

Max-Planck-Institut  
für Mathematik  
in den Naturwissenschaften  
Leipzig

Phylogenetic complexity of the  
3-Kimura model

by

*Mateusz Michałek and Emanuele Ventura*

Preprint no.: 27

2017





# Phylogenetic complexity of the Kimura 3-parameter model

Mateusz Michałek and Emanuele Ventura

## Abstract

In algebraic statistics, the Kimura 3-parameter model is one of the most interesting and classical phylogenetic models. We prove that the ideals associated to this model are generated in degree four, confirming a conjecture by Sturmfels and Sullivant.

2010 *Mathematics Subject Classification*. Primary 52B20, Secondary 14M25, 13P25

## 1 Introduction

The part of computational biology that models evolution and describes mutations in this process is called *phylogenetics* [40]. This is a fertile subject witnessing many connections to several parts of mathematics such as algebraic geometry [8, 23], combinatorics [4, 15, 34], and representation theory [9, 31]. The methods used in this context of research are powerful and do not only apply to biology, but are employed in several other fields [2] such as modeling changes of words in languages [21], literary studies [3] or linguistics itself [37] with ideas going back to Darwin [14].

A crucial object in phylogenetics is a *tree model*, which is a parametric family of probability distributions. It consists of a tree  $\mathcal{T}$ , a finite set of states  $S$  and a family  $\mathcal{M}$  of transition matrices, usually given by a linear subspace of all  $|S| \times |S|$  matrices. The case of particular interest is when  $S = \{A, C, G, T\}$ , where the basis elements correspond to the four nucleobases of DNA: adenine (A), cytosine (C), guanine (G), and thymine (T).

The models for which  $\mathcal{M}$  is a proper subspace of matrices reflect some symmetries among elements of  $S$ . These symmetries are usually encoded by the action of a finite group  $G$  on  $S$ . In these terms,  $\mathcal{M}$  can be regarded as the space of  $G$ -invariant matrices or tensors. Such models constitute a class of interest and they are called *equivariant* [18]. If  $G$  is the trivial group, we obtain the general Markov model, corresponding, on the algebraic geometry side, to secant varieties of Segre products. When the elements of  $S$  can be identified with those of  $G$ , the model is called *group-based*. Henceforth we assume  $G$  to be abelian.

The simplest among the equivariant, and group-based, models is the *Cavender-Farris-Neyman* model. This is the instance for  $S = G = \mathbb{Z}_2$ , the group with two elements. A good understanding of this model from the algebraic geometry point of view has led to tremendous advances in this field. Sturmfels and Sullivant [41, Theorem 28] showed that the algebraic varieties arising from it are defined by quadrics. Additionally, Buczyńska and Wiśniewski described many of its remarkable algebro-geometric properties [8]. Consequently, Sturmfels and Xu [44], and Manon [32] described the connections of the model to toric degenerations of moduli spaces of rank two vector bundles on marked curves of fixed genus. For more relations to conformal field theory, we refer to [29, 31].

The Cavender-Farris-Neyman model is the simplest among the *hyperbinary* models [6, Section

3], that are given by  $S = G = (\mathbb{Z}_2)^n$ . The most biologically meaningful example of those is the *Kimura 3-parameter* model; this corresponds to  $n = 2$ . In this case,  $S = \{A, C, G, T\}$ , and, moreover, the action of  $G$  reflects the pairing between purines (A, G) and pyrimidines (C, T). This model was introduced by Kimura [28] much before the setting above was developed. Using numerical experiments, Sturmfels and Sullivant conjectured that the ideals of the algebraic varieties associated to this model are generated by polynomials of degree at most four [41, Conjecture 30]. The confirmation of this conjecture is the main result of the present article. For any group  $G$ , Sturmfels and Sullivant defined the *phylogenetic complexity*  $\phi(G)$  of  $G$ .

**Definition 1.1 (Phylogenetic complexity [41]).** *Let  $K_{1,n}$  be the star with  $n$  leaves, and  $X(G, K_{1,n})$  the variety associated to the group-based model. Let  $\phi(G, K_{1,n})$  be the maximal degree of a generator in a minimal generating set of the ideal  $I(X(G, K_{1,n}))$ . The phylogenetic complexity  $\phi(G)$  of  $G$  is  $\sup_{n \in \mathbb{N}} \{\phi(G, K_{1,n})\}$ .*

In [35], it was shown that for any abelian group  $G$ , its phylogenetic complexity  $\phi(G)$  is finite. The main contribution of this article is a more detailed study of the phylogenetic complexity of  $G = \mathbb{Z}_2 \times \mathbb{Z}_2$ .

**Main Theorem.** *The phylogenetic complexity of the Kimura 3-parameter model  $\phi(\mathbb{Z}_2 \times \mathbb{Z}_2)$  equals four.*

For more interesting results on the Kimura 3-parameter model we refer to [9, 10, 11, 30].

### Algebraic varieties associated to a model.

We recall the explicit construction of the algebraic variety associated to a model. It is the Zariski closure of the locus of all probability distributions on the states of leaves allowed in the model.

A *representation* of a model on a tree  $\mathcal{T}$  is an association  $\mathcal{E} \rightarrow \mathcal{M}$  of transition matrices to edges  $\mathcal{E}$  of  $\mathcal{T}$ . The set of all representations is denoted by  $\text{Rep}(\mathcal{T})$ . (Here we do not mention the root distribution, since it does not affect the family of probability distributions we obtain.) To each vertex  $v$  of  $\mathcal{T}$  we associate an  $|S|$  dimensional vector space  $V_v$  with basis  $(v_s)_{s \in S}$ . We may regard an element of  $\mathcal{M}$  associated to an edge  $(v_1, v_2) = e \in \mathcal{E}$  as an element of the tensor product  $V_{v_1} \otimes V_{v_2}$ . We fix a representation  $M \in \text{Rep}(\mathcal{T})$  and an association  $\mathbf{s} : \mathcal{L} \rightarrow S$ . Here  $\mathcal{L}$  is the set of leaves, i.e. vertices of degree one, of  $\mathcal{T}$ . Following the usual Markov rule, we may compute the probability of  $\mathbf{s}$ :

$$P(M, \mathbf{s}) = \sum_{\substack{f: \mathcal{V} \rightarrow S \\ f|_{\mathcal{L}} = \mathbf{s}}} \prod_{(v_1, v_2) \in \mathcal{E}} \left( M((v_1, v_2)) \right)_{(f(v_1), f(v_2))},$$

where  $\mathcal{V}$  is the set of vertices of  $\mathcal{T}$ . We may identify  $\mathbf{s}$  with a basis element  $\bigotimes_{l \in \mathcal{L}} l_{\mathbf{s}(l)}$  of  $\bigotimes_{l \in \mathcal{L}} V_l$ . This provides the map:

$$\Psi : \text{Rep}(\mathcal{T}) \ni M \rightarrow \sum_{\mathbf{s}: \mathcal{L} \rightarrow S} P(M, \mathbf{s}) \bigotimes_{l \in \mathcal{L}} l_{\mathbf{s}(l)} \in \bigotimes_{l \in \mathcal{L}} V_l.$$

The image of this map is the family of probability distributions described by the model and its Zariski closure is the algebraic variety that represents the model. For group-based models, we denote this variety  $X(G, \mathcal{T})$ , where  $G$  is the group defining the model and  $\mathcal{T}$  is the tree as above.

### Earlier contributions.

Our proof of the main theorem relies on previous results by many authors that we now recall.

The first fundamental tool is the **D**iscrete **F**ourier **T**ransform. This is a linear change of coordinates, based on the representation theory of  $G$ . For special cases in phylogenetics, it was first used by Hendy and Penny [26], and by Erdős, Székely, and Steel [42]. In higher generality, it is treated in [33, 41]. For group-based models, the DFT turns  $\Psi$  into a monomial map, proving that the associated algebraic variety  $X(G, \mathcal{T})$  is a *toric variety*. This translates the classical algebraic problem of finding defining equations of a variety into a combinatorial one. For more information about toric methods we refer to [12, 25, 43].

Another key result is the reduction from arbitrary trees to the so-called stars or claw-trees  $K_{1,n}$ , i.e., trees with one inner vertex and  $n$  leaves. The general procedure for group-based models to obtain ideals arising from arbitrary trees, knowing the ideals for  $K_{1,n}$ , was discovered in [41]. Again, this turned out to be very influential, leading, on one hand, to the general constructions of *toric fiber products* [31, 45], and, on the other, to generalizations for equivariant models [18].

Combinatorial and computational methods in toric geometry are very well developed. As a starting point in our article we need to compute algebraic invariants of toric varieties embedded in very high dimensional ambient spaces. Here the computer algebra packages `Normaliz` [7], `4ti2` [47], along with previous computational results from [16] and [41] are used. In particular, Castenluovo-Mumford regularity plays a crucial role in the proof for  $n = 6$ . These classical invariants are briefly discussed in the Appendix 4, for the sake of completeness.

This work may be also seen in the framework of the *stabilisation* of equations of a family of algebraic varieties. Indeed, our proof not only bounds the degrees of the generators, but in principle provides an inductive procedure to obtain all generators in case of  $K_{1,n+1}$ , assuming the generators for  $K_{1,n}$  to be known. Finding equations of an infinite sequence of algebraic varieties, that come naturally in families, is an interesting current theme of research. This usually involves classical varieties such as secants of Segre varieties [19] and Grassmannians [20]. Indeed, the main result of Draisma and Eggermont in [17] shows that for equivariant models the associated algebraic variety can always be defined set-theoretically in some bounded degree, once  $G$  and  $S$  are both fixed. The fact that  $\phi(G)$  is finite constitutes the main result of [35]. Recently, another ideal-theoretic result was proved by Sam [38] showing that the ideal of  $k$ th secant variety of  $d$ th Veronese embeddings is generated in bounded degree that is independent of  $d$ . Interestingly, the ideal-theoretic generation in bounded degree for secants of Segre varieties and Grassmannians are still central open problems. Finiteness issues are strongly connected with the theory of twisted commutative algebras and  $\Delta$ -modules by Sam and Snowden [39], and the theory of noetherianity by Draisma and Kuttler [19], Hillar and Sullivant [27], and others.

Apart from beautiful results of existence, that are quite often non-constructive or very far from optimal, it is of interest finding an explicit description of phylogenetic algebraic varieties. One of the most well-known examples is the *salmon conjecture* [1], since the prize offered by Allman for the hypothetical solver would be a smoked Copper river salmon. It asks for the description of  $\sigma_4(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3)$ , the algebraic variety representing the general Markov model for  $|S| = 4$  and  $\mathcal{T} = K_{1,3}$ . The generators of the ideal are still unknown, however a set-theoretic description was found by Friedland and Gross [24]. More recently, Daleo and Hauenstein [13] gave a numerical proof of the salmon conjecture.

As far as we know, our result is the only ideal-theoretic description, apart from the Jukes-Cantor model, where  $|S| = 4$  and  $\mathcal{T}$  is an arbitrary tree.

## Plan of the article.

The whole article is devoted solely to the proof of the main theorem. In Section 2 we introduce the notation that is used throughout the proof. As the proof consists of several parts, some of them very technical, we present the overview of its structure in Section 3.1. The main result is established in Sections 3.2 and 3.3.

## 2 Preliminaries and notation

In this section we collect all the notation and terminology we will use in the rest of the paper. We divide this section into paragraphs to facilitate the reading.

### Groups.

Henceforth we set  $G = \mathbb{Z}_2 \times \mathbb{Z}_2$ , unless otherwise stated. We denote the elements of  $G$  by  $0, \alpha, \beta$ , and  $\gamma$ . To denote unknown elements of  $G$ , we use letters  $g, x, y, w, p, q \dots$ . We also refer to an unknown element, that is not relevant in a specific argument, with question mark “?”.

Apart from  $G$ , the most natural groups that enter the picture are the *symmetric group* on  $n$  leaves  $\mathfrak{S}_n$ , the *group of flows*  $\mathfrak{G}$ , and the *automorphism group*  $\text{Aut}(G)$ . The group of flows is the following.

**Definition 2.1 (Group of flows).** *Let  $G$  be an abelian group and  $n \in \mathbb{N}$ . The set of flows  $\mathfrak{G} = \{(g_1, \dots, g_n) \mid \sum g_i = 0\}$  of length  $n$  of  $G$  forms a group under the componentwise group operation. It is non-canonically isomorphic to the group  $G^{n-1}$ , the direct product of  $n - 1$  copies of  $G$ .*

The automorphism group of  $G$ ,  $\text{Aut}(G) \cong \mathfrak{S}_3$ , is the group of bijective group homomorphisms from  $G$  to itself. The automorphism of  $G$  specified by  $\alpha \mapsto \alpha, \beta \mapsto \gamma, \gamma \mapsto \beta$  is simply denoted by  $\beta \leftrightarrow \gamma$ ; similarly for all the other automorphisms of  $G$  having a non-trivial fixed element.

### The toric variety $X(G, K_{1,n})$ .

For any abelian group  $G$ , the variety  $X(G, K_{1,n})$  is a projective toric variety of dimension  $n(|G| - 1)$  living in  $\mathbb{P}^{|G|^{n-1}-1}$ , where the projective coordinates are in bijection with flows [33].

Let us recall here its corresponding polytope. Let  $M \cong \mathbb{Z}^{|G|}$  be the lattice whose basis corresponds to the elements of  $G$ . Consider  $M^n$  with the basis  $e_{(i,g)}$  indexed by pairs  $(i, g) \in [n] \times G$ . We define a map of sets from the group of flows to the lattice,  $\psi : \mathfrak{G} \rightarrow M^n$ , by  $\psi((g_1, \dots, g_n)) = \sum_{i=1}^n e_{(i, g_i)}$ . The vertices of the polytope of  $X(G, K_{1,n})$  are the images of the flows under the injective map  $\psi$ .

**Remark 2.2.** *The family of varieties  $X(G, K_{1,n})$  has a wealth of symmetries; the group  $\mathfrak{S}_n$ , the group of flows  $\mathfrak{G}$ , and the automorphism group  $\text{Aut}(G)$  all act on the ideals of these varieties.*

### Binomials, tables, and moves.

Ideals of toric varieties are binomial prime ideals. Thus they admit a minimal generating set of binomials. Binomials may be identified with a pair of tables of the same size,  $T_0$  and  $T_1$ , of elements of  $G$ , regarded up to row permutation; this is another natural group in this setting which we implicitly take into account. Indeed, a binomial is a pair of monomials and the variables correspond to rows. Given the number of leaves  $n$ , coordinates are in bijection with flows of length  $n$  of  $G$ . Hence rows are identified with flows of  $n$  elements in  $G$ . Columns are in bijection with the  $n$  leaves. From the definition of the toric ideals  $I(X(G, K_{1,n}))$  [41], it follows that a binomial belongs to

$I(X(G, K_{1,n}))$  if and only if the two tables representing it are *compatible*, i.e., for each  $i$ , the  $i$ th column of  $T_0$  and the  $i$ th column of  $T_1$  are equal as multisets. We index the columns of a given pair of tables  $T_0, T_1$ , with  $n$  columns, by integers  $1 \leq i \leq n$ . We refer to the element in the  $i$ th column of row  $r$  as  $r(i)$ .

Let  $T$  be any table of elements of  $G$ . The procedure consisting of selecting a subset of rows in  $T$  of cardinality at most  $d$ , and replacing it with a compatible set of rows is a *move of degree  $d$* . A binomial, represented by a pair of tables  $T_0, T_1$  of elements of  $G$ , is generated by binomials of degree at most  $d$  if and only if there exists a finite sequence of moves of degree  $d$  applied to  $T_0$  or  $T_1$  that transform  $T_0$  into  $T_1$ .

**Example 2.3.** Let  $T$  be the table

$$T = \begin{bmatrix} \alpha & \alpha & 0 & \dots & 0 \\ 0 & \beta & \dots & \dots & \dots \\ \gamma & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

The table  $T$  can be transformed by a move of degree three into the table

$$\tilde{T} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \gamma & \alpha & \dots & \dots & \dots \\ \alpha & \beta & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

Indeed, the set of the first three rows of  $T$  is compatible with the set of the first three rows of  $\tilde{T}$ . Note that if the rows in  $T$  are flows, then the rows of  $\tilde{T}$  are flows as well. The move described above is denoted by

$$\alpha\alpha + 0\beta + \gamma 0 = 00 + \gamma\alpha + \alpha\beta.$$

**Remark 2.4.** In the notation for moves, we do not use the indices of the columns involved in the move. Instead, the indices are always clear from the move itself. For instance, the move in Example 2.3 is in columns 1, 2. Also, note that, in general, the columns used for a move do not need to be consecutive.

**Remark 2.5.** The groups  $\mathfrak{S}_n$ , the group of flows  $\mathfrak{G}$ , and the automorphism group  $\text{Aut}(G)$  act on the equations of  $X(G, K_{1,n})$ , and hence on the tables. The group  $\mathfrak{S}_n$  acts permuting the columns of the pair of tables corresponding to a binomial in the ideal of the variety. The groups  $\mathfrak{G}$  and  $\text{Aut}(G)$  act on the entries of the tables in the natural way, i.e., by evaluation.

