3|\psi(A)| - 3 = 3|A| - 3$, implying $(0.313)p \leq |A| \leq \frac{p+3}{4}$, contradicting that $p \geq 14\,000$. Therefore, $|\psi(A) - \psi(A)| \leq 3|\psi(A)| - 4$, allowing us to apply the $3k - 4$ theorem (Theorem 2.2) with the sets $\psi(A)$, $-\psi(A)$. This, together with the remarks in the paragraph above Theorem 2.2, implies that $[-(|A|-1), (|A|-1)] \subseteq \psi(A) - \psi(A)$. Hence, $[-(|A|-1), (|A|-1)] \subseteq \psi'(A - A)$ and given the form of ψ', it follows that in $\mathbb{Z}/p\mathbb{Z}$, we have $[-(|A| - 1), (|A| - 1)] \subseteq A - A$. Since A being sum-free implies $(A - A) \cap A = \emptyset$, this forces $A \cap [-(|A|-1), (|A|-1)] = \emptyset$, i.e., $A \subseteq [\,|A|, p - |A|\,]$, which completes the proof. □

Acknowledgements. — The authors are very grateful to Tomasz Schoen for providing the original idea of the construction in Lemma 3.1 and for useful remarks. We also thank the anonymous referee very much for the insightful comments that helped us to improve this paper.

BIBLIOGRAPHY

[1] A. Baltz, P. Hegarty, J. Knape, U. Larsson & T. Schoen – "The structure of maximum subsets of $\{1,\dots,n\}$ with no solutions to $a+b=kc$", *Electron. J. Combin.* (2005), no. 12, Paper No. R19.

[2] P.-Y. Bienvenu, F. Hennecart & I. Shkredov – "A note on the set $A(A+A)$", *Mosc. J. Comb. Number Theory* **8** (2019), no. 2, p. 179–188.

[3] T. F. Bloom – "A quantitative improvement for Roth's theorem on arithmetic progressions", *J. Lond. Math. Soc. (2)* (2016), no. 93, p. 643–663.

[4] P. Candela & A. de Roton – "On sets with small sumset in the circle", *Q. J. Math.* **70** (2019), no. 1, p. 49–69.

[5] P. Candela, O. Serra & C. Spiegel – "A step beyond Freiman's theorem for set addition modulo a prime", *J. Théor. Nombres Bordeaux* **32** (2020), no. 1, p. 275–289.

[6] P. Candela & O. Sisask – "On the asymptotic maximal density of a set avoiding solutions to linear equations modulo a prime", *Acta Math. Hungar.* **132** (2011), no. 3, p. 223–243.

[7] F. R. K. Chung & J. L. Goldwasser – "Maximum subsets of $(0,1]$ with no solutions to $x+y=kz$", *Electron. J. Combin.* (1996), no. 1, p. 3, Research Paper 1.

[8] ———, "Integer sets containing no solutions to $x+y=3z$", in *The Mathematics of Paul Erdős* (R. Graham & J. Nesetřil, eds.), Springer, Berlin, 1997, p. 218–227.

[9] J.-M. Deshouillers & V. F. Lev – "A refined bound for sum-free sets in groups of prime order", *Bull. Lond. Math. Soc.* **40** (2008), no. 5, p. 863–875.

[10] G. Freiman – "Inverse problems in additive number theory. Addition of sets of residues modulo a prime", *Dokl. Akad. Nauk SSSR* **141** (1961), no. 3, p. 571–573.

[11] B. Green & I. Z. Ruzsa – "Sets with small sumset and rectification", *Bull. Lond. Math. Soc. (1)* **38** (2006), p. 43–52.

[12] D. J. Grynkiewicz – *Structural additive theory*, Developments in Mathematics, no. 30, Springer, Cham, 2013.

[13] V. F. Lev – "Distribution of points on arcs", *Integers* **5** (2005), no. 2, electronic.

[14] V. F. Lev & I. Shkredov – "Small doubling in prime-order groups: from 2.4 to 2.6", *J. Number Theory* **217** (2020), p. 278–291.

[15] M. Matolcsi & I. Z. Ruzsa – "Sets with no solutions to $x+y=3z$", *European J. Combin.* **34** (2013), no. 8, p. 1411–1414.

[16] A. Plagne & A. de Roton – "Maximal sets with no solution to $x+y=3z$", *Combinatorica* **36** (2016), no. 2, p. 229–248.

[17] Ø. J. Rødseth – "On Freiman's 2.4-theorem", *Skr. K. Nor. Vidensk. Selsk* (2006), no. 4, p. 11–18.

[18] K. F. Roth – "On certain sets of integers", *J. London Math. Soc.* **28** (1953), p. 104–109.
[19] T. Sanders – "On Roth's theorem on progressions", *Ann. of Math.* **174** (2011), p. 619–636.
[20] O. Serra & G. Zémor – "Large sets with small doubling modulo p are well covered by an arithmetic progression", *Ann. Inst. Fourier (Grenoble)* **59** (2009), no. 5, p. 2043–2060.
[21] G. Vosper – "The critical pairs of subsets of a group of prime order", *J. London Math. Soc.* **31** (1956), p. 200–205.

Bull. Soc. Math. France
149 (1), 2021, p. 179-233

MORITA EQUIVALENCES FOR CYCLOTOMIC HECKE ALGEBRAS OF TYPES B AND D

BY LOÏC POULAIN D'ANDECY & SALIM ROSTAM

ABSTRACT. — We give a Morita equivalence theorem for so-called cyclotomic quotients of affine Hecke algebras of types B and D, in the spirit of a classical result of Dipper–Mathas of type A for Ariki–Koike algebras. Consequently, the representation theory of affine Hecke algebras of types B and D reduces to the study of their cyclotomic quotients with eigenvalues in a single orbit under multiplication by q^2 and inversion. The main step in the proof consists in a decomposition theorem for generalisations of quiver Hecke algebras that recently appeared in the study of affine Hecke algebras of types B and D. This theorem reduces the general situation of a disconnected quiver with involution to a simpler setting. To be able to treat types B and D at the same time we unify the different definitions of quiver Hecke algebra for type B that exist in the literature.

Texte reçu le 23 septembre 2019, modifié le 4 septembre 2020, accepté le 20 octobre 2020.

LOÏC POULAIN D'ANDECY, Laboratoire de Mathématiques de Reims UMR 9008, Université de Reims Champagne-Ardenne, Moulin de la Housse BP 1039, 51100 Reims, France • *E-mail :* loic.poulain-dandecy@univ-reims.fr
SALIM ROSTAM, Univ Rennes, CNRS, IRMAR – UMR 6625, F-35000 Rennes, France • *E-mail :* salim.rostam@ens-rennes.fr

Mathematical subject classification (2010). — 20C08.

Key words and phrases. — Cyclotomic Hecke algebra, Morita equivalence, Quiver Hecke algebras, Representation theory.

The first author is supported by *Agence Nationale de la Recherche* through the JCJC project ANR-18-CE40-0001.

0037-9484/2021/179/$ 5.00
doi:10.24033/bsmf.2828

RÉSUMÉ (*Équivalences de Morita pour les algèbres de Hecke cyclotomiques de type B et D*). — Nous énonçons un théorème d'équivalence de Morita pour les quotients cyclotomiques des algèbres de Hecke affines de type B et D, suivant un résultat classique de Dipper–Mathas en type A pour les algèbres d'Ariki–Koike. Ainsi, la théorie des représentations des algèbres de Hecke affines de type B et D se réduit à l'étude de leurs quotients cyclotomiques où les valeurs propres sont dans une unique orbite pour la multiplication par q^2 et l'inversion. La preuve consiste notamment en un théorème de décomposition pour des généralisations d'algèbres de Hecke carquois introduites récemment dans l'étude des algèbres de Hecke affines de type B et D, ramenant la situation générale d'un carquois non connexe avec involution à un cadre plus simple. Pour traiter simultanément les deux types, nous unifions les différentes définitions d'algèbres de Hecke carquois pour le type B déjà existantes.

1. Introduction

Cyclotomic quotients of the affine Hecke algebra of type A, also known as Ariki–Koike algebras, have been extensively studied since their introduction by Broué–Malle [5] and Ariki–Koike [2]. Given a field K, a subset $I \subseteq K^\times$, an element $q \in K^\times$ and a finitely-supported family $\Lambda = (\Lambda_i)_{i\in I}$ of non-negative integers, the Ariki–Koike algebra $H^\Lambda(\mathfrak{S}_n)$ is defined by the generators $g_0, \dots, g_{n-1}$ and the relations

$$\begin{aligned} g_i g_j &= g_j g_i, && \text{for all } i,j \in \{0,\dots,n-1\}, |i-j|>1,\\ g_i g_{i+1} g_i &= g_{i+1} g_i g_{i+1}, && \text{for all } i \in \{1,\dots,n-2\},\\ g_0 g_1 g_0 g_1 &= g_1 g_0 g_1 g_0, &&\\ (g_i - q)(g_i + q^{-1}) &= 0, && \text{for all } i \in \{1,\dots,n-1\},\\ \prod_{i\in I} (g_0 - i)^{\Lambda_i} &= 0. && \end{aligned}$$

We note that Ariki–Koike algebras are quotients, by the last relation, of affine Hecke algebras of type A and that the study of their representations (for all choices of I and Λ) is equivalent to the study of finite-dimensional representations of affine Hecke algebras of type A.

By an important theorem of Dipper–Mathas [8], we know that it suffices to study Ariki–Koike algebras when the set I is q^2-*connected*, that is, in a single q^2-orbit (and even up to a scalar renormalisation of the generator g_0, when $I \subseteq \langle q^2 \rangle$). More precisely, if $I = \coprod_{j=1}^{d} I^{(j)}$ is the decomposition of I into q^2-connected sets, then we have a Morita equivalence

$$(\clubsuit) \qquad H^\Lambda(\mathfrak{S}_n) \overset{\text{Morita}}{\simeq} \bigoplus_{\substack{n_1,\dots,n_d \geq 0 \\ n_1+\cdots+n_d=n}} \bigotimes_{j=1}^{d} H^{\Lambda^{(j)}}(\mathfrak{S}_{n_j}),$$

where $\Lambda^{(j)}$ is the restriction of Λ to $I^{(j)}$. (Note that the assumption in [8] is slightly stronger than the one above, but in practice, it is this condition of q^2-connected sets that is used.) Hence, this Morita equivalence allows us to use results that are only known when the set I is q^2-connected, in particular, the celebrated Ariki's categorification theorem [1] that computes the decomposition numbers of Ariki–Koike algebras in terms of the canonical basis of a certain highest weight module over an affine quantum group.

Another way to obtain this Morita equivalence was given by the second author [21, §3.4], using the theory of quiver Hecke algebras. This is a family of graded algebras that was introduced independently by Khovanov–Lauda [15, 16] and Rouquier [22] a few years ago, in the context of categorification of quantum groups. If Γ is a quiver, we denote by $R_n(\Gamma)$ the associated quiver Hecke algebra (see §3.1). For a certain quiver Γ depending only on the order of q^2, Brundan–Kleshchev [6] and independently Rouquier [22] proved that a certain "cyclotomic" quotient of $R_n(\Gamma)$ is isomorphic to an Ariki–Koike algebra. This result is now a basic tool in the study of Ariki–Koike algebras and their degenerations, including the symmetric group and the classical Hecke algebra of type A. For instance, as consequences, first, the Ariki–Koike algebra inherits the grading of the cyclotomic quiver Hecke algebra and, second, it depends on q only through its order in $K^\times$. Now, if Γ is of the form $\Gamma = \coprod_{j=1}^{d} \Gamma^{(j)}$, where each $\Gamma^{(j)}$ is a full subquiver, it was shown in [20, §6] that we have a decomposition

$$(\spadesuit) \qquad R_n(\Gamma) \simeq \bigoplus_{\substack{n_1,\dots,n_d \geq 0 \\ n1+\cdots+n_d=n}} \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}} \left(\bigotimes_{j=1}^{d} R_{n_j}(\Gamma^{(j)}) \right).$$

This isomorphism of algebras is compatible with cyclotomic quotients and combining with the previous isomorphism of Brundan–Kleshchev and Rouquier allows to recover the Morita equivalence ($\clubsuit$). This Morita equivalence has been further generalised for the cyclotomic Hecke algebras of type $G(r,p,n)$ [11]. We also indicate the paper [12], where the Dipper–Mathas result is studied and derived again from the point of view of affine Hecke algebras, and where the question of a similar result for other affine Hecke algebras is evoked.

The main point of this paper is to prove a similar decomposition theorem for some generalisations of quiver Hecke algebras and, hence, obtain an analogue of the Dipper–Mathas Morita equivalence for cyclotomic quotients of affine Hecke algebras of types B and D. Such generalisations of quiver Hecke algebras were introduced by Varagnolo and Vasserot [24] (for type B) and together with Shan [23] (for type D), in the course of their proofs of conjectures by Kashiwara–Enomoto [9] and Kashiwara–Miemietz [14]. For certain subcategories of representations of affine Hecke algebras of types B and D, these algebras play a similar role to quiver Hecke algebras for affine Hecke algebras of

type A. Inspired by their results, the first author together with Walker [18, 19] obtained an isomorphism theorem *à la* Brundan–Kleshchev between cyclotomic quotients of affine Hecke algebras of types B and D and certain generalisations of cyclotomic quiver Hecke algebras.

The first step of this paper is to provide a definition of these generalisations of quiver Hecke algebras for type B, which encompasses all the slightly different versions previously defined. They are $\mathbb{Z}$-graded algebras and depend upon a quiver with an involution and certain weight functions on the vertices. As for the type A case, that is, for usual quiver Hecke alegbras, the algebra that we define admits a PBW basis, and this is a key ingredient to prove the decomposition theorem when the underlying quiver has several connected components. The point of having defined a new algebra in Section 4 is that we can now use the main results of [18, 19] at the same time. We deduce our main theorem for type B, Theorem 7.4, that we state now. Write $I \subseteq K^\times$ as $I = \coprod_{j=1}^{d} I^{(j)}$, such that each $I^{(j)}$ is q^2-connected *and* stable by scalar inversion. As in the type A case, for $\Lambda = (\Lambda_i)_{i\in I} \in \mathbb{N}^{(I)}$, we denote by $H^\Lambda(B_n)$ the quotient of the affine Hecke algebra of type B by the relation

$$\prod_{i\in I}(X_1 - i)^{\Lambda_i} = 0$$

(see §7.1 for a precise definition).

THEOREM. — *We have an (explicit) isomorphism*

$$H^\Lambda(B_n) \simeq \bigoplus_{\substack{n_1,\dots,n_d\geq 0\\ n_1+\cdots+n_d=n}} \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}}\left(\bigotimes_{j=1}^{d} H^{\Lambda^{(j)}}(B_{n_j})\right);$$

in particular, we have a Morita equivalence

$$H^\Lambda(B_n) \overset{Morita}{\simeq} \bigoplus_{\substack{n_1,\dots,n_d\geq 0\\ n_1+\cdots+n_d=n}} \bigotimes_{j=1}^{d} H^{\Lambda^{(j)}}(B_{n_j}).$$

We also deduce that a similar result holds for the cyclotomic quotient $H^\Lambda(D_n)$ of the affine Hecke algebra of type D. Some technicalities typical to the type D situation result in a formulation of the final result that is a bit more complicated than for type B in the theorem above, since in addition it involves a semi-direct product by powers of a cyclic group of order 2 (see Theorem 7.10).

One motivation for considering cyclotomic quotients of affine Hecke algebras is that the study of (finite-dimensional) representations of the affine Hecke algebra is equivalent to the study of representations of all their cyclotomic quotients. As a consequence of our main results, we obtain that, for affine Hecke algebras of types B and D, this study reduces to considering the algebras $H^\Lambda(B_n)$ and $H^\Lambda(D_n)$ when the set I is q^2-connected and stable by scalar

inversion (see Corollaries 7.5 and 7.11 for more details and a complete description of the finite number — up to four — of sets I to be considered). This generalises the classical reduction for the affine Hecke algebras of type A (for which it is enough to consider $I = q^{2\mathbb{Z}}$) induced by the Dipper–Mathas result.

Organisation of the paper. In Section 2, given an algebra A and a set of idempotents satisfying certain properties we prove a general decomposition theorem expressing A in terms of a direct sum involving matrix algebras on idempotent truncations (Corollary 2.7).

Let Γ be a (possible infinite) quiver with no 1-loops, let I be its vertex set and let $\alpha \subseteq \mathfrak{S}_n$ be a finite union of $\mathfrak{S}_n$-orbits. In Section 3, we recall the definition of the quiver Hecke algebra $R_\alpha(\Gamma)$. We then review the proof, based on the general theorem from Section 2, of the decomposition isomorphism of [20] when Γ has several connected components, generalising it to the case where Γ is not necessarily finite (as it is assumed in [20]). In §3.2.4, given a finitely-supported family Λ of non-negative integers we define the cyclotomic quotient $R_\alpha^\Lambda(\Gamma)$ of $R_\alpha(\Gamma)$ and give the corresponding isomorphism when Γ has several connected components.

Then we assume that Γ is endowed with an involution θ and let $\beta \subseteq I^n$ be an orbit for the action of the Weyl group B_n of type B and rank n. We begin Section 4 by defining the algebra $V_\beta(\Gamma, \lambda, \gamma)$ depending in addition on $\lambda \in \mathbb{N}^I$ and $\gamma \in K^I$ satisfying certain conditions. This algebra generalises the constructions of [24, 18, 19], see Remarks 4.5, 4.6 and 4.7, respectively. The algebra $V_\beta(\Gamma, \lambda, \gamma)$ is $\mathbb{Z}$-graded, and we prove in §4.2 that it admits a PBW basis, using a polynomial realisation (the calculations are postponed to Appendix A).

Section 5 is the heart of the paper. We prove a decomposition theorem, similar to (♠), for the algebra $V_\beta(\Gamma, \lambda, \gamma)$, when the quiver Γ is a disjoint union of θ-stable full subquivers $\Gamma = \coprod_{j=1}^d \Gamma^{(j)}$ (Theorem 5.1). As in Section 3, we first use the results of Section 2 and then prove that some idempotent truncation of $V_\beta(\Gamma, \lambda, \gamma)$ can be expressed as a tensor product on smaller algebras involving the quivers $\Gamma^{(j)}$. Note here a technical difficulty compared with the type A case: for $n_1 + \cdots + n_d = n$, the group $\mathfrak{S}_{n_1} \times \cdots \times \mathfrak{S}_{n_d}$ can be seen as a parabolic subgroup of $\mathfrak{S}_n$ for its standard Coxeter structure, but it is no longer the case for $B_{n_1} \times \cdots \times B_{n_d} \subseteq B_n$, although this is still a subgroup. We prove in §5.3 the cyclotomic analogue of the decomposition theorem (Corollary 5.5).

The shorter Section 6 is devoted to quiver Hecke algebras $W_\beta(\Gamma)$ for type D and their cyclotomic quotients $W_\beta^\Lambda(\Gamma)$. Using a result of [19] that expresses $W_\beta(\Gamma)$ as the subalgebra of fixed points of a certain involutive automorphism of $V_\beta(\Gamma, 0, 0)$ (Proposition 6.4), we manage to give a decomposition isomorphism for $W_\beta(\Gamma)$ and its cyclotomic quotient when the quiver Γ has several θ-stable full subquivers (Theorem 6.8).

Finally, in Section 7, we introduce the affine Hecke algebras $H(B_n)$ of type B and $H(D_n)$ of type D, together with their cyclotomic quotients $H^\Lambda(B_n)$ and $H^\Lambda(D_n)$. We then use the analogues of the Brundan–Kleshchev isomorphism theorem in types B and D from [18, 19] to deduce from our disjoint quiver isomorphisms the announced Morita equivalences: Theorem 7.4 for type B and Theorem 7.10 for type D.

2. Decomposition in matrix algebras on idempotent truncations

The results in this section, or some versions thereof, are probably known to specialists, but we could not find them in this precise form in the literature. So we state them in the form that we need and provide complete proofs. The framework presented here encompasses several cases of proved isomorphism theorems, such as in [13, 20].

Let A be a unitary algebra over a ring K. Let $\mathcal{I}$ be a complete (finite) set of orthogonal idempotents, that is,

- for all $e \in \mathcal{I}$ we have $e^2 = e$;
- for all $e, e' \in \mathcal{I}$, if $e \neq e'$ then $ee' = e'e = 0$;
- we have $1 = \sum_{e\in\mathcal{I}} e$.

For any $e \in \mathcal{I}$, let $\phi_e, \psi_e \in A$, such that

$$\phi_e\psi_e e = e, \tag{1a}$$

$$e\phi_e\psi_e = e. \tag{1b}$$

Remark 2.1. — Such elements necessarily exist, for instance $\phi_e = \psi_e = e$, for any $e \in \mathcal{I}$. However, obviously this will not lead to interesting results.

Lemma 2.2. — *For any $e \in \mathcal{I}$, the element $\psi_e e\phi_e$ is an idempotent.*

Proof. — Using (1a), we have

$$\begin{aligned}(\psi_e e\phi_e)^2 &= \psi_e e(\phi_e\psi_e e)\phi_e\\ &= \psi_e e^2\phi_e\\ &= \psi_e e\phi_e,\end{aligned}$$

as desired. □

Denote by $\mathcal{J}$ the image of the map $\mathcal{I} \to A$, $e \mapsto \psi_e e\phi_e$ and write $\mathcal{I}_\epsilon$ for the fibre of any element $\epsilon \in \mathcal{J}$. We have

$$\mathcal{I}_\epsilon = \{e \in \mathcal{I} : \psi_e e\phi_e = \epsilon\},$$

and

$$\bigsqcup_{\epsilon\in\mathcal{J}} \mathcal{I}_\epsilon = \mathcal{I}.$$

By Lemma 2.2, the set $\mathcal{J}$ consists of idempotents, however it is a priori not related to $\mathcal{I}$.

PROPOSITION 2.3. — *For any $\epsilon \in \mathcal{J}$ and any $e \in \mathcal{I}_\epsilon$, we have*

$$e\phi_e = \phi_e\epsilon, \tag{2a}$$

$$\epsilon\psi_e = \psi_e e. \tag{2b}$$

Proof. — We have

$$\psi_e e\phi_e = \epsilon, \tag{3}$$

and, thus, $(\phi_e\psi_e e)\phi_e = \phi_e\epsilon$. Using (1a) we obtain the first equality. We also obtain $\psi_e(e\phi_e\psi_e) = \epsilon\psi_e$ from (3), and, thus, by (1b) we obtain the second equality. □

PROPOSITION 2.4. — *For any $\epsilon \in \mathcal{J}$ and any $e \in \mathcal{I}_\epsilon$, we have*

$$\psi_e\phi_e\epsilon = \epsilon, \tag{4a}$$

$$\epsilon\psi_e\phi_e = \epsilon. \tag{4b}$$

Proof. — By (2a) we have $\phi_e\epsilon = e\phi_e$, and, thus,

$$\psi_e\phi_e\epsilon = \psi_e e\phi_e,$$

and we conclude that (4a) holds, since $\psi_e e\phi_e = \epsilon$ by definition of $\mathcal{I}_\epsilon$. Similarly, by (2b) we have

$$\epsilon\psi_e\phi_e = \psi_e e\phi_e = \epsilon,$$

and, thus, (4b) holds. □

If J is any finite set and B any K-algebra, we denote by $\mathrm{Mat}_J(B)$ the K-algebra of matrices with rows and columns indexed by J with entries in B.

