Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig

Gröbner bases for staged trees

by

Lamprini Ananiadi and Eliana Maria Duarte Gelvez

Preprint no.: 94 2019

GRÖBNER BASES FOR STAGED TREES

LAMPRINI ANANIADI AND ELIANA DUARTE

Abstract. In this article we consider the problem of finding generators of the toric ideal associated to a combinatorial object called a staged tree. Our main theorem states that toric ideals of staged trees that are balanced and stratified are generated by a quadratic Gröbner basis whose initial ideal is squarefree. The proof of this result is based on Sullivant's [17] toric fiber product construction.

1. Introduction

The study of toric ideals associated to statistical models was pioneered by the work of Diaconis and Sturmfels [3] who first used the generators of a toric ideal to formulate a sampling algorithm for discrete distributions. Since then, and with the subsequent work of [2, 17] and [7] the study of toric ideals of discrete statistical models has been an active area of research in Algebraic Statistics. The books by Sullivant [18, Chapter 9] and Aoki, Hara and Takemura [1] are good references to learn about the role of toric ideals in statistics. A recent introduction to the topic from the point of view of binomial ideals can be found in [12, Chapter 9], which also contains a thorough list of references of previous contributions to this topic.

In 2008, Smith and Anderson [14] introduced a new graphical discrete statistical model called a *staged tree model*. This model is represented by an event tree together with an equivalence relation on its vertices. Staged tree models are useful to represent conditional independence relations among events. In particular we can use staged tree models to represent some conditional independence statements between random variables. For example those coming from graphical models such as Bayesian networks and decomposable models. Hence any discrete Bayesian network or decomposable model is also a staged tree model [14]. There are two properties that make staged tree models more general than Bayesian networks or decomposable models. The first is that the state space of a staged tree model does not have to be a cartesian product. The second is that using staged tree models it is possible to represent extra context-specific conditional independence between events. The book of Collazo, Görgen and Smith [15] is a good reference to learn about these models.

In this article we define the toric ideal associated to a *staged tree* and study its properties from an algebraic and combinatorial point of view. We present Theorem 2.5 which states that toric ideals of staged trees that are *balanced* and *stratified* have quadratic Gröbner basis with squarefree initial ideal. We apply Theorem 2.5 in Section 5 to obtain Gröbner bases for toric ideals of staged tree models. Our results provide a new point of view on the construction of Gröbner bases for decomposable graphical models, some conditional independence models as well as the construction of Gröbner bases for staged tree models whose underlying tree is *asymmetric*.

This article is organized as follows. In Section 2 we define the toric ideal associated to a staged tree. In Section 3 we formulate a toric fiber product construction for balanced and stratified staged

Date: October 7, 2019.

Key words and phrases. Graphical models, toric ideals, staged trees, Markov bases, toric fiber products

2010 Mathematics Subject Classification: Primary: 113P10; Secondary: 13P25, 05E40.

trees. In Section 4 we prove our main result Theorem 2.5. Finally in Section 5 we apply our results to compute Gröbner bases for several statistical models.

2. Staged trees

We start by defining our two objects of interest: a staged tree and its associated toric ideal. Let $\mathcal{T} = (V, E)$ denote a directed rooted tree graph with vertex set V and edge set E. For $v, w \in V$ the directed edge in E from v to w is denoted by (v, w), the set of children of v is $\mathrm{ch}(v) = \{u \mid (v, u) \in E\}$, and the set of outgoing edges from v is $E(v) = \{(v, u) \mid u \in \mathrm{ch}(v)\}$. Given a set $\mathcal{L}$ of labels, to each $e \in E$ we associate a label from $\mathcal{L}$ via the rule $\theta : E \to \mathcal{L}$. We require that θ is surjective. For each vertex $v \in V$, we let $\theta_v := \{\theta(e) \mid e \in E(v)\}$ be the set of labels associated to v.

Definition 2.1. Let $\mathcal{L}$ be a set of labels. A tree $\mathcal{T} = (V, E)$ together with a labelling $\theta : E \to \mathcal{L}$ is a *staged tree* if: (1) for each $v \in V$, $|\theta_v| = |E(v)|$, and (2) for any two vertices $v, w \in V$ the sets θ_v, θ_w are either equal or disjoint.