We now introduce one of the most important concepts for our approach. Given a pair of flows, we define a *distance* between them, which will enable us to use an inductive procedure on tables. The distance we consider is the classical Hamming distance between two words.

**Definition 2.6 (Hamming distance).** Let  $r_0$  and  $r_1$  be two flows in  $\mathfrak{G}$ :

$$r_0 = (g_1 + a_1, \dots, g_n + a_n) \quad \text{and} \quad r_1 = (g_1, \dots, g_n).$$

Let  $I = \{\ell \in [n] \mid a_\ell \neq 0\}$  and  $J = \{\ell \in [n] \mid a_\ell = 0\}$ . The multiset  $\{a_\ell\}_{\ell \in I}$  constitutes the disagreement string  $a_{\ell_1} \dots a_{\ell_{|I|}}$  of the pair of flows  $r_0$  and  $r_1$ . The cardinality  $|I|$  is the Hamming distance between  $r_0$  and  $r_1$ . The multiset  $\{a_\ell\}_{\ell \in J}$  constitutes their agreement string. Up to the action of the group of flows  $\mathfrak{G}$  on both flows, we may assume that the group elements  $g_i = 0$  for all  $i$ .

**Remark 2.7 (Tables and Hamming distance).** *Given a pair of tables  $T_0, T_1$ , we “compare” them using the notion of Hamming distance as follows. Since the tables come with undistinguishable rows, we may choose as first rows of  $T_0$  and  $T_1$  two rows that minimize the Hamming distance among all the pairs of rows from  $T_0$  and  $T_1$ . After fixing the first row in  $T_0$  and in  $T_1$ , as described in Section 3.1, one of the techniques adopted in Sections 3.2 and 3.3 is as follows. With moves of degree at most four, we create another pair of rows with strictly smaller Hamming distance than the initial one.*

**Counting functions.**

We will make use of counting functions on the tables  $T_0$  and  $T_1$ . A counting function  $f$  on the columns of  $T_0$  has the same values as counting function on the columns of  $T_1$ , since the pairs of tables we are interested in are compatible, i.e., columnwise they are the same as multisets. Given  $x \in G$ , we denote by  $x_{i_1 \dots i_k}$  the number of copies  $x \in G$  appearing in the columns  $i_1, \dots, i_k$  in  $T_0$ , or in  $T_1$ .

**Example 2.8.** *The function  $\alpha_{12} - 2 \cdot 0_3$  counts the number of copies of  $\alpha$  in columns 1 and 2 minus two times the number of copies of 0 in column 3.*

From an algebraic point of view, a counting function defines a *grading* of the variables, that is a specialization of the multi-grading. Thus the fact that the counting function gives the same value on two tables is equivalent to the fact that the two corresponding monomials have the same degree with respect to the induced grading. Additionally, from the perspective of toric geometry, the counting function is induced by restricting the torus action to a special one-parameter subgroup.

**Group homomorphisms.**

We will make use of group homomorphisms in order to do counting arguments in a given pair of tables. We denote

$$\phi_g : \mathbb{Z}_2 \times \mathbb{Z}_2 \rightarrow \mathbb{Z}_2,$$

the group homomorphism given by the quotient map sending each element  $x \in G$  to its class modulo the subgroup generated by the element  $g \in G$ .

### 3 Complexity of the Kimura 3-parameter model

The aim of this section is to establish the phylogenetic complexity of the Kimura 3-parameter model. In Section 3.1, we discuss the structure of the proof, postponing the technical part of it to Sections 3.2 and 3.3.

#### 3.1 Main result and structure of the proof

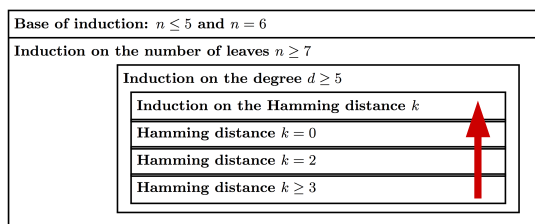
We proceed presenting our main result along with the outline of the plan of the proof strategy.

**Theorem 3.1.** *The phylogenetic complexity of the Kimura 3-parameter model  $\phi(\mathbb{Z}_2 \times \mathbb{Z}_2)$  equals four.*

The structure of the proof is presented in Figure 1. Our proof is an induction on the number of leaves  $n$ , i.e., the number of columns of the tables. The base of our induction is  $n = 3$ . The case



Figure 1: Matryoshka of the proof.



of  $n \leq 5$  leaves has been studied computationally. More precisely, for  $n = 3$  the result is presented in [41] and for  $n = 4$  it is computed in [16]. For  $n = 5$  we used the program featured in [16] to produce the vertices of the polytope. The computer algebra program `4ti2` [47] specialized for toric ideals was able to compute the Markov basis using a server equipped with a CPU 4 Intel-Xeon E7-8837/32 cores/2.67GHz and a memory of 1024Gb RAM.

**Proposition 3.2.** *The ideal  $I(X(G, K_{1,5}))$  is minimally generated by 22240 polynomials of degree at most four: 12960 quadrics, 2560 cubics, and 6720 quartics.*

The case  $n = 6$  is treated in Section 3.3.3. Methods similar to the general case  $n \geq 7$  and bounds on Castelnuovo-Mumford regularity obtained using `Normaliz` [7] allow us to reduce the problem to a computation handled with `4ti2`. From the computational point of view, it is interesting to note that we were not able to address the case  $n = 6$  only with computational tools. Based on our experiments with `4ti2`, we expect the computation to be not feasible: it would run for several years on a server of the same capability as the one mentioned above, and a memory of 1Tb RAM would not be sufficient to finish the computation.

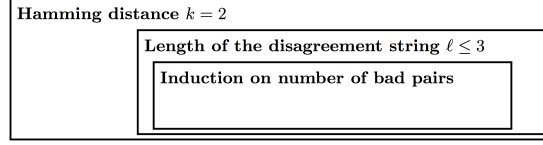
For  $n \geq 7$ , we have an induction on the degree  $d$  of the generators, i.e., the number of rows of the table. Inside a specific degree  $d$ , we have an induction on the Hamming distance  $k$  of two rows of the tables. The strategy in this inner induction on the Hamming distance  $k$  is the following. Suppose we have a binomial generator of degree  $d \geq 5$ . Hence, we have a pair of tables consisting of  $d$  rows each and with  $n \geq 7$  columns. Two rows have Hamming distance  $k$  and we reduce it to  $k = 0$ ; in other words, the given pair of tables is transformed into a pair of tables that have an identical row. This is a binomial which is a product of a binomial of degree  $d - 1$  and a variable. By induction on  $d$ , such a binomial can be generated in degree at most 4.

Hence the aim of the induction on the Hamming distance  $k$  is to reduce it to  $k = 0$ . In order to achieve this, we address the case  $k \geq 3$  into two separate propositions in Section 3.2; see Proposition 3.5 and Corollary 3.6, and Proposition 3.12. This reduces the proof to  $k = 2$ . Recall that there do not exist flows whose Hamming distance is  $k = 1$ , since they cannot disagree only in one entry.

We now discuss the strategy in case  $k = 2$ , the technical heart of the proof, which is tackled in Section 3.3. In spite of many symmetries, discussed in Section 2, there are several cases one has to consider: We identify ten cases, indexed by roman numerals, where the first two rows of the given pair of tables  $T_0, T_1$  have a disagreement string of length  $k = 2$ . Here we provide a uniform proof for three crucial cases: **Case I**, **II**, and **III**. As we show them simultaneously with the very same techniques, we refer to those as the “main case”. The rest of the cases is treated by reducing them to the main case.

For the proof in the main case, we look at the second rows of each of the tables  $T_0$  and  $T_1$ . Let

Figure 2: Zoom in of Hamming distance  $k = 2$  step.



$\ell$  denote the length of the disagreement string between those two, in columns *not* involving the first two. By Corollary 3.6, we are able to assume  $\ell \leq 3$  and, since  $n \geq 7$ , the length of the agreement string between the second row of  $T_0$  and the second row of  $T_1$ , outside columns 1 and 2, is at least  $n - 5 \geq 2$ . Since the columns are indistinguishable up to the action of  $\mathfrak{S}_n$ , we may assume that the columns  $n - 1$  and  $n$  are involved in the agreement string. Now the aim is to reduce to the situation in which no row has two nonzero entries in the columns  $n - 1$  and  $n$ : employing moves of degree at most four, we would like to eliminate all the strings which have nonzero entries on both columns  $n - 1$  and  $n$ . We call such strings *bad pairs*.

**Definition 3.3 (Bad pairs).** A bad pair is a string  $xy$ , where the elements  $x, y \in G$  are such that:

- (i) they are both nonzero;
- (ii)  $x$  is in column  $n - 1$  and  $y$  is in column  $n$ .

We now show that eliminating all the bad pairs we fall back to the case of  $n - 1$  leaves, which allows us to conclude, by the outermost induction.

**Theorem 3.4.** Suppose that a pair of compatible tables  $T_0, T_1$  with  $n \geq 7$  columns do not contain rows with bad pairs. Then the corresponding binomial is generated in degree at most  $\phi(G, K_{1, n-1})$ .

*Proof.* The assumption implies that for every row  $r$  of  $T_0$  and  $T_1$  we have either  $r(n - 1) = 0$  or  $r(n) = 0$ . Summing up the columns  $n - 1$  and  $n$ , we obtain two tables  $\tilde{T}_0$  and  $\tilde{T}_1$ . The crucial observation is that  $\tilde{T}_0$  and  $\tilde{T}_1$  are compatible tables with  $n - 1$  columns. Hence they correspond to a binomial in  $I(X(G, K_{1, n-1}))$ . This binomial is generated in degree at most  $\phi(G, K_{1, n-1})$  by definition. This implies that  $\tilde{T}_0$  and  $\tilde{T}_1$  can be transformed into each other by a finite sequence of moves of degree at most  $\phi(G, K_{1, n-1})$ . Each of these moves lifts to the tables  $T_0$  and  $T_1$ , transforming all their columns accordingly, except columns  $n - 1$  and  $n$ . Here the moves permute the pairs of elements, where each pair is formed by the two elements in columns  $n - 1$  and  $n$ , in a fixed row. These moves transform  $T_0, T_1$  into  $\hat{T}_0, \hat{T}_1$ . The latter need not be the same though; indeed, they may differ in columns  $n - 1$  and  $n$ . As in the proof of [35, Theorem 3.12], we make quadratic moves to adjust the elements in columns  $n - 1$  and  $n$ . These transform  $\hat{T}_0$  into  $\hat{T}_1$ . Hence the tables  $T_0, T_1$  are generated in degree at most  $\phi(G, K_{1, n-1})$ .  $\square$

### 3.2 Reduction of Hamming distance $\geq 3$

In this section, we start our reduction of the Hamming distance. More precisely, we assume the Hamming distance to be at least three and we prove that we can reduce it to two; the latter will be discussed in Section 3.3. We proceed analyzing the cases when the disagreement string is given by at least four entries.

**Proposition 3.5.** *The disagreement strings (i)  $\alpha\alpha\alpha\alpha$ , (ii)  $\alpha\alpha\beta\beta$ , (iii)  $\alpha\alpha\beta\gamma$ , and (iv)  $\alpha\alpha\alpha\beta$  can be reduced.*

*Proof.* (i). Consider the function  $0_{1234} - \alpha_{1234}$ . By the action of the group of flows  $\mathfrak{G}$ , we may assume that this counting function is nonpositive on both of the tables. Since the function is strictly positive in the first row of  $T_1$ , there exists a row  $r$  in  $T_1$  where there are strictly more copies of  $\alpha$  than copies of 0 in the columns 1, 2, 3, 4. On the other hand,  $r$  cannot contain  $\alpha\alpha$  in two of the columns 1, 2, 3, 4, since we would exchange those with the corresponding entries in the first row and this would decrease the Hamming distance. Thus  $r$  has one copy of  $\alpha$  and no copies of 0 in columns 1, 2, 3, 4. If the row  $r$  has both copies of  $\beta$  and  $\gamma$ , we would move the string  $\alpha\beta\gamma$  to the first row of  $T_1$ , reducing the Hamming distance. Whence we may assume that  $r$  contains the string  $\alpha\beta\beta\beta$  in columns 1, 2, 3, 4. Notice that in columns 2, 3, 4 of  $T_1$ , there are no strings of the form  $\alpha\alpha$  or  $\gamma\gamma$ , otherwise quadratic moves would decrease the Hamming distance. Additionally, in columns 2, 3, 4 there is no string of the form  $\alpha\gamma$ ; for this we can apply in  $T_1$  the cubic move  $0000 + \alpha\beta\beta\beta + ?\alpha\gamma = \alpha\beta\gamma 0 + 0\alpha 0\beta + ?0\beta$ . Now, we introduce the counting function  $0_{234} + \beta_{234} - \alpha_{234} - \gamma_{234}$  on  $T_1$ . By the previous discussion about the possible strings in columns 2, 3, 4, this function is at least one in every row of  $T_1$ . Consequently, there exists a row  $r'$  in  $T_0$  where this function is three. As a consequence, the row  $r'$  contains either the string  $\beta\beta$  or  $00$ . This would decrease the Hamming distance.

(ii). Consider the counting function  $0_{1234} - \alpha_{12} - \beta_{34}$ . By the action of the group of flows  $\mathfrak{G}$ , we may assume it is nonpositive on both of the tables. Since this function is strictly positive on the first row of  $T_1$ , there exists a row  $r$  in  $T_1$  where the function is strictly negative. Note that on the row  $r$ , one has  $\alpha_{12}, \beta_{34} \leq 1$ ; otherwise we would make a quadratic move, involving  $r$  and the first row of  $T_1$ , reducing the Hamming distance.

If in the row  $r$  we have  $\alpha_{12} = \beta_{34} = 1$ , then  $0_{1234} \leq 1$ , by the value of the counting function on  $r$ . Hence in the row  $r$ , there exists  $\gamma$ , which allows us to make a quadratic move reducing the Hamming distance. Without loss of generality, we have  $\alpha_{12} = 1, \beta_{34} = 0$ , and  $0_{1234} = 0$ . Thus the row  $r$  contains either the string  $\gamma\alpha\gamma\gamma$  or the string  $\alpha\gamma\gamma\gamma$ . In both cases, we exchange  $\gamma\gamma$  with the first row of  $T_1$  and we act with the flow  $(0, 0, \gamma, \gamma)$  on  $T_0$  producing  $\alpha\alpha\alpha\alpha$ , which is (i).

(iii). Consider the function  $0_{1234} - \alpha_{12} - \beta_3 - \gamma_4$ . By the action of the group of flows  $\mathfrak{G}$ , we may assume it is nonpositive on both of the tables. Therefore there exists a row  $r$  in  $T_1$  where the function is strictly positive. Note that on the row  $r$  one has  $\alpha_{12} \leq 1$ .

If in the row  $r$  we have  $\alpha_{12} = 1$  and  $\beta_3 = 1$ , then we may assume  $r$  contains the string  $\alpha x \beta y$  in columns 1, 2, 3, 4. We have  $x, y \neq \alpha, \beta, \gamma$ , as otherwise in each of these circumstances we would make a quadratic move between  $r$  and the first row of  $T_1$ , reducing the Hamming distance. Then the function is zero on  $r$ , which is not possible by assumption. Analogously, we may conclude when  $\alpha_{12} = 1$  and  $\gamma_4 = 1$ .

If in the row  $r$  we have  $\alpha_{12} = 1, \beta_3 = 0$ , and  $\gamma_4 = 0$ , then  $0_{1234} = 0$ . In this case we have

$\alpha_{34} = 0$ , because of a quadratic move between  $r$  and the first row of  $T_1$ . Hence the row  $r$  contains the string  $\alpha\gamma\beta$  in columns 1, 3, 4, which again would reduce the Hamming distance.

If in the row  $r$  we have  $\alpha_{12} = 0$ , then either  $\beta_3 = 1$  or  $\gamma_4 = 1$ . If  $\beta_3 = \gamma_4 = 1$ , then in columns 1, 2 the row  $r$  contains the string 00; indeed we cannot have copies of  $\alpha, \beta$  or  $\gamma$  by quadratic moves with the first row of  $T_1$ . This implies that the counting function  $\alpha_{12} + \beta_3 + \gamma_4 - 0_{1234}$  is zero on the row  $r$ , which is not possible by the assumption. If in the row  $r$  we have  $\beta_3 = 0$  and  $\gamma_4 = 1$ , then  $0_{1234} = 0$ . In the row  $r$  we can now exclude all the possible elements in each column by quadratic moves, obtaining the string  $\beta\alpha\gamma$  in columns 2, 3, 4. We exchange this string with the first row of  $T_1$ , reducing the Hamming distance. Analogously, if in the row  $r$  we have  $\beta_3 = 1$  and  $\gamma_4 = 0$ , we obtain  $\gamma\beta\alpha$  in columns 2, 3, 4, and we conclude in the same way.