DEFINITION 2.5. — For any $\epsilon \in \mathcal{J}$, we define the idempotent

$$\hat{\epsilon} := \sum_{e \in \mathcal{I}_\epsilon} e.$$

THEOREM 2.6. — *Let $\epsilon \in \mathcal{J}$. We have the following isomorphism of K-algebras:*

$$\hat{\epsilon}A\hat{\epsilon} \simeq \mathrm{Mat}_{\mathcal{I}_\epsilon}(\epsilon A\epsilon).$$

Proof. — We first prove that for any $e', e \in \mathcal{I}_\epsilon$, the maps

$$\begin{aligned} \theta_{e'e} : \quad e'Ae &\to \epsilon A\epsilon M_{e'e} \\ a &\mapsto \psi_{e'} a \phi_e M_{e'e}, \end{aligned}$$

and

$$\begin{aligned}\eta_{e'e}: \quad \epsilon A\epsilon M_{e'e} &\to e'Ae\\ aM_{e'e} &\mapsto \phi_{e'}a\psi_e,\end{aligned}$$

are well defined and inverse isomorphisms of K-modules. Here, we denoted by $M_{e'e} \in \mathrm{Mat}_{\mathcal{I}_\epsilon}(\epsilon A\epsilon)$ the matrix whose unique non-zero coefficient, which is 1, is at row e' and column e. The maps $\theta_{e'e}$ and $\eta_{e'e}$ are well defined by Proposition 2.3. Indeed, for any $a \in e'Ae$, then $a = e'ae$, and

$$\psi_{e'}a\phi_e = (\psi_{e'}e')a(e\phi_e) = (\epsilon\psi_{e'})a(\phi_e\epsilon) \in \epsilon A\epsilon,$$

so $\theta_{e'e}$ is well defined, and for any $a \in \epsilon A\epsilon$, then $a = \epsilon a\epsilon$, and

$$\phi_{e'}a\psi_e = (\phi_{e'}\epsilon)a(\epsilon\psi_e) = (e'\phi_{e'})a(\psi_e e) \in e'Ae,$$

so $\eta_{e'e}$ is well defined. Now for any $a \in e'Ae$, we have, using $a = e'ae$ and (1),

$$\begin{aligned}\eta_{e'e}\big(\theta_{e'e}(a)\big) &= \eta_{e'e}(\psi_{e'}a\phi_e M_{e'e})\\ &= \phi_{e'}(\psi_{e'}a\phi_e)\psi_e\\ &= (\phi_{e'}\psi_{e'}e')a(e\phi_e\psi_e)\\ &= e'ae\\ &= a.\end{aligned}$$

Moreover, for any $a \in \epsilon A\epsilon$, we have, using $a = \epsilon a\epsilon$ and Proposition 2.3,

$$\begin{aligned}\theta_{e'e}\big(\eta_{e'e}(aM_{e'e})\big) &= \theta_{e'e}(\phi_{e'}a\psi_e)\\ &= \psi_{e'}\phi_{e'}a\psi_e\phi_e M_{e'e}\\ &= (\psi_{e'}\phi_{e'}\epsilon)a(\epsilon\psi_e\phi_e)M_{e'e}\\ &= \epsilon a\epsilon\\ &= a.\end{aligned}$$

We now want to extend $\theta_{e'e}$ and $\eta_{e'e}$ to algebra isomorphisms. We have a direct sum decomposition

$$\hat{\epsilon}A\hat{\epsilon} = \bigoplus_{e',e\in\mathcal{I}_\epsilon} e'Ae. \tag{5}$$

We define two maps

$$\begin{aligned}\theta_\epsilon &: \hat{\epsilon}A\hat{\epsilon} \to \mathrm{Mat}_{\mathcal{I}_\epsilon}(\epsilon A\epsilon),\\ \eta_\epsilon &: \mathrm{Mat}_{\mathcal{I}_\epsilon}(\epsilon A\epsilon) \to \hat{\epsilon}A\hat{\epsilon},\end{aligned}$$

by

$$\begin{aligned}\theta_\epsilon &\coloneqq \bigoplus_{e',e\in\mathcal{I}_\epsilon} \theta_{e'e},\\ \eta_\epsilon &\coloneqq \bigoplus_{e',e\in\mathcal{I}_\epsilon} \eta_{e'e}.\end{aligned}$$

These two maps are inverse isomorphisms of K-modules. To prove that they are inverse isomorphisms of K-algebras, it suffices to prove that θ_ϵ is a morphism of K-algebras. Recalling the decomposition (5), it suffices to prove that

$$\theta_\epsilon(a_1a_2) = \theta_\epsilon(a_1)\theta_\epsilon(a_2), \tag{6}$$

for any $a_i \in e_i' A e_i$ for any $e_i \in \mathcal{I}_\epsilon$. If $e_1 \neq e_2'$, then the left-hand side is zero, and so is the right-hand one, since $M_{e_1'e_1} M_{e_2'e_2} = 0_{\mathrm{Mat}_{\mathcal{I}_\epsilon}(\epsilon A \epsilon)}$. Thus, we now assume that $e_1 = e_2'$. We have $a_1 = a_1 e_1$ and $a_1 a_2 = a_1 e_1 a_2 \in e_1' A e_2$; thus, using (1b) we obtain

$$\begin{aligned}
\theta_\epsilon(a_1a_2) &= \theta_{e_1'e_2}(a_1a_2) \\
&= \psi_{e_1'} a_1(e_1) a_2 \phi_{e_2} M_{e_1'e_2} \\
&= \psi_{e_1'} a_1(e_1\phi_{e_1}\psi_{e_1}) a_2 \phi_{e_2} M_{e_1'e_2} \\
&= (\psi_{e_1'} a_1 e_1 \phi_{e_1})(\psi_{e_1} a_2 \phi_{e_2}) M_{e_1'e_2} \\
&= \left(\psi_{e_1'} a_1 \phi_{e_1} M_{e_1'e_1}\right)\left(\psi_{e_1} a_2 \phi_{e_2} M_{e_1e_2}\right) \\
&= \theta_{e_1'e_1}(a_1)\theta_{e_1e_2}(a_2) \\
&= \theta_\epsilon(a_1)\theta_\epsilon(a_2).
\end{aligned}$$

This concludes the proof. □

Corollary 2.7. — *Assume that for all $\epsilon, \epsilon' \in \mathcal{J}$ we have*

$$\epsilon \neq \epsilon' \implies \hat{\epsilon} A \hat{\epsilon}' = \{0\}. \tag{7}$$

Then have the following isomorphism of K-algebras:

$$A \simeq \bigoplus_{\epsilon \in \mathcal{J}} \mathrm{Mat}_{\mathcal{I}_\epsilon}(\epsilon A \epsilon).$$

Proof. — The assumption (7) implies that

$$A \simeq \bigoplus_{\epsilon \in \mathcal{J}} \hat{\epsilon} A \hat{\epsilon}.$$

We now use the result of Theorem 2.6. □

3. Application to quiver Hecke algebras

We here review and generalise the decomposition theorem from [20, §6] to the case of a possibly infinite quiver. A careful analysis of the proofs in this section will be the starting point of several proofs later in the paper.

3.1. Definition. — Let Γ be a loop-free quiver, possibly infinite. We write I (or, respectively, A) for the vertex (or, respectively, arrow) set. We have a map $A \to I \times I$ given by $A \ni a \mapsto \big(o(a), t(a)\big) \in I \times I$. The loop-free condition says that for all $a \in A$, we have $o(a) \neq t(a)$. For any $i, j \in I$, we write $|i \to j|$ for the (finite) number of $a \in A$, such that $o(a) = i$ and $t(a) = j$. We also define $i \cdot j \coloneqq |i \to j| + |i \leftarrow j|$. (We warn the reader that the usual quantity is $-i \cdot j$.) For any $i, j \in I$, we define

$$d(i,j) \coloneqq \begin{cases} i \cdot j, & \text{if } i \neq j, \\ -2, & \text{otherwise.} \end{cases}$$

Let u, v be two indeterminates over K. For any $i, j \in I$, we define the polynomial $Q_{ij}(u,v) \in K[u,v]$ by

$$Q_{ij}(u,v) \coloneqq \begin{cases} (-1)^{|i \to j|}(u-v)^{i \cdot j}, & \text{if } i \neq j, \\ 0, & \text{otherwise,} \end{cases} \tag{8}$$

Note that

$$Q_{ij}(u,v) = Q_{ji}(v,u) = Q_{ij}(-v,-u). \tag{9}$$

Let $n \in \mathbb{N}$ and $\mathfrak{S}_n$ be the symmetric group on n letters. We denote by r_a the transposition $(a, a+1) \in \mathfrak{S}_n$, for any $a \in \{1, \dots, n-1\}$. We will consider the following two actions of $\mathfrak{S}_n$:

- The natural action on $\{1, \dots, n\}$, given by $r_a \cdot i \coloneqq r_a(i)$, for all $a \in \{1, \dots, n-1\}$ and $i \in \{1, \dots, n\}$.
- The action on I^n by place permutation, given by

$$r_a \cdot (\dots, i_a, i_{a+1}, \dots) \coloneqq (\dots, i_{a+1}, i_a, \dots), \tag{10}$$

for any $\boldsymbol{i} = (i_1, \dots, i_n) \in I^n$ and $a \in \{1, \dots, n-1\}$.

Let $\alpha \subseteq I^n$ be a finite $\mathfrak{S}_n$-stable subset, that is, a finite union of $\mathfrak{S}_n$-orbits.

Definition 3.1 (Khovanov–Lauda [15, 16], Rouquier [22]). — The *quiver Hecke algebra* associated with the quiver Γ and the finite stable $\mathfrak{S}_n$-subset $\alpha \subseteq I^n$, denoted by $R_\alpha(\Gamma)$, is the associative unitary K-algebra generated by elements

$$\{y_a\}_{1 \leq a \leq n} \cup \{\psi_b\}_{1 \leq b \leq n-1} \cup \{e(\boldsymbol{i})\}_{\boldsymbol{i} \in \alpha},$$

and relations, for any $\boldsymbol{i}, \boldsymbol{j} \in \alpha$ and $a, b \in \{1, \dots, n\}$,

$$\sum_{\boldsymbol{i} \in \alpha} e(\boldsymbol{i}) = 1, \quad e(\boldsymbol{i})e(\boldsymbol{j}) = \delta_{\boldsymbol{ij}} e(\boldsymbol{i}), \quad y_a y_b = y_b y_a, \quad y_a e(\boldsymbol{i}) = e(\boldsymbol{i}) y_a, \tag{11}$$

and

$$\psi_a e(\boldsymbol{i}) = e(r_a \cdot \boldsymbol{i})\psi_a, \tag{12}$$

$$(\psi_a y_b - y_{r_a(b)}\psi_a)e(\boldsymbol{i}) = \begin{cases} -e(\boldsymbol{i}), & \text{if } b = a \text{ and } i_a = i_{a+1}, \\ e(\boldsymbol{i}), & \text{if } b = a+1 \text{ and } i_a = i_{a+1}, \\ 0, & \text{otherwise,} \end{cases} \tag{13}$$

if $a \leq n-1$, and, finally,

$$\psi_a \psi_b = \psi_b \psi_a, \text{ if } |a-b| > 1, \tag{14}$$

$$\psi_a^2 e(\boldsymbol{i}) = Q_{i_a i_{a+1}}(y_a, y_{a+1})e(\boldsymbol{i}), \tag{15}$$

$$\begin{aligned}(\psi_{b+1}\psi_b\psi_{b+1} - \psi_b\psi_{b+1}\psi_b)e(\boldsymbol{i}) \\ = \begin{cases} \dfrac{Q_{i_b i_{b+1}}(y_b, y_{b+1}) - Q_{i_b i_{b+1}}(y_{b+2}, y_{b+1})}{y_b - y_{b+2}} e(\boldsymbol{i}), & \text{if } i_b = i_{b+2}, \\ 0, & \text{otherwise,} \end{cases}\end{aligned} \tag{16}$$

if $a \leq n-1$ and $b \leq n-2$.

We may form the direct sum $R_n(\Gamma) \coloneqq \bigoplus_\alpha R_\alpha(\Gamma)$, where α runs over all the orbits of I^n under the action of $\mathfrak{S}_n$. If Γ is finite, the direct sum is finite, and $R_n(\Gamma)$ is a unitary algebra, with unit $\sum_{\boldsymbol{i} \in I^n} e(\boldsymbol{i})$. Note that if $n = 0$, then $R_\alpha(\Gamma) = R_0(\Gamma) = K$.

PROPOSITION 3.2 ([15, 16, 22]). — *The algebra $R_\alpha(\Gamma)$ is endowed with the $\mathbb{Z}$-grading given by*

$$\begin{aligned} \deg e(\boldsymbol{i}) &= 0, \\ \deg y_a &= 2, \\ \deg \psi_b e(\boldsymbol{i}) &= d(i_b, i_{b+1}), \end{aligned}$$

for all $\boldsymbol{i} \in \alpha$ and $a, b \in \{1, \ldots, n\}$ with $b \leq n-1$.

For any $w \in \mathfrak{S}_n$, choose a reduced expression $w = r_{a_1} \cdots r_{a_k}$ and define $\psi_w \coloneqq \psi_{a_1} \cdots \psi_{a_k}$. Note that the element ψ_w may depend on the reduced expression chosen.

PROPOSITION 3.3 ([15, 16, 22]). — *The algebra $R_\alpha(\Gamma)$ is a free K-module, and*

$$\left\{ y_1^{a_1} \cdots y_n^{a_n} \psi_w e(\boldsymbol{i}) : a_i \in \mathbb{N}, w \in \mathfrak{S}_n, \boldsymbol{i} \in \alpha \right\},$$

is a K-basis.

REMARK 3.4. — We recall that there is a one-to-one correspondence between $\mathfrak{S}_n$-orbits $\alpha \subset I^n$ and maps $\hat{\alpha} : I \to \mathbb{N}$ of *weight* n, namely, such that $\sum_{i \in I} \hat{\alpha}(i) = n$ (the number $\hat{\alpha}(i)$ counts the number of occurrence of i in any element in the orbit α).

3.2. Disjoint union of quivers. — Let $d \in \mathbb{N}^*$. Like in [20, §6.1.3], we assume that the quiver Γ decomposes as a disjoint union of full subquivers

$$\Gamma = \bigsqcup_{j=1}^{d} \Gamma^{(j)} \ ,$$

where there are no arrows between $\Gamma^{(j)}$ and $\Gamma^{(j')}$ if $j \neq j'$. We denote by $I = \coprod_{j=1}^{d} I^{(j)}$ the subsequent partition of the vertex set. Note that $Q_{ii'} = 1$, whenever $i \in I^{(j)}$ and $i' \in I^{(j')}$ with $j \neq j'$.

Now we consider a special class of finite unions of $\mathfrak{S}_n$-orbits in I^n. We let G be a finite group acting on I and, for each $j \in \{1, \dots, d\}$, we assume that $I^{(j)}$ is stable under the action of G. We denote

$$G_n = G^n \rtimes \mathfrak{S}_n \ ,$$

the semi-direct product, where $\mathfrak{S}_n$ acts on place permutation on G^n.

The semi-direct product G_n acts naturally on I^n. For any $g = (g_1, \dots, g_n) \in G^n$ and $w \in \mathfrak{S}_n$, we have, for all $(i_1, \dots, i_n) \in I^n$,

$$(g, w) \cdot (i_1, \dots, i_n) = \big(g_1 \cdot i_{w^{-1}(1)}, \dots, g_n \cdot i_{w^{-1}(n)}\big) \ .$$

We fix $\alpha \subseteq I^n$ to be a G_n-orbit. Note that α is, indeed, a finite $\mathfrak{S}_n$-stable subset of I^n as in §3.1.

3.2.1. *Decomposition of orbits.* — For any $\boldsymbol{i} \in \alpha$ and $j \in \{1, \dots, d\}$, let $\boldsymbol{i}^{(j)}$ be the tuple obtained from $\boldsymbol{i}$ by removing the entries that are not in $I^{(j)}$. We denote by $n_j(\boldsymbol{i})$ the number of remaining entries, that is, the number of components of $\boldsymbol{i}^{(j)}$. It follows easily from the fact that each $I^{(j)}$ is stable under the action of G that:

(17) the tuple $(n_1(\boldsymbol{i}), \dots, n_d(\boldsymbol{i}))$ is the same for each $\boldsymbol{i} \in \alpha$.

Thus, we denote, for each $j \in \{1, \dots, d\}$, by $n_j(\alpha)$ the unique value of $n_j(\boldsymbol{i})$ for $\boldsymbol{i} \in \alpha$. We may simply write n_j instead of $n_j(\alpha)$, when α is clear from the context. Note that $n_1 + \cdots + n_d = n$.

We define

$$\alpha^{(j)} \coloneqq \big\{\boldsymbol{i}^{(j)} : \boldsymbol{i} \in \alpha\big\} \subseteq (I^{(j)})^{n_j}.$$

The set $\alpha^{(j)}$ is a finite $\mathfrak{S}_{n_j}$-stable subset of $(I^{(j)})^{n_j}$. We will see in (19) that it is, in fact, a G_{n_j}-orbit.

In addition to (17), we will need the following property of α.

Proposition 3.5. — *Recall that $\alpha \subseteq I^n$ is a G_n-orbit. We have:*

$$\alpha^{(1)} \times \cdots \times \alpha^{(d)} \subset \alpha \ , \tag{18}$$

where we use implicitly the natural identification (by concatenation) of $I^{n_1} \times \cdots \times I^{n_d}$ with a subset of I^n.

Proof. — Let us provide a proof that shows all the various elements explicitly. Since α is a G_n-orbit, it can be written in the form:

$$\alpha = \left\{ \left(g_1 \cdot i_{w^{-1}(1)}, \dots, g_n \cdot i_{w^{-1}(n)}\right) \mid g_1, \dots, g_n \in G\,,\ w \in \mathfrak{S}_n \right\},$$

for some element $(i_1, \dots, i_n) \in I^n$. By invariance under $\mathfrak{S}_n$, we can choose $(i_1, \dots, i_n)$ in an ordered form as follows:

$$\left(i_1^1, \dots, i_{n_1}^1, \dots\dots, i_1^d, \dots, i_{n_d}^d\right),$$

where $i_k^j \in I^{(j)}$ for all $j \in \{1, \dots, d\}$ and $k \in \{1, \dots, n_j\}$. Then it is clear that for each $j \in \{1, \dots, d\}$, we have simply

$$\alpha^{(j)} = \left\{ \left(g_1 \cdot i^j_{w^{-1}(1)}, \dots, g_{n_j} \cdot i^j_{w^{-1}(n_j)}\right) \mid g_1, \dots, g_{n_j} \in G\,,\ w \in \mathfrak{S}_{n_j} \right\}.$$

Property (18) can now immediately be checked. □

From the proof of the preceding proposition, it is easy to see that the map

$$\left\{G_n\text{-orbits of } I^n\right\} \longrightarrow \bigsqcup_{\substack{n_1,\dots,n_d \geq 0 \\ n_1+\cdots+n_d=n}} \prod_{j=1}^{d} \left\{G_{n_j}\text{-orbits of } \left(I^{(j)}\right)^{n_j}\right\}, \tag{19}$$

given by $\alpha \mapsto \left(\alpha^{(1)}, \dots, \alpha^{(d)}\right)$ is a bijection. The inverse map associates to $\left(\alpha^{(1)}, \dots, \alpha^{(d)}\right)$ the smallest G_n-stable subset in I^n containing $\alpha^{(1)} \times \cdots \times \alpha^{(d)}$.

Remark 3.6. — What we actually need for the results of this section is a subset α satisfying properties (17) and (18). However, since in the entire paper, we will use only G_n-orbits, we find it more convenient to start directly with G_n-orbits. In fact, we will only use the groups $G = \{1\}$ and $G = \mathbb{Z}/2\mathbb{Z}$ but considering an arbitrary finite group G does not lead to any complication.

Remark 3.7. —

- Let Ω be the set of G-orbits of I. Generalising Remark 3.4 it is easy to see that there is a one-to-one correspondence between G_n-orbits $\alpha \subseteq I^n$ and maps $\hat{\alpha} : \Omega \to \mathbb{N}$, such that $\sum_{\omega \in \Omega} \hat{\alpha}(\omega) = n$. If $\alpha \subseteq I^n$ is a G_n-orbit and $\omega \in \Omega$, then $\hat{\alpha}(\omega)$ counts the number of occurrence of the elements of ω in any element of α.
- For each $j = 1, \dots, d$, let $\Omega^{(j)}$ be the set of G-orbits of $I^{(j)}$. We have $\Omega = \coprod_{j=1}^{d} \Omega^{(j)}$. Then the bijection (19) in terms of maps simply associates to $\hat{\alpha} : \Omega \to \mathbb{N}$ the restrictions $\hat{\alpha}|_{\Omega^{(j)}} : \Omega^{(j)} \to \mathbb{N}$ to each $\Omega^{(j)}$.

Example 3.8. — Let us give an example of a subset α not satisfying property (18). Let $n = 2$ and $\alpha = \{(a, A), (A, a), (b, B), (B, b)\}$, where $a, b \in I^{(1)}$ and $A, B \in I^{(2)}$. Then α is a union of two $\mathfrak{S}_2$-orbits, and it satisfies (17). It does not satisfy (18). Indeed, we have $\alpha^{(1)} = \{a, b\}$ and $\alpha^{(2)} = \{A, B\}$ but, for example, $(a, B) \notin \alpha$.

3.2.2. *Decomposition along the connected components of the quiver.* — We keep $\alpha \subseteq I^n$ a G_n-orbit for some finite group G acting on each set $I^{(j)}$. We may (and we will) simply write n_j, instead of $n_j(\alpha)$.

For each $i \in I$, we set $p(i) = j \in \{1, \dots, d\}$ if $i \in I^{(j)}$. Then for each $\boldsymbol{i} = (i_1, \dots, i_n) \in I^n$, we define its *profile* by $p(\boldsymbol{i}) = \big(p(i_1), \dots, p(i_n)\big) \in \{1, \dots, d\}^n$. Let

$$\mathrm{Prof}^\alpha \coloneqq \{p(\boldsymbol{i})\,,\ \boldsymbol{i} \in \alpha\} \subseteq \{1, \dots, d\}^n$$

be the set of all profiles of elements of α. Note that (17) ensures that Prof^α is also a single orbit, now for the action of $\mathfrak{S}_n$ on $\{1, \dots, d\}^n$ by place permutation.

A natural element to consider in this orbit Prof^α is

$$\mathfrak{t}^\alpha \coloneqq (1, \dots, 1, 2, \dots, 2, \dots\dots, d, \dots, d),$$

where each $j \in \{1, \dots, d\}$ appears exactly n_j times. Then every element $\mathfrak{t} \in \mathrm{Prof}^\alpha$ can be reordered to obtain the distinguished element $\mathfrak{t}^\alpha$. More precisely, for any $\mathfrak{t} \in \mathrm{Prof}^\alpha$, the set of elements $w \in \mathfrak{S}_n$, such that $w \cdot \mathfrak{t} = \mathfrak{t}^\alpha$ forms a right coset in $\mathfrak{S}_n$ for the subgroup $\mathfrak{S}_{n_1} \times \cdots \times \mathfrak{S}_{n_d}$ (the stabiliser of $\mathfrak{t}^\alpha$). There is a unique minimal length element in this coset (see e.g. [10]), and we denote it by $\pi_\mathfrak{t}$. In particular, the element $\pi_\mathfrak{t}$ is the unique minimal length element of $\mathfrak{S}_n$, such that $\pi_\mathfrak{t} \cdot \mathfrak{t} = \mathfrak{t}^\alpha$.

For any $\mathfrak{t} \in \mathrm{Prof}^\alpha$, we define the idempotent

$$e(\mathfrak{t}) = \sum_{\substack{\boldsymbol{i} \in \alpha \\ p(\boldsymbol{i}) = \mathfrak{t}}} e(\boldsymbol{i}) \in R_\alpha(\Gamma),$$

and we set

$$\mathcal{I} \coloneqq \big\{e(\mathfrak{t}) : \mathfrak{t} \in \mathrm{Prof}^\alpha\big\}.$$

It is a complete set of orthogonal idempotents, and its cardinality is $\binom{n}{n_1, \dots, n_d}$. Then, for any $\mathfrak{t} \in \mathrm{Prof}^\alpha$, we fix a reduced expression $\pi_\mathfrak{t} = r_{a_1} \cdots r_{a_k}$ and define

$$\psi_\mathfrak{t} \coloneqq \psi_{a_1} \cdots \psi_{a_k} \in R_\alpha(\Gamma), \tag{20a}$$

$$\phi_\mathfrak{t} \coloneqq \psi_{a_k} \cdots \psi_{a_1} \in R_\alpha(\Gamma). \tag{20b}$$

In the following proposition, the grading on $\mathrm{Mat}_\mathcal{I}\big(e(\mathfrak{t}^\alpha) R_\alpha(\Gamma) e(\mathfrak{t}^\alpha)\big)$ is trivially induced from the grading on $e(\mathfrak{t}^\alpha) R_\alpha(\Gamma) e(\mathfrak{t}^\alpha)$ (a homogeneous element of degree N is a matrix where all coefficients are homogeneous elements of degree N).

Proposition 3.9. — *We have an isomorphism of graded algebras:*

$$R_\alpha(\Gamma) \simeq \mathrm{Mat}_{\binom{n}{n_1, \dots, n_d}}\big(e(\mathfrak{t}^\alpha) R_\alpha(\Gamma) e(\mathfrak{t}^\alpha)\big).$$

Proof. — The proof follows the same steps as in [20], and we only give a sketch and the precise references to [20]. First, we have that the data $\{e(\mathfrak{t}), \psi_{\mathfrak{t}}, \phi_{\mathfrak{t}}\}_{\mathfrak{t}\in \mathrm{Prof}^{\alpha}}$ in $R_{\alpha}(\Gamma)$ enters the general setting (1) of Section 2; namely, we have, for any $\mathfrak{t} \in \mathrm{Prof}^{\alpha}$ (see [20, Proposition 6.18]),

$$\phi_{\mathfrak{t}}\psi_{\mathfrak{t}}e(\mathfrak{t}) = e(\mathfrak{t})\phi_{\mathfrak{t}}\psi_{\mathfrak{t}} = e(\mathfrak{t}). \tag{21}$$

The main point to prove (21) is the following fact:

$$\psi_a^2 e(\mathfrak{t}) = e(\mathfrak{t}), \tag{22}$$

for any $a \in \{1, \dots, n-1\}$ and $\mathfrak{t} \in \mathrm{Prof}^{\alpha}$, such that $\mathfrak{t}_a \neq \mathfrak{t}_{a+1}$ (see [20, Lemma 6.15]). Similarly, we obtain, for any $\mathfrak{t} \in \mathrm{Prof}^{\alpha}$,

$$\psi_{\mathfrak{t}}e(\mathfrak{t})\phi_{\mathfrak{t}} = \psi_{\mathfrak{t}}\phi_{\mathfrak{t}}e(\mathfrak{t}^{\alpha}) = e(\mathfrak{t}^{\alpha}). \tag{23}$$

This last equality ensures that the set $\mathcal{J}$ in the notation of §2 is $\mathcal{J} = \{e(\mathfrak{t}^{\alpha})\}$. Since $\mathcal{J}$ is reduced to one element, we deduce that the assumption (7) is automatically satisfied, and we can use Corollary 2.7 to obtain the proposition. Finally, the fact that the isomorphism is homogeneous follows from $\deg \psi_{\mathfrak{t}}e(\mathfrak{t}) = \deg \phi_{\mathfrak{t}}e(\mathfrak{t}) = 0$, for any $\mathfrak{t} \in \mathrm{Prof}^{\alpha}$ (see [20, Remark 6.29]). □

Remark 3.10. — Similarly to (22), we have (see [20, Lemma 6.20])

$$y_a\phi_{\mathfrak{t}}e(\mathfrak{t}) = \phi_{\mathfrak{t}}y_{\pi_{\mathfrak{t}}(a)}e(\mathfrak{t}), \tag{24}$$

for any $a \in \{1, \dots, n-1\}$ and $\mathfrak{t} \in \mathrm{Prof}^{\alpha}$, such that $\mathfrak{t}_a \neq \mathfrak{t}_{a+1}$, and also (see [20, Lemma 6.15])

$$\psi_{a+1}\psi_a\psi_{a+1}e(\mathfrak{t}) = \psi_a\psi_{a+1}\psi_a e(\mathfrak{t}), \tag{25}$$

for any $a \in \{1, \dots, n-2\}$ and $\mathfrak{t} \in \mathrm{Prof}^{\alpha}$, such that $\mathfrak{t}_a \neq \mathfrak{t}_{a+2}$. In particular, (25) implies that the quantities $\psi_{\mathfrak{t}}e(\mathfrak{t})$ and $e(\mathfrak{t})\phi_{\mathfrak{t}}$ do not depend on the chosen reduced expression for $\pi_{\mathfrak{t}}$.