Conditions (1) and (2) in Definition 2.1 of a staged tree define an equivalence relation on the set of vertices of $\mathcal{T}$. Namely $v, w \in V$ are equivalent if and only if $\theta_v = \theta_w$. We refer to the partition induced by this equivalence relation on the set V as the set of stages of $\mathcal{T}$ and to a single element in this partition as a *stage*. We use $(\mathcal{T}, \theta)$ to denote a staged tree with labeling rule θ. For simplicity we will often drop the use of θ and write $\mathcal{T}$ for a staged tree.

To define the toric ideal associated to $(\mathcal{T}, \theta)$ we define two polynomial rings. The first ring is $\mathbb{R}[p]_{\mathcal{T}} := \mathbb{R}[p_\lambda \mid \lambda \in \Lambda]$ where Λ is the set of root-to-leaf paths in $\mathcal{T}$. The second ring is $\mathbb{R}[\Theta]_{\mathcal{T}} := \mathbb{R}[z, \mathcal{L}]$ where the labels in $\mathcal{L}$ are indeterminates together with a homogenizing variable z. For a directed or undirected path γ in $\mathcal{T}$, $E(\gamma)$ is the set of edges in γ.

Definition 2.2. The *toric staged tree ideal* associated to $(\mathcal{T}, \theta)$ is the kernel of the ring homomorphism $\varphi_{\mathcal{T}} : \mathbb{R}[p]_{\mathcal{T}} \to \mathbb{R}[\Theta]_{\mathcal{T}}$ defined as

$$p_\lambda \mapsto z \cdot \prod_{e \in E(\lambda)} \theta(e) \ . \tag{1}$$

If $n = |\mathcal{L}|$ is the number of distinct edge labels in $\mathcal{T}$, then $\ker(\varphi_{\mathcal{T}})$ is the defining ideal of the projective toric variety defined by the closure of the image of the monomial parameterization $\Phi_{\mathcal{T}} : (\mathbb{C}^*)^n \to \mathbb{P}^{|\Lambda|-1}$ given by $(\theta(e) \mid \theta(e) \in \mathrm{im}(\theta)) \mapsto \prod_{e \in E(\lambda)} \theta(e)$.

Example 2.3. The staged tree $\mathcal{T}_1$ in Figure 1 has label set $\mathcal{L} = \{s_0, \ldots, s_{13}\}$. Each vertex in $\mathcal{T}_1$ is denoted by a string of 0's and 1's and each edge has a label associated to it. The root-to-leaf paths in $\mathcal{T}$ are denoted by p_{ijkl} where $i, j, k, l \in \{0, 1\}$. A vertex in $\mathcal{T}_1$ represented with a blank circle indicates a stage consisting of a single vertex. We use colors in the vertices of $\mathcal{T}_1$ to indicate which vertices are in the same stage. For instance all the purple vertices, i.e. the set of vertices $\{000, 010, 100, 110\}$, are in the same stage and therefore they have the same set $\{s_{10}, s_{11}\}$ of associated edge labels. The map $\Phi_{\mathcal{T}}$ sends $(s_0, \ldots, s_{13})$ to

$$(s_0s_2s_6s_{10},\ s_0s_2s_6s_{11},\ s_0s_2s_7s_{12},\ s_0s_2s_7s_{13},\ s_0s_3s_8s_{10},\ s_0s_3s_8s_{11},\ s_0s_3s_9s_{12},\ s_0s_3s_9s_{13},$$
$$s_1s_4s_6s_{10},\ s_1s_4s_6s_{11},\ s_1s_4s_7s_{12},\ s_1s_4s_7s_{13},\ s_1s_5s_8s_{10},\ s_1s_5s_8s_{11},\ s_1s_5s_9s_{12},\ s_1s_5s_9s_{13}).$$

The toric ideal $\ker(\varphi_{\mathcal{T}})$ is generated by a quadratic Gröbner basis with squarefree initial ideal.

Example 2.4. Consider the two staged trees $\mathcal{T}_2, \mathcal{T}_3$ depicted in Figure 1. For the staged tree $\mathcal{T}_2$, $\ker(\varphi_{\mathcal{T}_2})$ is generated by a quadratic Gröbner basis with squarefree intial ideal. For $\mathcal{T}_3$, the ideal $\ker(\varphi_{\mathcal{T}_3})$ also has a Gröbner basis with squarefree initial ideal but its elements are of degree two and four.