(iv). Consider the counting function  $0_{1234} - \alpha_{123} - \beta_4$ . By the action of the group of flows  $\mathfrak{G}$ , we may assume it is nonpositive on the tables. Therefore there exists a row  $r$  in  $T_1$  where the function is strictly negative. Thus on the row  $r$  we have  $0_{1234} \leq 1$ , as  $\alpha_{123} \leq 1$ .

Suppose that in the row  $r$  we have  $0_{1234} = 1$ . Then  $\alpha_{123} = 1$  and  $\beta_4 = 1$ , by the assumption on the value of the counting function on  $r$ . In two of the columns 1, 2, 3 we cannot have  $\alpha$  or  $\beta$  by quadratic moves, involving  $r$  and the first row of  $T_1$ . Thus we have a copy of  $\gamma$ ; we now make a quadratic move between  $r$  and the first row of  $T_1$ , which decrease the Hamming distance.

Suppose that in the row  $r$  we have  $0_{1234} = 0$ . If in the row  $r$  we have  $\alpha_{123} = 0$ , then  $\beta_4 = 1$ . In columns 1, 2, 3 we cannot have  $\beta$ , as otherwise we would exchange the string  $\beta\beta$  with the first row of  $T_1$ , thus reducing the Hamming distance. Whence  $r$  contains the string  $\gamma\gamma\gamma\beta$  in columns 1, 2, 3, 4. If in the row  $r$  we have  $\alpha_{123} = 1$ , then  $\beta_4 = 0$ . In this situation, by the same argument,  $r$  contains the string  $\alpha\gamma\gamma\gamma$  (or  $\gamma\alpha\gamma\gamma$  or  $\gamma\gamma\alpha\gamma$ ). We claim that having the string  $\alpha\gamma\gamma\gamma$  can be reduced to the case of having the string  $\gamma\gamma\gamma\beta$  up to quadratic moves and group automorphism. Indeed, suppose we have the string  $\alpha\gamma\gamma\gamma$  in the row  $r$ . We exchange  $\gamma\gamma$  from  $r$  with 00 from the first row of  $T_1$  in columns 2, 3. We act with the flow  $(0, \gamma, \gamma, 0)$  on both tables and we transpose column 1 and column 4. Now the row  $r$  contains the string  $\gamma\gamma\gamma\beta$  in columns 1, 2, 3, 4.

By the previous discussion, it is enough to deal only with the string  $\gamma\gamma\gamma\beta$  in  $r$ . Consider the counting function  $\alpha_{123} + \beta_{123} - \gamma_{123} - 0_{123}$ . Note that this function has only odd values. We now show that the function cannot be positive on a row of  $T_1$ . Indeed, assume there is a row  $r'$  where the function takes a positive value. Then the row  $r'$  contains either  $\alpha\alpha, \beta\beta$  or  $\alpha\beta$  in columns 1, 2, 3. The first two cases are not possible, because we would exchange them with the string  $\gamma\gamma$  in the row  $r$ ; this would produce  $\alpha\alpha$  or  $\beta\beta$  in the row  $r$ , which we would exchange with 00 in the first row of  $T_1$ . We are left with the possibility of  $r'$  having  $\alpha\beta$  in columns 1, 2, 3. For this we apply in  $T_1$  the cubic move  $0000 + \gamma\gamma\gamma\beta + ?\alpha\beta = \gamma\gamma\beta\beta + 0\alpha 00 + ?0\gamma$ .

In conclusion, the counting function  $\alpha_{123} + \beta_{123} - \gamma_{123} - 0_{123}$  is strictly negative on every row of  $T_1$ . Since the value of this function on the first row of  $T_0$  is 3, there exists a row  $r''$  in  $T_0$  on which the function is  $-3$ . Thus in  $r''$  we have either 00 or  $\gamma\gamma$  in columns 1, 2, 3. In this case, we would exchange them with the first row of  $T_0$  reducing the Hamming distance.  $\square$

**Corollary 3.6.** *Suppose that a table  $T$  contains two rows  $r$  and  $r'$  having disagreement string of cardinality four. Then, using moves of degree at most three,  $T$  can be transformed in such a way that the disagreement string has cardinality at most three. Moreover, only the four columns of the disagreement string are involved in the reduction.*

*Proof.* Assume two rows  $r$  and  $r'$  do not agree on four elements. Up to the action of the group of flows  $\mathfrak{G}$  and  $\mathfrak{S}_4$ , the elements of  $r$  in the disagreement string can be set to be 0000; all the

possibilities for the elements of  $r'$  in the disagreement string are  $\alpha\alpha\alpha\alpha$ ,  $\alpha\alpha\beta\beta$ ,  $\alpha\alpha\beta\gamma$ , and  $\alpha\alpha\alpha\beta$ . By Proposition 3.5, these disagreement strings can be reduced. Hence, performing the moves in the proof of the Proposition 3.5, we transform the tables in such a way that the cardinality of the disagreement string is at most three.  $\square$

Now we deal with the disagreement string of length three,  $\alpha\beta\gamma$ . We begin with preparatory lemmas.

**Lemma 3.7.** *Suppose that the disagreement string between  $T_0$  and  $T_1$  is  $\alpha\beta\gamma$ , in columns 1, 2, 3. Then we may assume that there exists a row  $r'$  in  $T_0$  containing the string 00 in columns 1, 2, 3.*

*Proof.* We introduce the counting function  $0_{123} - \alpha_1 - \beta_2 - \gamma_3$ . By the action of the group of flows  $\mathfrak{G}$ , we may assume that the sum is nonnegative on  $T_0$ . Then there exists a row  $r'$  in  $T_0$  where the function is strictly positive.

If in the row  $r'$  we have  $0_{123} = 1$ , then  $\alpha_1 = \beta_2 = \gamma_3 = 0$ , by the assumption on the counting function evaluated at  $r'$ . By the action of the group of flows  $\mathfrak{G}$ , we may assume without loss of generality that  $r'$  contains the string  $0xy$  in columns 1, 2, 3. Then  $x \neq 0, \beta$  by assumption. Also,  $x \neq \gamma$ , as otherwise we would exchange the string  $0\gamma$  with  $\alpha\beta$  in the first row of  $T_0$ , reducing the Hamming distance between  $T_0$  and  $T_1$ . Hence  $x = \alpha$ . Similarly,  $y \neq 0, \gamma$  and  $y \neq \beta$ , as otherwise we exchange  $0\beta$  with  $\alpha\gamma$  in the first row of  $T_0$ . Hence  $r'$  contains the string  $0\alpha\alpha$  in columns 1, 2, 3, which we exchange with the first row of  $T_0$ .  $\square$

**Lemma 3.8.** *We may assume that the row  $r'$  of Lemma 3.7 in  $T_0$  contains the string  $00\gamma$  in columns 1, 2, 3. More generally, for every row  $r''$  containing the string 00 in columns 1, 2, 3, the nonzero element of  $r''$  in columns 1, 2, 3 coincides with the corresponding entry of the first row of  $T_0$ .*

*Proof.* The row  $r'$  contains a string with exactly two elements equal to 0 in the columns 1, 2 and 3. By the action of  $\mathfrak{S}_n$ , we may assume that  $r'$  contains the string  $00x$  in columns 1, 2, 3. Note that  $x \neq \alpha, \beta$ , as in both cases we make a quadratic move between  $r'$  and the first row of  $T_0$ , reducing the Hamming distance between  $T_0$  and  $T_1$ . Thus  $x = \gamma$ . By the action of the group of flows  $\mathfrak{G}$ , in every row  $r''$  containing the string 00 in columns 1, 2, 3, the nonzero entry coincides with the corresponding entry of the first row of  $T_0$ .  $\square$

**Lemma 3.9.** *Suppose that in  $T_0$  we have a row  $r'$  containing  $00\gamma$ . Then this is the only string that a row with 00 in columns 1, 2, 3 may contain.*

*Proof.* Since the row  $r'$  in  $T_0$  contains  $00\gamma$ , then it cannot contain another copy of  $\gamma$ , as we would exchange with the first row of  $T_0$ , thus reducing the Hamming distance. Hence  $r'$  contains  $00\gamma\alpha\beta$ , since it is a flow. Assume there exists another row  $r''$  containing a string with 00, different from  $00\gamma$ . By Lemma 3.8, the unique nonzero entry in columns 1, 2, 3 of  $r''$  agrees with the corresponding entry of the first row of  $T_0$ . Assume that  $r''$  contains  $0\beta 0$  in columns 1, 2, 3. Then we apply the cubic move  $\alpha\beta\gamma 00 + 00\gamma\alpha\beta + 0\beta 0 = 0\beta 00\beta + 0\beta\gamma\alpha 0 + \alpha 0\gamma$ , reducing the Hamming distance. For a row containing  $\alpha 00$  we conclude in the same way.  $\square$

**Lemma 3.10.** *As in the proof of Lemma 3.9, we assume that  $r'$  contains  $00\gamma\alpha\beta$  in columns 1, 2, 3, 4, 5. There exists a row  $r''$  in  $T_0$  such that  $r''(3) = 0$  and, moreover,  $r''$  contains the string  $\alpha\beta 0\alpha\beta$  in columns 1, 2, 3, 4, 5.*

*Proof.* Such a row  $r''$  exists in  $T_0$  by the compatibility of the two tables. The structure of  $T_0$  is:

$$T_0 = \begin{bmatrix} \alpha & \beta & \gamma & 0 & 0 & \dots & 0 \\ 0 & 0 & \gamma & \alpha & \beta & \dots & \dots \\ x & y & 0 & z & w & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

By Lemma 3.9, we have  $x, y \neq 0$ . Analogously, we have  $z, w \neq 0$  by applying Lemma 3.9, upon exchanging the string  $\alpha\beta\gamma 00$  in the first row with  $00\gamma\alpha\beta$  in the second row.

Note that  $x \neq \beta$  and  $y \neq \alpha$ , as otherwise, exchanging with the first row, in the first case with  $\alpha\gamma$  and in the second with  $\beta\gamma$ , we would reduce the Hamming distance; analogously,  $z \neq \beta$  and  $w \neq \alpha$ . Furthermore, by Lemma 3.8, we have  $x, y \neq \gamma$  as otherwise we would create the string  $\gamma 00$  and  $0\gamma 0$  respectively. Analogously  $z, w \neq \gamma$ . Hence the only remaining possibility is  $xy = \alpha\beta$  and  $zw = \alpha\beta$ .  $\square$

**Lemma 3.11.** *The counting function  $0_{12345} - \alpha_{14} - \beta_{25} - \gamma_3$  is at most  $-1$  on every row of  $T_0$ .*

*Proof.* For the sake of contradiction, suppose there exists a row  $r$  in  $T_0$ , where the counting function is nonnegative. In  $T_0$ , there exists a row  $r''$  with  $r''(3) = 0$ . By Lemma 3.10, the row  $r''$  contains the string  $\alpha\beta 0\alpha\beta$ .

If in the row  $r$  we have  $0_{12345} \geq 3$ , then  $r(3) \neq 0$ , again, by Lemma 3.10. Hence we have at least two differences with  $r''$  and we can make a quadratic move between  $r$  and  $r''$ . This reduces the Hamming distance. Thus on the row  $r$  one has  $0_{12345} \leq 2$ .

If  $0_{12345} = 2$  on  $r$ , we have the following possibilities:

- (i)  $r$  contains  $00\gamma xy$ ;
- (ii)  $r$  contains  $xyz00$ ;
- (iii)  $r$  contains  $x0yz0$ ;
- (iv)  $r$  contains  $x0y0z$ .

In case (i), we have  $x, y \neq 0$  by the assumption on the value of the counting function. Additionally,  $x, y \neq \gamma$ , as we would exchange the string  $00\gamma\gamma$  with the first row in  $T_0$ . Consider the differences between  $r$  and  $r''$ . If  $xy \neq \alpha\beta$ , then we can make a move involving column 3, at most one of columns 1, 2 and either column 4 or 5 between  $r$  and  $r''$ . This allows us to exchange  $\gamma$  in  $r$  with 0 in  $r''$ ; this contradicts Lemma 3.10. Hence  $xy = \alpha\beta$ , which on the other hand contradicts the nonnegativity of the counting function. Exchanging  $r'$ , the row appearing in Lemma 3.10 containing  $00\gamma\alpha\beta$ , with the first row of  $T_0$ , case (ii) is the same as case (i).

In case (iii),  $x \neq 0$  by the assumption on the value of the counting function. Moreover,  $x \neq \gamma$  since we would exchange  $\gamma 0$  in  $r$  with the string  $\alpha\beta$  in  $r''$  in columns 1, 2, contradicting Lemma 3.8. We also have  $x \neq \beta$ , because we could make a quadratic move in columns 1, 5 between  $\beta 0$  in  $r$  with  $0\beta$  in  $r'$ , obtaining the string  $\beta 0\gamma\alpha 0$  in  $r'$ . Now, we exchange in columns 1, 3, the string  $\beta\gamma$  in  $r'$  with  $\alpha 0$  in  $r''$ , which produces the string  $\alpha 00\alpha 0$ ; this reduces the Hamming distance. Finally, if  $x = \alpha$ , we exchange in columns 1, 2, 5, the string  $\alpha 00$  in  $r$  with  $\alpha\beta\beta$  in  $r''$ , obtaining  $\alpha 00\alpha 0$ , which again reduces the Hamming distance.

In case (iv),  $x \neq 0$  by the assumption on the value of the counting function. Additionally,  $x \neq \gamma$ , because otherwise we would exchange in columns 1, 2 the string  $\gamma 0$  in  $r$  with  $\alpha\beta$  in  $r''$ , thus

contradicting Lemma 3.10. Also,  $x \neq \beta$ , as we would make a quadratic move on columns 1, 2, 4 between  $r$  and  $r''$ , contradicting again Lemma 3.10. Analogously,  $z \neq 0, \alpha, \gamma$ . Hence  $r$  contains the string  $\alpha 0 y 0 \beta$ . We exchange in columns 1, 2, 4, 5 the string  $\alpha 0 0 \beta$  in  $r$  with  $0 0 \alpha \beta$  in  $r'$ , which produces  $0 0 y$  in  $r$ , which in turn implies  $y = \gamma$  by Lemma 3.8. This contradicts the nonnegativity of the counting function.

If  $0_{12345} = 1$ , by symmetry, we may assume  $r(2) = 0$  or  $r(3) = 0$ . If  $r(3) = 0$ , then by Lemma 3.10,  $r$  contains  $\alpha \beta 0 \alpha \beta$ , which contradicts the nonnegativity of the counting function. If  $r(2) = 0$ , then  $r$  contains  $x 0 y z t$ . Then  $z \neq 0$  by the assumption. Moreover,  $z \neq \gamma$ , as we would exchange  $r$  with  $r''$  in columns 2, 4, contradicting Lemma 3.10.

If  $z = \alpha$ , we now consider the value of  $x$ . We have  $x \neq 0$  by assumption. We have  $x \neq \alpha$  by assumption on the nonnegativity of the counting function. Moreover,  $x \neq \gamma$ , since otherwise we would exchange in columns 1, 2 the string  $\gamma 0$  in  $r$  with  $\alpha \beta$  in  $r''$  contradicting Lemma 3.10. Hence  $x = \beta$ , i.e.,  $r$  contains the string  $\beta 0 y \alpha t$ . Now,  $t \neq 0$ , by the assumption on the value of  $0_{12345}$ . Moreover  $t \neq \beta$ , by the assumption on the value of the counting function on  $r$ . Also notice that  $t \neq \gamma$ , as otherwise we exchange in columns 1, 5 the string  $\beta \gamma$  of  $r$  with  $\alpha 0$  of the first row of  $T_0$ , and then we exchange  $\beta \beta$  from the first row with  $0 0$  in  $r'$  reducing the Hamming distance. Therefore  $r$  contains the string  $\beta 0 y \alpha \alpha$ , which we exchange with the string  $\alpha \beta 0 \alpha \beta$  in  $r''$  in columns 1 and 5, contradicting Lemma 3.10.

If  $z = \beta$ , then  $r$  contains  $x 0 y \beta t$ . Furthermore,  $t \neq 0$  by assumption on the value of  $0_{12345}$ . Moreover,  $t \neq \alpha$  exchanging in columns 4, 5 the string  $\beta \alpha$  of  $r$  with  $\alpha \beta$  of  $r''$ , contradicting Lemma 3.10. Analogously, we would contradict Lemma 3.10 for  $t = \gamma$ , exchanging in columns 2, 4, 5, the string  $0 \beta \gamma$  in  $r$  with  $\beta \alpha \beta$  in  $r''$ . Hence  $r$  contains the string  $x 0 y \beta \beta$ . Here  $x \neq 0$ , by assumption. Moreover,  $x \neq \alpha$ , because of the nonnegativity of the counting function. Also,  $x \neq \gamma$ , because we would contradict Lemma 3.10, exchanging  $\gamma 0$  of  $r$  with  $\alpha \beta$  of  $r''$ . Therefore  $r$  contains  $\beta 0 y \beta \beta$ , but we exchange it with  $\alpha \beta 0 \alpha \beta$  in columns 1, 4 contradicting Lemma 3.10.