3.2.3. *Expression as a tensor product.* — We now want to write the algebra $e(\mathfrak{t}^{\alpha})R_{\alpha}(\Gamma)e(\mathfrak{t}^{\alpha})$ as a tensor product. Recall that α is a G_n-orbit and thus satisfies properties (17) and (18). We have already used the first property. The second will be explicitly used during the proof of the next result.

Note that, for any $j \in \{1, \dots, d\}$, the algebra $R_{\alpha^{(j)}}(\Gamma^{(j)})$ is well-defined since $\alpha^{(j)}$ consists of n_j-tuples of vertices $I^{(j)}$ of $\Gamma^{(j)}$ and is stable under permutations (see §3.2.1).

Theorem 3.11. — *We have an (explicit) isomorphism of graded algebras:*

$$e(\mathfrak{t}^{\alpha})R_{\alpha}(\Gamma)e(\mathfrak{t}^{\alpha}) \simeq R_{\alpha^{(1)}}(\Gamma^{(1)}) \otimes \cdots \otimes R_{\alpha^{(d)}}(\Gamma^{(d)}).$$

Proof. — We construct an algebra homomorphism f from the tensor product of $e(\mathfrak{t}^\alpha)R_\alpha(\Gamma)e(\mathfrak{t}^\alpha)$ as follows. For any $\boldsymbol{i}^{(j)} \in \alpha^{(j)} \subseteq (I^{(j)})^{n_j}$ with $j \in \{1, \dots, d\}$, we define

$$f\big(e(\boldsymbol{i}^{(1)}) \otimes \cdots \otimes e(\boldsymbol{i}^{(d)})\big) \coloneqq e(\boldsymbol{i}^{(1)}, \dots, \boldsymbol{i}^{(d)}).$$

Note that $\big(\boldsymbol{i}^{(1)}, \dots, \boldsymbol{i}^{(d)}\big) \in \alpha$ due to Proposition 3.5. Moreover, for any $j \in \{1, \dots, d\}$, we denote by $y_a^{(j)}$ and $\psi_b^{(j)}$ the generators of $R_{\alpha^{(j)}}(\Gamma^{(j)})$ in the tensor product and we define

$$f(y_a^{(j)}) \coloneqq e(\mathfrak{t}^\alpha) y_{n_1+\cdots+n_{j-1}+a} e(\mathfrak{t}^\alpha),$$
$$f(\psi_b^{(j)}) \coloneqq e(\mathfrak{t}^\alpha) \psi_{n_1+\cdots+n_{j-1}+b} e(\mathfrak{t}^\alpha),$$

for all $a, b \in \{1, \dots, n_j\}$ with $b \leq n_j - 1$. By [20, Lemma 6.24] the map f is, indeed, a homomorphism. Using the basis of Proposition 3.3, we can prove that f sends a basis onto a basis and is, thus, an isomorphism (see [20, Proposition 6.25]). Finally, the isomorphism f is clearly homogeneous. □

Combining Theorem 3.11 with Proposition 3.9, we obtain the main result of this section.

Corollary 3.12. — *We have an (explicit) isomorphism of graded algebras:*

$$R_\alpha(\Gamma) \simeq \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}} \left(\bigotimes_{j=1}^{d} R_{\alpha^{(j)}}(\Gamma^{(j)}) \right).$$

Remark 3.13. — If $\alpha = \coprod_{i=1}^{k} \alpha_i$ the decomposition of α into $\mathfrak{S}_n$-orbits, then we have $R_\alpha(\Gamma) = \bigoplus_{i=1}^{k} R_{\alpha_i}(\Gamma)$. So, of course, as far as the algebras $R_\alpha(\Gamma)$ are concerned, taking α a single $\mathfrak{S}_n$-orbit would be enough. However, we really needed a more general setting, since we will later apply the above results for orbits $\alpha \subset I^n$ of the Weyl group of type B.

We now show how to recover [20, Theorem 6.26], with the difference that the result that we obtain here is also valid if the quiver Γ is infinite.

Corollary 3.14. — *We have an (explicit) isomorphism of graded algebras:*

$$R_n(\Gamma) \simeq \bigoplus_{\substack{n_1,\dots,n_d \geq 0 \\ n_1+\cdots+n_d=n}} \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}} \left(\bigotimes_{j=1}^{d} R_{n_j}(\Gamma^{(j)}) \right).$$

Proof. — We write $I^n/\mathfrak{S}_n$ to denote the $\mathfrak{S}_n$-orbits in I^n. We apply the isomorphism of Corollary 3.12 in each term in the right-hand side of the equality

$R_n(\Gamma) = \bigoplus_\alpha R_\alpha(\Gamma)$, where α runs over $I^n/\mathfrak{S}_n$ (so here we use the situation $G = \{1\}$). Recalling the 1:1-correspondence in (19), we obtain

$$\begin{aligned}
R_n(\Gamma) &\simeq \bigoplus_{\alpha\in I^n/\mathfrak{S}_n} R_\alpha(\Gamma)\\
&\simeq \bigoplus_{\alpha\in I^n/\mathfrak{S}_n} \mathrm{Mat}_{\binom{n}{n_1(\alpha),\dots,n_d(\alpha)}}\left(\bigotimes_{j=1}^d R_{\alpha^{(j)}}(\Gamma^{(j)})\right)\\
&\simeq \bigoplus_{\substack{n_1,\dots,n_d\geq 0\\ n_1+\cdots+n_d=n}} \bigoplus_{\substack{\alpha\in I^n/\mathfrak{S}_n\\ n_j(\alpha)=n_j}} \mathrm{Mat}_{\binom{n}{n_1(\alpha),\dots,n_d(\alpha)}}\left(\bigotimes_{j=1}^d R_{\alpha^{(j)}}(\Gamma^{(j)})\right)\\
&\simeq \bigoplus_{\substack{n_1,\dots,n_d\geq 0\\ n_1+\cdots+n_d=n}} \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}}\left(\bigoplus_{\substack{\alpha\in I^n/\mathfrak{S}_n\\ n_j(\alpha)=n_j}} \bigotimes_{j=1}^d R_{\alpha^{(j)}}(\Gamma^{(j)})\right)\\
&\simeq \bigoplus_{\substack{n_1,\dots,n_d\geq 0\\ n_1+\cdots+n_d=n}} \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}}\left(\bigoplus_{\alpha^{(1)}\in I^{n_1}/\mathfrak{S}_{n_1}} \cdots \bigoplus_{\alpha^{(d)}\in I^{n_d}/\mathfrak{S}_{n_d}} \bigotimes_{j=1}^d R_{\alpha^{(j)}}(\Gamma^{(j)})\right)\\
&\simeq \bigoplus_{\substack{n_1,\dots,n_d\geq 0\\ n_1+\cdots+n_d=n}} \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}}\left(\bigotimes_{j=1}^d R_{n_j}(\Gamma^{(j)})\right),
\end{aligned}$$

as desired. □

3.2.4. *Cyclotomic case.* — We keep the above setting with the quiver Γ, its full sub-quivers $\Gamma^{(j)}$ and a G_n-orbit α. In addition, let $\Lambda = (\Lambda_i)_{i\in I}$ be a finitely-supported family of non-negative integers.

Definition 3.15 ([22, 6]). — The *cyclotomic* quiver Hecke algebra $R^\Lambda_\alpha(\Gamma)$ is the quotient of the quiver Hecke algebra $R_\alpha(\Gamma)$ by the two-sided ideal $\mathfrak{I}^\Lambda_\alpha$ generated by the relations

$$y_1^{\Lambda_{i_1}} e(\boldsymbol{i}) = 0, \tag{26}$$

for all $\boldsymbol{i} = (i_1, \dots, i_n) \in \alpha$.

Since the above relations are homogeneous, the cyclotomic quiver Hecke algebras is graded, as in Proposition 3.2. Note that if $\Lambda_i = 0$ for all i, then

$$R^\Lambda_\alpha(\Gamma) = \begin{cases} \{0\}, & \text{if } n \geq 1,\\ K, & \text{if } n = 0. \end{cases}$$

As in [20, §6.4.1], we want to state Corollaries 3.12 and 3.14 in the cyclotomic setting. First, for any $j \in \{1, \dots, d\}$, let $\Lambda^{(j)}$ be the restriction of Λ to $I^{(j)}$.

THEOREM 3.16. — *We have an (explicit) isomorphism of graded algebras:*

$$R_\alpha^\Lambda(\Gamma) \simeq \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}} \left(\bigotimes_{j=1}^d R_{\alpha^{(j)}}^{\Lambda^{(j)}}(\Gamma^{(j)}) \right) .$$

Proof. — The proof is similar to that of [20, Theorem 6.30]. We provide the details, since it will be used later in the paper.

Note that $\bigotimes_{j=1}^d R_{\alpha^{(j)}}^{\Lambda^{(j)}}(\Gamma^{(j)})$ is the quotient of $\bigotimes_{j=1}^d R_{\alpha^{(j)}}(\Gamma^{(j)})$ by the two-sided ideal

$$\mathfrak{I}_{\alpha,\otimes}^\Lambda \coloneqq \langle 1 \otimes \cdots \otimes 1 \otimes \mathfrak{I}_{\alpha^{(j)}}^{\Lambda^{(j)}} \otimes 1 \otimes \cdots \otimes 1 \,,\ j = 1, \dots, d \rangle$$

generated by the ideals $\mathfrak{I}_{\alpha^{(j)}}^{\Lambda^{(j)}}$ in position j in the tensor product. We will identify the algebra $\bigotimes_{j=1}^d R_{\alpha^{(j)}}(\Gamma^{(j)})$ with the algebra $e(\mathfrak{t}^\alpha) R_\alpha(\Gamma) e(\mathfrak{t}^\alpha)$ due to the explicit isomorphism given in the proof of Theorem 3.11. With this identification, the ideal $\mathfrak{I}_{\alpha,\otimes}^\Lambda$ is generated by the elements

$$y_b^{\Lambda_{i_b}} e(\boldsymbol{i}) \,,$$

where $\boldsymbol{i} \in \alpha$ is of profile $\mathfrak{t}^\alpha$, and b is of the form $b = n_1 + \cdots + n_{j-1} + 1$ for $j \in \{1, \dots, d\}$.

Now let θ be the isomorphism of Proposition 3.9 and η its inverse. For convenience, we denote during the proof $N \coloneqq \binom{n}{n_1,\dots,n_d}$. We will prove the following two inclusions:

$$\theta\left(\mathfrak{I}_\alpha^\Lambda\right) \subseteq \mathrm{Mat}_N\left(\mathfrak{I}_{\alpha,\otimes}^\Lambda\right), \tag{27a}$$

$$\mathfrak{I}_\alpha^\Lambda \supseteq \eta\left(\mathrm{Mat}_N\left(\mathfrak{I}_{\alpha,\otimes}^\Lambda\right)\right). \tag{27b}$$

Let $\mathfrak{t}', \mathfrak{t} \in \mathrm{Prof}^\alpha$. First recall that for $h \in e(\mathfrak{t}') R_\alpha(\Gamma) e(\mathfrak{t})$, we have

$$\theta(h) = \psi_{\mathfrak{t}'} h \phi_{\mathfrak{t}} M_{\mathfrak{t}'\mathfrak{t}} \in \mathrm{Mat}_N\left(e(\mathfrak{t}^\alpha) R_\alpha(\Gamma) e(\mathfrak{t}^\alpha)\right).$$

while for $h \in e(\mathfrak{t}^\alpha) R_\alpha(\Gamma) e(\mathfrak{t}^\alpha)$, we have

$$\eta\left(h M_{\mathfrak{t}'\mathfrak{t}}\right) = \phi_{\mathfrak{t}'} h \psi_{\mathfrak{t}} \,,$$

where the elements $\phi_{\mathfrak{t}}, \psi_{\mathfrak{t}}$ were introduced in (20).

- Let $\boldsymbol{i} \in \alpha$ of profile $\mathfrak{t}$. By (21), (12) and (24) we have:

  $$y_1^{\Lambda_{i_1}} e(\boldsymbol{i}) = y_1^{\Lambda_{i_1}} e(\boldsymbol{i}) e(\mathfrak{t}) = y_1^{\Lambda_{i_1}} e(\boldsymbol{i}) \phi_{\mathfrak{t}} \psi_{\mathfrak{t}} e(\mathfrak{t}) = \phi_{\mathfrak{t}} y_{\pi_{\mathfrak{t}}(1)}^{\Lambda_{i_1}} e(\pi_{\mathfrak{t}} \cdot \boldsymbol{i}) \psi_{\mathfrak{t}} e(\mathfrak{t}).$$

  Thus, to prove (27a) it suffices to show that

  $$\theta\left(y_{\pi_{\mathfrak{t}}(1)}^{\Lambda_{i_1}} e(\pi_{\mathfrak{t}} \cdot \boldsymbol{i})\right) \in \mathrm{Mat}_N\left(\mathfrak{I}_{\alpha,\otimes}^\Lambda\right) .$$

  By definition of $\pi_{\mathfrak{t}}$, we have that $\boldsymbol{i}' \coloneqq \pi_{\mathfrak{t}} \cdot \boldsymbol{i}$ has profile $\mathfrak{t}^\alpha$, and, therefore, $y_{\pi_{\mathfrak{t}}(1)}^{\Lambda_{i_1}} e(\boldsymbol{i}') \in e(\mathfrak{t}^\alpha) R_\alpha(\Gamma) e(\mathfrak{t}^\alpha)$. Let $b \coloneqq \pi_{\mathfrak{t}}(1)$, so that we have $i_1 = i'_b$,

and, moreover, by [20, Proposition 6.7], the element b is of the form $n_1 + \cdots + n_{j-1} + 1$. We conclude that

$$\theta\big(y_{\pi_{\mathfrak{t}}(1)}^{\Lambda_{i_1}} e(\pi_{\mathfrak{t}} \cdot \boldsymbol{i})\big) = y_b^{\Lambda_{i'_b}} e(\boldsymbol{i}') M_{\mathfrak{t}^\alpha \mathfrak{t}^\alpha} \in \mathrm{Mat}_N\left(\mathfrak{I}^\Lambda_{\alpha,\otimes}\right) \ .$$

- Let $\boldsymbol{i} \in \alpha$ with profile $\mathfrak{t}^\alpha$ and let $b = n_1 + \cdots + n_{j-1} + 1$ with $j \in \{1, \ldots, d\}$, such that $n_j \neq 0$. Let us prove that

$$\eta\big(y_b^{\Lambda_{i_b}} e(\boldsymbol{i}) M_{\mathfrak{t}'\mathfrak{t}}\big) \in \mathfrak{I}^\Lambda_\alpha.$$

Since $M_{\mathfrak{t}'\mathfrak{t}} = M_{\mathfrak{t}'\mathfrak{t}''} M_{\mathfrak{t}''\mathfrak{t}}$ for any $\mathfrak{t}''$, it is enough to prove it for a single value of $\mathfrak{t}'$. So without loss of generality, since $n_j \neq 0$, we can assume that $\mathfrak{t}'$ starts with j, so that $\pi_{\mathfrak{t}'}(1) = b$. We conclude that

$$\eta\big(y_b^{\Lambda_{i_b}} e(\boldsymbol{i}) M_{\mathfrak{t}'\mathfrak{t}}\big) = \phi_{\mathfrak{t}'} y_b^{\Lambda_{i_b}} e(\boldsymbol{i}) \psi_{\mathfrak{t}} = y_1^{\Lambda_{i_b}} e(\pi_{\mathfrak{t}'}^{-1} \cdot \boldsymbol{i}) \phi_{\mathfrak{t}'} \psi_{\mathfrak{t}} \in \mathfrak{I}^\Lambda_\alpha,$$

since, if we denote $\boldsymbol{i}' = \pi_{\mathfrak{t}'}^{-1} \cdot \boldsymbol{i}$, then we have $i'_1 = i_b$.

This concludes the proof of (27) showing that we have

$$\theta\left(\mathfrak{I}^\Lambda_\alpha\right) = \mathrm{Mat}_N\left(\mathfrak{I}^\Lambda_{\alpha,\otimes}\right) \ .$$

Thus, we can deduce the isomorphism of Theorem 3.16 from Corollary 3.12. □

Remark 3.17. —

- We saw that if $\Lambda^{(j)} \equiv 0$ on $\Gamma^{(j)}$ for some j, then, if moreover $\alpha^{(j)} \neq \emptyset$ (that is, if $n_j(\alpha) \neq 0$), we have $R^{\Lambda^{(j)}}_{\alpha^{(j)}}(\Gamma^{(j)}) = \{0\}$ from the defining relations. So, in turn, Theorem 3.16 implies that $R^\Lambda_\alpha(\Gamma) = \{0\}$.
- The conclusion of the preceding item can, in fact, be seen more directly. Indeed, the cyclotomic relations in $R^\Lambda_\alpha(\Gamma)$ imply that $e(\boldsymbol{i}) = 0$ for all $\boldsymbol{i} \in \alpha$ with $i_1 \in \Gamma^{(j)}$. So we have that the idempotent $e(\mathfrak{t})$ is 0 for any profile $\mathfrak{t}$ starting with j (and at least one profile like this exists in Prof^α when $n_j(\alpha) \neq 0$). Since:

$$\psi_{\mathfrak{t}} e(\mathfrak{t}) \phi_{\mathfrak{t}} = e(\mathfrak{t}^\alpha) \quad \text{and} \quad \phi_{\mathfrak{t}} e(\mathfrak{t}^\alpha) \psi_{\mathfrak{t}} = e(\mathfrak{t}) \ ,$$

it follows immediately that if $n_j(\alpha) \neq 0$, then all idempotents $e(\mathfrak{t})$ are 0, and, in turn, all idempotents $e(\boldsymbol{i})$, $\boldsymbol{i} \in \alpha$ are 0, which shows that $R^\Lambda_\alpha(\Gamma) = \{0\}$.

As in Corollary 3.14, we deduce the following corollary.

Corollary 3.18. — *We have an (explicit) isomorphism of graded algebras:*

$$R^\Lambda_n(\Gamma) \simeq \bigoplus_{\substack{n_1,\ldots,n_d \geq 0 \\ n_1 + \cdots + n_d = n}} \mathrm{Mat}_{\binom{n}{n_1,\ldots,n_d}} \left(\bigotimes_{j=1}^d R^{\Lambda^{(j)}}_{n_j}(\Gamma^{(j)}) \right) .$$

REMARK 3.19. — It follows from Remark 3.17 that we can assume that Λ is supported on all components of Γ, that is, $\Lambda^{(j)} \not\equiv 0$ for all $j \in \{1, \dots, d\}$. In other words, we can replace from the beginning Γ by $\tilde{\Gamma}$, where we removed the components $\Gamma^{(j)}$, such that $\Lambda^{(j)} \equiv 0$. In particular, we have $R_n^{\Lambda}(\Gamma) = R_n^{\Lambda_{|\tilde{I}}}(\tilde{\Gamma})$, where $\tilde{I}$ denotes the vertex set of $\tilde{\Gamma}$. We could have done that, but it turned out to not be really necessary to state Theorem 3.16 or Corollary 3.18. For example, in Corollary 3.18, if $\Lambda^{(j)} \equiv 0$ for some j, then all the summands with $n_j \neq 0$ are $\{0\}$ and can, thus, be removed from the direct sum.

4. Interpolating quiver Hecke algebras for type B

The aim of this section is to unite the definitions of quiver Hecke algebras for type B that are introduced in [24] by Varagnolo and Vasserot and in [18, 19] by the first author and Walker.

4.1. Definition. — Let Γ be a quiver as in §3.1. We also adopt the notation of this section. Let θ be an involution of Γ, that is, the map θ is an involution on both sets I and A and satisfies

$$\theta(o(a)) = t(\theta(a)), \tag{28}$$

for all $a \in A$. Note the following consequence: for any $i, j \in I$, we have $|i \to j| = |\theta(j) \to \theta(i)|$ and, thus,

$$i \cdot j = \theta(i) \cdot \theta(j). \tag{29}$$

It follows from the definition (8) of the polynomials Q_{ij} and from (28), again, that

$$Q_{ij}(u, v) = Q_{\theta(j)\theta(i)}(u, v), \tag{30}$$

for any $i, j \in I$.

Let B_n be the group of signed permutations of $\{\pm 1, \dots, \pm n\}$, that is, the group of permutations π of $\{\pm 1, \dots, \pm n\}$ satisfying $\pi(-i) = -\pi(i)$ for all $i \in \{1, \dots, n\}$. We have a natural isomorphism $B_n \simeq (\mathbb{Z}/2\mathbb{Z})^n \rtimes \mathfrak{S}_n$. In particular, we are in the setting of §3.2 with $G = \mathbb{Z}/2\mathbb{Z}$, which acts on I via the canonical surjection $G \twoheadrightarrow \langle\theta\rangle$. We have a natural inclusion $\mathfrak{S}_n \subseteq B_n$, where r_a is identified with $(a, a+1)(-a, -a-1)$ for all $a \in \{1, \dots, n-1\}$. We see B_n as a Weyl group of type B by adding the generator $r_0 \coloneqq (-1, 1)$. The action of B_n on I^n is given by (10) and

$$r_0 \cdot (i_1, \dots, i_n) \coloneqq (\theta(i_1), i_2, \dots, i_n),$$

for any $\boldsymbol{i} = (i_1, \dots, i_n) \in I^n$. Let $\beta \subseteq I^n$ be a B_n-orbit. In particular, the set β is a finite $\mathfrak{S}_n$-stable subset of I^n.

REMARK 4.1. — The result of Remark 3.7 can here be written as follows. There is a one-to-one correspondence between B_n-orbits $\beta \subset I^n$ and maps $\hat{\beta} : I \to \mathbb{N}$, such that $\hat{\beta} = \hat{\beta} \circ \theta$ and $\frac{1}{2}\sum_{\substack{i\in I\\ \theta(i)\neq i}} \hat{\beta}(i) + \sum_{\substack{i\in I\\ \theta(i)=i}} \hat{\beta}(i) = n$ (the number $\hat{\beta}(i)$ counts the number of occurrence of both i and $\theta(i)$ in any element in the orbit β). See also [18, Remark 2.5].

Let $\lambda \in \mathbb{N}^I$ and $\gamma \in K^I$. Define

$$d(i) \coloneqq \begin{cases} \lambda_i + \lambda_{\theta(i)}, & \text{if } \gamma_i = 0, \\ -2, & \text{otherwise.} \end{cases}$$

For any $i \in I$, we make the following assumptions:

$$\theta(i) \neq i \implies \gamma_i = 0, \tag{31a}$$

$$\gamma_i = 0 \implies [\theta(i) \neq i \text{ or } d(i) = 0]. \tag{31b}$$

Note that γ is θ-invariant, that is, we have

$$\gamma_{\theta(i)} = \gamma_i, \text{ for all } i \in I. \tag{32}$$

REMARK 4.2. —
- Condition (31b) may seem strong; without it in §A.1 we encounter useless complications for our means (see also Remark A.1).
- Similarly, one could consider a more general definition than the one below. As, for example, in [22, §3.2], we could remove any reference to a quiver and start only with a family of polynomials associated to the set I with involution θ (namely, $Q_{ij}[u, v]$ and a polynomial replacing $(-1)^{\lambda_{\theta(i_1)}} y_1^{d(i_1)}$ in the definition below). Then one should look for conditions ensuring the existence of a polynomial representation. We do not pursue this direction to avoid adding another layer of technicalities.

DEFINITION 4.3. — The algebra $V_\beta(\Gamma, \lambda, \gamma)$ is the unitary associative K-algebra generated by elements

$$\{y_a\}_{1\leq a\leq n} \cup \{\psi_b\}_{0\leq b\leq n-1} \cup \{e(\boldsymbol{i})\}_{\boldsymbol{i}\in\beta},$$

with the relations (11)–(16) of Section 3 involving all the generators but ψ_0, together with

$$\psi_0 e(\boldsymbol{i}) = e(r_0 \cdot \boldsymbol{i})\psi_0, \tag{33}$$

$$\psi_0\psi_b = \psi_b\psi_0, \quad \text{for all } b \in \{2, \dots, n-1\}, \tag{34}$$

$$(\psi_0 y_1 + y_1\psi_0)e(\boldsymbol{i}) = 2\gamma_{i_1} e(\boldsymbol{i}), \tag{35}$$

$$\psi_0 y_a = y_a\psi_0, \quad \text{for all } a \in \{2, \dots, n\}, \tag{36}$$

$$\psi_0^2 e(\boldsymbol{i}) = \begin{cases} (-1)^{\lambda_{\theta(i_1)}} y_1^{d(i_1)} e(\boldsymbol{i}), & \text{if } \gamma_{i_1} = 0, \\ 0, & \text{otherwise,} \end{cases} \tag{37}$$

$$(38)\quad \left((\psi_0\psi_1)^2-(\psi_1\psi_0)^2\right)e(\boldsymbol{i}) = \begin{cases} (-1)^{\lambda_{\theta(i_1)}}\dfrac{(-y_1)^{d(i_1)}-y_2^{d(i_1)}}{y_1+y_2}\psi_1 e(\boldsymbol{i}), & \text{if } \gamma_{i_1}=0 \text{ and } \theta(i_1)=i_2, \\ \gamma_{i_2}\dfrac{Q_{i_2i_1}(y_1,-y_2)-Q_{i_2i_1}(y_1,y_2)}{y_1y_2}\left(y_1\psi_0-\gamma_{i_1}\right)e(\boldsymbol{i}) & \text{otherwise,} \end{cases}$$

for all $\boldsymbol{i}\in\beta$.