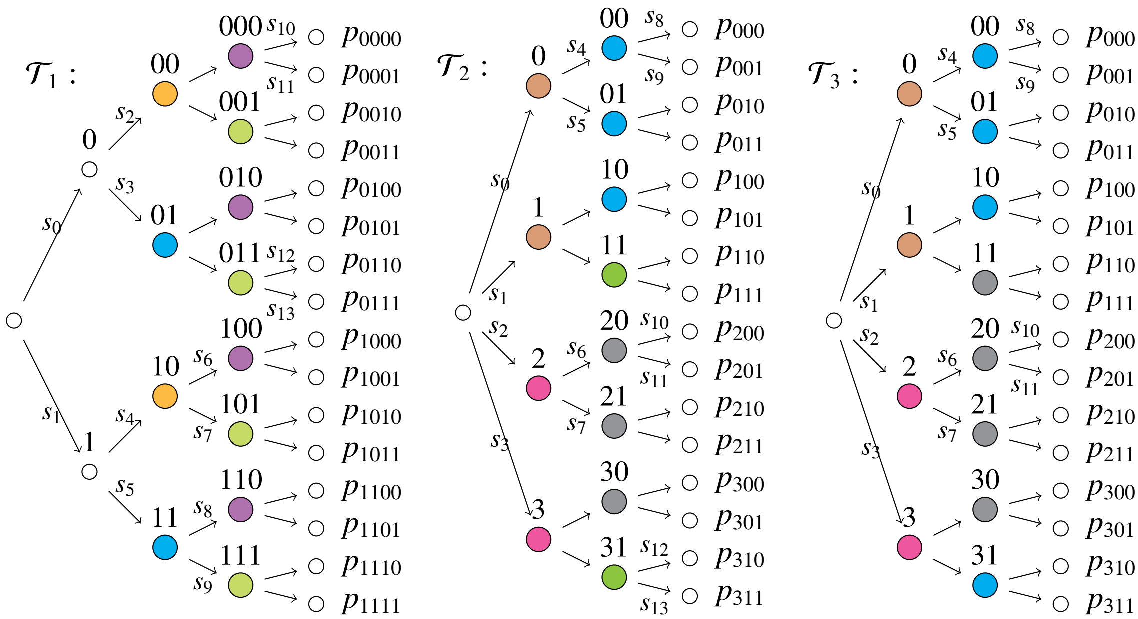

FIGURE 1. Three examples of staged trees. In each tree two vertices with the same color are in the same stage.

We are interested in relating the combinatorial properties of the staged tree $(\mathcal{T}, \theta)$ with the properties of the toric ideal $\ker(\varphi_{\mathcal{T}})$. Before we dive into the combinatorics of staged trees we present our main Theorem 2.5. In its statement we use the notion of *balanced* staged tree and of *stratified* staged tree.

Theorem 2.5. *If* $(\mathcal{T}, \theta)$ *is a balanced and stratified staged tree then* $\ker(\varphi_{\mathcal{T}})$ *is generated by a quadratic Gröbner basis with squarefree initial ideal.*

We clarify that the conditions of $(\mathcal{T}, \theta)$ being balanced and stratified in Theorem 2.5 are sufficient for $\ker(\varphi_{\mathcal{T}})$ to have a quadratic Gröbner basis but are not necessary. In the examples of staged trees in Figure 1, all of the trees $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$ are stratified but only $\mathcal{T}_1$ is balanced. Even though $\mathcal{T}_2$ is not balanced, it has a quadratic Gröbner basis with squarefree initial terms.

Definition 2.6. Let $\mathcal{T}$ be a tree. For $v \in V$, the *level* of v is denoted by $\ell(v)$ and it is equal to the number of edges in the unique path from the root of $\mathcal{T}$ to v. If all the leaves in $\mathcal{T}$ have the same level then the level of $\mathcal{T}$ is equal to the level of any of its leaves. The staged tree $(\mathcal{T}, \theta)$ is *stratified* if all its leaves have the same level and if every two vertices in the same stage have the same level.