If  $0_{12345} = 0$ , then  $\alpha_{14} = \beta_{25} = \gamma_3 = 0$ , by the assumption on the nonnegativity of the function on  $r$ . Thus  $r$  contains  $x y z t w$  different from  $\alpha \beta 0 \alpha \beta$  in columns 1, 2, 4, 5. Hence we have two identical differences between  $r$  and  $r''$ , which allow to make a quadratic move, contradicting Lemma 3.10.  $\square$

**Proposition 3.12.** *The disagreement string  $\alpha \beta \gamma$  can be reduced.*

*Proof.* By Lemma 3.11, the counting function  $0_{12345} - \alpha_{14} - \beta_{25} - \gamma_3$  is at most  $-1$  on every row of  $T_0$ . As a consequence, there exists a row  $r$  in  $T_1$ , where the function is at most  $-2$ . By the value of the counting function on the row  $r$ , the entries in  $r$  must agree in two, three, four or five entries with  $\alpha \beta \gamma \alpha \beta$ .

If  $r$  agrees in five entries, it contains  $\alpha \beta \gamma \alpha \beta$ . We exchange  $\alpha \beta \gamma$  with  $0 0 0$  in the first row of  $T_1$ , which reduces the Hamming distance between  $T_0$  and  $T_1$ . If  $r$  agrees in four entries, we denote by  $x$  the element where  $r$  does not agree with  $\alpha \beta \gamma \alpha \beta$ . If  $x \neq r(3)$ , then we would have either the string  $\alpha \beta \gamma$  or  $\gamma \alpha \beta$ , which is also in table  $T_0$ ; this reduces the Hamming distance. Suppose  $r$  contains  $\alpha \beta x \alpha \beta$ . If  $x = 0$ , the table  $T_0$  contains the same flow. If  $x = \alpha$  or  $\beta$ , we exchange  $\alpha \alpha$  or  $\beta \beta$  with  $0 0$  in the first row of  $T_1$ .

If  $r$  agrees with  $\alpha \beta \gamma \alpha \beta$  in three entries, we denote by  $xy$  the remaining two. First, note that if  $xy$  are in columns 1, 2 or in columns 4, 5, we exchange  $\alpha \beta \gamma$  or  $\gamma \alpha \beta$  with  $0 0 0$  in the first row of  $T_1$ ; this decreases the Hamming distance.

Assume that both of  $x$  and  $y$  are in columns 1, 2, 3. If  $r$  contains  $x \beta y \alpha \beta$ , then  $x \neq \alpha, \beta$ , because otherwise we would exchange the string  $\alpha \alpha$  or  $\beta \beta$  with the first row of  $T_1$  reducing the Hamming distance. Whence  $x = 0, \gamma$ . Moreover  $y \neq \gamma$ , by definition. Additionally,  $y \neq \beta$ , because we would

move  $\beta\beta$  to the first row of  $T_1$ , reducing the Hamming distance. It follows that  $y = 0, \alpha$ . On the other hand,  $xy \neq 00$ , since the counting function  $0_{12345} - \alpha_{14} - \beta_{25} - \gamma_3$  is at most  $-2$  on  $r$ . Furthermore,  $x + y \neq \beta$ , as otherwise we would exchange  $x\beta y$  with  $000$  in the first row of  $T_1$ , reducing the Hamming distance between  $T_0$  and  $T_1$ . Hence  $r$  contains either  $\gamma\beta 0\alpha\beta$  or  $0\beta\alpha\alpha\beta$ . For the first, we exchange in columns 2, 3, 5, the string  $\beta 0\beta$  with  $000$  in the first row of  $T_1$ , and we exchange  $\alpha\beta 0\alpha\beta$  in  $T_0$  with the first row of  $T_0$ . For the second, we exchange  $0\beta\alpha\alpha\beta$  with the first row of  $T_1$  and  $\alpha\beta 0\alpha\beta$  in  $T_0$  with the first row of  $T_0$ , which reduces the Hamming distance.

If  $r$  contains  $\alpha xy\alpha\beta$ , then applying the automorphism  $\alpha \leftrightarrow \beta$  and a transposition between columns 1 and 2, we are in the case when the row  $r$  contains  $x\beta y\alpha\beta$ .

If  $x, y$  are both in columns 3, 4, 5, we apply analogous moves as the ones featured above. Then we may assume that  $x$  is either in column 1 or 2, and  $y$  is either in column 4 or 5. In all these cases, we have  $x = 0$  and  $y = 0$ , as all the other possibilities are excluded by exchanging with the first row of  $T_1$ . The fact that  $x = y = 0$  contradicts the value of the counting function on  $r$ .

If  $r$  agrees with  $\alpha\beta\gamma\alpha\beta$  in two entries, we have  $0_{12345} = 0$  on  $r$ , since the value of the counting function  $0_{12345} - \alpha_{14} - \beta_{25} - \gamma_3$  on  $r$  is at most  $-2$ . In columns 1, 2, 3, there is at least one entry  $x$  which does not agree with the corresponding entry in  $\alpha\beta\gamma$ , because otherwise we would move  $\alpha\beta\gamma$  to the first row of  $T_1$ , reducing the Hamming distance. Denoting the elements where they do not agree by  $x, y, z$ , the strings that  $r$  may contain are:  $\alpha xy\alpha z$ ,  $\alpha\beta xyz$ , and  $\alpha xyz\beta$ . Note that these are all the possible, as the remaining ones are resolved in the same way upon exchanging the string  $\alpha\beta\gamma 00$  in the first row with  $00\gamma\alpha\beta$  in the second row of  $T_0$ . If  $r$  contains  $\alpha xy\alpha z$ , then we exchange the string  $\alpha\alpha$  of  $r$  in columns 1, 4 with  $00$  in  $T_1$ . We now exchange the string  $\alpha\beta 0\alpha\beta$  of  $r''$  with the first row in  $T_0$ ; these two rows have lower Hamming distance. If  $r$  contains  $\alpha\beta x$  in columns 1, 2, 3, then  $x \neq \gamma$ , by the counting function. Moreover,  $x \neq 0$  since  $0_{12345} = 0$ . Hence  $x = \alpha$  or  $\beta$ . Now we exchange  $\alpha\alpha$  or  $\beta\beta$  with  $00$  in the first row of  $T_1$  reducing the Hamming distance. If  $r$  contains  $\alpha xyz\beta$ , by definition or by quadratic moves we can exclude the cases  $x = \alpha, \beta, 0$ , and  $y = \alpha, \gamma, 0$ . Hence  $r$  contains  $\alpha\gamma\beta$ , which we exchange with the first row of  $T_1$ , decreasing the Hamming distance.  $\square$

The preceding results of this section show the following corollary.

**Corollary 3.13.** *The Hamming distance of two flows can be reduced to at most two.*

### 3.3 The disagreement string $\alpha\alpha$

In this section, we proceed in the case of the disagreement string  $\alpha\alpha$ .

$$T_0 - T_1 = \begin{bmatrix} \alpha & \alpha & 0 & \dots & 0 \\ 0 & x & \dots & \dots & \dots \\ y & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \alpha & z & \dots & \dots & \dots \\ w & \alpha & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}. \quad (1)$$

Let us denote the row in  $T_0$  starting with the string  $0x$  by  $r_{0x}$  and the row in  $T_1$  starting with the string  $\alpha z$  by  $r_{\alpha z}$ . After fixing the first rows and the first two columns, we make moves of degree at most four on the rest of tables in such a way that the number of agreements in  $r_{0x}$  and  $r_{\alpha z}$  is maximized.

**Remark 3.14.** *Corollary 3.6 ensures that, after possibly making moves of degree at most four, the rows  $r_{0x}$  and  $r_{\alpha z}$  in  $T_0$  and  $T_1$  respectively, agree in at least  $n - 5$  entries. Up to the action of  $\mathfrak{S}_n$  on the  $n$  leaves, and hence on the columns, these are the last  $n - 5$  columns.*



**Definition 3.15.** *The string in the last  $n - 5$  columns of the rows  $r_{0x}$  and  $r_{\alpha z}$  is the agreement string between  $r_{0x}$  and  $r_{\alpha z}$ . Up to the action of the group of flows  $\mathfrak{G}$ , these entries are zeros.*

Our aim is to prove the following three crucial cases, which we refer to as the *main case*:

$$\begin{aligned} \text{Case I} &: x = \beta, y = \alpha, z = \beta, w = 0; \\ \text{Case II} &: x = \beta, y = \alpha, z = \beta, w = \beta; \\ \text{Case III} &: x = \beta, y = \beta, z = \beta, w = \beta. \end{aligned} \tag{*}$$

In Section 3.3.1, we reduce any other possible case to one of the above.

### 3.3.1 Reduction to the main case

Up to the action of the group of flows  $\mathfrak{G}$ , there are at least as many copies of 0 as copies of  $\alpha$  in the first two columns of  $T_0$ . Up to the action of  $\text{Aut}(G)$ , we may assume  $x = \beta$ . We will show that all cases can be resolved, by reducing to the main case (\*).

We first collect a useful lemma which we will use to resolve easily some of the cases.

**Lemma 3.16.** *If in table  $T_1$  in (1) we have  $\{z, w\} = \{\beta, \gamma\}$ , then the corresponding cases can be reduced. If in table  $T_0$  in (1) we have  $\{x, y\} = \{\beta, \gamma\}$ , then the corresponding cases can be reduced.*

*Proof.* If  $\{z, w\} = \{\beta, \gamma\}$ , then in  $T_1$  we have either the cubic move  $00 + \alpha\beta + \gamma\alpha = \alpha\alpha + 0\beta + \gamma 0$  or  $00 + \alpha\gamma + \beta\alpha = \alpha\alpha + \beta 0 + \gamma 0$ . The second sentence is the symmetric version of the first: acting with the flow  $(\alpha, \alpha, 0, \dots, 0) \in \mathfrak{G}$  on the tables, we produce the same tables as in the first statement.  $\square$

We now analyze all the possible cases. We refer to the tables  $T_0$  and  $T_1$  in (1).

**Case  $y = \alpha$ .** In this case, the table  $T_0$  has the form:

$$T_0 = \begin{bmatrix} \alpha & \alpha & 0 & \dots & 0 \\ 0 & \beta & \dots & \dots & \dots \\ \alpha & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

We may have  $z = 0, \beta, \gamma$ .

$z = \beta$ .

Here,  $w = \gamma$  is reduced by Lemma 3.16. Hence we have  $w = 0$  (**Case I**) or  $w = \beta$  (**Case II**).

$z = 0$ .

Here,  $w = 0$  (**Case X**),  $w = \beta$  (**Case VII**),  $w = \gamma$  (**Case VI**).

$z = \gamma$ .

Here,  $w = 0$  (**Case IV**),  $w = \gamma$  (**Case V**),  $w = \beta$  is resolved by Lemma 3.16.

**Case  $y = \beta$ .** In this case, the table  $T_0$  has the form:

$$T_0 = \begin{bmatrix} \alpha & \alpha & 0 & \dots & 0 \\ 0 & \beta & \dots & \dots & \dots \\ \beta & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

We may have  $z = \beta, 0, \gamma$ .

$z = \beta$ .

Here,  $w = 0$  (which is **Case II** by acting with the flow  $(\alpha, \alpha, 0, \dots, 0) \in \mathfrak{G}$  and  $\gamma \leftrightarrow \beta$ ),  $w = \beta$  (**Case III**),  $w = \gamma$  resolved by Lemma 3.16.

$z = 0$ .

Here,  $w = 0$  (**Case IX**),  $w = \beta$  (which is **Case II** by acting with the flow  $(\alpha, \alpha, 0, \dots, 0) \in \mathfrak{G}$ , transposing and  $\gamma \leftrightarrow \beta$ ),  $w = \gamma$  (which is **Case V** by acting the flow  $(\alpha, \alpha, 0, \dots, 0) \in \mathfrak{G}$  and transposition).

$z = \gamma$ .

Here,  $w = 0$  (which is **Case V** by acting with the flow  $(\alpha, \alpha, 0, \dots, 0) \in \mathfrak{G}$ ),  $w = \gamma$  (**Case VIII**).

We now reduce all the cases to the main case  $(\star)$ , postponing its proof for the moment, as this requires more technical results.

#### Cases IV and V.

In this case we have:

$$T_0 - T_1 = \begin{bmatrix} \alpha & \alpha & 0 & \dots & 0 \\ 0 & \beta & \dots & \dots & \dots \\ \alpha & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \alpha & \gamma & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

We may assume we do not have strings  $\gamma\gamma, 00, \gamma 0, 0\gamma$  in columns 1, 2 of  $T_0$ ; this is shown by the same arguments in the proof of Lemma 3.22. Hence the counting function  $\alpha_{12} + \beta_{12} - 0_{12} - \gamma_{12}$  is nonnegative on every row of  $T_0$ . On the other hand, in the table  $T_1$ , in columns 1, 2 we do not have the string  $\alpha\beta$ , as we would reduce this case with a cubic move. In the same columns of  $T_1$ , the string  $\alpha\alpha$  would decrease the Hamming distance. Moreover, the string  $\beta\beta$  is reduced by the cubic move  $\alpha\gamma + 0\alpha + \beta\beta = 0\beta + \beta\gamma + \alpha\alpha$ , and  $\beta\alpha$  is reduced by the cubic move  $00 + \alpha\gamma + \beta\alpha = \alpha\alpha + \beta 0 + 0\gamma$ . This is a contradiction and thus it shows the reduction.

#### Case VI.

In this case we have:

$$T_0 - T_1 = \begin{bmatrix} \alpha & \alpha & 0 & \dots & 0 \\ 0 & \beta & \dots & \dots & \dots \\ \alpha & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \alpha & 0 & \dots & \dots & \dots \\ \gamma & \alpha & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

In columns 1, 2 in  $T_1$ , the string  $\alpha\beta$  is resolved by Lemma 3.16. The string  $\alpha\gamma$  in columns 1, 2 of  $T_1$  is **Case V**. Since we cannot have the string  $\alpha\alpha$  in columns 1, 2 of  $T_1$ , the counting function  $\alpha_1 - 0_2$  is nonpositive in every row of  $T_1$ . Thus there exists a row  $r$  in  $T_0$  with  $r(2) = 0$  and  $r(1) \neq \alpha$ . Hence  $r(1) = \beta$ . Acting by the flow  $(\alpha, \alpha, 0, \dots, 0)$  and transposition we reduce to **Case V**.

#### Case VII.

In this case we have:

$$T_0 - T_1 = \begin{bmatrix} \alpha & \alpha & 0 & \dots & 0 \\ 0 & \beta & \dots & \dots & \dots \\ \alpha & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \alpha & 0 & \dots & \dots & \dots \\ \beta & \alpha & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

We exclude the string  $\alpha\beta$  in columns 1, 2 in  $T_1$ , since it is **Case II**. We also exclude  $\alpha\gamma$  by Lemma 3.16. As in **Case VI**, there exists a row  $r$  in  $T_0$  such that  $r(1) = \beta$  and  $r(2) = 0$ . Now, by acting with the flow  $(\alpha, \alpha, 0, \dots, 0) \in \mathfrak{G}$ , making a transposition and applying the group automorphism  $\gamma \leftrightarrow \beta$ , we reduce to **Case II**.

### Case VIII.

In this case we have:

$$T_0 - T_1 = \begin{bmatrix} \alpha & \alpha & 0 & \dots & 0 \\ 0 & \beta & \dots & \dots & \dots \\ \beta & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \alpha & \gamma & \dots & \dots & \dots \\ \gamma & \alpha & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

We may exclude in columns 1, 2 in  $T_0$  the string 00. Also, we exclude the string  $\gamma\gamma$  by the quartic move  $\alpha\alpha + 0\beta + \beta 0 + \gamma\gamma = 00 + \alpha\gamma + \gamma\alpha + \beta\beta$ . Moreover, in columns 1, 2 in  $T_0$ , notice that we can exclude the strings  $0\gamma$  and  $\gamma 0$  by Lemma 3.16. Hence the counting function  $\alpha_{12} + \beta_{12} - 0_{12} - \gamma_{12}$  is nonnegative on every row of  $T_0$ . On the other hand, in  $T_1$  we may reduce the string  $\alpha\alpha$ ,  $\alpha\beta$  and  $\beta\alpha$  by Lemma 3.16. Finally, we are able to reduce the string  $\beta\beta$  by the quartic move  $00 + \alpha\gamma + \gamma\alpha + \beta\beta = \gamma\gamma + \beta 0 + 0\beta + \alpha\alpha$ . This is a contradiction and thus it shows the reduction.