It is clear that the fraction in the first line of the right hand side in (38) is a polynomial in y_1, y_2. Then we note that the second line in the right-hand side of (38) is 0, when $\gamma_{i_2}=0$ or when $i_1=i_2$ (recalling (8)), and is a polynomial in y_1, y_2 when $\gamma_{i_1}=0$. So for the second line, if $\gamma_{i_1}\neq 0\neq\gamma_{i_2}$ and $i_1\neq i_2$, then by (31a) we have $\theta(i_1)=i_1$ and $\theta(i_2)=i_2$, and, thus, we can use (9) and (30), so that

$$\begin{aligned}\frac{Q_{i_1i_2}(u,-v)-Q_{i_1i_2}(u,v)}{uv} &= \frac{Q_{i_1i_2}(v,-u)-Q_{\theta(i_2)\theta(i_1)}(u,v)}{uv} \\ &= \frac{Q_{i_1i_2}(v,-u)-Q_{i_2i_1}(u,v)}{uv} \\ &= \frac{Q_{i_1i_2}(v,-u)-Q_{i_1i_2}(v,u)}{uv},\end{aligned}$$

is a polynomial.

Finally, note that when $n=0$ then $V_\beta(\Gamma,\lambda,\gamma)=K$.

Remark 4.4. — Since β is a finite $\mathfrak{S}_n$-stable subset of I^n, we can also consider the algebra $R_\beta(\Gamma)$ as defined in §3.1. The subalgebra of $V_\beta(\Gamma,\lambda,\gamma)$ generated by all the generators but ψ_0 is an obvious quotient of $R_\beta(\Gamma)$ (see also Corollary 4.11).

Remark 4.5. — If θ has no fixed point in I, then $V_\beta(\Gamma,\lambda,\gamma)$ is exactly the algebra defined in [24]. In this case, by (31a) we necessarily have $\gamma_i=0$ for any i, and (31b) is automatically satisfied. In particular, in (38), the second line is always zero in this situation.

Remark 4.6. — Assume that K is a field of characteristic different from 2 and let $p,q\in K^\times$ with $q^2\neq 1\neq p^2$. Let $\theta: K^\times\to K^\times$ be the scalar inversion. For any $x\in K^\times$, we define the set $I_x := \{x^\epsilon q^{2l} : \epsilon\in\{\pm1\}, l\in\mathbb{Z}\}$. Let $x_1,\dots,x_k\in K^\times$, such that the sets I_{x_a} are pairwise disjoint. Let Γ be the quiver with vertices $I := \coprod_{a=1}^k I_{x_a}$ and arrows between v and q^2v for all $v\in I$. Finally, let λ be the indicator function of $P := \{\pm p\}\cap I$ and define $\gamma_i := 1$ if $\theta(i)=i$ and $\gamma_i := 0$ otherwise (thus, (31) is satisfied). Then $V_\beta(\Gamma,\lambda,\gamma)$ is exactly the algebra $V_{\boldsymbol{x}}^{I^\beta}$ defined in [18]. This is, together with the next remark, the situation relevant for the applications to affine Hecke algebras, see Section 7.

REMARK 4.7. — The algebra of [19, §3.1] is obtained with the same choice of Γ, θ as in the preceding remark, together with $\gamma_i := 0$ and $\lambda_i := 0$ for all i. In particular, Condition (31b) is satisfied, since $d(i) = 0$ for all $i \in I$. We will come back to this particular situation in Section 6.

The algebra $V_\beta(\Gamma, \lambda, \gamma)$ is endowed with the $\mathbb{Z}$-grading given by

$$\deg e(\boldsymbol{i}) = 0, \tag{39a}$$

$$\deg y_a = 2, \tag{39b}$$

$$\deg \psi_0 e(\boldsymbol{i}) = d(i_1), \tag{39c}$$

$$\deg \psi_b e(\boldsymbol{i}) = d(i_b, i_{i+1}). \tag{39d}$$

The homogeneity of the defining relations that do not involve ψ_0 is as in Section 3, the other ones being a simple calculation. For (35), note that if $\gamma_{i_1} = 0$, there is nothing to check, and if $\gamma_{i_1} \neq 0$, then by definition we have $d(i_1) = -2$ and, thus, $\deg \psi_0 y_1 e(\boldsymbol{i}) = \deg y_1 \psi_0 e(\boldsymbol{i}) = 0$. To check the last relation, let us write $i_1 i_2$ instead of $\boldsymbol{i}$, and even a instead of i_a and $\bar{a}$ instead of $\theta(i_a)$. We have

$$\begin{aligned}\big((\psi_0\psi_1)^2 - (\psi_1\psi_0)^2\big)\, e(12) &= \psi_0\psi_1\psi_0\psi_1 e(12) - \psi_1\psi_0\psi_1\psi_0 e(12) \\ &= \psi_0 e(1\bar{2})\psi_1 e(\bar{2}1)\psi_0 e(21)\psi_1 e(12) \\ &\quad - \psi_1 e(\bar{2}\bar{1})\psi_0 e(2\bar{1})\psi_1 e(\bar{1}2)\psi_0 e(12).\end{aligned} \tag{40}$$

We have:

$$\deg \psi_0 e(1\bar{2}) = \deg \psi_0 e(12) = d(1),$$
$$\deg \psi_0 e(21) = \deg \psi_0 e(2\bar{1}) = d(2).$$

Moreover, by (29) we have

$$\deg \psi_1 e(\bar{2}1) = d(\bar{2}, 1) = d(1, \bar{2}) = d(\bar{1}, 2) = \deg \psi_1 e(\bar{1}2),$$
$$\deg \psi_1 e(12) = d(1, 2) = d(2, 1) = d(\bar{2}, \bar{1}) = \deg \psi_1 e(\bar{2}\bar{1}).$$

Thus, the quantity $\big((\psi_0\psi_1)^2 - (\psi_1\psi_0)^2\big)\, e(\boldsymbol{i})$ is homogeneous of degree

$$d(i_1) + d(i_2) + d\big(i_1, i_2\big) + d\big(i_1, \theta(i_2)\big).$$

A quick calculation now shows that the last relation is homogeneous (note that in the first case, we have $\gamma_{i_2} = 0$ by (32)).

4.2. Basis theorem. — We now want to give an analogue of the basis theorem Proposition 3.3 for quiver Hecke algebras. As in [15, 16, 22], we will construct a polynomial realisation of $V_\beta(\Gamma, \lambda, \gamma)$. Let $(P_{ij}(u, v))_{i,j \in I}$ be a family of polynomials satisfying

$$P_{ij}(u, v) = P_{ij}(-v, -u), \tag{41a}$$

$$P_{ij}(u, v) = P_{\theta(j)\theta(i)}(u, v), \tag{41b}$$

and such that

$$P_{ij}(u,v)P_{ji}(v,u) = Q_{ij}(u,v). \tag{42}$$

Note that $P_{ij}(u,v) \coloneqq (u-v)^{|j\to i|}$ if $i \neq j$ and $P_{ij}(u,v) \coloneqq 0$ if $i = j$ is an example of such a family, by (28). Now let $(\alpha_i(y))_{i\in I}$ be a family of polynomials, such that

$$\alpha_{\theta(i)}(y)\alpha_i(-y) = (-1)^{\lambda_{\theta(i)}} y^{d(i)}, \quad \text{if } \gamma_i = 0, \tag{43}$$

$$\alpha_i(y) = 0, \quad \text{otherwise.} \tag{44}$$

Note that if $\gamma_i = 0$, we can just set $\alpha_i(y) \coloneqq y^{\lambda_{\theta(i)}}$. We now consider the sum of polynomial algebras $K[x,\beta] \coloneqq \bigoplus_{\boldsymbol{i}\in\beta} K[x_1,\dots,x_n]\mathbf{1}_{\boldsymbol{i}}$, where $\mathbf{1}_{\boldsymbol{i}}$ denotes the unit of the summand corresponding to $\boldsymbol{i}$, so that

$$\begin{aligned} f\mathbf{1}_{\boldsymbol{i}} &= \mathbf{1}_{\boldsymbol{i}} f, \quad && \text{for all } f \in K[x_1,\dots,x_n] \text{ and } \boldsymbol{i} \in \beta, \\ \mathbf{1}_{\boldsymbol{i}}\mathbf{1}_{\boldsymbol{j}} &= \delta_{\boldsymbol{ij}}\mathbf{1}_{\boldsymbol{i}}, && \text{for all } \boldsymbol{i},\boldsymbol{j} \in \beta. \end{aligned}$$

The Weyl group B_n acts on $K[x_1,\dots,x_n]$ by

$${}^w f(x_1,\dots,x_n) \coloneqq f\big(w^{-1}\cdot(x_1,\dots,x_n)\big)$$

for any $w \in B_n$ and $f \in K[x_1,\dots,x_n]$, where the action of the generator r_0 on $(x_1,\dots,x_n)$ is by multiplying x_1 by -1, and the action of the generator r_a, $a = 1,\dots,n-1$, on $(x_1,\dots,x_n)$ is by exchanging x_a and x_{a+1}. The action of B_n on $K[x_1,\dots,x_n]$ extends by linearity to $K[x,\beta]$ by setting $w\star f\mathbf{1}_{\boldsymbol{i}} \coloneqq {}^w f\mathbf{1}_{w\cdot\boldsymbol{i}}$ for any $\boldsymbol{i} \in \beta$.

We now consider the linear action of $V_\beta(\Gamma,\lambda,\gamma)$ on $K[x,\beta]$ given on the generators by

$$\begin{aligned} e(\boldsymbol{j})\cdot f\mathbf{1}_{\boldsymbol{i}} &\coloneqq \delta_{\boldsymbol{ij}} f\mathbf{1}_{\boldsymbol{i}} = \delta_{\boldsymbol{ij}}\mathbf{1}_{\boldsymbol{i}} f, \\ y_a\cdot f\mathbf{1}_{\boldsymbol{i}} &\coloneqq x_a f\mathbf{1}_{\boldsymbol{i}} = x_a\mathbf{1}_{\boldsymbol{i}} f, \\ \psi_b\cdot f\mathbf{1}_{\boldsymbol{i}} &\coloneqq \delta_{i_b,i_{b+1}}\frac{{}^{r_b}f - f}{x_b - x_{b+1}}\mathbf{1}_{\boldsymbol{i}} + P_{i_b,i_{b+1}}(x_{b+1},x_b)\,{}^{r_b}f\mathbf{1}_{r_b\cdot\boldsymbol{i}} \\ &= \big(\delta_{i_b,i_{b+1}}(x_b - x_{b+1})^{-1}(r_b - 1) + P_{i_b,i_{b+1}}(x_{b+1},x_b)r_b\big)\star f\mathbf{1}_{\boldsymbol{i}}, \\ \psi_0\cdot f\mathbf{1}_{\boldsymbol{i}} &\coloneqq \left(\gamma_{i_1}\frac{f - {}^{r_0}f}{x_1} + \alpha_{i_1}(x_1)\,{}^{r_0}f\right)\mathbf{1}_{r_0\cdot\boldsymbol{i}} \\ &= \big(\gamma_{i_1}x_1^{-1}(1 - r_0) + \alpha_{i_1}(x_1)r_0\big)\star f\mathbf{1}_{\boldsymbol{i}}, \end{aligned}$$

for any $\boldsymbol{i},\boldsymbol{j} \in \beta$ and $f \in K[x_1,\dots,x_n]$.

LEMMA 4.8. — *The previous action is well defined.*

The proof of Lemma 4.8 is given in Appendix A. For each $w \in B_n$, we now fix a reduced expression $w = r_{a_1}\cdots r_{a_k}$ and define $\psi_w \coloneqq \psi_{a_1}\cdots\psi_{a_k} \in V_\beta(\Gamma,\lambda,\gamma)$. Note that the element ψ_w may depend on the reduced expression chosen.

THEOREM 4.9. — *The algebra $V_\beta(\Gamma,\lambda,\gamma)$ is a free K-module, and*

$$\{y_1^{a_1}\cdots y_n^{a_n}\psi_w e(\boldsymbol{i}) : a_i \in \mathbb{N}, w \in B_n, \boldsymbol{i} \in \beta\},$$

is a K-basis.

Proof. — As in [15, 16, 22], successively applying the defining relations of $V_\beta(\Gamma,\lambda,\gamma)$ we can see that the above family is a spanning set, and, hence, it remains to prove that it is linearly independent. For any $b \in \{0,\dots,n-1\}$, $\boldsymbol{i} \in \beta$ and $f \in K[x_1,\dots,x_n]$, we can write

$$\psi_b \cdot f\mathbf{1}_{\boldsymbol{i}} = \left(A_{\boldsymbol{i}}^{r_b} r_b + A_{\boldsymbol{i}}^{1,r_b}\right) \star f\mathbf{1}_{\boldsymbol{i}},$$

where $A_{\boldsymbol{i}}^{r_b}, A_{\boldsymbol{i}}^{1,r_b} \in K(x_1,\dots,x_n)$ with $A_{\boldsymbol{i}}^{r_b}$ non-zero (recall that $P_{ij} \neq 0$ if $i \neq j$). If $<$ is the Bruhat order on B_n, we deduce that for each $w \in B_n$, we can write

$$\psi_w \cdot f\mathbf{1}_{\boldsymbol{i}} = \left(A_{\boldsymbol{i}}^{w} w + \sum_{w'<w} A_{\boldsymbol{i}}^{w',w} w'\right) \star f\mathbf{1}_{\boldsymbol{i}},$$

where $A_{\boldsymbol{i}}^{w}, A_{\boldsymbol{i}}^{w',w} \in K(x_1,\dots,x_n)$ with $A_{\boldsymbol{i}}^{w}$ non-zero. Thus,

$$y_1^{a_1}\cdots y_n^{a_n}\psi_w \cdot f\mathbf{1}_{\boldsymbol{i}} = \left(A_{\boldsymbol{i}}^{w} x_1^{a_1}\cdots x_n^{a_n} w + \sum_{w'<w} A_{\boldsymbol{i}}^{w',w} x_1^{a_1}\cdots x_n^{a_n} w'\right) \star f\mathbf{1}_{\boldsymbol{i}},$$

for any $a_1,\dots,a_n \in \mathbb{N}$. We now use the following basic Lemma 4.10 from field theory and note that the elements of B_n induce distinct field homomorphisms of $K(x_1,\dots,x_n)$.

LEMMA 4.10 (Dedekind). — *If $u_1,\dots,u_n : F \to G$ are distinct field homomorphisms, then they form a linearly independent family over G.*

So we can use reverse induction in the Bruhat order to show that the images of the basis elements are linearly independent in $\mathrm{End}_K(K[x,\beta])$ and, thus, conclude the proof. □

As a corollary, we obtain the sequel to Remark 4.4.

COROLLARY 4.11. — *The subalgebra of $V_\beta(\Gamma,\lambda,\gamma)$ generated by all generators but ψ_0 is isomorphic to $R_\beta(\Gamma)$.*

5. Disjoint quiver isomorphism

Let Γ be a quiver with an involution θ and $\lambda \in \mathbb{N}^I, \gamma \in K^I$ as in §4.1. Let d be a positive integer and write $\Gamma = \coprod_{j=1}^{d} \Gamma^{(j)}$, such that

- Each $\Gamma^{(j)}$ is a full subquiver of Γ.
- Each $\Gamma^{(j)}$ is stable under θ.

We write $I = \coprod_{j=1}^{d} I^{(j)}$ as the corresponding partition of the vertex set of Γ. Recall that $B_n \simeq G^n \rtimes \mathfrak{S}_n$ with $G = \mathbb{Z}/2\mathbb{Z}$ acting on I via $G \twoheadrightarrow \langle\theta\rangle$. In particular, each $I^{(j)}$ for $j \in \{1, \dots, d\}$ is stable under the action of G, so that we are in the setting of §3.2.

Let β be a B_n-orbit in I^n. As explained in §3.2, both properties (17) and (18) are satisfied. In particular, for any $j \in \{1, \dots, d\}$, we have an integer $n_j(\beta) = n_j$, and we have a B_{n_j}-orbit $\beta^{(j)} \subseteq (I^{(j)})^{n_j}$.

For any $j \in \{1, \dots, d\}$, we define $\lambda^{(j)} \in \mathbb{N}^{I^{(j)}}$ (or, respectively, $\gamma^{(j)} \in K^{I^{(j)}}$) to be the restriction of λ (or, respectively, γ) to $I^{(j)}$.

THEOREM 5.1. — *We have an (explicit) isomorphism of graded algebras*

$$V_\beta(\Gamma, \lambda, \gamma) \simeq \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}} \left(\bigotimes_{j=1}^{d} V_{\beta^{(j)}} \left(\Gamma^{(j)}, \lambda^{(j)}, \gamma^{(j)} \right) \right).$$

As in §3, we will first apply the result of §2 and then prove an isomorphism with a tensor product. Parts 5.1 and 5.2 are devoted to the proof of Theorem 5.1, which is a direct consequence of (45) and Proposition 5.2.

5.1. Fixing the profile. — As defined in §3.2.2, to each $\boldsymbol{i} \in \beta$ we associate its profile $p(\boldsymbol{i}) \in \{1, \dots, d\}^n$, and we write $\mathrm{Prof}^\beta \subseteq \{1, \dots, d\}^n$ to denote the set of all profiles of elements of β. Any element of Prof^β can be reordered so that we obtain

$$\mathfrak{t}^\beta = (1, \dots, 1, \dots, d, \dots, d),$$

where each $j \in \{1, \dots, d\}$ appears exactly n_j times. To any $\mathfrak{t} \in \mathrm{Prof}^\beta$, we define the idempotent

$$e(\mathfrak{t}) \coloneqq \sum_{\substack{\boldsymbol{i} \in \beta \\ p(\boldsymbol{i}) = \mathfrak{t}}} e(\boldsymbol{i}) \in V_\beta(\Gamma, \lambda, \gamma),$$

and we define

$$\mathcal{I} \coloneqq \{e(\mathfrak{t}) : \mathfrak{t} \in \mathrm{Prof}^\beta\}.$$

It is a complete set of orthogonal idempotents, and its cardinality is exactly $\binom{n}{n_1,\dots,n_d}$. Since any reduced expression in $\mathfrak{S}_n$ in the generators $r_1, \dots, r_{n-1}$ is also reduced in B_n for these same generators, the definitions (20) make sense in $V_\beta(\Gamma, \lambda, \gamma)$ for any $\mathfrak{t} \in \mathrm{Prof}^\beta$. Moreover, since the defining relations of $R_\beta(\Gamma)$ are also satisfied in $V_\beta(\Gamma, \lambda, \gamma)$, we deduce that equations (21) and (23) are still satisfied in $V_\beta(\Gamma, \lambda, \gamma)$, and, thus, as in §3.2.2, we conclude that

$$V_\beta(\Gamma, \lambda, \gamma) \simeq \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}} \left(e(\mathfrak{t}^\beta) V_\beta(\Gamma, \lambda, \gamma) e(\mathfrak{t}^\beta) \right). \tag{45}$$

5.2. Embedding the tensor product. — The aim of this section is to prove the following proposition.

PROPOSITION 5.2. — *We have an (explicit) isomorphism of graded algebras*

$$e(\mathfrak{t}^\beta)V_\beta(\Gamma,\lambda,\gamma)e(\mathfrak{t}^\beta)\simeq\bigotimes_{j=1}^{d}V_{\beta^{(j)}}\left(\Gamma^{(j)},\lambda^{(j)},\gamma^{(j)}\right).$$

5.2.1. *Images of the generators.* — Set $n=n_1+\cdots+n_d$. We start by defining a map from the set of generators of the algebra $\bigotimes_{j=1}^{d}V_{\beta^{(j)}}\left(\Gamma^{(j)},\lambda^{(j)},\gamma^{(j)}\right)$ to $e(\mathfrak{t}^\beta)V_\beta(\Gamma,\lambda,\gamma)e(\mathfrak{t}^\beta)$.

Let $j\in\{1,\dots,d\}$. We denote $\psi_0^{(j)},\dots,\psi_{n_j-1}^{(j)}$, $y_1^{(j)},\dots,y_{n_j}^{(j)}$, $e(\boldsymbol{i}^j)$ with $\boldsymbol{i}^j\in\beta^{(j)}$, the generators of $V_{\beta^{(j)}}\left(\Gamma^{(j)},\lambda^{(j)},\gamma^{(j)}\right)$. Then we consider the map

$$e(\boldsymbol{i}^1)\otimes\cdots\otimes e(\boldsymbol{i}^d)\mapsto e(\boldsymbol{i}^1,\dots,\boldsymbol{i}^d)\,, \tag{46}$$

$$\psi_0^{(j)}\mapsto e(\mathfrak{t}^\beta)\psi_{n_1+\cdots+n_{j-1}}\dots\psi_1\psi_0\psi_1\dots\psi_{n_1+\cdots+n_{j-1}}e(\mathfrak{t}^\beta)\,, \tag{47}$$

$$\psi_a^{(j)}\mapsto e(\mathfrak{t}^\beta)\psi_{n_1+\cdots+n_{j-1}+a}e(\mathfrak{t}^\beta)\,,\quad a=1,\dots,n_j-1\,, \tag{48}$$

$$y_b^{(j)}\mapsto e(\mathfrak{t}^\beta)y_{n_1+\cdots+n_{j-1}+b}e(\mathfrak{t}^\beta)\,,\quad b=1,\dots,n_j\,, \tag{49}$$

where each $\boldsymbol{i}^j\in\beta^{(j)}$ and $(\boldsymbol{i}^1,\dots,\boldsymbol{i}^d)$ is simply the concatenation. Note that $(\boldsymbol{i}^1,\dots,\boldsymbol{i}^d)\in\beta$, since β is a B_n-orbit, using Proposition 3.5. Moreover, the profile of $(\boldsymbol{i}^1,\dots,\boldsymbol{i}^d)$ is $\mathfrak{t}^\beta$ and thus $e(\boldsymbol{i}^1,\dots,\boldsymbol{i}^d)e(\mathfrak{t}^\beta)=e(\mathfrak{t}^\beta)e(\boldsymbol{i}^1,\dots,\boldsymbol{i}^d)=e(\boldsymbol{i}^1,\dots,\boldsymbol{i}^d)$. By convention, $n_1+\cdots+n_{j-1}=0$ if $j=1$ (and $\psi_0^{(1)}\mapsto\psi_0$). Note also that the Formula (46) extended by linearity gives the image of an idempotent $e(\boldsymbol{i}^j)\in V_{\beta^{(j)}}\left(\Gamma^{(j)},\lambda^{(j)},\gamma^{(j)}\right)$:

$$e(\boldsymbol{i}^j)\mapsto\sum_{\substack{j'=1\\ j'\neq j}}^{d}\sum_{\boldsymbol{i}^{j'}\in\beta^{(j')}}e(\boldsymbol{i}^1,\dots,\boldsymbol{i}^d)\,. \tag{50}$$

Equivalently, the image of $e(\boldsymbol{i}^j)$ is the sum of the idempotents $e(\boldsymbol{i})$, where the sum is taken over $\boldsymbol{i}\in\beta$, such that the profile of $\boldsymbol{i}$ is $\mathfrak{t}^\beta$, and, moreover, $(i_{n_1+\cdots+n_{j-1}+1},\dots,i_{n_1+\cdots+n_j})=\boldsymbol{i}^j$.

We will prove that the map given in (46)–(49) extends to an homomorphism of graded algebras denoted ρ and that ρ is bijective.

5.2.2. *Grading.* — We check that the map given in (46)–(49) preserves the grading given in (39). For the images of the idempotents and of the generators $y_b^{(j)}$, there is nothing to check.

Let $\boldsymbol{i}^j\in\beta^{(j)}$ and $\boldsymbol{i}\in\beta$, such that $(i_{n_1+\cdots+n_{j-1}+1},\dots,i_{n_1+\cdots+n_j})=\boldsymbol{i}^j$. Let $a\in\{1,\dots,n_j-1\}$. On the one hand, we have $\deg\psi_a^{(j)}e(\boldsymbol{i}^j)=d(i_a^j,i_{a+1}^j)$. On

the other hand, we have

$$\deg \psi_{n_1+\cdots+n_{j-1}+a} e(\boldsymbol{i}) = d(i_{n_1+\cdots+n_{j-1}+a}, i_{n_1+\cdots+n_{j-1}+a+1}) = d(i^j_a, i^j_{a+1}) \ .$$

Finally, on the one hand, we have $\deg \psi_0^{(j)} e(\boldsymbol{i}^j) = d(i^j_1)$. On the other hand, we claim that we have

$$\deg \psi_k \dots \psi_0 \dots \psi_k e(\boldsymbol{j}) = d(j_{k+1}) \ ,$$

for any $k \geq 0$ and any $\boldsymbol{j} \in \beta$, such that j_{k+1} is not in the same component as $j_1, \dots, j_k$ for the decomposition of the quiver $\Gamma = \coprod_{j=1}^d \Gamma^{(j)}$. Taking $k = n_1 + \cdots + n_{j-1}$ and $\boldsymbol{j} = \boldsymbol{i}$ this concludes the verification.

To prove the claim we use induction on k. For $k = 0$, this is the definition of the degree of $\psi_0 e(\boldsymbol{j})$. For $k > 0$, we have $\deg \psi_k e(\boldsymbol{j}) = j_k \cdot j_{k+1} = |j_k \to j_{k+1}| + |j_k \leftarrow j_{k+1}| = 0$ by assumption on $\boldsymbol{j}$. Similarly, $\deg \psi_k e(\boldsymbol{j}') = 0$, where $\boldsymbol{j}' = r_{k-1} \dots r_0 \dots r_{k-1} r_k(\boldsymbol{j})$, since $(j'_k, j'_{k+1}) = (\theta(j_{k+1}), j_k)$. It remains to use the induction hypothesis, namely that $\deg \psi_{k-1} \dots \psi_0 \dots \psi_{k-1} e(r_k(\boldsymbol{j})) = d(j_{k+1})$, which is valid because $r_k(\boldsymbol{j})$ has j_{k+1} in position k.

5.2.3. *Bijectivity.* — We assume for a moment that the map given in (46)–(49) extends to an algebra homomorphism. We denote this map by ρ and prove here that ρ is bijective.

For any $j \in \{1, \dots, d\}$, we write $B^{(j)} \coloneqq B_{n_j}$ and rename its generators to $r_0^{(j)}, \dots, r_{n_j-1}^{(j)}$. We recall the following fact.