It is easy to check that the trees in Figure 1 are stratified. Namely we only need to verify that every two vertices with the same color are also in the same level. Notice that the combinatorial condition of $(\mathcal{T}, \theta)$ being stratified translates into the algebraic condition that the map $\varphi_{\mathcal{T}}$ is squarefree.

We now turn our attention to the definition of a balanced staged tree. This definition is formulated in terms of polynomials associated to each vertex of the tree. We proceed to explain their notation and basic properties.

Definition 2.7. Let $(\mathcal{T}, \theta)$ be a staged tree, $v \in V$, and $\mathcal{T}_v$ the subtree of $\mathcal{T}$ rooted at v. The tree $\mathcal{T}_v$ is a staged tree with the induced labelling from $\mathcal{T}$. Let Λ_v be the set of v-to-leaf paths in $\mathcal{T}$ and define

$$t(v) := \sum_{\lambda \in \Lambda_v} \prod_{e \in E(\lambda)} \theta(e).$$

When v is the root of $\mathcal{T}$, the polynomial $t(v)$ is called the *interpolating polynomial* of $\mathcal{T}$. Two staged trees $(\mathcal{T}, \theta)$ and $(\mathcal{T}, \theta')$ with the same label set $\mathcal{L}$ are *polynomially equivalent* if their interpolating polynomials are equal.

The interpolating polynomial of a staged tree is an important tool in the study of the statistical properties of staged tree models. This polynomial was defined by Görgen and Smith in [10] and further studied by Görgen et al. in [9]. Although these two articles are written for a statistical audience, we would like to emphasize that their symbolic algebra approach to the study of statistical models proves to be very important for the use of these models in practice. We will define the statistical model associated to a staged tree and connect Theorem 2.5 to other results in Algebraic Statistics in Section 5.

If $(\mathcal{T}, \theta)$ is a staged tree, the polynomials $t(\cdot)$ satisfy a recursive relation. This relation is useful to prove statements about the algebraic and combinatorial properties of $\mathcal{T}$. We state this property as a lemma.

Lemma 2.8 ([9, Theorem 1])**.** *Let $(\mathcal{T}, \theta)$ be a staged tree, $v \in V$ and $\mathrm{ch}(v) = \{v_0, \ldots, v_k\}$. Then the polynomial $t(v)$ admits the recursive representation $t(v) = \sum_{i=0}^{k} \theta(v, v_i) t(v_i)$.*

Example 2.9. Consider the staged tree $\mathcal{T}_1$ in Figure 1. If v and w are orange and blue vertices in $\mathcal{T}_1$ respectively and r is the root of $\mathcal{T}_1$ then $t(v) = s_6(s_{10} + s_{11}) + s_7(s_{12} + s_{13})$, $t(w) = s_8(s_{10} + s_{11}) + s_9(s_{12} + s_{13})$, and $t(r) = (s_0 s_2 + s_1 s_4) t(v) + (s_0 s_3 + s_1 s_5) t(w)$.

Definition 2.10. Let $(\mathcal{T}, \theta)$ be a staged tree and v, w be two vertices in the same stage with $\mathrm{ch}(v) = \{v_0, \ldots, v_k\}$ and $\mathrm{ch}(w) = \{w_0, \ldots, w_k\}$. After a possible reindexing of the elements in $\mathrm{ch}(w)$ we may assume that $\theta(v, v_i) = \theta(w, w_i)$ for all $i \in \{0, \ldots, k\}$. The vertices v, w satisfy condition $(\star)$ if

$$t(v_i) t(w_j) = t(w_i) t(v_j) \text{ in } \mathbb{R}[\Theta]_{\mathcal{T}}, \text{ for all } i \neq j \in \{0, \ldots, k\}. \tag{2}$$

The staged tree $(\mathcal{T}, \theta)$ is *balanced* if every pair of vertices in the same stage satisfy condition $(\star)$.

Example 2.11. The two staged trees $\mathcal{T}_2, \mathcal{T}_3$ in Figure 1 are not balanced. The two pink vertices in $\mathcal{T}_2$ do not satisfy condition $(\star)$ because $(s_{10} + s_{11})(s_{12} + s_{13}) \neq (s_{10} + s_{11})^2$. By a similar argument we can check that $\mathcal{T}_3$ is also not balanced.