### Case IX.

In this case we have:

$$T_0 - T_1 = \begin{bmatrix} \alpha & \alpha & 0 & \dots & 0 \\ 0 & \beta & \dots & \dots & \dots \\ \beta & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \alpha & 0 & \dots & \dots & \dots \\ 0 & \alpha & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

Analogously to the proof of Lemma 3.22, we exclude  $\gamma\gamma, 0\gamma, \gamma 0, 00$  in columns 1, 2 of  $T_0$ . So the counting function  $\alpha_{12} + \beta_{12} - 0_{12} - \gamma_{12}$  is nonnegative on every row of  $T_0$ . On the other hand, in columns 1, 2 of  $T_1$ , the strings  $\alpha\beta, \beta\alpha$  correspond to the case for  $z = \beta$  and  $w = 0$  in tables (1), which were previously done. Thus there exists a row  $r$  such that  $r(1) = \beta$  and  $r(2) = \beta$  by the positivity of the counting function in  $T_0$  and  $T_1$ . Exchanging the string 00 in the first row with the string  $\beta\beta$  in  $r$ , acting by  $\alpha\alpha$  on both  $T_0$  and  $T_1$ , applying the automorphisms  $\gamma \leftrightarrow \alpha$  and  $\gamma \leftrightarrow \beta$  we obtain **Case III**.

### Case X.

In this case we have:

$$T_0 - T_1 = \begin{bmatrix} \alpha & \alpha & 0 & \dots & 0 \\ 0 & \beta & \dots & \dots & \dots \\ \alpha & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \alpha & 0 & \dots & \dots & \dots \\ 0 & \alpha & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

In  $T_1$ , in columns 1, 2 we can exclude  $\alpha\beta$ , because it is **Case I**. The string  $\alpha\gamma$  reduces to **Case IV**. As usual, the string  $\alpha\alpha$  is excluded. Hence the counting function  $\alpha_1 - 0_2$  is nonpositive in every row of  $T_1$ . Hence there exists a row  $r$  in  $T_0$  such that  $r(1) \neq \alpha$  and  $r(2) = 0$ . The possible values of  $r(1)$  are either  $\gamma$  or  $\beta$ , since for  $r(1) = 0$  we have an immediate reduction. For  $r(1) = \gamma$  we apply Lemma 3.16 and  $r(1) = \beta$  is **Case IX**.

### 3.3.2 Preliminary Lemmas

We are now ready to present our preliminary lemmas, that are devised to tackle the main case ( $\star$ ). As they will be used very often, we give them specific reference names in order to facilitate the reading.

**Lemma 3.17 (Difference Lemma).** *Suppose we have the table  $T$  whose first three rows are  $r_1, r_2, r_3$ :*

$$T = \begin{bmatrix} q & q & \dots \\ x & y & \dots \\ z & x & \dots \\ \dots & \dots & \dots \end{bmatrix},$$

where  $q, x, y, z \in G$  and  $x \neq y, z$ . If one of the following holds:

- (i)  $z \neq y$  and  $r_2(i) - r_3(i)$  is  $x - y$  or  $x - z$  for some  $i > 2$ ; or
- (ii)  $z = y$ ,  $q \neq y$  and  $r_2(i) - r_3(i)$  is  $x - y$  or  $x - q$  for some  $i > 2$ ,

then we can transform the row  $r_1$  to a row starting with the string  $xx$ .

*Proof.* When the difference  $r_2(i) - r_3(i) = x - y$  in both (i) and (ii), we make the quadratic move  $xyw + zx(w + x - y) = xx(w + x - y) + zyw$ , which exchanges the corresponding entries in rows  $r_2$  and  $r_3$ , thus creating a row starting with the string  $xx$ . Analogously for the case (i), when the difference is  $r_2(i) - r_3(i) = x - z$ . In (ii), when the difference is  $r_2(i) - r_3(i) = x - q$ , we make the cubic move  $qq + xyw + yx(w + x - q) = xx + qy(w + x - q) + yqw$ .  $\square$

**Remark 3.18.** *Note that the Difference Lemma 3.17 distinguishes one group element in each table in each of the crucial cases **Case I**, **Case II**, and **Case III**. In all the cases, these are  $\gamma$  in  $T_0$  and  $\beta$  in  $T_1$ . In particular, if the second and third row differ on some index  $i > 2$ , then their difference must be equal to the distinguished element.*

Although basic, the Difference Lemma 3.17 will be used very frequently. We apply it following the observation above. Indeed, our aim will be often to produce a row starting with a string of type  $xx$  and conclude by induction. To this end, after identifying the situation described in Lemma 3.17, if  $r_2(i) \neq r_3(i)$ , then we will be able to immediately infer *what can be* the element  $r_2(i) - r_3(i) \in G$ ; to exclude all the other possible values we apply the Difference Lemma 3.17, obtaining a row starting with the string  $xx$ . This will be useful to decrease the given Hamming distance and conclude by induction on the degree.

**Lemma 3.19 (Standard Lemma).** *Let  $T$  be a table and suppose there is an element  $y \in G$  in some row  $r$  with  $r(n-1) = 0$  and  $r(n) = 0$ . Suppose there is a row  $r'$  of  $T$  with  $r'(n-1) = x$  and  $r'(n) = x$ , where  $0 \neq x \in G$ , and a row  $r''$  with the element  $y + x$  in the same column as  $y$ . Then we can exchange  $y$  and  $y + x$  (and appropriate entries in columns  $n-1$  and  $n$ ). The same statement holds when  $y$  is a string of elements of  $G$ .*

*Proof.* Let us consider the entries  $r''(n-1) = u$  and  $r''(n) = v$ . If  $u = x$  or  $v = x$ , then we make a quadratic move putting  $y + x$  and  $x$  in the row  $r$ . If  $u = 0$  or  $v = 0$ , then we move the string  $xx$  to the row  $r$ , and finally we exchange  $x$  with  $0$  and  $y$  with  $y + x$ . If  $u = v$  are equal, then we move the string  $xx$  in the row  $r''$  exchanging it with the string  $uv$ , thus we exchange  $y$  with  $y + x$  and  $0$  with  $x$ . Hence, we may assume that  $u \neq v$  and they are both different from  $0$  and  $x$ . Hence, the sum  $u + v + x = 0$ . Thus, we may exchange  $y$  with  $y + x$  and  $00$  with  $uv$ . The last statement is shown using the same arguments. This completes the proof.  $\square$

We now record some technical results on the main case  $(\star)$ . Note that in the main case we have  $x = z = \beta$ .

**Lemma 3.20.** *If  $r_{0\beta}(i) = r_{\alpha\beta}(i)$  for some  $i > 2$  then we may assume that both are equal to  $0 \in G$ . In particular, both rows have  $0$  on the agreement string.*

*Proof.* Without loss of generality, let us assume  $i = 3$ . If  $r_{0\beta}(i) = r_{\alpha\beta}(i) = \beta$ , then a quadratic move allows us to produce the string  $0\beta\beta 0$  in both tables. If  $r_{0\beta}(i) = r_{\alpha\beta}(i) = \gamma$ , then in both tables we obtain the string  $\alpha\beta\gamma 0$  by quadratic moves again. If  $r_{0\beta}(i) = r_{\alpha\beta}(i) = \alpha$ , then in both tables we obtain  $\alpha 0\alpha 0$ . The last string is obtained in  $T_1$  by quadratic moves, and in  $T_0$  by the following moves:

- (i) in **Case I** and **Case II**, by the cubic move  $\alpha\alpha 0 + 0\beta\alpha + \alpha 0 = \alpha 0\alpha + \alpha\beta 0 + 0\alpha$ ;
- (ii) in **Case III**, by two quadratic moves, upon exchanging  $0\beta$  with  $\beta 0$ .

$\square$

**Remark 3.21.** *We observe that in **Case I** and **Case III**, the tables  $T_0$  and  $T_1$  are in “symmetry”. More precisely, the fixed entries in table  $T_1$  can be obtained from the ones in  $T_0$ , by acting with the flow  $(\alpha, \alpha, 0, \dots, 0) \in \mathfrak{G}$  and applying the automorphism  $\beta \leftrightarrow \gamma$  of  $G$ , that exchanges  $\beta$  and  $\gamma$ . In particular, if we can prove a statement for  $T_0$  then a “symmetric” statement holds for  $T_1$ .*

**Lemma 3.22.** *We may assume that no row in  $T_0$  contains in columns 1, 2 any string of the form  $\gamma\gamma, 0\gamma, \gamma 0, 00$ . Analogously, no row in  $T_1$  contains in columns 1, 2 any of the strings of the form  $\gamma\gamma, \alpha\gamma, \gamma\alpha, \alpha\alpha$ .*

*Proof.* In all the cases, one can obtain either  $00$  in  $T_0$  or  $\alpha\alpha$  in  $T_1$ . For  $T_0$ , these are:  $\alpha\alpha + 0\beta + \alpha 0 + 0\gamma = 00 + \alpha\gamma + 0\alpha + \alpha\beta$ ,  $\alpha\alpha + 0\beta + \gamma 0 = 00 + \gamma\alpha + \alpha\beta$ ,  $\alpha\alpha + 0\beta + \beta 0 + \gamma\gamma = 00 + \gamma\alpha + \alpha\gamma + \beta\beta$ . The statement for  $T_1$  readily follows by Remark 3.21.  $\square$

**Lemma 3.23.** *For any row  $r_{xy}$  in  $T_0$  differing from  $r_{0\beta}$  on some column index  $i > 2$  not by  $\gamma$ , we may assume  $xy = \alpha\alpha, \beta\beta, \alpha\beta$  or  $\beta\alpha$ . Analogously, in  $T_1$ , if  $r_{xy}$  differs from  $r_{\alpha\beta}$  on some column index  $i > 2$  not by  $\beta$ , then  $xy = 00, \beta\beta, \beta 0$  or  $0\beta$ .*

*Proof.* In **Case I** and **Case II**, by the Difference Lemma 3.17, and a quadratic move with  $r_{0\beta}$  or with  $r_{\alpha 0}$  in  $T_0$ , we may assume  $x + y = 0$  or  $x + y = \gamma$ . The result follows by Lemma 3.22.

In **Case III** we exclude  $x + y = \alpha$ . Indeed, if  $xy = \alpha 0$  or  $xy = 0\alpha$ , then we are in **Case II** (more precisely, for  $0\alpha$  we also need to exchange the two columns to reduce to **Case II**). If  $xy = \beta\gamma$  or  $\gamma\beta$ , by the quadratic moves  $0\beta w + \beta\gamma(w + \alpha) = 0\gamma(w + \alpha) + \beta\beta w$  or  $0\beta w + \gamma\beta(w + \beta) = \beta\beta(w + \beta) + \gamma 0 w$  we produce  $0\gamma$  or  $\gamma 0$  and apply Lemma 3.16. Remark 3.21 gives the symmetric statement for  $T_1$ .  $\square$

**Lemma 3.24.** *If there exists an index  $j$  such that  $r_{0\beta}(j) = \beta$  in  $T_0$ , then we may assume that  $r_{\alpha\beta}(j) = 0$  in  $T_1$ . Analogously, if there exists an index  $j$  such that  $r_{\alpha\beta}(j) = \gamma$  in  $T_1$ , then we may assume that  $r_{0\beta}(j) = 0$ .*

*Proof.* Assume  $r_{0\beta}(3) = \beta$ . Suppose  $r_{\alpha\beta}(3) = \alpha$  or  $\gamma$  in  $T_1$ . Then there exists a row  $r$  in  $T_1$  with  $r(3) = \beta$ . The row  $r$  contains the string  $xy\beta$  in columns 1, 2, 3 for some  $x, y \in G$ . Let us determine the possible values of  $r(2) = y$ . If  $y = \beta$  we would have the string  $0\beta\beta$  in both of the tables. By Lemma 3.23,  $y = 0$ . Whence the counting function  $\beta_3 - 0_2$  is nonpositive on every row of  $T_1$ . It follows that in  $T_0$  there exists a row  $r'$  with  $r'(2) = 0$  and  $r'(3) \neq \beta$ . By Lemma 3.23, we have  $r'(3) = \alpha$ . For  $r_{\alpha\beta}(3) = \alpha$ , by quadratic moves, we obtain  $\alpha 0\alpha 0$  in both tables. Now consider the case  $r_{\alpha\beta}(3) = \gamma$ . In  $T_1$  for every row  $r$  with  $r(2) = 0$  and  $r(3) = \beta$  (likewise the  $r$  above), we have  $r(1) = \beta$ . Indeed  $r''(1)$  is either 0 or  $\beta$  by Lemma 3.23. On the other hand,  $r''(1) \neq 0$  because otherwise we would produce the string  $0\beta\beta$  in  $T_1$ , which is also in  $T_0$ . Hence  $r''(1) = \beta$ . Since in  $T_0$  we have the row  $r'$  with  $r'(3) = \alpha$ , there exists a row  $r''$  in  $T_1$  with  $r''(3) = \alpha$ . If  $r''(2) = 0$  we are done, as we produce  $\alpha 0\alpha$  in both tables. For  $r''(2) = \alpha$  we have the cubic move in  $T_1$ ,  $\alpha\beta\gamma + \beta 0\beta + \alpha\alpha = \alpha 0\alpha + \beta\alpha\gamma + \beta\beta$ . For  $r''(2) = \beta$ , we have the quartic move in  $T_1$ ,  $000 + \alpha\beta\gamma + \beta 0\beta = \alpha 0\alpha + \beta\beta 0 + 0\beta\beta + 0\gamma$ . For  $r''(2) = \gamma$ , we have the quartic move in  $T_1$ ,  $000 + \alpha\beta\gamma + \beta 0\beta + \gamma\alpha = \alpha 0\alpha + \beta\beta 0 + 0\gamma\gamma + 0\beta$ .  $\square$

### 3.3.3 The case of $n = 6$ leaves

After having set up the cornerstone of our approach, we are ready to first establish the case of  $n = 6$  leaves. Let  $P$  be the lattice polytope of the Kimura 3-parameter model for  $n = 6$  leaves. Here we are in the setting of polytopes. To be consistent with standard terminology, binomials in the ideal of the Kimura 3-parameter model are identified with relations among lattice points, which in turn are naturally identified with variables. The minimal generating relations among the vertices of the polytope  $P$  constitute a *Markov basis*. The degree of an element of a Markov basis is the total degree of the corresponding binomial in the standard grading. The degree of the corresponding table is the number of rows. Only in this section, given a Markov basis element  $B$ , which we think of as a binomial, we introduce the notation  $\deg(B)$  to denote its degree.

As recalled in Section 2, the polytope  $P$  is 18 dimensional. Following the notation of Section 2, a generating set of the full lattice  $M^6$  is  $e_{(i,g)} \in [6] \times G$ . However, our lattice is a sublattice of  $M^6$ . Since we have the six linear relations  $e_{(i,0)}^* + e_{(i,\alpha)}^* + e_{(i,\beta)}^* + e_{(i,\gamma)}^* = 1$  for  $1 \leq i \leq 6$  satisfied by the vertices of the polytope, we can choose the elements  $e_{(i,\alpha)}, e_{(i,\beta)}, e_{(i,\gamma)}$  for  $1 \leq i \leq 6$  to serve as a basis of the 18-dimensional lattice of interest.

**Proposition 3.25.** *The polytope  $P$  defines an 18 dimensional projectively normal (in particular, Cohen-Macaulay) toric variety in  $\mathbb{P}^{1023}$ . Its Hilbert series is  $Hs(t) = \frac{N(t)}{(1-t)^{19}}$ , where*

$$\begin{aligned} N(t) = & t^{15} + 1005t^{14} + 230763t^{13} + 11423223t^{12} \\ & + 197336781t^{11} + 1476133641t^{10} + 5369113631t^9 \\ & + 10097960379t^8 + 10077653595t^7 + 5323111487t^6 \\ & + 1442513865t^5 + 187603341t^4 + 10384023t^3 + 198795t^2 + 1005t + 1. \end{aligned}$$

*Its Hilbert polynomial is*

$$\begin{aligned} H(t) = & \frac{22261501}{4168212048000}t^{18} + \frac{799045380}{4168212048000}t^{17} + \frac{13381457673}{4168212048000}t^{16} + \frac{138721353336}{4168212048000}t^{15} \\ & + \frac{995839168812}{4168212048000}t^{14} + \frac{5247736051320}{4168212048000}t^{13} + \frac{21011354421226}{4168212048000}t^{12} + \frac{65366574541632}{4168212048000}t^{11} \end{aligned}$$

$$\begin{aligned}
& + \frac{160636901283573}{4168212048000} t^{10} + \frac{316408365264420}{4168212048000} t^9 + \frac{507035368484229}{4168212048000} t^8 + \frac{671227146881928}{4168212048000} t^7 \\
& + \frac{744003206327314}{4168212048000} t^6 + \frac{695859081785280}{4168212048000} t^5 + \frac{545170528162872}{4168212048000} t^4 + \frac{340981469563104}{4168212048000} t^3 \\
& + \frac{151089754960800}{4168212048000} t^2 + \frac{38894674089600}{4168212048000} t + 1.
\end{aligned}$$

In particular, the Markov basis has elements of degree at most 16.