Lemma 5.3. — *We have an injective group homomorphism*

$$B^{(1)} \times \cdots \times B^{(d)} \to B_n$$
$$(w_1, \dots, w_d) \mapsto \overline{w}_1 \dots \overline{w}_d$$

given on the generators by, for $j \in \{1, \dots, d\}$,

$$r_0^{(j)} \mapsto r_{n_1+\cdots+n_{j-1}} \dots r_1 r_0 r_1 \dots r_{n_1+\cdots+n_{j-1}} \ ,$$
$$r_a^{(j)} \mapsto r_{n_1+\cdots+n_{j-1}+a} \ , \quad a = 1, \dots, n_j - 1 \ .$$

By convention, $n_1 + \cdots + n_{j-1} = 0$ if $j = 1$ (and $r_0^{(1)} \mapsto r_0$). Moreover, any d-tuple of reduced expressions is sent onto a reduced expression in B_n.

Proof. — Recall that $B_n = \langle r_0, \dots, r_{n-1} \rangle$ is the group of signed permutations of $\{\pm 1, \dots, \pm n\}$, with $r_0 = (1, -1)$ and $r_a = (a, a+1)(-a, -a-1)$ for $a = 1, \dots, n-1$. Let $t_1 \coloneqq r_0$ and $t_{a+1} \coloneqq r_a t_a r_a$, for $a = 1, \dots, n-1$. The element t_a corresponds to the transposition $(-a, a)$.

For any $i \in \{1, \dots, n\}$ and $a \in \{1, \dots, i\}$, we set $r_a \dots r_{i-1} = 1$ if $a = i$ by convention. It is easy to see (for example, [17, Figure 9]) that:

$$B_n = \bigsqcup_{a=1}^{n} r_a \dots r_{n-1} B_{n-1} \sqcup \bigsqcup_{a=1}^{n} t_a r_a \dots r_{n-1} B_{n-1} \ .$$

So, if we define, for $i \in \{1, \dots, n\}$,

$$R^{(i)} := \{t_a^{\epsilon} r_a \dots r_{i-1} \mid a \in \{1, \dots, i\}\,,\ \epsilon \in \{0,1\}\}\ ;$$

then we have that

$$\{u_n \dots u_1 \mid u_i \in R^{(i)}\ \} \tag{51}$$

forms a complete set of pairwise distinct elements of B_n. Moreover, this set consists of reduced expressions in terms of the generators $r_0, r_1, \dots, r_{n-1}$, since the polynomial $\sum_k a_k t^k$, where a_k records the number of elements in (51) written as a product of k generators, is easily found to be $\prod_{i=1}^{n} \frac{1-t^{2i}}{1-t}$, which is the Poincaré polynomial $\sum_{w \in B_n} t^{\ell(w)}$ of the Coxeter group of type B_n (see, for instance, [4, Theorem 7.1.5]).

Now, to prove the lemma we note that the subgroup permuting only the numbers $\pm 1, \dots, \pm n_1$ is isomorphic to $B^{(1)}$, the subgroup permuting only the numbers $\pm(n_1+1), \dots, \pm(n_1+n_2)$ is isomorphic to $B^{(2)}$ and so on. These subgroups commute, and, therefore, we have an embedding of $B^{(1)} \times \dots \times B^{(d)}$ inside B_n (although not as a parabolic subgroup). It is straightforward to see that this corresponds to the embedding described at the level of the generators in the lemma.

For the statement about the reduced expressions, let us first recall that the length function of the Coxeter group B_n can be expressed in terms of inversions as follows (see, for example, [4, §8.1]):

$$\ell(\pi) = \sharp\{1 \le i < j \le n \mid \pi(i) > \pi(j)\} + \sharp\{1 \le i \le j \le n \mid \pi(-i) > \pi(j)\}\ .$$

Using the notations of the lemma we obtain that $\ell(\overline{w}_1 \dots \overline{w}_d) = \ell(\overline{w}_1) + \dots + \ell(\overline{w}_d)$, since $\overline{w}_1$ permutes only the numbers $\pm 1, \dots, \pm n_1$, $\overline{w}_2$ permutes only the numbers $\pm(n_1+1), \dots, \pm(n_1+n_2)$ and so on. So it remains to show that a reduced expression in $B^{(j)}$, $j = 1, \dots, d$, is sent to a reduced expression in B_n.

Let $j \in \{1, \dots, d\}$. We claim that it is enough to show our assertion for a single reduced expression for each element of $B^{(j)}$. Indeed, the number of occurrences of r_0 in different reduced expressions of a same element remains constant (due to the homogeneity in r_0 of the braid relations of B_n), and therefore, the number of generators in the images of these different reduced expressions is also constant. So, if one of these images is reduced, they are all reduced.

Finally, to conclude the proof of the lemma, we observe that the set of reduced expressions of the form (51) in $B^{(j)}$ is sent to expressions of the same form in B_n, which are, therefore, also reduced. □

To prove that ρ is bijective, we first use that we know a basis of $\bigotimes_{j=1}^{d} V_{\beta^{(j)}}\big(\Gamma^{(j)}, \lambda^{(j)}, \gamma^{(j)}\big)$ by Theorem 4.9. A basis element is of the form

$$\bigotimes_{j=1}^{d} (y_1^{(j)})^{a_1^{(j)}} \dots (y_{n_j}^{(j)})^{a_{n_j}^{(j)}} \psi_{w_j}^{(j)} e(\boldsymbol{i}^j), \tag{52}$$

where $a_1^{(j)}, \dots a_{n_j}^{(j)} \in \mathbb{N}$, $\boldsymbol{i}^j \in \beta^{(j)}$ and $w_j \in B^{(j)}$. Note that we have fixed a reduced expression for each element $w_j \in B^{(j)}$, for each $j = 1, \dots, d$, in order to define $\psi_{w_j}^{(j)}$.

On the other hand, we also know a basis of $e(\mathfrak{t}^\beta)V_\beta(\Gamma, \lambda, \gamma)e(\mathfrak{t}^\beta)$ again by Theorem 4.9. Indeed, note that $e(\boldsymbol{i})e(\mathfrak{t}^\beta) = e(\boldsymbol{i})$, if the profile of $\boldsymbol{i}$ is $\mathfrak{t}^\beta$ and $e(\boldsymbol{i})e(\mathfrak{t}^\beta) = 0$ otherwise. Moreover, $\psi_w e(\boldsymbol{i}) = e(w{\cdot}\boldsymbol{i})\psi_w$. So it is straightforward to conclude that a basis element of $e(\mathfrak{t}^\beta)V_\beta(\Gamma, \lambda, \gamma)e(\mathfrak{t}^\beta)$ is of the form

$$y_1^{a_1} \dots y_n^{a_n} \psi_w e(\boldsymbol{i})\,, \tag{53}$$

where $a_1, \dots, a_n \in \mathbb{N}$, $\boldsymbol{i} \in \beta$ with profile $\mathfrak{t}^\beta$, and w is in the subgroup of B_n isomorphic to $B^{(1)} \times \dots \times B^{(d)}$ from Lemma 5.3 (the stabiliser of $\mathfrak{t}^\beta$). We must fix reduced expressions for such w in order to define ψ_w. We fix them as the images of the reduced expressions of elements $B^{(1)} \times \dots \times B^{(d)}$ chosen in the preceding paragraph. That we can do so is the last statement in Lemma 5.3.

Finally, the image of a basis element (52) under the homomorphism ρ is

$$y_1^{b_1} \dots y_n^{b_n} \psi_{\overline{w}_1} \cdots \psi_{\overline{w}_d} e(\boldsymbol{i}^1, \dots, \boldsymbol{i}^d), \tag{54}$$

where $b_{n_1+\dots+n_{j-1}+k} = a_k^{(j)}$, and the notation $\overline{w}_j$ comes from Lemma 5.3. The concatenation $(\boldsymbol{i}^1, \dots, \boldsymbol{i}^d)$ has the profile $\mathfrak{t}^\beta$, since each $\boldsymbol{i}^j \in \beta^{(j)}$, and due to our choice of reduced expressions, we have $\psi_{\overline{w}_1} \cdots \psi_{\overline{w}_d} = \psi_{\overline{w}_1 \cdots \overline{w}_d}$. So we conclude that the element (54) is of the form (53). Further, it follows immediately that this way we can obtain all the basis elements of $e(\mathfrak{t}^\beta)V_\beta(\Gamma, \lambda, \gamma)e(\mathfrak{t}^\beta)$. We conclude that the homomorphism ρ sends a basis onto a basis and is, thus, bijective.

5.2.4. *Homomorphism property.* — To finish the proof of Proposition 5.2, it remains to check that the map defined in (46)–(49) extends to an algebra homomorphism. It is possible but quite lengthy to check explicitly that all defining relations are preserved. Instead, we will use the polynomial representation introduced in §4.2. We keep the use of the notations introduced in §4.2.

From the proof of Theorem 4.9, we see that the action of the algebra $V_\beta(\Gamma, \lambda, \gamma)$ on $K[x, \beta]$ is faithful, or in other words, we have an embedding of $V_\beta(\Gamma, \lambda, \gamma)$ in $\mathrm{End}_K(K[x, \beta])$. Therefore, if we denote $\phi\big(e(\mathfrak{t}^\beta)\big)$ the image of $e(\mathfrak{t}^\beta)$ by this embedding, we obtain an embedding of the algebra $e(\mathfrak{t}^\beta)V_\beta(\Gamma, \lambda, \gamma)e(\mathfrak{t}^\beta)$ in $\mathrm{End}_K\big(\phi\big(e(\mathfrak{t}^\beta)\big)K[x, \beta]\big)$. We immediately have:

$$\phi\big(e(\mathfrak{t}^\beta)\big)K[x, \beta] = \bigoplus_{\substack{\boldsymbol{i} \in \beta \\ p(\boldsymbol{i}) = \mathfrak{t}^\beta}} K[x_1, \dots, x_n]\mathbf{1}_{\boldsymbol{i}}\ . \tag{55}$$

On the other hand, we also have an embedding of $\bigotimes_{j=1}^{d} V_{\beta^{(j)}}\left(\Gamma^{(j)},\lambda^{(j)},\gamma^{(j)}\right)$ in $\mathrm{End}_K(\bigotimes_{j=1}^{d} K[x,\beta^{(j)}])$ and the natural identification:

$$\bigotimes_{j=1}^{d} K[x,\beta^{(j)}] = \bigotimes_{j=1}^{d} \bigoplus_{\boldsymbol{i}^j\in\beta^{(j)}} K[x_1^{(j)},\dots,x_{n_j}^{(j)}]\mathbf{1}_{\boldsymbol{i}^j} \cong \bigoplus_{\substack{\boldsymbol{i}\in\beta\\ p(\boldsymbol{i})=\mathfrak{t}^\beta}} K[x_1,\dots,x_n]\mathbf{1}_{\boldsymbol{i}} \ . \tag{56}$$

The identification simply maps $f_1\mathbf{1}_{\boldsymbol{i}^1}\otimes\cdots\otimes f_d\mathbf{1}_{\boldsymbol{i}^d}$ to $f_1\dots f_d\mathbf{1}_{(\boldsymbol{i}^1,\dots,\boldsymbol{i}^d)}$.

Through the identifications that we just made, both algebras related by the map in (46)–(49) are seen as algebras of endomorphisms of the same space, in (55) and (56). So in order to check the homomorphism property it is enough to check that both sides of Formulas (46)–(49) are, in fact, the same elements in the endomorphism algebra.

This verification follows immediately for (46)–(47) and (49). For the image of $\psi_0^{(j)}$, we proceed as follows. First, it is convenient to choose a polynomial representation as in §4.2 for which $P_{ij}(u,v) := (u-v)^{|j\to i|}$, if $i\neq j$ and $P_{ij}(u,v) := 0$ if $i=j$.

Let $\boldsymbol{i}\in\beta$, such that $p(\boldsymbol{i})=\mathfrak{t}^\beta$. This means that $\boldsymbol{i}=(\boldsymbol{i}^1,\dots,\boldsymbol{i}^d)$, where $\boldsymbol{i}^j\in\beta^{(j)}$. Fix $j\in\{1,\dots,d\}$ and set for brevity $k=n_1+\cdots+n_{j-1}$. Through the identifications explained above, the action of $\psi_0^{(j)}$ is given by:

$$f\mathbf{1}_{\boldsymbol{i}} \mapsto \left(\gamma_{i_{k+1}}\frac{f-{}^{r_0^{(j)}}f}{x_{k+1}}+\alpha_{i_{k+1}}(x_{k+1})\,{}^{r_0^{(j)}}f\right)\mathbf{1}_{r_0^{(j)}\cdot\boldsymbol{i}} \ ,$$

where we recall that $r_0^{(j)}=r_k\dots r_1r_0r_1\dots r_k$ acts on $\boldsymbol{i}$ simply by replacing i_{k+1} by $\theta(i_{k+1})$.

On the other hand, we need to calculate the action of $\psi_k\dots\psi_1\psi_0\psi_1\dots\psi_k$. We note that, with our choice of $P_{ij}(u,v)$, we have that $P_{ij}(u,v)=1$, if one index is among $\{i_1,\dots,i_k\}$, and the other is i_{k+1} or $\theta(i_{k+1})$. Indeed, i_{k+1} is not in the same connected component of the quiver as $i_1,\dots,i_k$, since $p(\boldsymbol{i})=\mathfrak{t}^\beta$. This is also true for $\theta(i_{k+1})$, since θ keeps the set $I^{(j)}$ stable.

Then the calculation is made in three steps, corresponding, respectively, to the action of $\psi_1\dots\psi_k$, the action of ψ_0 and the action of $\psi_k\dots\psi_1$:

$$\begin{aligned} f\mathbf{1}_{\boldsymbol{i}} &\mapsto {}^{r_1\dots r_k}f\mathbf{1}_{r_1\dots r_k\cdot\boldsymbol{i}}\\ &\mapsto \left(\gamma_{i_{k+1}}\frac{{}^{r_1\dots r_k}f-{}^{r_0r_1\dots r_k}f}{x_1}+\alpha_{i_{k+1}}(x_1)\,{}^{r_0r_1\dots r_k}f\right)\mathbf{1}_{r_0r_1\dots r_k\cdot\boldsymbol{i}}\\ &\mapsto \left(\gamma_{i_{k+1}}\frac{f-{}^{r_0^{(j)}}f}{x_{k+1}}+\alpha_{i_{k+1}}(x_{k+1})\,{}^{r_0^{(j)}}f\right)\mathbf{1}_{r_0^{(j)}\cdot\boldsymbol{i}} \ . \end{aligned}$$

This concludes the verification of the homomorphism property and the proof of Proposition 5.2.

5.3. Cyclotomic quotients. — As in §3.2.4, let $\Lambda = (\Lambda_i)_{i\in I}$ be a finitely-supported family of non-negative integers. In the same way as [24, 18, 19], we define the cyclotomic quotient of the algebra $V_\beta(\Gamma, \lambda, \gamma)$.

Definition 5.4. — We define the algebra $V_\beta^\Lambda(\Gamma, \lambda, \gamma)$ as the quotient of $V_\beta(\Gamma, \lambda, \gamma)$ by the two-sided ideal $\mathfrak{J}_\beta^\Lambda$ generated by the relations

$$y_1^{\Lambda_{i_1}} e(\boldsymbol{i}) = 0\,, \quad \text{for all } \boldsymbol{i} = (i_1, \dots, i_n) \in \beta.$$

The above relations are homogeneous, so that $V_\beta^\Lambda(\Gamma, \lambda, \gamma)$ is graded. Note that if $\Lambda_i = 0$ for all i, then

$$V_\beta^\Lambda(\Gamma, \lambda, \gamma) = \begin{cases} \{0\}, & \text{if } n \geq 1, \\ K, & \text{if } n = 0. \end{cases}$$

As in §3.2.4, for any $j \in \{1, \dots, d\}$ let $\Lambda^{(j)}$ be the restriction of Λ to the vertex set $I^{(j)}$ of $\Gamma^{(j)}$.

Corollary 5.5. — *We have an (explicit) isomorphism of graded algebras:*

$$V_\beta^\Lambda(\Gamma, \lambda, \gamma) \simeq \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}} \left(\bigotimes_{j=1}^d V_{\beta^{(j)}}^{\Lambda^{(j)}} \left(\Gamma^{(j)}, \lambda^{(j)}, \gamma^{(j)} \right) \right).$$

Proof. — Recall that the algebra $R_\beta(\Gamma)$ is isomorphic to a subalgebra of $V_\beta(\Gamma, \lambda, \gamma)$ (see Corollary 4.11). Moreover, if ϑ denotes the isomorphism of Theorem 5.1, its restriction to $R_\beta(\Gamma)$ is by construction the isomorphism of Corollary 3.12. Therefore, it follows immediately that the calculations made in the proof of Theorem 3.16 can be repeated verbatim here. They show that if we denote by $\mathfrak{J}_{\beta,\otimes}^\Lambda$ the ideal of $\bigotimes_{j=1}^d V_{\beta^{(j)}}(\Gamma^{(j)}, \lambda^{(j)}, \gamma^{(j)})$, such that the quotient is $\bigotimes_{j=1}^d V_{\beta^{(j)}}^{\Lambda^{(j)}}(\Gamma^{(j)}, \lambda^{(j)}, \gamma^{(j)})$ (see the proof of Theorem 3.16), we have

$$\vartheta(\mathfrak{J}_\beta^\Lambda) = \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}} \left(\mathfrak{J}_{\beta,\otimes}^\Lambda \right) .$$

This concludes the proof. □

We define $V_n^\Lambda(\Gamma, \lambda, \gamma) := \bigoplus_\beta V_\beta^\Lambda(\Gamma, \lambda, \gamma)$, where the direct sum is over the B_n-orbits β in I^n. As in Corollary 3.14, using the bijection (19) we deduce the following corollary. Note that we now use (19) with $G = \mathbb{Z}/2\mathbb{Z}$.

Corollary 5.6. — *We have an (explicit) isomorphism of graded algebras:*

$$V_n^\Lambda(\Gamma, \lambda, \gamma) \simeq \bigoplus_{\substack{n_1,\dots,n_d \geq 0 \\ n_1+\cdots+n_d=n}} \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}} \left(\bigotimes_{j=1}^d V_{n_j}^{\Lambda^{(j)}}(\Gamma^{(j)}, \lambda^{(j)}, \gamma^{(j)}) \right).$$

Remark 5.7. — As in Remark 3.19, we deduce that we can assume that Λ is supported on all components of Γ.

6. Quiver Hecke algebras for type D

To fit with the setting of [19], we now assume that K is a field with $char(K) \neq 2$.

Let Γ be a quiver with an involution θ as in §4.1 and let β be a B_n-orbit in I^n. As before, let $\Lambda = (\Lambda_i)_{i\in I}$ be a finitely-supported family of non-negative integers.

In this section, as in Remark 4.7, we consider the situation $\lambda_i = \gamma_i = 0$ for all $i \in I$, and simply denote by $V_\beta(\Gamma) = V_\beta(\Gamma, 0, 0)$ the resulting algebra, defined in Section 4.1 (note that Conditions (31) are satisfied with this choice of λ and γ). The defining relations (33)–(38) (those involving the generator ψ_0) simply become:

$$\psi_0 e(\boldsymbol{i}) = e(r_0 \cdot \boldsymbol{i})\psi_0, \tag{57}$$

$$\psi_0\psi_b = \psi_b\psi_0, \quad \text{for all } b \in \{2, \dots, n-1\}, \tag{58}$$

$$\psi_0 y_1 = -y_1\psi_0, \tag{59}$$

$$\psi_0 y_a = y_a\psi_0, \quad \text{for all } a \in \{2, \dots, n\}, \tag{60}$$

$$\psi_0^2 = 1\,, \tag{61}$$

$$(\psi_0\psi_1)^2 = (\psi_1\psi_0)^2\ . \tag{62}$$

So we can immediately see that we have an homogeneous involutive algebra automorphism ι of $V_\beta(\Gamma)$ given on the generators by:

$$\iota(\psi_0) = -\psi_0 \quad \text{and} \quad \iota(X) = X \quad \text{for } X \in \{\psi_1, \dots, \psi_{n-1}, y_1, \dots, y_n\} \cup \{e(\boldsymbol{i})\}_{\boldsymbol{i}\in\beta}\ . \tag{63}$$

Note that ι is the identity map if $n = 0$. We denote by $V_\beta(\Gamma)^\iota$ the fixed-point subalgebra of $V_\beta(\Gamma)$, that is, $V_\beta(\Gamma)^\iota = \{x \in V_\beta(\Gamma) \mid \iota(x) = x\}$. The subalgebra $V_\beta(\Gamma)^\iota$ is a graded subalgebra of $V_\beta(\Gamma)$, since ι is homogeneous.

Cyclotomic quotients. We recall that $V_\beta^\Lambda(\Gamma)$ is the quotient of $V_\beta(\Gamma)$ by the two-sided ideal $\mathfrak{J}_\beta^\Lambda$ generated by

$$y_1^{\Lambda_{i_1}} e(\boldsymbol{i}) = 0\,, \quad \text{for all } \boldsymbol{i} \in \beta.$$

These relations are homogeneous, so that the algebra $V_\beta^\Lambda(\Gamma)$ inherits the grading of $V_\beta(\Gamma)$. The same formulas as in (63) define an homogeneous involutive algebra automorphism of $V_\beta^\Lambda(\Gamma)$, and we make the slight abuse of notation of keeping the name ι for this automorphism. The fixed-point subalgebra is denoted $V_\beta^\Lambda(\Gamma)^\iota$.

6.1. Definition and main property of $W_\delta(\Gamma)$. — We recall some definitions and the results we need from [19].

If $n \geq 2$, we identify the Weyl group D_n of type D as the subgroup of B_n generated by $s_0 \coloneqq r_0r_1r_0$, $s_1 \coloneqq r_1$, …, $s_{n-1} \coloneqq r_{n-1}$. The convention we need

here is that $D_n = \{1\}$ if $n \in \{0, 1\}$. The group D_n then acts on I^n, if $n \geq 2$, by

$$\begin{aligned} s_0 \cdot (i_1, i_2, \dots, i_n) &= (\theta(i_2), \theta(i_1), i_3, \dots, i_n)\,, \\ s_a \cdot (\dots, i_a, i_{a+1}, \dots) &= (\dots, i_{a+1}, i_a, \dots) \quad a = 1, \dots, n-1. \end{aligned}$$

Let δ be a finite subset of I^n stable by D_n, that is, a finite union of D_n-orbits.

Definition 6.1. — Let $n \geq 2$. The algebra $W_\delta(\Gamma)$ is the unitary associative K-algebra generated by elements

$$\{y_a\}_{1 \leq a \leq n} \cup \{\psi_b\}_{1 \leq b \leq n-1} \cup \{\Psi_0\} \cup \{e(\boldsymbol{i})\}_{\boldsymbol{i} \in \delta},$$

with the relations (11)–(16) of Section 3 involving all the generators but Ψ_0, together with

$$\Psi_0 e(\boldsymbol{i}) = e(s_0 \cdot \boldsymbol{i})\Psi_0, \tag{64}$$

$$\Psi_0 \psi_b = \psi_b \Psi_0, \quad \text{for all } b \in \{1, \dots, n-1\} \text{ with } b \neq 2\,, \tag{65}$$

$$(\Psi_0 y_a + y_{r_1(a)}\Psi_0)e(\boldsymbol{i}) = \begin{cases} e(\boldsymbol{i}) & \text{if } \theta(i_1) = i_2, \\ 0 & \text{otherwise,} \end{cases} \quad \text{for } a \in \{1, 2\}, \tag{66}$$

$$\Psi_0 y_a = y_a \Psi_0, \quad \text{for all } a \in \{3, \dots, n\}, \tag{67}$$

$$\Psi_0^2 e(\boldsymbol{i}) = Q_{\theta(i_1), i_2}(-y_1, y_2)e(\boldsymbol{i}), \tag{68}$$

$$\begin{aligned} &(\Psi_0\psi_2\Psi_0 - \psi_2\Psi_0\psi_2)e(\boldsymbol{i}) \\ &\quad = \begin{cases} \dfrac{Q_{\theta(i_1), i_2}(-y_1, y_2) - Q_{\theta(i_1), i_2}(y_3, y_2)}{y_1 + y_3} e(\boldsymbol{i}), & \text{if } \theta(i_1) = i_3, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \tag{69}$$

for all $\boldsymbol{i} \in \delta$.

By convention, we set $W_\delta(\Gamma) = R_\delta(\Gamma)$ if $n \in \{0, 1\}$. Explicitly, $W_\delta(\Gamma) = K$ if $n = 0$, and $W_\delta(\Gamma) = \sum_{\boldsymbol{i} \in \delta} K[y_1]e(\boldsymbol{i})$ if $n = 1$. This choice for $n \in \{0, 1\}$ is important for the statements of the results in the next section.

Remark 6.2. — With the choices of Γ, θ and the notations of Remark 4.6, the algebra $W_\delta(\Gamma)$ is exactly the algebra $W_{\boldsymbol{x}}^\delta$ defined in [19].

The algebra $W_\delta(\Gamma)$ is $\mathbb{Z}$-graded with

$$\begin{aligned} \deg e(\boldsymbol{i}) &= 0, \\ \deg y_a &= 2, \\ \deg \Psi_0 e(\boldsymbol{i}) &= d(\theta(i_1), i_2), \\ \deg \psi_b e(\boldsymbol{i}) &= d(i_b, i_{b+1}). \end{aligned}$$

DEFINITION 6.3. — The cyclotomic quotient $W_\delta^\Lambda(\Gamma)$ is the quotient of the algebra $W_\delta(\Gamma)$ by the relations

$$y_1^{\Lambda_{i_1}} e(\boldsymbol{i}) = 0\,, \quad \text{for all } \boldsymbol{i} \in \delta.$$

The algebra $W_\delta^\Lambda(\Gamma)$ inherits the grading from $W_\delta(\Gamma)$, since the additional relations are homogeneous. If $\Lambda_i = 0$ for all i, then

$$W_\delta^\Lambda(\Gamma) = \begin{cases} \{0\}, & \text{if } n \geq 1, \\ K, & \text{if } n = 0. \end{cases}$$

Fixed-point isomorphism. Let β be a B_n-orbit in I^n. Note that β is a finite union of D_n-orbits, so that both algebras $V_\beta(\Gamma)$ and $W_\beta(\Gamma)$ are defined.