Although condition $(\star)$ seems to be algebraic and hard to check, in many cases it is very combinatorial. To formulate a precise instance where this is true we need the following definition.

Definition 2.12. Let $(\mathcal{T}, \theta)$ be a staged tree. We say that two vertices $v, w \in V$ are in the same *position* if they are in the same stage and $t(v) = t(w)$.

The notion of position for vertices in the same stage was formulated in [14]. Intuitively it means that if we regard the subtrees $\mathcal{T}_v$ and $\mathcal{T}_w$ as representing the unfolding of a sequence of events, then the future of v and w is essentially the same. In the next lemma we use positions of vertices to provide a sufficient condition on a stratified staged tree $(\mathcal{T}, \theta)$ to be balanced.

Lemma 2.13. *Let $(\mathcal{T}, \theta)$ be a stratified staged tree. Suppose that every two vertices in $\mathcal{T}$ that are in the same stage are also in the same position. Then $(\mathcal{T}, \theta)$ is balanced.*

Proof. Following Definition 2.10 it suffices to prove that any two vertices in the same position satisfy condition $(\star)$. Let v, w be two vertices in the same position, by definition this means that they are in the same stage and $t(v) = t(w)$. Let $\mathrm{ch}(v) = \{v_0, \dots, v_k\}$ and $\mathrm{ch}(w) = \{w_0, \dots, w_k\}$, after possibly permuting the subindices of elements in $\mathrm{ch}(w)$, we may assume $\theta(v, v_i) = \theta(w, w_i)$. Using Lemma 2.8 we write

$$t(v) = t(w) \;\Leftrightarrow\; \sum_{i=0}^{k} \theta(v, v_i)t(v_i) = \sum_{i=0}^{k} \theta(w, w_i)t(w_i) \;\Leftrightarrow\; \sum_{i=0}^{k} \theta(v, v_i)(t(v_i) - t(w_i)) = 0.$$

Since $(\mathcal{T}, \theta)$ is stratified, the variables appearing in the polynomials $t(v_i), t(w_i)$ are disjoint from the set of variables $\{\theta(v, v_0), \dots, \theta(v, v_k)\}$. Thus $t(v_i) = t(w_i)$ for all $i \in \{0, \dots, k\}$. It follows that for all $i, j \in \{0, \dots, k\}$ the equality $t(v_i)t(w_j) = t(w_i)t(v_j)$ is true. Hence $(\mathcal{T}, \theta)$ is balanced.

□

Example 2.14. The staged tree $\mathcal{T}_1$ in Figure 1 is balanced. This can be readily checked by noting that the blue vertices are in the same position and that the same is true for the orange vertices. The tree $\mathcal{T}$ in Figure 3 is an example of a balanced staged tree in which the blue vertices are not in the same position.

3. Toric fiber products for staged trees

In this section we define a toric fiber product for staged trees. We then use these results in Section 4 to prove Theorem 2.5. We start with a review of toric fiber products following [17].

Given a positive integer m, set $[m] = \{1, 2, \dots, m\}$. Let r be a positive integer and let s and t be two vectors of positive integers in $\mathbb{Z}^r_{>0}$. Consider the homogeneous, multigraded polynomial rings

$$\mathbb{K}[x] := \mathbb{K}[x^i_j \mid i \in [r], j \in [s_i]] \quad \text{and} \quad \mathbb{K}[y] := \mathbb{K}[y^i_k \mid i \in [r], k \in [t_i]]$$

with the same multigrading

$$\deg(x^i_j) = \deg(y^i_k) = \boldsymbol{a}^i \in \mathbb{Z}^d.$$

Denote by $\mathcal{A} = \{\boldsymbol{a}^1, \dots, \boldsymbol{a}^r\}$ the set of all multidegrees of these variables and assume that there exists a vector $w \in \mathbb{Q}^d$ such that $\langle w, \boldsymbol{a}^i \rangle = 1$ for any $\boldsymbol{a}^i \in \mathcal{A}$. If $I \subseteq \mathbb{K}[x]$ and $J \subseteq \mathbb{K}[y]$ are homogeneous ideals, then the quotient rings $R = \mathbb{K}[x]/I$ and $S = \mathbb{K}[y]/J$ are also multigraded rings. Let

$$\mathbb{K}[z] := \mathbb{K}[z^i_{jk} \mid i \in [r], j \in [s_i], k \in [t_i]]$$

and consider the ring homomorphism

$$\phi_{I,J} : \mathbb{K}[z] \to R \otimes_{\mathbb{K}} S$$
$$z^i_{jk} \mapsto \overline{x^i_j} \otimes \overline{y^i_k},$$

where $\overline{x^i_j}$ and $\overline{y^i_k}$ are the equivalence classes of x^i_j and y^i_k respectively.