Let us consider the following two codimension two faces of  $P$ :

- (i)  $\tilde{P}$  contains points corresponding to flows that have 0 or  $\alpha$  on the sixth leaf. This is the intersection of  $P$  with the linear subspace  $e_{(6,\beta)}^* = e_{(6,\gamma)}^* = 0$ .
- (ii)  $\tilde{P}'$  contains points corresponding to flows that do not have  $\gamma$  on the sixth leaf and on the fifth leaf. This is the intersection of  $P$  with the linear subspace  $e_{(5,\gamma)}^* = e_{(6,\gamma)}^* = 0$ .

The Hilbert series of (i) is  $Hs(t) = \frac{\tilde{N}(t)}{(1-t)^{17}}$ , where

$$\begin{aligned}
\tilde{N}(t) &= t^{13} + 1007t^{12} + 107752t^{11} + 2813176t^{10} \\
&+ 26622909t^9 + 109147219t^8 + 211160560t^7 + 199302992t^6 \\
&+ 91202787t^5 + 19336749t^4 + 1724040t^3 + 54360t^2 + 495t + 1.
\end{aligned}$$

The Hilbert series of (ii) is  $Hs(t) = \frac{\tilde{N}'(t)}{(1-t)^{17}}$ , where

$$\begin{aligned}
\tilde{N}'(t) &= 3t^{13} + 2253t^{12} + 211288t^{11} + 5060488t^{10} \\
&+ 44891401t^9 + 174437831t^8 + 321990512t^7 + 291183248t^6 \\
&+ 127959653t^5 + 26052683t^4 + 2223560t^3 + 66520t^2 + 559t + 1.
\end{aligned}$$

In particular, the Markov basis in both cases has elements of degree at most 14.

*Proof.* The computation of Hilbert series and verification of normality were obtained using Normaliz [7]. The statements about the degree of Markov basis are a consequence of well-known theorems on regularity of normal toric varieties, see Appendix 4.  $\square$

**Lemma 3.26.** *The following three codimension three faces  $P_1, P_2, P_3$  of  $P$  have Markov basis with elements of degree at most four:*

- (i)  $P_1$  contains points corresponding to flows that have 0 on the sixth leaf. This is the intersection of  $P$  with the linear subspace  $e_{(6,\alpha)}^* = e_{(6,\beta)}^* = e_{(6,\gamma)}^* = 0$  and is isomorphic to the Kimura 3-parameter model polytope for five leaves.
- (ii)  $P_2$  contains points corresponding to flows that do not have  $\beta$  or  $\gamma$  on the sixth leaf and do not have  $\gamma$  on the fifth leaf. This is the intersection of  $P$  with the linear subspace  $e_{(5,\gamma)}^* = e_{(6,\beta)}^* = e_{(6,\gamma)}^* = 0$ .
- (iii)  $P_3$  contains points corresponding to flows that do not have  $\gamma$  on the fourth, the fifth and the sixth leaf. This is the intersection of  $P$  with the linear subspace  $e_{(4,\gamma)}^* = e_{(5,\gamma)}^* = e_{(6,\gamma)}^* = 0$ .

*Proof.* We employed 4ti2 [47] to compute explicitly the Markov basis in all three cases. More specifically, for  $P_2$  we obtained 47112 relations: 36840 quadrics, 2304 cubics, and 7968 quartics. For  $P_3$ , we obtained 57058 relations: 48600 quadrics, 2176 cubics, and 6282 quartics.  $\square$

**Remark 3.27.** *The polytopes  $P_1, P_2$  and  $P_3$  are not isomorphic, although they have the same dimension. One can easily see that  $P_1, P_2, P_3$  have 256, 384, 432 vertices respectively. Similarly,  $\tilde{P}$  and  $\tilde{P}'$  have 512 and 576 vertices respectively.*

Let us consider a Markov basis element  $B$  of  $P$ . We show that one of the following holds:

- (i)  $B$  has either degree less than or equal to four;
- (ii)  $B$  has  $\deg(B) > 16$ , which is not possible by Proposition 3.25;
- (iii)  $B$  is a Markov basis element of  $\tilde{P}$  or  $\tilde{P}'$  of degree at least 15, which is not possible by Proposition 3.25;
- (iv)  $B$  is a Markov basis element for a polytope isomorphic to  $P_1, P_2$  or  $P_3$  (in this case, it has degree at most four by Lemma 3.26).

**Proposition 3.28.** *Any Markov basis element  $B$  for  $P$  has degree at most four.*

*Proof.* It is enough to restrict to the main case  $(\star)$ . We first prove two claims, Claim (i) and (ii). Claim (i): For any row  $r$  of  $T_0$  distinct from the first one, for any pair of indices  $2 < i < j \leq 6$ , we have that either  $\phi_\gamma(r_{0\beta}(i)) = \phi_\gamma(r(i))$  or  $\phi_\gamma(r_{0\beta}(j)) = \phi_\gamma(r(j))$ . The analogous statement holds for  $T_1$ , with the group homomorphism  $\phi_\gamma$  replaced by  $\phi_\beta$ .

*Proof of Claim (i).* Suppose the statement is not true for some pair of indices  $i, j$ . If  $r_{0\beta}(i) - r(i) = r_{0\beta}(j) - r(j)$ , then we can make a quadratic move on  $i, j$ , and conclude using the Difference Lemma 3.17. Thus, without loss of generality, we may assume  $r_{0\beta}(i) - r(i) = \alpha$  and  $r_{0\beta}(j) - r(j) = \beta$ . If there exists another index  $2 < k \leq 6$  such that  $r_{0\beta}(k) - r(k) \neq 0$ , then we can make a move on a subset of  $\{i, j, k\}$  and, again, conclude by the means of the Difference Lemma 3.17. In conclusion,  $\sum_{l=3}^6 r_{0\beta}(l) = \alpha + \beta + \sum_{l=3}^6 r(l)$ . As  $r$  and  $r_{0\beta}$  are flows and  $r_{0\beta}(1) + r_{0\beta}(2) = \beta$ , this contradicts Lemma 3.23, which prescribes the first two columns of a row differing *not* by  $\gamma$  with  $r_{0\beta}$ .  $\square$

By Proposition 3.5 and Lemma 3.20, we may assume that  $r_{0\beta}(6) = r_{\alpha\beta}(6) = 0$ , as the disagreement string between the two rows has length at most three, outside the first two columns.

Claim (ii): There exists at most one index  $i > 2$  such that  $\phi_\gamma(r_{0\beta}(i)) \neq 0$ .

*Proof of Claim (ii).* As the number of such indices must be odd it is enough to prove that not all  $r_{0\beta}(3), r_{0\beta}(4), r_{0\beta}(5)$  are equal to  $\alpha$  or  $\beta$ . Not all can be equal to  $\beta$  since, by Lemma 3.24, that would contradict the fact that  $r_{\alpha\beta}$  is a flow. Say  $r_{0\beta}(3) = \beta$  and  $r_{0\beta}(4) = r_{0\beta}(5) = \alpha$ . Then we have  $r_{\alpha\beta}(3) = 0$  by Lemma 3.24 and thus  $r_{\alpha\beta}(4) + r_{\alpha\beta}(5) = \gamma$ . However, we may exclude  $\{r_{\alpha\beta}(4), r_{\alpha\beta}(5)\} = \{\alpha, \beta\}$  by Lemma 3.20 and we may exclude  $\{r_{\alpha\beta}(4), r_{\alpha\beta}(5)\} = \{0, \gamma\}$  by Lemma 3.24.  $\square$

To continue our proof, we need to introduce some terminology, which we will use only here. A column index  $1 \leq i \leq 6$  is of type:

- (a) if all elements of  $G$  appear in the corresponding  $i$ th column of  $T_0$  (and of  $T_1$ );
- (b) if exactly three elements of  $G$  appear in the  $i$ th column;
- (c) if exactly two elements of  $G$  appear in the  $i$ th column;
- (d) if exactly one element of  $G$  appears in the  $i$ th column.



**Step 0:** We suppose that all columns are of type (a).

By Claim (ii), there exists one index  $j > 2$  such that  $\phi_\gamma(r_{0\beta}(j)) \neq 0$ . For  $i > 2$ ,  $i \neq j$  there must exist at least two rows  $r_{i,1}, r_{i,2}$  such that  $r_{i,1}(i) = r_{0\beta}(i) + \alpha$  and  $r_{i,2}(i) = r_{0\beta}(i) + \beta$ . Note that  $r_{i,1}, r_{i,2}$  are not the first row. Further, for the index  $j$  there must exist one row  $r_{j,1}$  different from the first one such that  $r_{j,1}(j) = r_{0\beta}(j) + \alpha$  or  $r_{j,1}(j) = r_{0\beta}(j) + \beta$ . All these rows are distinct by Claim (i). Hence, we obtain seven rows; we call them *difference rows* for  $T_0$ . Note that the *difference rows* for  $T_0$  may only have  $\alpha$  and  $\beta$  in columns 1, 2 by Lemma 3.23. Analogously, we obtain at least seven *difference rows* in  $T_1$ , with copies of 0 or  $\beta$  in columns 1 and 2.

If there exist *difference rows* in  $T_0$  and  $T_1$  with  $\beta\beta$  in the first two columns, then we obtain the string  $\beta\beta 0$  in both tables and we conclude by induction on the degree of  $B$ .

Thus suppose that there is no string  $\beta\beta$  in columns 1, 2 of  $T_1$ . It follows that there must be at least seven copies of 0 in columns 1, 2 in the *difference rows* of  $T_1$ . Consequently, there are at least nine copies of 0 in columns 1, 2 in  $T_1$ . By Lemma 3.22, there is no string 00 in columns 1, 2 in  $T_0$ , and the *difference rows* for  $T_0$  do not have copies of 0 in columns 1, 2. In conclusion, we have at least this amount of distinct rows in  $T_0$ :

- (i) three, that are the first ones;
- (ii) seven, that are the *difference rows*;
- (iii) seven, that contain copies of 0 in column 1 or 2;
- (iv) two, that have  $\gamma$  in column 1 or 2.

Then, we have  $\deg(B) > 18$ . This is impossible for a Markov basis element by Proposition 3.25.

**Step 1:** We suppose that there exists exactly one column of type (b) and all others are of type (a). We may proceed as before, however we obtain only six *difference rows* in the case when the column of type (b) has column index  $3 \leq i \leq 6$ . In the case when the column index of the column of type (b) is either 1 or 2, we obtain seven *difference rows*, but we cannot assume that there exists an additional row with  $\gamma$  in the same column index of the column of type (b). In either of these cases, we have  $\deg(B) \geq 3 + 2 \times 6 + 2 = 17$ , that contradicts Proposition 3.25.

**Step 2:** We suppose that there exist exactly two columns of type (b) (resp. one column of type (c)). Here, we obtain five *difference rows*. However,  $B$  represents a Markov element for  $\tilde{P}'$  (resp.  $\tilde{P}$ ), whose ideals have regularity 14; see Appendix 4 for the definition of the associated ideal. We obtain the bound  $\deg B \geq 3 + 2 \times 5 + 2 = 15 > 14$  which contradicts Proposition 3.25.

**Step 3:** We suppose there exist either:

- (i) three columns of type (b)), or
- (ii) one column of type (b) or (c) and one column of type (c), or
- (iii) one column of type (d).

In such cases we conclude by Lemma 3.26. □

### 3.3.4 Proof of the main case

In this last part, we finish our proof dealing with the main case ( $\star$ ). This will be done uniformly, i.e., with the same arguments in all the three instances of the main case and only technical details differ. Here the number of leaves is  $n \geq 7$ . The outline is as follows:

- (i) We show that, if  $r_{0\beta}(i) = r_{\alpha\beta}(i)$ , then we have  $r_{0\beta}(i) = r_{\alpha\beta}(i) = 0$ ;
- (ii) Among the pairs of tables we consider (tables where we have fixed the first two entries of the rows  $r_{0\beta}$  and  $r_{\alpha\beta}$  and performed moves of degree at most four so that  $r_{0\beta}$  and  $r_{\alpha\beta}$  have the agreement string as large as possible) using at most moves of degree four, we attain the situation where the number of *bad pairs*, i.e., strings  $xy$ , with  $x, y \neq 0$ , in columns  $n - 1$  and  $n$  is as small as possible;
- (iii) We show that we can kill all the *bad pairs*, i.e., we can make moves of degree at most four killing all of them. Summing up the two columns indexed by  $n - 1$  and  $n$  allows us to conclude by induction on the number of leaves  $n$ ; see Theorem 3.4.

We are now ready to establish the main case in the following lemmas.

**Lemma 3.29.** *We may assume that no rows in  $T_0$  has the string  $\alpha\alpha$  or  $\beta\beta$  in columns  $n - 1$  and  $n$ . Analogously, no row in  $T_1$  has the string  $\alpha\alpha$  or  $\gamma\gamma$  in columns  $n - 1$  and  $n$ .*

*Proof.* In such a case we make a quadratic move in columns  $n - 1$  and  $n$  and we conclude by applying the Difference Lemma 3.17.  $\square$

**Lemma 3.30.** *We may assume that no row in  $T_0$  has the string  $\alpha\beta$  or  $\beta\alpha$  in columns  $n - 1$  and  $n$ . Analogously, no row in  $T_1$  has the string  $\alpha\gamma$  or  $\gamma\alpha$  in columns  $n - 1$  and  $n$ .*

*Proof.* Let  $r$  be such a row with such a string in columns  $n - 1$  and  $n$  of  $T_0$ . If for some other column index  $i > 2$ , we have  $r(i) \neq r_{0\beta}(i)$  then we may exchange  $i$  and a nonempty subset of elements under the agreement string. Then we conclude by applying the Difference Lemma 3.17. As  $r$  is a flow, we have  $r(1) + r(2) = \alpha$ . This contradicts Lemma 3.23.  $\square$

**Lemma 3.31.** *We may assume that under the agreement string no row in  $T_0$  has  $\gamma\gamma$ . Analogously, no row in  $T_1$  has  $\beta\beta$ .*

*Proof.* Let  $r$  be a row in  $T_0$  with  $\gamma\gamma$  under the agreement string. We first claim we may assume that  $r_{\alpha\beta}$  does not have  $\gamma$  in any column. For the sake of contradiction, suppose  $r_{\alpha\beta}(i) = \gamma$  for some column index  $i$ . Whence, by Lemma 3.24, we have  $r_{0\beta}(i) = 0$ . By compatibility of the tables  $T_0$  and  $T_1$ , there exists a row  $r'$  in  $T_0$  with  $r'(i) = \gamma$ . By the Standard Lemma 3.19, we can make a move to obtain  $r_{0\beta}(i) = \gamma$  and conclude by applying Lemma 3.20.

We divide the rest of the proof into two steps according to whether or not there exists  $\beta$  in  $r_{0\beta}$ .

**Step 1:** Suppose there exists another  $\beta$  in  $r_{0\beta}$ . The tables  $T_0$  and  $T_1$  are the following:

$$T_0 - T_1 = \begin{bmatrix} \alpha & \alpha & 0 & \dots & 0 & 0 \\ 0 & \beta & \beta & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \gamma & \gamma \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ \alpha & \beta & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

By Lemma 3.23, the counting function  $\alpha_3 + \beta_3 - 0_{12} - \gamma_{12}$  is nonnegative on  $T_0$ . Let  $r$  be a row in  $T_1$ , where the function is strictly positive. We now exclude the case  $r(3) = \beta$ . Indeed, in this case, if  $r(2) = \beta$ , then we obtain  $0\beta\beta$  in both tables. If  $r(2) = \alpha$ , by the positivity of the counting function on  $r$ , we have  $r(1) = \beta$  and we may perform a quadratic move to obtain the string  $0\beta\beta$ . Whence  $r(3) = \alpha$ .