We recall the following results from [19]. Note that they were proved for a particular choice of Γ and θ (the one relevant for the next section). However, the proof does not depend on this choice and can be repeated verbatim in our general setting.

PROPOSITION 6.4 ([19]). — *(i) The algebra $W_\beta(\Gamma)$ is isomorphic to the subalgebra $V_\beta(\Gamma)^\iota$ of $V_\beta(\Gamma)$.*

(ii) Assume that Λ satisfies $\Lambda_{\theta(i)} = \Lambda_i$ for all $i \in I$. The cyclotomic quotient $W_\beta^\Lambda(\Gamma)$ is isomorphic to $V_\beta^\Lambda(\Gamma)^\iota$.

In both cases, an isomorphism is given by $\Psi_0 \mapsto \psi_0\psi_1\psi_0$ and $X \mapsto X$ for all the generators X but Ψ_0.

REMARK 6.5. — Note that it is assumed in [19] that $n \geq 2$. With our conventions, the statements are also true for $n \in \{0, 1\}$, in which cases the verification is straightforward.

REMARK 6.6. — Recall the defining relations (57), (59) and (61) of $V_\beta^\Lambda(\Gamma)$. Conjugating the cyclotomic relations of $V_\beta^\Lambda(\Gamma)$ by ψ_0, we obtain $y_1^{\Lambda_{i_1}} e(r_0 \cdot \boldsymbol{i}) = 0$ for any $\boldsymbol{i} \in \beta$. From this remark, it is easy to see that we have, in fact, $V_\beta^\Lambda(\Gamma) = V_\beta^{\widetilde{\Lambda}}(\Gamma)$, where $\widetilde{\Lambda}$ is now given by $\widetilde{\Lambda}_i = \min\{\Lambda_i, \Lambda_{\theta(i)}\}$. This phenomenon does not necessarily also occur in $W_\beta^\Lambda(\Gamma)$ (where ψ_0 is not present), and this explains the assumptions on Λ in Proposition 6.4(ii).

We note that the isomorphisms given in the preceding proposition are isomorphisms of graded algebras. Indeed, in $V_\beta(\Gamma)$, we have $\deg \psi_0 = 0$, and so it is straightforward to check that the given map is homogeneous.

From Proposition 6.4(i) and Corollary 4.11, we immediately obtain the following statement.

COROLLARY 6.7. — *The subalgebra of $W_\beta(\Gamma)$ generated by all generators but Ψ_0 is isomorphic to $R_\beta(\Gamma)$.*

Semi-direct product. In this paragraph, assume that $n \geq 1$. Since ι is involutive, the vector space $V_\beta(\Gamma)$ decomposes as

$$V_\beta(\Gamma) = V_\beta(\Gamma)^\iota \oplus V_\beta(\Gamma)^- \ ,$$

where $V_\beta(\Gamma)^-$ is the eigenspace of ι for the eigenvalue -1. Moreover, the generator ψ_0 is invertible (in fact, $\psi_0^2 = 1$) and satisfies $\iota(\psi_0) = -\psi_0$. So the multiplication by ψ_0 provides an isomorphism of vector spaces between $V_\beta(\Gamma)^\iota$ and $V_\beta(\Gamma)^-$, so that $V_\beta(\Gamma)^-$ can be written as $V_\beta(\Gamma)^\iota\psi_0$. Working out the multiplication in $V_\beta(\Gamma)$

$$(x + y\psi_0)(x' + y'\psi_0) = xx' + y\psi_0 y'\psi_0 + (y\psi_0 x'\psi_0 + xy')\psi_0 \ ,$$

we obtain as a standard consequence that $V_\beta(\Gamma)$ is isomorphic to the semi-direct product $V_\beta(\Gamma)^\iota \rtimes C_2$, where the action of the cyclic group C_2 of order 2 on $V_\beta(\Gamma)^\iota$ is by conjugation by ψ_0. Recall that as a vector space $V_\beta(\Gamma)^\iota \rtimes C_2$ is the tensor product $V_\beta(\Gamma)^\iota \otimes K[C_2]$, and the multiplication is given by

$$(x \otimes \psi_0^\epsilon)(x' \otimes \psi_0^{\epsilon'}) = (x\psi_0^\epsilon x'\psi_0^\epsilon) \otimes \psi_0^{\epsilon+\epsilon'} \ .$$

Then we formulate the preceding standard facts taking into account Proposition 6.4. First we explicitly give the automorphism of $W_\beta(\Gamma)$ induced by conjugation by ψ_0 in $V_\beta(\Gamma)$. We denote this automorphism of order 2 by π. It is given on the generators by:

$$\pi : \quad \Psi_0 \mapsto \psi_1 \, , \quad \psi_1 \mapsto \Psi_0 \, , \quad y_1 \mapsto -y_1 \, , \quad e(\boldsymbol{i}) \mapsto e(r_0 \cdot \boldsymbol{i}) \ , \tag{70}$$

and the identity on all the other generators. As a consequence of Proposition 6.4 together with the preceding discussion, we conclude that

$$V_\beta(\Gamma) \simeq W_\beta(\Gamma) \rtimes \langle \pi \rangle \ ,$$

and similarly, for Λ as in Proposition 6.4(ii),

$$V_\beta^\Lambda(\Gamma) \simeq W_\beta^\Lambda(\Gamma) \rtimes \langle \pi \rangle \ , \tag{71}$$

where we still denote by π the automorphism of order 2 of $W_\beta^\Lambda(\Gamma)$ given by the same Formulas (70). This is, indeed, an automorphism, since Λ satisfies the assumption of Proposition 6.4(ii).

With these descriptions as semi-direct products, the involution ι on $V_\beta(\Gamma)$ (and on $V_\beta^\Lambda(\Gamma)$) is simply given by:

$$\iota\big(x \otimes \pi^\epsilon\big) = (-1)^\epsilon x \otimes \pi^\epsilon \ , \tag{72}$$

where $\epsilon \in \{0, 1\}$ and $x \in W_\beta(\Gamma)$ (or $x \in W_\beta^\Lambda(\Gamma)$).

6.2. Disjoint quiver isomorphism for $W_\delta(\Gamma)$. — Now let d be a positive integer and assume that the quiver Γ admits a decomposition $\Gamma = \coprod_{j=1}^{d} \Gamma^{(j)}$ as in §5. Let β be a B_n-orbit in I^n. As in §5, for any $j \in \{1, \dots, d\}$, we have an integer $n_j(\beta) = n_j$ and a B_{n_j}-orbit $\beta^{(j)}$ in $(I^{(j)})^{n_j}$.

If $n_j(\beta) = 0$ for some $j \in \{1, \dots, d\}$, then consider $\tilde{\Gamma}$ the quiver where we removed the component $\Gamma^{(j)}$. It follows immediately from the definitions that $W_\beta(\Gamma)$ is the same algebra as $W_\beta(\tilde{\Gamma})$. So we lose no generality by assuming that $n_j(\beta) \neq 0$ for all $j \in \{1, \dots, d\}$.

Fixed points of tensor products. Since $n_j(\beta) \geq 1$ for all $j \in \{1, \dots, d\}$, in the preceding section, we have $V_{\beta^{(j)}}(\Gamma^{(j)}) \simeq W_{\beta^{(j)}}(\Gamma^{(j)}) \rtimes C_2$ for all j. Hence,

$$\bigotimes_{j=1}^{d} V_{\beta^{(j)}}(\Gamma^{(j)}) \simeq \Big(\bigotimes_{j=1}^{d} W_{\beta^{(j)}}(\Gamma^{(j)})\Big) \rtimes C_2^d \ ,$$

where C_2^d acts on the tensor product by the automorphism π from (70) on each factor.

We would like to describe the fixed points of $\bigotimes_{j=1}^{d} V_{\beta^{(j)}}(\Gamma^{(j)})$ for the involutive automorphism $\iota^{\otimes}$ given by the tensor product of ι for each factor. From Formula (72), we can immediately see that

$$\Big(\bigotimes_{j=1}^{d} V_{\beta^{(j)}}(\Gamma^{(j)})\Big)^{\iota^{\otimes}} \simeq \Big(\bigotimes_{j=1}^{d} W_{\beta^{(j)}}(\Gamma^{(j)})\Big) \rtimes C_2^{d-1} \ , \tag{73}$$

where C_2^{d-1} is seen as the subgroup of "even" elements of C_2^d, namely,

$$C_2^{d-1} = \{(\pi^{\epsilon_1}, \dots, \pi^{\epsilon_d}) \in C_2^d \quad \text{such that } \epsilon_1 + \cdots + \epsilon_d = 0 \pmod 2\} \ . \tag{74}$$

Disjoint quiver isomorphism. We can now formulate the main result of this section. Recall that $n_j(\beta) \neq 0$ for all $j \in \{1, \dots, d\}$.

THEOREM 6.8. — *We have (explicit) isomorphisms of graded algebras:*

$$W_\beta(\Gamma) \simeq \mathrm{Mat}_{\binom{n}{n_1, \dots, n_d}} \left(\Big(\bigotimes_{j=1}^{d} W_{\beta^{(j)}}(\Gamma^{(j)})\Big) \rtimes C_2^{d-1} \right) \ , \tag{75}$$

and, assuming $d > 1$,

$$W_\beta^\Lambda(\Gamma) \simeq \mathrm{Mat}_{\binom{n}{n_1, \dots, n_d}} \left(\Big(\bigotimes_{j=1}^{d} W_{\beta^{(j)}}^{\widetilde{\Lambda}^{(j)}}(\Gamma^{(j)})\Big) \rtimes C_2^{d-1} \right) \ , \tag{76}$$

where $\widetilde{\Lambda} = (\widetilde{\Lambda}_i)_{i \in I}$ is defined by $\widetilde{\Lambda}_i \coloneqq \min\{\Lambda_i, \Lambda_{\theta(i)}\}$.

Note that in both formulas above, the group C_2^{d-1} is as given in (74). Moreover, the semi-direct product in Formula (76) is well defined, since each $\widetilde{\Lambda}^{(j)}$ satisfies the condition $\widetilde{\Lambda}_i^{(j)} = \widetilde{\Lambda}_{\theta(i)}^{(j)}$ of Proposition 6.4(ii) (see (71)).

REMARK 6.9. — The reader may have noticed that the assumptions $d > 1$ and $n_j(\beta) \neq 0$ (which do not reduce the generality, as explained above) were not present in the preceding section for the type B in Theorem 5.1 and Corollary 5.5. Indeed, those statements are more uniform in the sense that they are also valid as they are, even if some $n_j(\beta)$ are 0 or if $d = 1$. In particular, for $d = 1$, we do *not* necessarily have $W_\beta^\Lambda(\Gamma) = W_\beta^{\widetilde{\Lambda}}(\Gamma)$ (cf. Remark 6.6).

Proof. — • Recall from Theorem 5.1 that we have an isomorphism between $V_\beta(\Gamma)$ and the algebra $\mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}}\big(\bigotimes_{j=1}^d V_{\beta^{(j)}}(\Gamma^{(j)})\big)$. This isomorphism was obtained with the following two steps:

$$V_\beta(\Gamma) \simeq \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}}\big(e(\mathfrak{t}^\beta)V_\beta(\Gamma)e(\mathfrak{t}^\beta)\big) \quad \text{and}$$

$$\bigotimes_{j=1}^d V_{\beta^{(j)}}(\Gamma^{(j)}) \simeq e(\mathfrak{t}^\beta)V_\beta(\Gamma)e(\mathfrak{t}^\beta)\ .$$

For the first isomorphism, see §5.1, the construction of the idempotent $e(\mathfrak{t}^\beta)$ does not involve ψ_0, and neither does the construction of the matrix units (that is, the construction of the elements $\psi_\mathfrak{t}$ and $\phi_\mathfrak{t}$ given by Formulas (20)). So we immediately deduce how the automorphism ι of $V_\beta(\Gamma)$ behaves with respect to this isomorphism; namely, we have

$$V_\beta(\Gamma)^\iota \simeq \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}}\big(e(\mathfrak{t}^\beta)V_\beta(\Gamma)^\iota e(\mathfrak{t}^\beta)\big)\ .$$

According to Formula (73) (that we can use since $n_j(\beta) \neq 0$), to prove (75) it only remains to show that

$$e(\mathfrak{t}^\beta)V_\beta(\Gamma)^\iota e(\mathfrak{t}^\beta) \simeq \Big(\bigotimes_{j=1}^d V_{\beta^{(j)}}(\Gamma^{(j)})\Big)^{\iota^\otimes}\ .$$

So if we denote by ρ the isomorphic map from $\bigotimes_{j=1}^d V_{\beta^{(j)}}(\Gamma^{(j)})$ to $e(\mathfrak{t}^\beta)V_\beta(\Gamma)e(\mathfrak{t}^\beta)$, it remains to check that

$$\rho \circ \iota^\otimes = \iota \circ \rho\ .$$

This is immediately verified from Formulas (46)–(49) giving the map ρ in the proof of Proposition 5.2. Moreover, the isomorphism (75) is graded, since it is the restriction of a graded isomorphism (to a graded subalgebra).

• To prove (76) we start exactly as in the proof of Corollary 5.5; namely, we repeat the calculations in the proof of Theorem 3.16. We can do so, since $R_\beta(\Gamma)$ is a subalgebra of $W_\beta(\Gamma)$ by Corollary 6.7.

Let ϑ denote the isomorphism in (75) and let $\mathfrak{K}^\Lambda_\beta$ denote the ideal of $W_\beta(\Gamma)$ giving the cyclotomic quotient $W^\Lambda_\beta(\Gamma)$. The proofs of Corollary 5.5 and Theorem 3.16 show that

$$\vartheta(\mathfrak{K}^\Lambda_\beta) = \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}}(\mathfrak{K}^\Lambda_{\beta,\otimes}),$$

where $\mathfrak{K}^\Lambda_{\beta,\otimes}$ is the ideal of $\big(\bigotimes_{j=1}^d W_{\beta^{(j)}}(\Gamma^{(j)})\big) \rtimes C_2^{d-1}$ generated by the elements

$$y_b^{\Lambda_{i_b}} e(\boldsymbol{i}) \,, \tag{77}$$

where $\boldsymbol{i}$ is of profile $\mathfrak{t}^\beta$, and b is of the form $b = n_1 + \cdots + n_{j-1} + 1$ for $j \in \{1, \dots, d\}$. Note that, as in the proof of Theorem 3.16, we abuse notations slightly: if $\boldsymbol{i} = (\boldsymbol{i}^1, \dots, \boldsymbol{i}^d)$ with $\boldsymbol{i}^k \in \beta^{(k)}$, we identify $y_b^{\Lambda_{i_b}} e(\boldsymbol{i}) \in W_\beta(\Gamma)$ with the element of $\bigotimes_{j=1}^d W_{\beta^{(j)}}(\Gamma^{(j)})$, which is $e(\boldsymbol{i}^k)$ in the k-th factor with $k \neq j$ and $(y_1^{(j)})^{\Lambda_{(\boldsymbol{i}^j)_1}} e(\boldsymbol{i}^j)$ in the j-th factor (where $y_1^{(j)}$ denotes the generator y_1 of $W_{\beta^{(j)}}(\Gamma^{(j)})$).

Contrary to the types A and B, we need to show something more here to prove (76). In particular, we cannot consider the semi-direct product $\big(\bigotimes_{j=1}^d W^{\Lambda^{(j)}}_{\beta^{(j)}}(\Gamma^{(j)})\big) \rtimes C_2^{d-1}$, since the elements $\Lambda^{(j)}$ do not necessarily satisfy the stability condition of Proposition 6.4(ii). Thus, let $\mathfrak{K}^{\widetilde{\Lambda}}_{\beta,\otimes}$ be the ideal of $\big(\bigotimes_{j=1}^d W_{\beta^{(j)}}(\Gamma^{(j)})\big) \rtimes C_2^{d-1}$ generated by the elements

$$y_b^{\widetilde{\Lambda}_{i_b}} e(\boldsymbol{i}) \,, \tag{78}$$

where $\boldsymbol{i} \in \beta$ is of profile $\mathfrak{t}^\beta$, and b is of the form $b = n_1 + \cdots + n_{j-1} + 1$ for $j \in \{1, \dots, d\}$, and where $\widetilde{\Lambda}$ is defined in Theorem 6.8. We will show that

$$\mathfrak{K}^\Lambda_{\beta,\otimes} = \mathfrak{K}^{\widetilde{\Lambda}}_{\beta,\otimes} \,.$$

First, since $\widetilde{\Lambda}_i \leq \Lambda_i$ for all $i \in I$, we have $\mathfrak{K}^\Lambda_{\beta,\otimes} \subset \mathfrak{K}^{\widetilde{\Lambda}}_{\beta,\otimes}$. For the reverse inclusion, take an element $y_b^{\widetilde{\Lambda}_{i_b}} e(\boldsymbol{i})$ as in (77). If $\widetilde{\Lambda}_{i_b} = \Lambda_{i_b}$, then $y_b^{\widetilde{\Lambda}_{i_b}} e(\boldsymbol{i}) \in \mathfrak{K}^\Lambda_{\beta,\otimes}$, and, thus, we assume that $\widetilde{\Lambda}_{i_b} = \Lambda_{\theta(i_b)}$. Let $\xi \in C_2^{d-1}$, such that the component of ξ in position j is π. Such an element exists, since we assumed that $d > 1$. Then, using Formulas (70) for the action of π on $W_{\beta^{(j)}}(\Gamma^{(j)})$, we have, where $\boldsymbol{i}' \in \beta$ of profile $\mathfrak{t}^\beta$ is such that $i'_b = \theta(i_b)$,

$$\xi \cdot \big(y_b^{\widetilde{\Lambda}_{i_b}} e(\boldsymbol{i})\big) = (-y_b)^{\widetilde{\Lambda}_{i_b}} e(\boldsymbol{i}') = (-y_b)^{\Lambda_{\theta(i_b)}} e(\boldsymbol{i}') = (-y_b)^{\Lambda_{i'_b}} e(\boldsymbol{i}') \,.$$

Since the action of ξ is invertible, we thus deduce that $y_b^{\widetilde{\Lambda}_{i_b}} e(\boldsymbol{i}) \in \mathfrak{K}^\Lambda_{\beta,\otimes}$. Finally, we show that all elements in (78) are in $\mathfrak{K}^\Lambda_{\beta,\otimes}$, and thus $\mathfrak{K}^{\widetilde{\Lambda}}_{\beta,\otimes} \subset \mathfrak{K}^\Lambda_{\beta,\otimes}$. This concludes the proof. □

We define $W_n(\Gamma) \coloneqq \bigoplus_\delta W_\delta(\Gamma)$, where δ runs over all the orbits of I^n under the action of D_n, and, similarly, $W_n^\Lambda(\Gamma) = \bigoplus_\delta W_\delta^\Lambda(\Gamma)$. In the type D situation, the statements below are less clean than those of Corollary 3.14 or Corollary 5.6. Nevertheless, they still allow to explicitly reduce the study of $W_n(\Gamma)$ and $W_n^\Lambda(\Gamma)$ to the situation of a quiver with a single component.

For $(n_1, \dots, n_d) \in (\mathbb{Z}_{\geq 0})^d$, we denote by $l(n_1, \dots, n_d)$ the number of its non-zero components. Assume that $n \geq 1$ to avoid a trivial situation.

COROLLARY 6.10. — *We have (explicit) isomorphisms of graded algebras:*

$$W_n(\Gamma) \simeq \bigoplus_{\substack{n_1,\dots,n_d\geq 0\\ n_1+\cdots+n_d=n}} \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}} \left(\Big(\bigotimes_{\substack{j=1\\ n_j\neq 0}}^{d} W_{n_j}(\Gamma^{(j)}) \Big) \rtimes C_2^{l(n_1,\dots,n_d)-1} \right) ,$$

$$W_n^\Lambda(\Gamma) \simeq \bigoplus_{\substack{n_1,\dots,n_d\geq 0\\ n_1+\cdots+n_d=n}} \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}} \big(\mathcal{W}(n_1, \dots, n_d) \big) ,$$

where:

- *If $l(n_1, \dots, n_d) = 1$ then $\mathcal{W}(n_1, \dots, n_d) \coloneqq W_{n_j}^{\Lambda^{(j)}}(\Gamma^{(j)})$, where j is the component such that $n_j = n$.*
- *If $l(n_1, \dots, n_d) > 1$, then*

$$\mathcal{W}(n_1, \dots, n_d) \coloneqq \Big(\bigotimes_{\substack{j=1\\ n_j\neq 0}}^{d} W_{n_j}^{\widetilde{\Lambda}^{(j)}}(\Gamma^{(j)}) \Big) \rtimes C_2^{l(n_1,\dots,n_d)-1} .$$

Proof. — We write $W_n(\Gamma) = \bigoplus_\beta W_\beta(\Gamma)$ and $W_n^\Lambda(\Gamma) = \bigoplus_\beta W_\beta^\Lambda(\Gamma)$, where β runs over all the orbits of I^n under the action of B_n. We note that, if some $n_j(\beta)$ are equal to 0 then, as explained at the beginning of this section, we can remove the corresponding components of Γ to obtain another quiver $\tilde{\Gamma}$ for which the assumptions of Theorem 6.8 are satisfied. Then the proof is a repetition of the proof of Corollary 3.14, using Theorem 6.8 for each orbit β. □

REMARK 6.11. — As in Remarks 3.19 and 5.7, we deduce that we can assume that Λ is supported on all the components of Γ.

7. Morita equivalence for cyclotomic quotients of affine Hecke algebras of types B and D

In this section, we will combine our previous results Corollaries 5.6 and 6.10 with [18, 19] to obtain Morita equivalence theorems for cyclotomic quotients of affine Hecke algebras of types B and D. We emphasize that these Morita

equivalences will be deduced from isomorphisms. As they combine the isomorphisms of [18, 19] with those of the previous sections, these isomorphisms can be written down explicitly, even though they are rather complicated.

Recall that K is a field with characteristic different from 2. Let $p, q \in K \setminus \{0\}$ such that $q^2 \neq 1$. As in Remark 4.6, for any $x \in K \setminus \{0\}$, we define the set

$$I_x \coloneqq \{x^\epsilon q^{2l} : \epsilon \in \{\pm 1\}, l \in \mathbb{Z}\} \ .$$

Then we take $d \geq 1$ and $x_1, \ldots, x_d \in K^\times$, such that the sets $I^{(j)} \coloneqq I_{x_j}$ are pairwise disjoint, and we set

$$I \coloneqq \coprod_{j=1}^{d} I_{x_j} \ .$$

The quiver Γ with involution that we will be considering in this section is the following:

- The vertex set of Γ is I as above.
- There is an arrow starting from v and pointing to $q^2 v$ for all $v \in I$. These are all arrows.
- The involution θ on I is the scalar inversion $\theta(x) = x^{-1}$ for all $x \in I$.

The partition $I = \coprod_{j=1}^{d} I^{(j)}$ induces a decomposition of Γ into full subquivers $\Gamma = \coprod_{j=1}^{d} \Gamma^{(j)}$ as in Section 5, in particular each $\Gamma^{(j)}$ is stable under the scalar inversion θ. We also choose a finitely supported family $\Lambda = (\Lambda_i)_{i \in I}$ of non-negative integers. Finally, we let L be a free $\mathbb{Z}$-module of rank n with basis $\{\epsilon_i\}_{i=1,\ldots,n}$:

$$L \coloneqq \bigoplus_{i=1}^{n} \mathbb{Z}\epsilon_i \ .$$

7.1. Morita equivalence for cyclotomic quotients of affine Hecke algebras of type B. — We set

$$\alpha_0 \coloneqq 2\epsilon_1 \quad \text{and} \quad \alpha_i \coloneqq \epsilon_{i+1} - \epsilon_i \ , \ i = 1, \ldots, n-1 \ .$$

For $n \geq 1$, the Weyl group B_n of type B acts on L by

$$\begin{aligned} r_0(\epsilon_1) &= -\epsilon_1, \\ r_0(\epsilon_i) &= \epsilon_i \quad \text{if } i > 1, \\ r_a(\epsilon_i) &= \epsilon_{r_a(i)}, \end{aligned}$$

for $i = 1, \ldots, n-1$ and $a = 1, \ldots, n-1$.

We denote $q_0 \coloneqq p$ and $q_i \coloneqq q$ for $i = 1, \ldots, n-1$. The affine Hecke algebra $\widehat{H}(B_n)$ is the unitary K-algebra generated by elements

$$g_0, g_1, \ldots, g_{n-1} \quad \text{and} \quad X^x, \quad x \in L \ .$$

The defining relations are $X^0 = 1$, $X^x X^{x'} = X^{x+x'}$ for any $x, x' \in L$, and the characteristic equations for the generators g_i:

$$g_i^2 = (q_i - q_i^{-1})g_i + 1 \quad \text{for } i \in \{0, \dots, n-1\}\,, \tag{79}$$

with the braid relations of type B

$$g_0 g_1 g_0 g_1 = g_1 g_0 g_1 g_0 \tag{80}$$

$$g_i g_{i+1} g_i = g_{i+1} g_i g_{i+1} \quad \text{for } i \in \{1, \dots, n-2\}, \tag{81}$$

$$g_i g_j = g_j g_i \quad \text{for } i, j \in \{0, \dots, n-1\} \text{ such that } |i-j| > 1, \tag{82}$$

together with

$$g_i X^x - X^{r_i(x)} g_i = (q_i - q_i^{-1}) \frac{X^x - X^{r_i(x)}}{1 - X^{-\alpha_i}}\,,$$

for any $x \in L$ and $i = 0, 1, \dots, n-1$. Note that the right-hand side is a well-defined element, since there exists $k \in \mathbb{Z}$, such that $r_i(x) = x - k\alpha_i$. Note also that $\widehat{H}(B_0) = K$.