Definition 3.1. The *toric fiber product* of I and J is $I \times_{\mathcal{A}} J := \ker(\phi_{I,J})$.

Let $(\mathcal{T}, \theta)$ be a staged tree with root r. We recursively define an indexing of the vertices in $V \setminus \{r\}$. This identifies each vertex in $V \setminus \{r\}$ with a unique index. From this point on we refer to an element in V via its index $\boldsymbol{a}$ or by r in case the vertex is the root. The children of r are indexed by $\{0, 1, \dots, k\}$. For $\boldsymbol{a} \in V \setminus \{r\}$, we index the children of $\boldsymbol{a}$ as follows. If $E(\boldsymbol{a}) = \emptyset$ then $\boldsymbol{a}$ is a leaf of the tree and we are done. If $|E(\boldsymbol{a})| = j + 1$, index the children of $\boldsymbol{a}$ by $\boldsymbol{a}0, ..., \boldsymbol{a}j$. This way each vertex in V is indexed by a finite sequence of nonnegative integers

$$\boldsymbol{a} = a_1 a_2 \cdots a_\ell,$$

where ℓ is the level of $\boldsymbol{a}$. All vertices of the trees in Figure 1 are indexed following this rule. In Figure 1 the index of each vertex is displayed immediately above each vertex and on the side for the leaves. We denote by $\boldsymbol{i}_{\mathcal{T}}$ the set of indices of the leaves in $\mathcal{T}$.

Definition 3.2. Let $(\mathcal{T},\theta)$ be a staged tree. If every leaf in $\mathcal{T}$ has level one then we call $\mathcal{T}$ a *one-level tree*. We reserve for it the special notation $(\mathcal{B},t)$ where t is the set of edge labels of $\mathcal{B}$.

Definition 3.3. Let $(\mathcal{T},\theta)$ be a staged tree and $G=\{G_1,\dots,G_r\}$ be a partition of the set of leaves $\boldsymbol{i}_{\mathcal{T}}$. For each $i\in[r]$, let $(\mathcal{B}_i,t^{(i)})$ be a one-level tree as in Definition 3.2 with label set $t^{(i)}$. We define the *gluing component* $\mathcal{T}_G$ associated to $\mathcal{T}$ and G as the disjoint union of $(\mathcal{B}_i,t^{(i)})$, namely

$$\mathcal{T}_G=\bigsqcup_{i\in[r]}(\mathcal{B}_i,t^{(i)}).$$

We denote by $[\mathcal{T},\mathcal{T}_G]$ the tree obtained by gluing $\mathcal{B}_i$ to the leaf $\boldsymbol{a}$ for all $\boldsymbol{a}\in G_i$ and all $i\in[r]$.

Remark 3.4. The labelling in $[\mathcal{T},\mathcal{T}_G]$ is inherited from the labellings of $\mathcal{T}$ and $\mathcal{T}_G$. With this labelling $[\mathcal{T},\mathcal{T}_G]$ is a staged tree. Moreover, $\boldsymbol{i}_{[\mathcal{T},\mathcal{T}_G]}=\{\boldsymbol{a}k\,|\,\boldsymbol{a}\in G_i,k\in\boldsymbol{i}_{\mathcal{B}_i},i\in[r]\}$. The stages in $[\mathcal{T},\mathcal{T}_G]$ are the ones inherited from $\mathcal{T}$ union the new stages determined by G. This means that two vertices $\boldsymbol{a},\boldsymbol{b}\in\boldsymbol{i}_{\mathcal{T}}$ are in the same stage in $[\mathcal{T},\mathcal{T}_G]$ provided $\boldsymbol{a},\boldsymbol{b}\in G_i$.