Let  $r'$  be a row in  $T_0$  with  $r'(3) = \alpha$ . By the Standard Lemma 3.19, we can make a move between  $r'$  and  $r_{0\beta}$  involving this entry. In particular, if  $r(1) + r(2) = \gamma$ , then we make a quadratic move between  $r$  and  $r_{\alpha\beta}$  on first two entries and conclude by Lemma 3.20. Thus  $r(1) = r(2) = \beta$ . Let  $r''$  be a row in  $T_1$  with  $r''(3) = \beta$ . We finish the proof of **Step 1** by proving that we can always obtain  $0\beta\beta$  in  $T_1$ . First, suppose  $r''(j) = \alpha$  for  $j = 1$  or  $2$ . Then we may exchange  $r''$  with  $r$  on column indices  $j$  and  $3$ , obtaining a row  $\tilde{r}$  such that  $\tilde{r}(1) + \tilde{r}(2) = \gamma$  and  $\tilde{r}(3) = \beta$ . Then we can make a quadratic move between  $\tilde{r}$  and  $r_{\alpha\beta}$  to obtain  $0\beta\beta$  in both tables. Also, notice that  $r''(2) \neq \beta$  as this immediately leads to  $0\beta\beta$  in both tables. If  $r''(1) = \gamma$  we may exchange  $r''$  and  $r_{\alpha\beta}$  on column indices  $1$  and  $3$ , obtaining  $0\beta\beta$  in both tables. If  $r''(1) = r''(2) = 0$ , we can make a quadratic move between  $r''$  and  $r$ . Similarly, if  $r''(1) = 0$  and  $r''(2) = \gamma$  we can make an exchange with  $r_{\alpha\beta}$ . If  $r''(1) = \beta$  and  $r''(2) = 0$  we first exchange it with  $r_{\alpha\beta}$  on column indices  $2, 3$ , then we apply  $\alpha 0\beta + \beta\beta\alpha = \beta 0\alpha + \alpha\beta\beta$ . Finally, if  $r''(1) = \beta$  and  $r''(2) = \gamma$ , we apply the cubic move

$$000 + \beta\beta\alpha + \beta\gamma\beta = 0\beta\beta + \beta\gamma 0 + \beta 0\alpha.$$

**Step 2:** Suppose there is no  $\beta$  in  $r_{0\beta}$ ; without loss of generality we may assume we have  $\alpha$  and  $\gamma$  in columns  $3, 4$ . In column  $3$ , in row  $r_{\alpha\beta}$  of  $T_1$  we cannot have  $\alpha$  by Lemma 3.20; moreover, we cannot have  $\beta$  by the Standard Lemma 3.19 applied to table  $T_0$ , as we would produce  $\beta$  in the row  $r_{0\beta}$ , contradicting Lemma 3.20. Thus we have  $0$  in column  $3$  in the row  $r_{\alpha\beta}$ , since  $\gamma$  is excluded in row  $r_{\alpha\beta}$  by the claim in the very first part of the proof. Since the disagreement string has length at most three by Corollary 3.6, we have the following tables  $T_0$  and  $T_1$ :

$$T_0 - T_1 = \begin{bmatrix} \alpha & \alpha & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \beta & \alpha & \gamma & x & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \gamma & \gamma \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \alpha & \beta & 0 & y & z & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

Furthermore  $\{y, z\} = \{\alpha, \beta\}$ . By the disagreement string length, we have  $x = 0$ . Consider the group morphism  $\phi_\gamma : G \rightarrow \mathbb{Z}_2$  and apply it to columns  $3, 4, 5$ . Note that the evaluation of  $r_{0\beta}$  under  $\phi_\gamma$  in column indices  $3, 4, 5$  is the  $0/1$  vector  $(1, 0, 0)$ . We claim that no row of  $T_0$  can differ by more than one element with respect to  $r_{0\beta}$  in column indices  $3, 4, 5$ . Indeed, suppose a row  $r$  in  $T_1$  differs on  $i, j \in \{3, 4, 5\}$ . Then  $r(i) + r(j) - r_{0\beta}(i) - r_{0\beta}(j) \in \{0, \gamma\}$ . Thus, by the Standard Lemma 3.19, we can make a quadratic move on  $i, j$  and conclude by Difference Lemma 3.17. By double counting, there must exist a row  $r'$  in  $T_1$  such that  $r'(3) \in \{\alpha, \beta\}$  and  $r'(4), r'(5) \in \{0, \gamma\}$ . By a quadratic move and the claim at the very first part of the proof, we may assume  $r'(4) = r'(5)$ . Now we can make a quadratic move between  $r'$  and  $r_{\alpha\beta}$  involving the entry in column  $3$  and the entry in either column  $4$  or  $5$ . However, we may conclude as in the first part of **Step 2**.  $\square$

**Lemma 3.32.** *We may assume that under the agreement string no row in  $T_0$  has  $\alpha\gamma$  or  $\gamma\alpha$  (resp.  $\beta\gamma$  or  $\gamma\beta$ ).*

*Proof. Step 0:* Assume that there exists  $\beta$  in  $r_{0\beta}$  and  $\gamma$  in  $r_{\alpha\beta}$ ; without loss of generality we may assume that they are in columns  $3, 4$ . In this case the tables are:

$$T_0 - T_1 = \begin{bmatrix} \alpha & \alpha & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & \beta & \beta & 0 & \dots & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \alpha(\text{resp. } \beta) & \gamma \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ \alpha & \beta & 0 & \gamma & \dots & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

Let  $r$  be the row in  $T_0$  that contains the string  $\alpha\gamma$  (resp.  $\beta\gamma$ ) under the agreement string. By Lemma 3.20 and Lemma 3.23, we see that  $r(4) = 0$ . Let  $r'$  be a row in  $T_0$  such that  $r'(4) = \gamma$ . By Lemma 3.20, we can exclude  $\gamma$  under the agreement string. Furthermore, performing a quadratic move, we notice that if  $r'$  has  $00$  under the agreements string, we could reduce  $\alpha\gamma$  (resp.  $\beta\gamma$ ) to  $\alpha 0$  (resp.  $\beta 0$ ), contradicting the minimality of the number of *bad pairs*. Also,  $r'$  cannot have  $0\beta$  or  $\beta 0$  (resp.  $0\gamma$  or  $\gamma 0$ ) under the agreement string, as we could exchange it with  $\alpha\gamma$  (resp.  $\beta\gamma$ ) and conclude as before. Thus, under the agreement string,  $r'$  has either the string  $0\alpha$  or  $\alpha 0$  (resp.  $0\beta$  or  $\beta 0$ ). Now, Lemma 3.20 and Lemma 3.23 allow us to conclude that  $r'(3) = \beta$ . Hence, the counting function  $\beta_3 - \gamma_4$  is strictly positive in  $T_0$ . Let  $r''$  be a row in  $T_1$  such that  $r''(3) = \beta$  and  $r''(4) \neq \gamma$ . By Lemma 3.20, we may exclude  $\alpha$  in column 4 in  $r''$ . Consequently, by Lemma 3.23,  $r''$  has either  $0$  or  $\beta$  in column 2. If  $r''(2) = \beta$ , we obtain the same string  $0\beta\beta 0$  in both tables. If  $r''(2) = 0$ , we obtain the string  $\alpha 0\beta\gamma 0$  in  $T_1$ ; we now show we may also obtain it in  $T_0$ . We discuss this according to the three crucial cases:

- (i) **Case I and II:** We apply the move  $\alpha\alpha 00 + 0\beta\beta 000 + \alpha 0 + ??\beta\gamma xy = \alpha 0\beta\gamma + \alpha\beta\beta 0xy + 0\alpha + ??0000$ , where  $xy$  is under the agreement string and  $x + y = \alpha$ . (resp. We consider the first two entries of  $r'$ , which by Lemma 3.23 could be:  $\alpha\alpha, \alpha\beta, \beta\alpha, \beta\beta$ . The last three allow to obtain  $\alpha\beta 0\gamma$  in both tables. As  $r'$  must agree on all nonspecified entries with  $r_{0\beta}$  this contradicts the fact that  $r'$  is a flow.);
- (ii) **Case III:** We apply the move  $\alpha\alpha 00 + 0\beta\beta 000 + \beta 0 + ??\beta\gamma xy = \alpha 0\beta\gamma + \beta\alpha\beta 0xy + 0\beta + ??0000$  where  $x + y = \alpha$ . (resp. We proceed as before, noting that we do not use the third row, except for  $\beta\alpha$ , in which case we obtain  $\beta\alpha 0\gamma$  in both tables).

**Step 1:** Assume there exists  $\beta$  in  $r_{0\beta}$  and no  $\gamma$  in  $r_{\alpha\beta}$ . The tables are:

$$T_0 - T_1 = \begin{bmatrix} \alpha & \alpha & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \beta & \beta & x & y & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \alpha(\text{resp. } \beta) & \gamma \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \alpha & \beta & 0 & \alpha & \beta & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

As the disagreement string is of length at most three, we must have  $x = y$ . Further, by Lemma 3.20  $x = y = 0$  or  $x = y = \gamma$ . Consider the group morphism  $\phi_\gamma : G \rightarrow \mathbb{Z}_2$ . We claim that after applying  $\phi_\gamma$  to column indices 3, 4, 5, no row can differ on more than one index from  $\phi_\gamma((\beta, x, x)) = (1, 0, 0)$ . Indeed, if a row  $\tilde{r}$  differs on two indices  $i, j$ , then, by the Difference Lemma 3.17, we may assume  $r_{0\beta}(i) = \tilde{r}(i) + \alpha$  and  $r_{0\beta}(j) = \tilde{r}(j) + \beta$ . The rows  $r$  and  $\tilde{r}$  must differ by  $\alpha$  either in column index  $i$  or  $j$ , and by  $\beta$  on the other. In particular, by reducing the number of *bad pairs*  $\alpha\gamma$  (resp.  $\beta\gamma$ ) under the agreement string, we exclude the situation when  $\tilde{r}$  has  $00$  under the agreement string. By the Difference Lemma 3.17, we also know that  $\gamma$  does not appear in  $\tilde{r}$  under the agreement string. In the same way, if  $\alpha$  or  $\beta$  appears under the agreement string, we may exchange it along with the index  $i$  or  $j$ , again contradicting Difference Lemma 3.17. By double counting, there exists a row  $\tilde{r}'$

in  $T_1$  such that  $\phi_\gamma((\tilde{r}'(3, 4, 5))) = (1, 0, 0)$ . In particular, there exist two indices such that we can make a quadratic move between  $\tilde{r}'$  and  $r_{\alpha\beta}$ . This either contradicts Lemma 3.24 or one decreases the Hamming distance.

**Step 2:** Assume there is no  $\beta$  in  $r_{0\beta}$  and there exists  $\gamma$  in  $r_{\alpha\beta}$ . The tables are:

$$T_0 - T_1 = \begin{bmatrix} \alpha & \alpha & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \beta & \alpha & \gamma & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \alpha(\text{resp. } \beta) & \gamma \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \alpha & \beta & x & y & \gamma & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

As before  $x = y$  equals  $\beta$  or  $0$ . Further  $r(5) = 0$ . We apply  $\phi_\beta$  to column indices  $3, 4, 5$ .

We claim we may assume that no row  $\tilde{r}$  in  $T_0$  differs from  $\phi_\beta((\alpha, \gamma, 0)) = (1, 1, 0)$  on more than one index. For the sake of the contradiction, suppose there exists  $\tilde{r}$  in  $T_0$  differing on  $i$  and  $j$ . If  $r_{0\beta}(i) - \tilde{r}(i) = r_{0\beta}(j) - \tilde{r}(j)$ , then we make a quadratic move between  $r_{0\beta}$  and  $\tilde{r}$  on  $i, j$ . If the difference equals  $\alpha$ , we conclude by the Difference Lemma 3.17. Thus we assume the difference equals  $\gamma$ . If  $5 \in \{i, j\}$  we conclude by Lemma 3.20. Hence,  $\{i, j\} = \{3, 4\}$ ; on the other hand, this reduces the Hamming distance. Consequently we have  $r_{0\beta}(i) - \tilde{r}(i) = \alpha$  and  $r_{0\beta}(j) - \tilde{r}(j) = \gamma$ . Notice that we cannot have  $r(i) - \tilde{r}(i) = r(j) - \tilde{r}(j) = \alpha$ , thus at least one difference must be equal to  $\gamma$ . Hence, we exclude  $00$  in  $\tilde{r}$  under the agreement string, as then we could reduce the number of  $\alpha\gamma$  (resp.  $\beta\gamma$ ) under the agreement string. Further,  $\alpha$  and  $\beta$  also cannot appear under the agreement string, as otherwise we may conclude by the Difference Lemma 3.17. Whence  $\tilde{r}$  has  $0\gamma$  or  $\gamma 0$  under the agreement string. By Lemma 3.20, we have  $j \neq 5$ . Let  $\tilde{r}'$  be a row of  $T_0$  with  $\tilde{r}'(5) = \gamma$ . As before, we conclude that  $\tilde{r}'$  has  $\alpha 0$  or  $0\alpha$  under the agreement string (resp.  $0\beta$  or  $\beta 0$ ), and  $\tilde{r}'(3) = \alpha$ ,  $\tilde{r}'(4) = \gamma$ . We now exclude the case  $i = 5$ , i.e.,  $\tilde{r}(5) = \alpha$ . In such a case, we could exchange  $r$  and  $\tilde{r}$  on column 5 and under the agreement string; then with  $\tilde{r}'$  on column indices 5 and  $j$ ; finally with  $r_{0\beta}$  on column indices 5 and the last entry to conclude by Lemma 3.20. (Resp. We apply the relation on 5 and the agreement string  $000 + \alpha 0\gamma + \gamma 0\beta = \gamma 0\gamma + 00\beta + \alpha 00$ .)

In conclusion, our discussion leads to  $\{i, j\} = \{3, 4\}$  and  $\tilde{r}(5) = 0$ . However, we may exchange  $\tilde{r}$  with  $\tilde{r}'$  on 5 and  $j$ . Consequently we exchange with  $r_{0\beta}$  on 5 and under the agreement string to conclude by Lemma 3.20. This concludes the verification of our claim.

By the claim, there must exist a row  $r''$  in  $T_1$ , such that  $\phi_\beta(r''((3, 4, 5))) = (1, 1, 0)$ . On two of these indices,  $r''$  differs from  $r_{\alpha\beta}$  by the same element: either  $\alpha$  or  $\gamma$ . We can make a quadratic move on these two column indices and conclude by Difference Lemma 3.17.

**Step 3:** Assume there is no  $\beta$  in  $r_{0\beta}$  and no  $\gamma$  in  $r_{\alpha\beta}$ . The tables are:

$$T_0 - T_1 = \begin{bmatrix} \alpha & \alpha & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \beta & \alpha & \gamma & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \alpha(\text{resp. } \beta) & \gamma \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \alpha & \beta & x & y & z & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

Suppose  $x = \beta$ . Let  $\tilde{r}$  be a row of  $T_0$  with  $\tilde{r}(3) = \beta$ . As in the previous steps, we may assume that  $\tilde{r}$  has  $0\alpha$  or  $\alpha 0$  (resp.  $0\beta$  or  $\beta 0$ ) under the agreement string and  $\tilde{r}(i) = r_{0\beta}(i)$  for  $4 \leq i \leq n-3$ . By Lemma 3.23, we have  $\tilde{r}(1, 2) = \alpha\alpha$  or  $\beta\beta$  (resp.  $\tilde{r}(1, 2) = \alpha\beta$  or  $\beta\alpha$ ); we may obtain  $0\beta\beta$  in both tables by the move:  $\alpha\alpha 0 + 0\beta\alpha\gamma 00 + (\alpha\beta/\beta\alpha)\beta\gamma 0\beta = 0\beta\beta + (\alpha\beta/\beta\alpha)0\gamma 00 + \alpha\alpha\alpha\gamma 0\beta$ . However,  $\beta\beta$  easily leads to  $0\beta\beta$  in both tables by the cubic move  $\alpha\alpha 0 + 0\beta\alpha + \beta\beta\beta = 0\beta\beta + \beta\alpha 0 + \alpha\beta\alpha$  in

$T_0$ . Furthermore, we may assume that  $\beta\beta$  does not appear on column indices 1, 2 in any row in  $T_0$ , otherwise we would exchange with  $\tilde{r}$  obtaining  $\beta\beta\beta$  in columns 1, 2, 3. It follows that  $\alpha_{12} - 0_3 - \beta_3$  is positive on  $T_0$ . However, a positive row in  $T_1$  contradicts Lemma 3.23.

Thus we may assume  $x = 0$ . Without loss of generality  $\{y, z\} = \{\alpha, \beta\}$ . We apply the homomorphism  $\phi_\gamma$  to column indices 3, 4, 5. We prove that no row may differ on two indices from  $\phi_\gamma(r_{0\beta}(3, 4, 5)) = (1, 0, 0)$  in  $T_0$ . This is analogous to **Step 1**. Whence there exists a row  $\tilde{r}$  in  $T_1$ , such that  $\phi_\gamma(\tilde{r}((3, 4, 5))) = (1, 0, 0)$ . We may assume  $\tilde{r}(4) = \tilde{r}(5)$ , as otherwise we can make a quadratic move on column indices 4, 5 and conclude by previous steps. However, in such a case we may exchange  $\tilde{r}$  with  $r_{0\beta}$  (on column index 3 and on column index either 4 or 5), conclude by Lemma 3.20 or reduce to the first part of this step, where we assume  $x = \beta$ .  $\square$

**Lemma 3.33.** *We may assume that under the agreement string no row in  $T_1$  has  $\alpha\beta$  or  $\beta\alpha$  (resp.  $\beta\gamma$  or  $\gamma\beta$ ).*

*Proof.* Let us act on tables  $T_0, T_1$  by the flow  $(\alpha, \alpha, 0, \dots, 0) \in \mathfrak{G}$  and then apply the group automorphism  $\beta \leftrightarrow \gamma$ . This translates **Case I** and **Case III** to **Case III** and **Case I** of Lemma 3.32 respectively; cf. Remark 3.21. However, **Case II** is not transformed to the previous cases, due to the rows  $r_{\alpha 0}$  in  $T_0$  and  $r_{\beta\alpha}$  in  $T_1$ . We note that in **Steps 1, 2, and 3** of Lemma 3.32 we are only using the rows  $r_{0\beta}$  in  $T_0$  and  $r_{\alpha\beta}$  in  $T_1$  that still appear after translating **Case II**.