Let $X_i \coloneqq X^{\epsilon_i}$ for $i = 1, \dots, n$. An equivalent presentation of the algebra $\widehat{H}(B_n)$ is with generators

$$g_0, g_1, \dots, g_{n-1}, X_1^{\pm 1}, \dots, X_n^{\pm 1}\,,$$

and defining relations (79)–(82) together with

$X_i X_j = X_j X_i$ for $i, j \in \{1, \dots, n\}$,

$g_0 X_1^{-1} g_0 = X_1$,

$g_i X_i g_i = X_{i+1}$ for $i \in \{1, \dots, n-1\}$,

$g_i X_j = X_j g_i$ for $i \in \{0, \dots, n-1\}$ and $j \in \{1, \dots, n\}$ such that $j \neq i, i+1$.

Definition 7.1. — The cyclotomic quotient $H^\Lambda(B_n)$ of type B associated with $\Lambda = (\Lambda_i)_{i \in I}$ is the quotient of the algebra $\widehat{H}(B_n)$ over the relation

$$\prod_{i \in I} (X_1 - i)^{\Lambda_i} = 0\,.$$

Note that if $\Lambda_i = 0$ for all i, then

$$H^\Lambda(B_n) = \begin{cases} \{0\}, & \text{if } n \geq 1, \\ K, & \text{if } n = 0. \end{cases}$$

We recall the main result of [18, 19] concerning $H^\Lambda(B_n)$.

Theorem 7.2. — *Let λ, γ be as in Remarks 4.6 and 4.7 if $p^2 \neq 1$ and $p^2 = 1$, respectively. The algebras $H^\Lambda(B_n)$ and $V_n^\Lambda(\Gamma, \lambda, \gamma)$ are (explicitly) isomorphic.*

Remark 7.3. — Theorem 7.2 is proven for $n \geq 1$ but is also trivially true for $n = 0$.

We now state the first main application of the results of the preceding sections.

THEOREM 7.4. — *We have an (explicit) isomorphism of algebras:*

$$H^{\Lambda}(B_n) \simeq \bigoplus_{\substack{n_1,\dots,n_d\geq 0\\ n_1+\cdots+n_d=n}} \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}} \left(\bigotimes_{j=1}^{d} H^{\Lambda^{(j)}}(B_{n_j}) \right) .$$

In particular, $H^{\Lambda}(B_n)$ is Morita equivalent to

$$\bigoplus_{\substack{n_1,\dots,n_d\geq 0\\ n_1+\cdots+n_d=n}} \left(\bigotimes_{j=1}^{d} H^{\Lambda^{(j)}}(B_{n_j}) \right) .$$

Proof. — Note that the statement is true if $n = 0$, and, thus, we now assume $n \geq 1$. Let us first assume that $p^2 \neq 1$. Let λ be the indicator function of $\{\pm p\} \cap I$ and $(\gamma_i)_{i\in I}$ be given by $\gamma_i = \begin{cases} 1, & \text{if } \theta(i) = i, \\ 0, & \text{otherwise,} \end{cases}$ as in Remark 4.6. By Theorem 7.2, we have an isomorphism $H^{\Lambda}(B_n) \simeq V_n^{\Lambda}(\Gamma, \lambda, \gamma)$. For any $j \in \{1, \dots, d\}$, the restrictions $\lambda^{(j)}$ and $\gamma^{(j)}$ of λ and γ, respectively, to $I^{(j)}$ satisfy, by Corollary 5.6,

$$V_n^{\Lambda}(\Gamma, \lambda, \gamma) \simeq \bigoplus_{\substack{n_1,\dots,n_d\geq 0\\ n_1+\cdots+n_d=n}} \mathrm{Mat}_{\binom{n}{n_1,\dots,n_d}} \left(\bigotimes_{j=1}^{d} V_{n_j}^{\Lambda^{(j)}}(\Gamma^{(j)}, \lambda^{(j)}, \gamma^{(j)}) \right) .$$

Since $\lambda^{(j)}$ and $\gamma^{(j)}$ are still of the above form with respect to the quiver $\Gamma^{(j)}$, by Theorem 7.2 we have $V_{n_j}^{\Lambda^{(j)}}(\Gamma^{(j)}, \lambda^{(j)}, \gamma^{(j)}) \simeq H^{\Lambda^{(j)}}(B_{n_j})$ for any n_j. We, thus, deduce the isomorphism of the theorem. We deduce the statement of Morita equivalence, since $\mathrm{Mat}_N(A)$ and A are Morita equivalent for any algebra A and $N \in \mathbb{N}^*$. The case $p^2 = 1$ is similar, still by Theorem 7.2. □

We obtain the following corollary.

COROLLARY 7.5. — *To study an arbitrary cyclotomic quotient of the affine Hecke algebra $\widehat{H}(B_n)$ it is enough to consider cyclotomic quotients given by a relation*

$$\prod_{\substack{\epsilon\in\{\pm 1\}\\ l\in\mathbb{Z}}} (X_1 - x^{\epsilon} q^{2l})^{m_{\epsilon,l}} = 0 ,$$

for any finitely supported family of non-negative integers $(m_{\epsilon,l})_{\epsilon\in\{\pm 1\},l\in\mathbb{Z}}$, where $x \in K^{\times}$ satisfies one of the following four cases:

(**a**) $x = 1$ (**b**) $x = q$ (**c**) $x = p$ (**d**) $x \notin \pm q^{\mathbb{Z}} \cup \pm p^{\pm 1} q^{2\mathbb{Z}}$.

Proof. — We sketch a proof, in the same spirit as in the introduction of [18]. By Theorem 7.4 it is clear that it suffices to consider cyclotomic quotients given by a relation

$$\prod_{i\in I_x}(X_1-i)^{\Lambda_i}=0,$$

where $I_x=\{x^\epsilon q^{2l}:\epsilon\in\{\pm1\},l\in\mathbb{Z}\}$ with $x\in K^\times$ and $\Lambda=(\Lambda_i)_{i\in I_x}$ is a finitely supported family of non-negative integers. By Theorem 7.2 and Remark 4.6 this cyclotomic quotient is determined by:

- The quiver Γ with vertex set I_x, arrows $v\to q^2v$ for all $v\in I_x$ and involution $\theta:v\mapsto v^{-1}$ on I_x
- The set $\{\pm p\}\cap I_x$.

A first distinction arises when looking at the number of connected components of Γ. It has exactly one (or two) connected component(s), when $x^2\in q^{2\mathbb{Z}}$ (or $x^2\notin q^{2\mathbb{Z}}$).

The first case, $x^2\in q^{2\mathbb{Z}}$, is equivalent to $x\in\pm q^{\mathbb{Z}}$. We can switch between x and $-x$ by the variable change $X_i\leftarrow -X_i$ for all $i\in\{1,\dots,n\}$, replacing I_x by $-I_x=I_{-x}$ and $\Lambda=(\Lambda_i)_{i\in I_x}$ by $\Lambda'=(\Lambda'_i)_{i\in I_{-x}}$ given by $\Lambda'_i:=\Lambda_{-i}$ for all $i\in I_{-x}$. Thus, it suffices to consider $x\in q^{\mathbb{Z}}$, but now a simple shift of Λ (that is, setting $\Lambda'_i=\Lambda_{iq^{2N}}$ for appropriate N) shows that it suffices to consider the cases $x=1$ (this is case (**a**)) or $x=q$ (this is case (**b**)), according to the parity of the power of q.

We now consider the case $x^2\notin q^{2\mathbb{Z}}$, that is, $x\notin\pm q^{\mathbb{Z}}$. If $\{\pm p\}\cap I_x=\emptyset$, then $x\notin\pm q^{\mathbb{Z}}\cup\pm p^{\pm1}q^{2\mathbb{Z}}$, and all these choices of x lead to isomorphic algebras, since, moreover, θ has no fixed points (if $x^{\pm1}q^{2k}$ is fixed by θ, then $x^2\in q^{4\mathbb{Z}}$, and, thus, $x\in\pm q^{2\mathbb{Z}}$). This is the case (**d**). Now, if $\{\pm p\}\cap I_x\neq\emptyset$, using the variable change $X_i\leftarrow -X_i$ for all $i\in\{1,\dots,n\}$, we can always assume that $p\in I_x$, that is, $x\in p^{\pm1}q^{2\mathbb{Z}}$. In fact, it suffices to consider $x\in pq^{2\mathbb{Z}}$, since the variable change $g_0\leftarrow -g_0$ exchanges p and p^{-1}. This case reduces to $x=p$ by shifting Λ as above, and this is case (**c**). □

Remark 7.6. — We make additional final remarks on the four cases (**a**)–(**d**) to be considered.

- Cases (**a**) and (**b**) correspond to a quiver with a single connected component (an infinite oriented line or a finite oriented polygon, depending on whether or not q is a root, of unity). This quiver is stable by the involution θ, and then Case (**a**) corresponds to θ having a fixed point, while Case (**b**) generically corresponds to the situation where there is no fixed point. This latter situation cannot occur if the number of vertices is finite and odd, that is, Case (**b**) is not present (or more precisely, is not necessary, since it is equivalent to Case (**a**)) when q^2 is an odd root of unity.

- Cases (**c**) and (**d**) (generically) correspond to a quiver with two identical connected components (two infinite oriented lines or two finite oriented polygons depending on whether or not q is a root of unity), which are exchanged by the involution θ. Then Case (**c**) corresponds to the situation where one of the special values $\pm p^{-1}$ is present, while Case (**d**) corresponds to the situation where no such values occur. We see that Case (**c**) is not necessary (more precisely, it reduces to one of Cases (**a**) or (**b**)) when p^2 is a power of q^2.
- To summarise, there are at least two cases to consider in general: (**a**) and (**d**), while the additional two cases (**b**) and (**c**) are to be considered or not, depending on p and q.

7.2. Morita equivalence for cyclotomic quotients of affine Hecke algebras of type D. — Let $n \geq 2$. We set

$$\alpha'_0 = \epsilon_1 + \epsilon_2 \quad \text{and} \quad \alpha'_i = \epsilon_{i+1} - \epsilon_i\,,\ i = 1, \dots, n-1\ .$$

The Weyl group D_n of type D acts on L by

$$\begin{aligned} s_0(\epsilon_1) &= -\epsilon_2, \\ s_0(\epsilon_2) &= -\epsilon_1, \\ s_0(\epsilon_i) &= \epsilon_i, \quad \text{if } i > 2, \\ s_a(\epsilon_i) &= \epsilon_{r_a(i)}, \end{aligned}$$

for $i = 1, \dots, n-1$ and $a = 1, \dots, n-1$.

The affine Hecke algebra $\widehat{H}(D_n)$ is the unitary K-algebra generated by elements

$$\{g_i\}_{1 \leq i \leq n-1} \cup \{G_0\} \cup \{X^x\}_{x \in L}.$$

The defining relations are $X^0 = 1$, $X^x X^{x'} = X^{x+x'}$ for any $x, x' \in L$, and the characteristic equations for the generators g_i and G_0:

$$\begin{aligned} g_i^2 &= (q - q^{-1})g_i + 1 \qquad \text{for } i \in \{1, \dots, n-1\}, \\ G_0^2 &= (q - q^{-1})G_0 + 1, \end{aligned} \tag{83}$$

with the braid relations of type D

$$G_0 g_2 G_0 = g_2 G_0 g_2, \tag{84}$$

$$G_0 g_i = g_i G_0 \qquad \text{for } i \in \{1, \dots, n-1\} \setminus \{2\}, \tag{85}$$

$$g_i g_{i+1} g_i = g_{i+1} g_i g_{i+1} \qquad \text{for } i \in \{1, \dots, n-2\}, \tag{86}$$

$$g_i g_j = g_j g_i \qquad \text{for } i, j \in \{1, \dots, n-1\} \text{ such that } |i - j| > 1, \tag{87}$$

together with

$$g_i X^x - X^{s_i(x)} g_i = (q - q^{-1}) \frac{X^x - X^{s_i(x)}}{1 - X^{-\alpha_i'}},$$
$$G_0 X^x - X^{s_0(x)} G_0 = (q - q^{-1}) \frac{X^x - X^{s_0(x)}}{1 - X^{-\alpha_0'}},$$

for any $x \in L$ and $i = 1, \dots, n-1$. Note that the right-hand sides are well-defined elements, since for any $i \in \{0, \dots, n-1\}$, there exists $k \in \mathbb{Z}$, such that $s_i(x) = x - k\alpha_i'$.

An equivalent presentation of the algebra $\widehat{H}(D_n)$ is with generators (where again $X_i \coloneqq X^{\epsilon_i}$)

$$\{g_i\}_{1 \le i \le n-1} \cup \{G_0\} \cup \{X_i^{\pm 1}\}_{1 \le i \le n}$$

and defining relations (83)–(87) together with

$X_i X_j = X_j X_i$ for $i, j \in \{1, \dots, n\}$,
$G_0 X_1^{-1} G_0 = X_2$,
$G_0 X_i = X_i G_0$ for $i \in \{3, \dots, n-1\}$,
$g_i X_i g_i = X_{i+1}$ for $i \in \{1, \dots, n-1\}$,
$g_i X_j = X_j g_i$ for $i \in \{1, \dots, n-1\}$ and $j \in \{1, \dots, n\}$ such that $j \neq i, i+1$.

By convention, we set that $\widehat{H}(D_n)$ coincides with the usual affine Hecke algebra of type A_n if $n \in \{0, 1\}$, that is, we have $\widehat{H}(D_0) = K$ and $\widehat{H}(D_1) = K[X_1^{\pm 1}]$.

Definition 7.7. — The cyclotomic quotient $H^\Lambda(D_n)$ of type D associated with $\Lambda = (\Lambda_i)_{i \in I}$ is the quotient of the algebra $\widehat{H}(D_n)$ over the relation

$$\prod_{i \in I} (X_1 - i)^{\Lambda_i} = 0 \ .$$

Note that if $\Lambda_i = 0$ for all i, then

$$H^\Lambda(D_n) = \begin{cases} \{0\}, & \text{if } n \ge 1, \\ K, & \text{if } n = 0. \end{cases}$$

We recall the main result of [19] concerning $H^\Lambda(D_n)$. Recall that the quiver Γ was defined at the beginning of Section 7.

Theorem 7.8. — *The algebras $H^\Lambda(D_n)$ and $W_n^\Lambda(\Gamma)$ are (explicitly) isomorphic.*

Remark 7.9. — Theorem 7.8 is proven for $n \ge 2$, but with our conventions it follows immediately that it remains true for $n \in \{0, 1\}$.

Expression as a semi-direct product. We assume here that $n \geq 1$. Assuming $p^2 = 1$, we can now see $\widehat{H}(D_n)$ as a subalgebra of $\widehat{H}(B_n)$. Namely, we have an inclusion (see, for instance, [19, §2.3]) $\widehat{H}(D_n) \subseteq \widehat{H}(B_n)$, given on the generators by

$$G_0 \mapsto g_0 g_1 g_0, \quad g_i \mapsto g_i, \quad X_j^{\pm 1} \mapsto X_j^{\pm 1},$$

for any $i \in \{1, \dots, n-1\}$ and $j \in \{1, \dots, n\}$. Another way to see $\widehat{H}(D_n)$ as a subalgebra of $\widehat{H}(B_n)$ is to write $\widehat{H}(D_n)$ as the subalgebra of fixed points of $\widehat{H}(B_n)$ under the involution η given by

$$g_0 \mapsto -g_0, \quad g_i \mapsto g_i, \quad X_j^{\pm 1} \mapsto X_j^{\pm 1},$$

for each $i \in \{1, \dots, n-1\}$ and $j \in \{1, \dots, n\}$ (note that since $p^2 = 1$, the defining relation for the generator g_0 is $g_0^2 = 1$). In particular, as in §6.1, we have a vector space decomposition $\widehat{H}(B_n) = \widehat{H}(D_n) \oplus \widehat{H}(D_n) g_0$ and, thus, an isomorphism of algebras

$$\widehat{H}(B_n) \simeq \widehat{H}(D_n) \rtimes C_2.$$

Note that the action of the generator of C_2 on the generating set of $\widehat{H}(D_n)$ is given by

$$\begin{aligned} &G_0 \mapsto g_1, \qquad g_1 \mapsto G_0, \qquad g_i \mapsto g_i, \\ &X_1 \mapsto X_1^{-1}, \quad X_1^{-1} \mapsto X_1, \quad X_j^{\pm 1} \mapsto X_j^{\pm 1}, \end{aligned}$$

for all $i \in \{2, \dots, n-1\}$ and $j \in \{2, \dots, n\}$.

The involution η on $\widehat{H}(B_n)$ is compatible with the cyclotomic quotient $H^{\Lambda}(B_n)$. Now, if Λ satisfies the stability condition of Proposition 6.4(ii) (which is here $\Lambda_{i^{-1}} = \Lambda_i$ for all $i \in I$), the previous action of C_2 on $\widehat{H}(D_n)$ is compatible with the cyclotomic quotient $H^{\Lambda}(D_n)$, and, as above, we have

$$H^{\Lambda}(B_n) \simeq H^{\Lambda}(D_n) \rtimes C_2.$$

Morita equivalence theorem. Let $n_1, \dots, n_d \geq 1$. If Λ satisfies $\Lambda_{i^{-1}} = \Lambda_i$ for all $i \in I$, the previous action of C_2 on $H^{\Lambda}(D_n)$ extends to a (diagonal) action of C_2^d on $\bigotimes_{j=1}^{d} H^{\Lambda^{(j)}}(D_{n_j})$. As in §6.2, we restrict this action to the subgroup C_2^{d-1} of even elements given in (74). Recall also the definition of $\widetilde{\Lambda} = (\widetilde{\Lambda}_i)_{i \in I}$ given in Theorem 6.8.

We now state the second main application of the paper. As in Corollary 6.10, for any $(n_1, \dots, n_d) \in (\mathbb{Z}_{\geq 0})^d$, we denote by $l(n_1, \dots, n_d)$ the number of its non-zero components.

THEOREM 7.10. — *We have an (explicit) isomorphism of algebras:*

$$H^{\Lambda}(D_n) \simeq \bigoplus_{\substack{n_1, \dots, n_d \geq 0 \\ n_1 + \dots + n_d = n}} \mathrm{Mat}_{\binom{n}{n_1, \dots, n_d}} \Big(\mathcal{H}(n_1, \dots, n_d) \Big)$$

Where:

- *If* $l(n_1,\dots,n_d)=1$, *then* $\mathcal{H}(n_1,\dots,n_d) \coloneqq H^{\Lambda^{(j)}}(D_{n_j})$, *where* j *is the component, such that* $n_j=n$.
- *If* $l(n_1,\dots,n_d)>1$, *then*

$$\mathcal{H}(n_1,\dots,n_d) \coloneqq \Big(\bigotimes_{\substack{j=1\\ n_j\neq 0}}^{d} H^{\widetilde{\Lambda}^{(j)}}(D_{n_j})\Big) \rtimes C_2^{l(n_1,\dots,n_d)-1} \ .$$

In particular, $H^{\Lambda}(D_n)$ *is Morita equivalent to*

$$\bigoplus_{\substack{n_1,\dots,n_d\geq 0\\ n_1+\cdots+n_d=n}} \mathcal{H}(n_1,\dots,n_d) \ .$$

Proof. — We argue as in the proof of Theorem 7.4, using Corollary 6.10 and Theorem 7.8. Note that the isomorphism of [19] is compatible with the semi-direct product, since the involution ι (or the element ψ_0) of $V_n^{\Lambda}(\Gamma)$ is sent to the involution η (or the element g_0) of $H^{\Lambda}(B_n)$ by the isomorphism of [19]. □

We obtain the following corollary. We note that the situation is a little bit more intricate than for type B because of the presence of semi-direct products with products of groups C_2. So below, it is implicit that it is enough to consider some special cyclotomic quotients, up to the application of standard Clifford theory to deal with the semi-direct products.

Corollary 7.11. — *To study an arbitrary cyclotomic quotient of the affine Hecke algebra* $\widehat{H}(D_n)$, *it is enough to consider cyclotomic quotients given by a relation*

$$\prod_{\substack{\epsilon\in\{\pm 1\}\\ l\in\mathbb{Z}}} (X_1 - x^{\epsilon}q^{2l})^{m_{\epsilon,l}} = 0 \ ,$$

for any finitely supported family of non-negative integers $(m_{\epsilon,l})_{\epsilon\in\{\pm1\},l\in\mathbb{Z}}$, *where* $x\in K^{\times}$ *satisfies one of the following three cases:*

$$\textbf{(a)}\ x=1 \quad \textbf{(b)}\ x=q \quad \textbf{(c)}\ x\notin \pm q^{\mathbb{Z}} \ .$$

Proof. — We sketch a proof in the same spirit as in the introduction of [19]. We deduce from Theorem 7.10 that it suffices to study the cyclotomic quotients of $\widehat{H}(D_n)$ given by a relation

$$\prod_{i\in I_x} (X_1 - i)^{\Lambda_i},$$

where I_x and Λ are as in the proof of Corollary 7.5. By Theorem 7.8 this cyclotomic quotient is only determined by the quiver Γ and its involution θ as defined in the proof of Corollary 7.5. In particular, looking at the number of connected components of Γ we still have the two cases $x\in\pm q^{\mathbb{Z}}$ (which give

cases (**a**) and (**b**)) and $x \notin \pm q^{\mathbb{Z}}$ (which is case (**c**)). In the latter case, all the choices of x lead to isomorphic algebras, since θ has no fixed points. □

REMARK 7.12. — We make an additional final remark on the three cases (**a**)–(**c**) to be considered, similarly to Remark 7.6. Cases (**a**) and (**b**) correspond to a quiver with a single connected component (an infinite oriented line or a finite oriented polygon, depending on whether or not q is a root of unity), while (**c**) corresponds to a quiver with two identical connected components exchanged by the involution θ. Case (**a**) corresponds to θ having a fixed point, while Case (**b**) generically corresponds to the situation where there is no fixed point. As before, when q^2 is an odd root of unity, Case (**b**) is not necessary, since it is equivalent to Case (**a**).

Appendix A. Polynomial realisation

Here, we prove Lemma 4.8. In this Appendix, for any $f \in K[x,\beta]$ we also systematically write f for the element of $\mathrm{End}_K(K[x,\beta])$ given by left multiplication and use concatenation to denote the composition in $\mathrm{End}_K(K[x,\beta])$. In particular, for any $w \in B_n$ and $f \in K[x]$, we have $wf = ({}^w f)w$ inside $\mathrm{End}_K(K[x,\beta])$.

We now define some elements of $\mathrm{End}_K(K[x,\beta])$ by

$$\begin{aligned}
\varphi(e(\boldsymbol{i})) &= \mathbf{1}_{\boldsymbol{i}},\\
\varphi(y_a e(\boldsymbol{i})) &= x_a \mathbf{1}_{\boldsymbol{i}},\\
\varphi(\psi_b e(\boldsymbol{i})) &= \big(\delta_{i_b,i_{b+1}}(x_b - x_{b+1})^{-1}(r_b - 1) + P_{i_b,i_{b+1}}(x_{b+1},x_b)r_b\big)\mathbf{1}_{\boldsymbol{i}},\\
\varphi(\psi_0 e(\boldsymbol{i})) &= \big(\gamma_{i_1} x_1^{-1}(1 - r_0) + \alpha_{i_1}(x_1)r_0\big)\mathbf{1}_{\boldsymbol{i}},
\end{aligned} \tag{88}$$

for any $a \in \{1,\dots,n\}$ and $b \in \{1,\dots,n-1\}$, and extend these formulas to $\varphi(X)$ for $X \in \{y_1,\dots,y_n,\psi_0,\dots,\psi_{n-1}\}$ by $\varphi(X) = \sum_{\boldsymbol{i}\in\beta} \varphi(Xe(\boldsymbol{i}))$.

We will prove that φ extends to an algebra homomorphism $\varphi : V_\beta(\Gamma,\lambda,\gamma) \to \mathrm{End}_K(K[x,\beta])$, which will imply Lemma 4.8. Indeed, the map φ is the homomorphism associated with the action defined in §4.2. To prove that φ extends to an algebra homomorphism, we check the defining relations of $V_\beta(\Gamma,\lambda,\gamma)$. Recall that $P_{i,j} = 0$ when $i = j$, so that

$$\varphi(\psi_b e(\boldsymbol{i})) = \begin{cases} (x_b - x_{b+1})^{-1}(r_b - 1)\mathbf{1}_{\boldsymbol{i}}, & \text{if } i_b = i_{b+1},\\ P_{i_b,i_{b+1}}(x_{b+1},x_b)r_b\mathbf{1}_{\boldsymbol{i}}, & \text{otherwise.}\end{cases}$$

Moreover, by (31a) and (44) we have

$$\varphi(\psi_0 e(\boldsymbol{i})) = \begin{cases} \alpha_{i_1}(x_1)r_0\mathbf{1}_{\boldsymbol{i}}, & \text{if } \gamma_{i_1} = 0,\\ \gamma_{i_1}x_1^{-1}(1 - r_0)\mathbf{1}_{\boldsymbol{i}}, & \text{otherwise.}\end{cases}$$

The relations that do not involve ψ_0 are satisfied, since the action is the same as in [22, Proposition 3.12]. Relations (33), (34) and (36) follow immediately.