We relate $\ker(\varphi_{[\mathcal{T},\mathcal{T}_G]})$ to the toric fiber product of the two ideals $\ker(\varphi_{\mathcal{T}})$ and the zero ideal $\langle 0\rangle$. Let $\mathcal{T},G,\mathcal{T}_G$ and $[\mathcal{T},\mathcal{T}_G]$ be as in Definition 3.3. First we associate to $\mathcal{T}_G$ the ring $\mathbb{R}[p]_{\mathcal{T}_G}:=\mathbb{R}[p^i_k\mid k\in\boldsymbol{i}_{\mathcal{B}_i},i\in[r]]$ and the ring map

$$\begin{aligned}\varphi_{\mathcal{T}_G}:\mathbb{R}[p]_{\mathcal{T}_G}&\to\mathbb{R}[\Theta]_{\mathcal{T}_G}\\ p^i_k&\mapsto t^{(i)}_k.\end{aligned}$$

Since there is a one-to-one correspondence between the variables p^i_k and $t^{(i)}_k$, we see that $\varphi_{\mathcal{T}_G}$ is an isomorphism. In particular, $\ker(\varphi_{\mathcal{T}_G})=\langle 0\rangle$. Second, using G we regroup the variables in $\mathbb{R}[p]_{\mathcal{T}}$ by

$$\mathbb{R}[p]_{\mathcal{T}}=\mathbb{R}[p^i_{\boldsymbol{j}}\mid\boldsymbol{j}\in G_i,\ i\in[r]].$$

We define the multigrading on the polynomial rings $\mathbb{R}[p]_{\mathcal{T}}$ and $\mathbb{R}[p]_{\mathcal{T}_G}$ as

$$\deg(p^i_{\boldsymbol{j}})=\deg(p^i_k)=\boldsymbol{e}_i,\ \text{ for }\boldsymbol{j}\in G_i,\ k\in\boldsymbol{i}_{\mathcal{B}_i},\ i\in[r].$$

Here $\boldsymbol{e}_i$ is the ith standard unit vector in $\mathbb{Z}^r$. If $\mathcal{A}$ is the set of all these multidegrees, then $\mathcal{A}$ is linearly independent as it is a collection of standard unit vectors in $\mathbb{Z}^r$. We say a homogeneous polynomial in the ring $\mathbb{R}[p]_{\mathcal{T}}$ or $\mathbb{R}[p]_{\mathcal{T}_G}$ is *$\mathcal{A}$-graded* whenever it is homogeneous with respect to the multigrading determined by $\mathcal{A}$.

Fix $R=\mathbb{R}[p]_{\mathcal{T}}/\ker(\varphi_{\mathcal{T}})$, $S=\mathbb{R}[p]_{\mathcal{T}_G}/\ker(\varphi_{\mathcal{T}_G})$ and let $\mathbb{R}[p]_{[\mathcal{T},\mathcal{T}_G]}=\mathbb{R}[p^i_{\boldsymbol{j}k}\mid\boldsymbol{j}\in G_i,\ k\in\boldsymbol{i}_{\mathcal{B}_i},\ i\in[r]]$. Consider the ring homomorphism

$$\begin{aligned}\psi:\mathbb{R}[p]_{[\mathcal{T},\mathcal{T}_G]}&\to R\otimes_{\mathbb{R}}S\\ p^i_{\boldsymbol{j}k}&\mapsto\overline{p^i_{\boldsymbol{j}}}\otimes\overline{p^i_k},\quad\text{ for }\boldsymbol{j}\in G_i,\ k\in\boldsymbol{i}_{\mathcal{B}_i}\ i\in[r].\end{aligned}\tag{3}$$

The ideal $\ker(\psi)=\ker(\varphi_{\mathcal{T}})\times_{\mathcal{A}}\langle 0\rangle$ is the toric fiber product of $\ker(\varphi_{\mathcal{T}})$ and $\langle 0\rangle$.