Thus, we only need to conclude in **Case II** and **Step 0**, i.e., there exists  $\beta$  in  $r_{0\beta}$  and  $\gamma$  in  $r_{\alpha\beta}$ . Without loss of generality, we may assume that they are in columns 3, 4. The tables are:

$$T_0 - T_1 = \begin{bmatrix} \alpha & \alpha & 0 & 0 & \dots & 0 & 0 \\ 0 & \beta & \beta & 0 & \dots & 0 & 0 \\ \alpha & 0 & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \alpha & \beta & 0 & \gamma & \dots & 0 & 0 \\ \beta & \alpha & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \alpha(\text{resp. } \gamma) & \beta \end{bmatrix}.$$

Let  $r$  be the row in  $T_1$  with a bad pair of the form  $\alpha\gamma$  (resp.  $\beta\gamma$ ). First we exclude  $r(3) = \alpha, \beta, \gamma$  by quadratic exchange with  $r_{\alpha\beta}$ , and the Difference Lemma 3.17 and Lemma 3.20. Let  $\tilde{r}$  be the row in  $T_1$  such that  $\tilde{r}(3) = \beta$ . We note that if  $\tilde{r}(n-1) = \tilde{r}(n) = 0$  then, exchanging with  $r$  we could reduce the number of bad pairs. Moreover, by Lemma 3.20 we know that  $\tilde{r}(n-1), \tilde{r}(n) \neq \beta$ . Furthermore, as we already know that  $r(3)$  must be equal to zero, we have  $\tilde{r}(n-1) + \tilde{r}(n) \neq r(n-1) + r(n)$ . Thus, we must have  $\{\tilde{r}(n-1), \tilde{r}(n)\} = \{0, \alpha\}$  (resp.  $\{0, \gamma\}$ ). Note that  $\tilde{r}(2) = \beta$  gives  $0\beta\beta$  in both tables, thus we may assume  $\tilde{r}(2) = 0$ , by Lemma 3.23. Moreover, we have  $\tilde{r}(4) = \gamma$  and hence the counting function  $\beta_3 - \gamma_4$  is negative on  $T_1$ . Let  $r'$  be the row in  $T_0$  on which the function is negative, i.e.,  $r'(4) = \gamma$  and  $r'(3) \neq \beta$ . Now,  $r'(3) \neq \alpha$ , as otherwise we exchange  $r'$  and  $r_{0\beta}$  and conclude by Lemma 3.20. Thus, by Lemma 3.23, we have  $r'(1), r'(2) \in \{\alpha, \beta\}$ . If  $r'(2) = \beta$  then we obtain  $\alpha\beta 0\gamma$  in both tables, thus we may assume  $r'(2) = \alpha$ . We may obtain the flow  $0\alpha\beta\gamma$  in  $T_0$ , by exchanging  $r'$  and  $r_{0\beta}$ , and  $\alpha 0\beta\gamma$  in  $T_0$ , by exchanging with  $r_{\alpha 0}$ . We finish the proof by showing that we may obtain the latter in  $T_1$ , by the quadratic move  $\alpha\beta 0\gamma + ?0\beta = \alpha 0\beta\gamma + ?\beta 0$ .  $\square$

## 4 Appendix

We present known algebraic results for algebras over monoids that are cones over normal lattice polytopes. Much more information can be found in [5, 22, 36, 43, 46].

Let  $M$  be a lattice and  $P \subset \{1\} \times M \subset \mathbb{Z} \times M$  be a *normal* lattice polytope generating the

ambient lattice. Let  $C(P) \subset \mathbb{Z} \times M$  be the cone over  $P$ . The cone  $C(P)$ , equipped with addition, has a natural structure of a graded monoid, with the grading induced by the first coordinate. The algebraic properties of the graded algebra  $\mathbb{C}[C(P)]$  are strongly related to combinatorial properties of  $P$ .

**Proposition 4.1.** *The function  $H_P : \mathbb{Z}_{\geq 0} \rightarrow \mathbb{Z}$  defined by  $H_P(n) = |nP \cap \{n\} \times M|$  is a polynomial known as Ehrhart polynomial. For all  $n \geq 0$ , it coincides with the Hilbert function (and hence with the Hilbert polynomial) of the algebra  $\mathbb{C}[C(P)]$ . Moreover, it satisfies the Ehrhart reciprocity, i.e.  $|H_P(-n)| = |\text{int}(nP) \cap \{n\} \times M|$  for  $n > 0$ , where  $\text{int}$  denotes the interior points of the polytope.*

We immediately see that the polynomial  $H_P(n)$  may agree with the Hilbert function even for negative  $n$ . This happens if and only if  $H_P(n) = 0$ , as the algebra is positively graded.

**Definition 4.2 (a-invariant, Hilbert regularity).** *The a-invariant  $a(A)$  of an algebra  $A$  is the largest integer  $a$  such that the Hilbert function differs from the Hilbert polynomial. Hilbert regularity equals the a-invariant plus one.*

**Corollary 4.3.** *The a-invariant of  $\mathbb{C}[C(P)]$  is always negative. It equals  $-n$  for the smallest  $n \in \mathbb{Z}_{>0}$  such that  $nP$  contains an interior point.*

**Proposition 4.4.** *If  $\dim P = d$ , then  $\sum_{j=0}^{\infty} H_P(j)t^j = h(t)/(1-t)^{d+1}$  for some polynomial  $h$ . The a-invariant of  $\mathbb{C}[C(P)]$  equals  $\deg h - d - 1$ .*

We note that  $d + 1 - \deg h$  is the smallest dilation of  $P$  that contains an interior lattice point.

**Proposition 4.5 (Hochster's Theorem).** *The algebra  $\mathbb{C}[C(P)]$  is Cohen-Macaulay.*

Throughout the article we were interested in generators of the ideal  $I$  such that  $\mathbb{C}[C(P)] = \mathbb{C}[x_p : p \in P \cap M]/I = S/I$ . These are usually very hard to understand even for specific instances. However, there is an algebraic invariant that bounds their degree, known as *Castelnuovo-Mumford regularity*, or simply, the *regularity*.

**Definition 4.6 (Castelnuovo-Mumford regularity).** *For an  $S$ -module  $M$  its regularity  $\text{reg}(M)$  is defined as*

$$\text{reg}(M) = \max\{j - i : b_{ij} \neq 0\},$$

where

$$0 \leftarrow M \leftarrow \bigoplus_j S(-j)^{b_{0j}} \leftarrow \dots \leftarrow \bigoplus_j S(-j)^{b_{ij}} \leftarrow \dots \leftarrow 0$$

is the minimal free resolution of  $M$ .

As  $I$  is an  $S$  module, its regularity in particular bounds the degree of generators; this is the case  $i = 0$  in the definition. It can be seen that  $\text{reg}(\mathbb{C}[C(P)])$  is the maximal degree of standard monomials under rev-lex in generic coordinates. Hence  $\text{reg}(I)$  bounds the degree of such a Gröbner basis, as  $\text{reg}(S/I) + 1 = \text{reg}(I)$ . The following proposition relates both notions of regularity introduced above.

**Proposition 4.7.**  *$a(M) \leq \text{reg}(M) - \text{depth}(M)$  and equality holds if  $M$  is Cohen-Macaulay. In particular,  $\text{reg}(\mathbb{C}[C(P)]) = \deg h$  and  $I$  is generated in degree at most  $1 + \deg h$ .*

#### Acknowledgements.

Mateusz Michałek was supported by Polish National Science Centre grant no. 2015/19/D/ST1/01180, the Foundation for Polish Science (FNP) and is a member of AGATES group. The authors acknowledge the kind hospitality of UC Berkeley and FU Berlin, where this research was in part conducted.

## References

- [1] Elizabeth S. Allman, *Open Problem: Determine the Ideal Defining  $\sigma_4(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3)$* , Available on-line (<http://www.dms.uaf.edu/~eallman/Papers/salmonPrize.pdf>), 2010.
- [2] Quentin Atkinson and Russell D. Gray, *Curious Parallels and Curious Connections—Phylogenetic Thinking in Biology and Historical Linguistics*, *Systematic biology* **54**(4) (2005): 513–526.
- [3] Adrian C. Barbrook et al., *The Phylogeny of the Canterbury Tales*, *Nature* **394**(6696)(1998), 839.
- [4] Louis J. Billera, Susan P. Holmes, and Karen Vogtmann, *Geometry of the space of phylogenetic trees*. *Adv. in Appl. Math.*, **27**(4):733–767, 2001.
- [5] Winfried Bruns and Joseph Gubeladze, *Polytopes, Rings, and K-Theory*, Springer Monographs in Mathematics, Springer, 2009.
- [6] Weronika Buczyńska, Maria Donten-Bury, and Jarosław A. Wiśniewski, *Isotropic models of evolution with symmetries*, *Contemporary Mathematics* **496** (2009), 111–132.
- [7] Winfried Bruns, Richard Sieg, Tim Römer, and Christof Söger, *Normaliz*, <http://www.home.uni-osnabrueck.de/wbruns/normaliz/> (2001).
- [8] Weronika Buczyńska and Jarosław A. Wiśniewski, *On geometry of binary symmetric models of phylogenetic trees*, *J. Eur. Math. Soc.* **9**(3) (2007), 609–635.
- [9] Marta Casanellas, Jesús Fernández-Sánchez and Mateusz Michałek, *Complete intersection for equivariant models*, arXiv preprint arXiv:1512.07174 (2015).
- [10] Marta Casanellas and Jesús Fernández-Sánchez, *Relevant phylogenetic invariants of evolutionary models*, *J. Math. Pures Appl.*, **96** (2011) 207–229.
- [11] Marta Casanellas and Jesús Fernández-Sánchez, *Geometry of the Kimura 3-parameter model*, *Adv. in Appl. Math.* **41** (2008) 265–292.
- [12] David A. Cox, John B. Little, and Henry K. Schenck, *Toric varieties*, American Mathematical Soc., 2011.
- [13] Noah S. Daleo and Jonathan D. Hauenstein, *Numerically deciding the arithmetically Cohen-Macaulayness of a projective scheme*, *J. Symbolic Comp.*, **72** (2016), 128–146.
- [14] Charles Darwin, *The Descent of Man and Selection in Relation to Sex*, J. Murray London, 1871.
- [15] Maria Donten-Bury, *Phylogenetic Invariants for  $\mathbb{Z}_3$  Scheme-Theoretically*, *Annals of Combinatorics* **20**(3) (2016), 549–568.
- [16] Maria Donten-Bury and Mateusz Michałek, *Phylogenetic invariants for group-based models*, *Journal of Algebraic Statistics* **3** (2012), no. 1, 44–63.
- [17] Jan Draisma and Rob H. Eggermont, *Finiteness results for Abelian tree models*, *J. Eur. Math. Soc. (JEMS)* **17** (2015), no. 4, 711–738.
- [18] Jan Draisma and Jochen Kuttler, *On the ideals of equivariant tree models*, *Mathematische Annalen* **344**(3) (2009), 619–644.
- [19] Jan Draisma and Jochen Kuttler, *Bounded-rank tensors are defined in bounded degree*, *Duke Math. J.* **163** (2014), no. 1, 35–63.
- [20] Jan Draisma and Rob H. Eggermont, *Plücker varieties and higher secants of Sato Grassmannian*, *Reine Angew. Math.*, 25 pp., to appear, 2015.
- [21] Andreas Dress, Stefan Grünewald, and Zhenbing Zeng, *A cognitive network for oracle bone characters related to animals*, *Int. J. Mod. Phys. B*, **30**(4) (2016).



- [22] David Eisenbud, *Commutative Algebra: With a View Toward Algebraic Geometry*, Vol. 150 of Graduate Texts in Mathematics, Springer, 1995.
- [23] Nicholas Eriksson, Kristian Ranestad, Bernd Sturmfels, and Seth Sullivant, *Phylogenetic algebraic geometry*, Projective Varieties with Unexpected Properties; Siena, Italy (2004), 237–256.
- [24] Shmuel Friedland and Elizabeth Gross, *A proof of the set-theoretic version of the salmon conjecture*, J. Algebra, **356** (2012), 374–379.
- [25] William Fulton, *Introduction to toric varieties*, Annals of Mathematics Studies, vol. 131, Princeton University Press, Princeton, NJ, 1993, The William H. Roever Lectures in Geometry.
- [26] Michael Hendy and David Penny, *A framework for the quantitative study of evolutionary trees*, Systematic Zoology **38** (1989), 297–309.
- [27] Cristopher J. Hillar and Seth Sullivant, *Finite Gröbner bases in infinite dimensional polynomial rings and applications*, Adv. Math. **221**(1) (2012), 1–25.
- [28] Motoo Kimura, *Estimation of evolutionary distances between homologous nucleotide sequences*, Proc. Natl. Acad. Sci. U.S.A., **78**(1):454–458, 1981.
- [29] Kaie Kubjas and Christopher Manon, *Conformal blocks, Berenstein–Zelevinsky triangles, and group-based models* J. Algebr. Comb. **40**(3) (2014), 861–886.
- [30] Kaie Kubjas, *Hilbert polynomial of the Kimura 3-parameter model*, J. Algebraic Stat. **3**(1) (2012), 64–69.
- [31] Christopher Manon, *Coordinate rings for the moduli stack of  $\mathfrak{sl}_2(\mathbb{C})$  quasi-parabolic principal bundles on a curve and toric fiber products*, Journal of Algebra **365** (2012), 163–183.
- [32] Christopher Manon, *The Algebra of  $\mathbf{SL}_3(\mathbb{C})$  Conformal Blocks*, Transform. Groups **18**(4) (2013), 1165–1187.
- [33] Mateusz Michałek, *Geometry of phylogenetic group-based models*, Journal of Algebra **339**(1) (2011), 339–356.
- [34] Mateusz Michałek, *Constructive degree bounds for group-based models*, Journal of Combinatorial Theory, Series A **120**(7) (2013): 1672–1694.
- [35] Mateusz Michałek and Emanuele Ventura, *Phylogenetic complexity and combinatorics of tables*, Algebra Number Theory **11**(1) (2017), 235–252.
- [36] Ezra Miller and Bernd Sturmfels, *Combinatorial Commutative Algebra*, Vol. 227 of Graduate Texts in Mathematics, Springer, 2004.
- [37] Richard T. O’Grady et al., *Genes and tongues*, Science **243**(4899) (1989), 1651.
- [38] Steven V Sam, *Ideals of bounded rank symmetric tensors are generated in bounded degree*, Invent. Math., appeared online, 1–21, 2016.
- [39] Steven V Sam and Andrew Snowden, *GL-equivariant modules over polynomial rings in infinitely many variables*, Trans. Amer. Math. Soc., **368**(2):1097–1158, 2016.
- [40] Charles Semple and Mike Steel, *Phylogenetics*, Oxford University Press, 2003.
- [41] Bernd Sturmfels and Seth Sullivant, *Toric ideals of phylogenetic invariants*, J. Comput. Biology **12** (2005), 204–228.
- [42] László A. Székely, Mike A. Steel, and Péter L. Erdős, *Fourier calculus on evolutionary trees*, Adv. in Appl. Math., **14**(2):200–210, 1993.
- [43] Bernd Sturmfels, *Gröbner bases and convex polytopes*, University Lecture Series, vol. 8, American Mathematical Society, 1996.

- [44] Bernd Sturmfels and Zhiqiang Xu, *Sagbi bases of Cox-Nagata rings*, Journal of the European Mathematical Society **12** (2010), 429–459.
- [45] Seth Sullivant, *Toric fiber products*, Journal of Algebra **316**(2) (2007), 560 – 577.
- [46] Wolmer Vasconcelos, *Computational Methods in Commutative Algebra and Algebraic Geometry*, Vol. 2 of Algorithms and Computation in Mathematics, Springer, 2004.
- [47] 4ti2 team, *4ti2—A software package for algebraic, geometric and combinatorial problems on linear spaces*, [www.4ti2.de](http://www.4ti2.de).

**Authors' addresses:**

Mateusz Michałek, Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

[mateusz.michalek@mis.mpg.de](mailto:mateusz.michalek@mis.mpg.de)

Institute of Mathematics of Polish Academy of Sciences, Warsaw, Poland

[mmichalek@impan.pl](mailto:mmichalek@impan.pl)

Emanuele Ventura, Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

[emanuele.ventura@mis.mpg.de](mailto:emanuele.ventura@mis.mpg.de)