To simplify the notation, for any $v \in V_\beta(\Gamma, \lambda, \gamma)$, we also write v' instead of $\varphi(v)$. Note that the composition operation in $\mathrm{End}_K(K[x,\beta])$ is denoted as a simple multiplication. For example, $\psi_0' x_1$ means composition of the multiplication by x_1 with the operator $\phi(\psi_0)$. Concerning (35), we have

$$\begin{aligned}(\psi_0'y_1' + y_1'\psi_0')e(\boldsymbol{i})' &= \psi_0' x_1 \mathbf{1}_{\boldsymbol{i}} + x_1\big(\gamma_{i_1} x_1^{-1}(1-r_0) + \alpha_{i_1}(x_1) r_0\big)\mathbf{1}_{\boldsymbol{i}} \\ &= \big[\big(\gamma_{i_1} x_1^{-1}(1-r_0)x_1 + \alpha_{i_1}(x_1) r_0 x_1\big) \\ &\qquad + \big(\gamma_{i_1}(1-r_0) + x_1\alpha_{i_1}(x_1) r_0\big)\big]\mathbf{1}_{\boldsymbol{i}} \\ &= \big[\gamma_{i_1}(1+r_0) - x_1\alpha_{i_1}(x_1) r_0 + \gamma_{i_1}(1-r_0) + x_1\alpha_{i_1}(x_1) r_0\big]\mathbf{1}_{\boldsymbol{i}} \\ &= 2\gamma_{i_1}\mathbf{1}_{\boldsymbol{i}} = \varphi\big(2\gamma_{i_1} e(\boldsymbol{i})\big)\ .\end{aligned}$$

For (37), if $\gamma_i = 0$, then $\gamma_{\theta(i)} = 0$ by (32), and we have, noting that $\mathbf{1}_{\boldsymbol{j}} r_0 = r_0 \mathbf{1}_{r_0\cdot \boldsymbol{j}}$ inside $\mathrm{End}_K(K[x,\beta])$,

$$\begin{aligned}{\psi_0'}^2 e(\boldsymbol{i})' &= \psi_0'\alpha_{i_1}(x_1) r_0 \mathbf{1}_{\boldsymbol{i}} \\ &= \alpha_{\theta(i_1)}(x_1) r_0 \alpha_{i_1}(x_1) r_0 \mathbf{1}_{\boldsymbol{i}} \\ &= \alpha_{\theta(i_1)}(x_1)\alpha_{i_1}(-x_1)\mathbf{1}_{\boldsymbol{i}} \\ &= (-1)^{\lambda_{\theta(i_1)}} x_1^{d(i_1)}\mathbf{1}_{\boldsymbol{i}} \\ &= \varphi\big((-1)^{\lambda_{\theta(i_1)}} y_1^{d(i_1)} e(\boldsymbol{i})\big)\,,\end{aligned}$$

by (43), and if $\gamma_{i_1} \neq 0$, then $\gamma_{\theta(i_1)} \neq 0$, and we have

$$\begin{aligned}{\psi_0'}^2 e(\boldsymbol{i})' &= \psi_0'\gamma_{i_1} x_1^{-1}(1-r_0)\mathbf{1}_{\boldsymbol{i}} \\ &= \gamma_{\theta(i_1)}\gamma_{i_1}\left(x_1^{-1}(1-r_0)\right)^2 \mathbf{1}_{\boldsymbol{i}} \\ &= 0.\end{aligned}$$

It remains to check (38). As in (40), we write $i_1 i_2$ and even 12, instead of $\boldsymbol{i}$, and $\bar{a}$ instead of $\theta(i_a)$. We have, using (33),

$$(\psi_0'\psi_1')^2 e(12)' = (\psi_0'\mathbf{1}_{1\bar{2}})(\psi_1'\mathbf{1}_{\bar{2}1})(\psi_0'\mathbf{1}_{21})(\psi_1'\mathbf{1}_{12}), \tag{89a}$$

$$(\psi_1'\psi_0')^2 e(12)' = (\psi_1'\mathbf{1}_{\bar{2}\bar{1}})(\psi_0'\mathbf{1}_{2\bar{1}})(\psi_1'\mathbf{1}_{\bar{1}2})(\psi_0'\mathbf{1}_{12}). \tag{89b}$$

A.1. Case $\boldsymbol{\gamma_{i_1} = 0 = \gamma_{i_2}}$. — First, recall that by (32) we know that if $\gamma_{i_1} = 0$ and $\theta(i_1) = i_2$, then $\gamma_{i_2} = 0$. Thus, we want to prove that

$$\big((\psi_0'\psi_1')^2 - (\psi_1'\psi_0')^2\big)\mathbf{1}_{\boldsymbol{i}} = \begin{cases}(-1)^{\lambda_{\theta(i_1)}}\frac{(-y_1')^{d(i_1)} - {y_2'}^{d(i_1)}}{y_1' + y_2'}\psi_1'\mathbf{1}_{\boldsymbol{i}}, & \text{if } \theta(i_1) = i_2, \\ 0, & \text{otherwise.}\end{cases} \tag{90}$$

Since $\gamma_{i_1} = \gamma_{\theta(i_1)} = \gamma_{i_2} = \gamma_{\theta(i_2)} = 0$, for any $a, b \in \{1, 2, \bar{1}, \bar{2}\}$ the element ψ_0 acts on $\mathbf{1}_{ab}$ as $\alpha_a(x_1) r_0$.

Assume that $\theta(i_1) = i_1$ and $\theta(i_2) = i_2$. By (31b) we have $d(i_1) = d(i_2) = 0$, and, thus, (90) becomes

$$\big((\psi_0'\psi_1')^2 - (\psi_1'\psi_0')^2\big)\,\mathbf{1}_{\boldsymbol{i}} = 0. \tag{91}$$

Since $d(i_1) = d(i_2) = 0$, by (43) we can assume $\alpha_{i_1}(y) = \alpha_{i_2}(y) = 1$, and, thus, ψ_0 acts on $\mathbf{1}_{ab}$, as r_0 for any a, b. Hence, the same calculation as in [19, §3.1] proves that (91) is satisfied. In the opposite case, if $\theta(i_1) \neq i_1$ and $\theta(i_2) \neq i_2$, we know by the proof of [24, Proposition 7.4] that (90) holds.

Thus, we now assume that $\theta(i_1) = i_1$ and $\theta(i_2) \neq i_2$; in particular, $i_1 \neq i_2$ and $\theta(i_1) \neq i_2$. As above, we have $d(i_1) = 0$, and, thus, ψ_0 acts on $\mathbf{1}_{1a}$ as r_0. We obtain from (89), omitting the idempotents,

$$\begin{aligned}(\psi_0'\psi_1')^2 &= r_0 P_{\bar{2}\bar{1}}(x_2, x_1) r_1 \alpha_2(x_1) r_0 P_{\bar{1}2}(x_2, x_1) r_1 \\ &= P_{\bar{2}\bar{1}}(x_2, -x_1) r_0 \alpha_2(x_2) r_1 r_0 P_{\bar{1}2}(x_2, x_1) r_1 \\ &= P_{\bar{2}\bar{1}}(x_2, -x_1) \alpha_2(x_2) P_{\bar{1}2}(-x_1, -x_2) r_0 r_1 r_0 r_1,\end{aligned}$$

and

$$\begin{aligned}(\psi_1'\psi_0')^2 &= P_{\bar{2}\bar{1}}(x_2, x_1) r_1 \alpha_2(x_1) r_0 P_{\bar{1}\bar{2}}(x_2, x_1) r_1 r_0 \\ &= P_{\bar{2}\bar{1}}(x_2, x_1) \alpha_2(x_2) r_1 r_0 P_{\bar{1}\bar{2}}(x_2, x_1) r_1 r_0 \\ &= P_{\bar{2}\bar{1}}(x_2, x_1) \alpha_2(x_2) P_{\bar{1}\bar{2}}(x_1, -x_2) r_1 r_0 r_1 r_0,\end{aligned}$$

and, thus, $(\psi_0'\psi_1')^2 = (\psi_1'\psi_0')^2$ as desired, where we used $\bar{1} = 1$ and (41). The case $\theta(i_1) \neq i_1$ and $\theta(i_2) = i_2$ is similar.

Remark A.1. — (See Remark 4.2.) Without condition (31b), we have to choose another, more complicated, relation (38), if we want it to be compatible with the action on polynomials.

A.2. Case $\boldsymbol{\gamma_{i_1} = 0 \neq \gamma_{i_2}}$. — We want to prove that

$$\big((\psi_0'\psi_1')^2 - (\psi_1'\psi_0')^2\big)\,\mathbf{1}_{\boldsymbol{i}} = \gamma_{i_2} \frac{Q_{i_2 i_1}(y_1', -y_2') - Q_{i_2 i_1}(y_1', y_2')}{y_2'} \psi_0' \mathbf{1}_{\boldsymbol{i}},$$

that is,

$$\big((\psi_0'\psi_1')^2 - (\psi_1'\psi_0')^2\big)\,\mathbf{1}_{\boldsymbol{i}} = \gamma_{i_2} \frac{Q_{i_2 i_1}(x_1, -x_2) - Q_{i_2 i_1}(x_1, x_2)}{x_2} \alpha_{i_1}(x_1) r_0 \mathbf{1}_{\boldsymbol{i}}.$$

By (31a) we have $\theta(i_2) = i_2$. Note that $\gamma_{i_1} = 0 \neq \gamma_{i_2}$ implies $i_1 \neq i_2$. By (89) we have, omitting the idempotents,

$$\begin{aligned}(\psi_0'\psi_1')^2 &= \alpha_1(x_1) r_0 P_{\bar{2}1}(x_2, x_1) r_1 \gamma_2 x_1^{-1} (1 - r_0) P_{12}(x_2, x_1) r_1 \\ &= \alpha_1(x_1) P_{\bar{2}1}(x_2, -x_1) \gamma_2 x_2^{-1} r_0 r_1 (1 - r_0) P_{12}(x_2, x_1) r_1 \\ &= \alpha_1(x_1) P_{\bar{2}1}(x_2, -x_1) \gamma_2 x_2^{-1} \big[P_{12}(-x_1, x_2) r_0 r_1 - P_{12}(-x_1, -x_2) r_0 r_1 r_0\big] r_1 \\ &= \gamma_2 x_2^{-1} \alpha_1(x_1) P_{\bar{2}1}(x_2, -x_1) \big[P_{12}(-x_1, x_2) r_0 r_1 - P_{12}(-x_1, -x_2) r_0 r_1 r_0\big] r_1,\end{aligned}$$

and

$$\begin{aligned}(\psi_1'\psi_0')^2 &= P_{\bar{2}\bar{1}}(x_2,x_1)r_1\gamma_2x_1^{-1}(1-r_0)P_{\bar{1}2}(x_2,x_1)r_1\alpha_1(x_1)r_0\\ &= P_{\bar{2}\bar{1}}(x_2,x_1)\gamma_2x_2^{-1}r_1(1-r_0)P_{\bar{1}2}(x_2,x_1)\alpha_1(x_2)r_1r_0\\ &= P_{\bar{2}\bar{1}}(x_2,x_1)\gamma_2x_2^{-1}\big[P_{\bar{1}2}(x_1,x_2)r_1-P_{\bar{1}2}(x_1,-x_2)r_1r_0\big]\alpha_1(x_2)r_1r_0\\ &= P_{\bar{2}\bar{1}}(x_2,x_1)\gamma_2x_2^{-1}\alpha_1(x_1)\big[P_{\bar{1}2}(x_1,x_2)r_1-P_{\bar{1}2}(x_1,-x_2)r_1r_0\big]r_1r_0\\ &= \gamma_2x_2^{-1}\alpha_1(x_1)P_{\bar{2}\bar{1}}(x_2,x_1)\big[P_{\bar{1}2}(x_1,x_2)-P_{\bar{1}2}(x_1,-x_2)r_1r_0r_1\big]r_0.\end{aligned}$$

Thus, recalling $\bar{2}=2$ and using the properties (9), (30), (41) and (42) for the families P and Q, we have

$$\begin{aligned}(\psi_0'\psi_1')^2-(\psi_1'\psi_0')^2 &= \gamma_2x_2^{-1}\alpha_1(x_1)\big[P_{\bar{2}1}(x_2,-x_1)P_{12}(-x_1,x_2)\\ &\qquad -P_{\bar{2}\bar{1}}(x_2,x_1)P_{\bar{1}2}(x_1,x_2)\big]r_0\\ &= \gamma_2x_2^{-1}\big[Q_{21}(x_2,-x_1)-Q_{2\bar{1}}(x_2,x_1)\big]\alpha_1(x_1)r_0\\ &= \gamma_2x_2^{-1}\big[Q_{21}(x_1,-x_2)-Q_{21}(x_1,x_2)\big]\alpha_1(x_1)r_0,\end{aligned}$$

as desired.

A.3. Case $\gamma_{i_1}\neq 0=\gamma_{i_2}$. — We want to prove that

$$\big((\psi_0'\psi_1')^2-(\psi_1'\psi_0')^2\big)\,\mathbf{1}_{\boldsymbol{i}}=0.$$

Similarly to §A.2 we have $\theta(i_1)=i_1\neq i_2$. By (89) we have, omitting the idempotents,

$$\begin{aligned}(\psi_0'\psi_1')^2 &= \gamma_1x_1^{-1}(1-r_0)P_{\bar{2}1}(x_2,x_1)r_1\alpha_2(x_1)r_0P_{12}(x_2,x_1)r_1\\ &= \gamma_1x_1^{-1}\left[P_{\bar{2}1}(x_2,x_1)-P_{\bar{2}1}(x_2,-x_1)r_0\right]\alpha_2(x_2)P_{12}(x_1,-x_2)r_1r_0r_1\\ &= \gamma_1x_1^{-1}\alpha_2(x_2)\big[P_{\bar{2}1}(x_2,x_1)P_{12}(x_1,-x_2)\\ &\qquad -P_{\bar{2}1}(x_2,-x_1)P_{12}(-x_1,-x_2)r_0\big]r_1r_0r_1\\ &= \gamma_1x_1^{-1}\alpha_2(x_2)P_{\bar{2}1}(x_2,x_1)P_{12}(x_1,-x_2)(1-r_0)r_1r_0r_1\end{aligned}$$

by (41), and

$$\begin{aligned}(\psi_1'\psi_0')^2 &= P_{\bar{2}1}(x_2,x_1)r_1\alpha_2(x_1)r_0P_{12}(x_2,x_1)r_1\gamma_1x_1^{-1}(1-r_0)\\ &= \gamma_1x_1^{-1}\alpha_2(x_2)P_{\bar{2}1}(x_2,x_1)P_{12}(x_1,-x_2)r_1r_0r_1(1-r_0),\end{aligned}$$

Thus $(\psi_0'\psi_1')^2=(\psi_1'\psi_0')^2$ as desired.

A.4. Case $\gamma_{i_1} \neq 0 \neq \gamma_{i_2}$. — We want to prove that (recalling from (8) that $Q_{ii} = 0$)

$$\begin{aligned}&\big((\psi_0'\psi_1')^2 - (\psi_1'\psi_0')^2\big)\,\mathbf{1_i}\\ &\quad= \begin{cases}\gamma_{i_2}\dfrac{Q_{i_2i_1}(y_1', -y_2') - Q_{i_2i_1}(y_1', y_2')}{y_1'y_2'}\,(y_1'\psi_0' - \gamma_{i_1})\,\mathbf{1_i}, & \text{if } i_1 \neq i_2,\\ 0, & \text{otherwise,}\end{cases}\end{aligned}$$

that is, since ψ_0 acts on $\mathbf{1_i}$ as $\gamma_{i_1}x_1^{-1}(1-r_0)$ (recalling that $\theta(i_1) = i_1$ by (31a)),

$$\begin{aligned}&\big((\psi_0'\psi_1')^2 - (\psi_1'\psi_0')^2\big)\,\mathbf{1_i}\\ &\quad= \begin{cases}\gamma_{i_1}\gamma_{i_2}\dfrac{Q_{i_2i_1}(x_1, x_2) - Q_{i_2i_1}(x_1, -x_2)}{x_1x_2}r_0\mathbf{1_i}, & \text{if } i_1 \neq i_2,\\ 0, & \text{otherwise.}\end{cases}\end{aligned}$$

The next result is an easy calculation.

LEMMA A.2. — *Let P be a polynomial in x_1, x_2 and let $w \in \langle r_0, r_1\rangle$. Then*

$$\begin{aligned}&x_1^{-1}(1-r_0)Pw - Pwx_1^{-1}(1-r_0)\\ &\quad= \big(x_1^{-1} - {}^w x_1^{-1}\big)\,Pw + {}^w x_1^{-1}Pwr_0 - x_1^{-1\,r_0}Pr_0w\,,\end{aligned}$$

in $\mathrm{End}_K(K[x,\beta])$.

By (31a) we have $\theta(i_2) = i_2$. If $i_1 \neq i_2$, we obtain from (89)

$$\begin{aligned}\psi_1'\psi_0'\psi_1'\mathbf{1}_{12} &= P_{21}(x_2,x_1)r_1\gamma_2x_1^{-1}(1-r_0)P_{12}(x_2,x_1)r_1\mathbf{1}_{12}\\ &= \gamma_2x_2^{-1}P_{21}(x_2,x_1)r_1(1-r_0)P_{12}(x_2,x_1)r_1\mathbf{1}_{12}\\ &= \gamma_2x_2^{-1}P_{21}(x_2,x_1)r_1\big[P_{12}(x_2,x_1) - P_{12}(x_2,-x_1)r_0\big]r_1\mathbf{1}_{12}\\ &= \gamma_2x_2^{-1}P_{21}(x_2,x_1)\big[P_{12}(x_1,x_2)r_1 - P_{12}(x_1,-x_2)r_1r_0\big]r_1\mathbf{1}_{12}\\ &= \gamma_2x_2^{-1}P_{21}(x_2,x_1)\big[P_{12}(x_1,x_2) - P_{12}(x_1,-x_2)r_1r_0r_1\big]\mathbf{1}_{12}.\end{aligned}$$

Since $\psi_0'\mathbf{1}_{12} = \gamma_1x_1^{-1}(1-r_0)\mathbf{1}_{12}$, we can apply Lemma A.2 for the above two summands. We obtain that the second summand will vanish in $\big((\psi_0'\psi_1')^2 - (\psi_1'\psi_0')^2\big)\mathbf{1}_{12}$, since $x_1^{-1} \in K(x_1)$ is invariant under $r_1r_0r_1$, and $P_{21}(x_2,x_1)P_{12}(x_1,-x_2) \in K[x_1,x_2]$ is invariant under r_0 by (41). Thus, we only consider the first summand, which is equal to $\gamma_2x_2^{-1}Q_{21}(x_2,x_1)$, and we obtain, omitting the idempotents and using (9) and (30),

$$\begin{aligned}(\psi_0'\psi_1')^2 - (\psi_1'\psi_0')^2 &= \gamma_1\gamma_2x_1^{-1}x_2^{-1}\big[Q_{21}(x_2,x_1) - Q_{21}(x_2,-x_1)\big]r_0\\ &= \gamma_1\gamma_2x_1^{-1}x_2^{-1}\big[Q_{21}(x_1,x_2) - Q_{21}(x_1,-x_2)\big]r_0,\end{aligned}$$

as desired.

Finally, assume that $i_1 = i_2$. We have

$$\begin{aligned}(\psi_0'\psi_1')^2 - (\psi_0'\psi_1')^2 \\ = \gamma_1^2\big[x_1^{-1}(1-r_0)(x_1-x_2)^{-1}(r_1-1)x_1^{-1}(1-r_0)(x_1-x_2)^{-1} \\ - (x_1-x_2)^{-1}(r_1-1)x_1^{-1}(1-r_0)(x_1-x_2)^{-1}x_1^{-1}(1-r_0)\big] = 0,\end{aligned}$$

since this is just the braid relation for the divided difference operators $\partial_0 \coloneqq x_1^{-1}(1-r_0)$ and $\partial_1 \coloneqq (x_1-x_2)^{-1}(r_1-1)$ (see [3, 7]).

Acknowledgements. — The authors would like to thank Ruari Walker for many interesting discussions initiating this work. The second author would like to thank Ruslan Maksimau for explaining a proof of Proposition 3.3. The authors are very grateful to an anonymous referee for many useful suggestions.

BIBLIOGRAPHY

[1] S. Ariki – "On the decomposition numbers of the Hecke algebra of $G(m,1,n)$", *J. Math. Kyoto Univ.* **36** (1996), p. 789–808.

[2] S. Ariki & K. Koike – "A Hecke algebra of $\mathbb{Z}/n\mathbb{Z} \wr \mathfrak{S}_n$ and construction of its irreducible representations", *Adv. Math.* **106** (1994), p. 216–243.

[3] I. N. Bernstein, I. M. Gel'fand & I. S. Gel'fand – "Schubert cells and cohomology of the spaces G/P", *Russian Mathematical Surveys* **28** (1973), p. 1–26.

[4] A. Björner & F. Brenti – *Combinatorics of Coxeter groups*, Graduate Texts in Mathematics, no. 231, Springer-Verlag, 2005.

[5] M. Broué & G. Malle – "Zyklotomische Heckealgebren", in *Représentations unipotentes génériques et blocs des groupes réductifs finis*, Astérisque, no. 212, 1993.

[6] J. Brundan & A. Kleshchev – "Blocks of cyclotomic Hecke algebras and Khovanov–Lauda algebras", *Invent. Math.* **178** (2009), no. 3, p. 451–484.

[7] M. Demazure – "Invariants symétriques entiers des groupes de Weyl et torsion", *Invent. Math.* **21** (1973), p. 287–301.

[8] R. Dipper & A. Mathas – "Morita equivalences of Ariki–Koike algebras", *Math. Z.* **240** (2002), p. 579–610.

[9] N. Enomoto & M. Kashiwara – "Symmetric crystals and affine Hecke algebras of type B", *Proc. Japan Acad. Ser. A* **82** (2006), no. 8, p. 131–136.

[10] M. Geck & G. Pfeiffer – *Characters of finite Coxeter groups and Iwahori–Hecke algebras*, London Math. Soc. Monographs, New Series, no. 21, Oxford University Press, New York, 2000.

[11] J. Hu & A. Mathas – "Morita equivalences of cyclotomic Hecke algebras of type $G(r,p,n)$", *J. Reine. Angew Math.* **628** (2009), p. 169–194.

[12] J. Hu & K. Zhou – "On Dipper–Mathas's Morita equivalences", *Colloquium Mathematicum* **149** (2017), p. 103–123.
[13] N. Jacon & L. Poulain d'Andecy – "An isomorphism theorem for Yokonuma–Hecke algebras and applications to link invariants", *Math. Z.* **283** (2016), no. 1-2, p. 301–338.
[14] M. Kashiwara & V. Miemietz – "Crystals and affine Hecke algebras of type D", *Proc. Japan Acad. Ser. A Math. Sci.* **83** (2007), no. 7, p. 135–139.
[15] M. Khovanov & A. D. Lauda – "A diagrammatic approach to categorification of quantum groups I", *Represent. Theory* **13** (2009), p. 309–347.
[16] ———, "A diagrammatic approach to categorification of quantum groups II", *Trans. Amer. Math. Soc.* **363** (2011), p. 2685–2700.
[17] O. Ogievetsky & L. Poulain d'Andecy – "Alternating subgroups of Coxeter groups and their spinor extensions", *J. Pure Appl. Algebra* **217** (2013), no. 11, p. 2198–2211.
[18] L. Poulain d'Andecy & R. Walker – "Affine Hecke algebras and generalisations of quiver Hecke algebras for type B", *Proc. Edinburgh Math. Soc.* **63** (2020), no. 2, p. 531–578.
[19] ———, "Affine Hecke algebras of type D and generalisations of quiver Hecke algebras", *J. Algebra* **552** (2020), p. 1–37.
[20] S. Rostam – "Cyclotomic Yokonuma–Hecke algebras and cyclotomic quiver Hecke algebras", *Adv. Math.* **311** (2017), p. 662–729.
[21] ———, "Cyclotomic quiver Hecke algebras and Hecke algebra of $G(r,p,n)$", *Trans. Amer. Math. Soc.* **371** (2019), p. 3877–3916.
[22] R. Rouquier – "2-Kac–Moody algebras", arXiv:0812.5023.
[23] P. Shan, M. Varagnolo & E. Vasserot – "Canonical bases and affine Hecke algebras of type D", *Adv. Math.* **1** (2011), no. 227, p. 267–291.
[24] M. Varagnolo & E. Vasserot – "Canonical bases and affine Hecke algebras of type B", *Invent. Math.* **183** (2011), no. 3, p. 593–693.

Bull. Soc. Math. France
149 (1), 2021, p. 235

ERRATUM ON THE PAPER *NON-COMPACT FORM OF THE ELEMENTARY DISCRETE INVARIANT*

BY RAPHAËL FINO

The statements of Section 4 *Steenrod operations* (except for Lemma 4.1 and the indirect part of Corollary 4.2) are untrue or unproven (this does not affect the main result of the paper).

The reason is as follows. The homomorphism

$$S^l \times \mathrm{Id}^{\times i} : \mathrm{Ch}(X_K^{i+1}) \to \mathrm{Ch}(X_K^{i+1})$$

(which would be better denoted by $S^l \otimes \mathrm{Id}^{\otimes i}$), used to define the cycle $\rho_{i,j,l} \in \mathrm{Ch}(X_K^i)$ at the beginning of Section 4.1, does not exist in general at the level of the Chow group $\mathrm{Ch}(X^{i+1})$. Indeed, over the base field F, one does not have $\mathrm{Ch}(X^{i+1}) \simeq \mathrm{Ch}(X)^{\otimes i+1}$ in general (whereas it becomes the case when passing to a splitting field K since the variety X_K is cellular).

Thus, the rationality of the cycle $\rho_{i,j}$ may not necessarily imply the rationality of the cycle $\rho_{i,j,l}$ (as wrongly suggested in the first sentence of the erroneous proof of Corollary 4.2). Therefore, the direct part of Corollary 4.2 and Propositon 4.4 are untrue or unproven (hence so are Examples 4.3 and 4.5). Proposition 4.6 is also untrue or unproven (hence so are Example 4.7 and Remark 4.8) since the same mistake has been made at the beginning of its erroneous proof: for the aforementioned reason, the rationality of the cycle $\rho_{i,j}$ and identity (12) may not necessarily imply the rationality of (13).

As a consequence, we do not obtain new restrictions on the possible values of the elementary discrete invariant.

The rest of the paper is totally independent from Section 4.

Texte reçu le 2 mai 2020, modifié le 10 juillet 2020, accepté le 10 octobre 2020.

RAPHAËL FINO, Instituto de Matemáticas, Ciudad Universitaria, UNAM, DF 04510, México
• *E-mail :* `fino@im.unam.mx` • *Url :* `http://www.matem.unam.mx/fino`

0037-9484/2021/235/$ 5.00
doi:10.24033/bsmf.2829