Proposition 3.5. *Let $\mathcal{T},G,\mathcal{T}_G$ and $[\mathcal{T},\mathcal{T}_G]$ be as in Definition 3.3. Suppose that* $\ker(\varphi_{\mathcal{T}})$ *is homogeneous with respect to the multigrading given by $\mathcal{A}$. Then*

$$\ker(\varphi_{[\mathcal{T},\mathcal{T}_G]})=\ker(\varphi_{\mathcal{T}})\times_{\mathcal{A}}\langle 0\rangle.$$

Proof. For each $\boldsymbol{j}k \in \boldsymbol{i}_{[\mathcal{T},\mathcal{T}_G]}$, the map $\varphi_{[\mathcal{T},\mathcal{T}_G]}$ can be rewritten as

$$p_{\boldsymbol{j}k} \mapsto z \cdot \prod_{e \in E(\lambda_{\boldsymbol{j}k})} \theta(e) = z \cdot \left(\prod_{e \in E(\lambda_{\boldsymbol{j}})} \theta(e) \right) \theta(\boldsymbol{j}, \boldsymbol{j}k)$$

where $\theta(\boldsymbol{j}, \boldsymbol{j}k) = t_k^{(i)}$. Written in this factored form we see that $\varphi_{[\mathcal{T},\mathcal{T}_G]}$ factors trough ψ. Since $\ker(\varphi_{\mathcal{T}})$ is $\mathcal{A}$-homogeneous we conclude $\ker(\varphi_{[\mathcal{T},\mathcal{T}_G]}) = \ker(\varphi_{\mathcal{T}}) \times_{\mathcal{A}} \langle 0 \rangle$.

□

We make several remarks on the scope of Proposition 3.5 via the next set of examples.

Definition 3.6. Let $(\mathcal{T}, \theta)$ be a stratified staged tree of level m. For $1 \leq q \leq m$ we define $V_{\leq q} := \cup_{i=0}^{q} V_i$ where $V_q := \{v \in V \,|\, \ell(v) = q\}$ and $E_{\leq q} := \{(v, w) \in E \,|\, \ell(v) \leq q, \ell(w) \leq q\}$. The staged tree $(\mathcal{T}^{(q)}, \theta)$ is defined to be the subtree $\mathcal{T}^{(q)}$ of $\mathcal{T}$ with vertex set $V_{\leq q}$, edge set $E_{\leq q}$ and labeling induced from $\mathcal{T}$.

Example 3.7. Consider the staged tree $\mathcal{T}_1$ in Figure 1 and let $\mathcal{T} = \mathcal{T}_1^{(3)}$ as in Definition 3.6. Then $\mathcal{T}$ is a staged tree with label set $\{s_0, \ldots, s_9\}$. Fix $G = \{\{000, 010, 100, 110\}, \{001, 011, 101, 111\}\}$ and $\mathcal{T}_G = (\mathcal{B}_1, \{s_{10}, s_{11}\}) \sqcup (\mathcal{B}_2, \{s_{12}, s_{13}\})$. With this choice of $\mathcal{T}, G$ and $\mathcal{T}_G$ we see that $\mathcal{T}_1 = [\mathcal{T}, \mathcal{T}_G]$. Now $\mathbb{R}[p]_{\mathcal{T}} = \mathbb{R}[p_{\boldsymbol{a}}^i \,|\, \boldsymbol{a} \in G_i, i \in \{1, 2\}]$, hence $\deg(p_{000}^1, p_{010}^1, p_{100}^1, p_{110}^1) = \boldsymbol{e}_1$ and $\deg(p_{001}^2, p_{011}^2, p_{101}^2, p_{111}^2) = \boldsymbol{e}_2$ so $\mathcal{A} = \{\boldsymbol{e}_1, \boldsymbol{e}_2\} \subset \mathbb{Z}^2$ is of full rank. The ideal

$$\ker(\varphi_{\mathcal{T}}) = \langle p_{000}^1 p_{101}^2 - p_{100}^1 p_{011}^2, p_{010}^1 p_{111}^2 - p_{110}^1 p_{011}^2 \rangle$$

is $\mathcal{A}$-graded. Hence by Proposition 3.5 $\ker(\varphi_{\mathcal{T}_1}) = \ker(\varphi_{\mathcal{T}}) \times_{\mathcal{A}} \langle 0 \rangle$.

Example 3.8. Let $\mathcal{T}_2$ be the staged tree from Figure 1. We proceed in a similar fashion as in Example 3.7. Set $\mathcal{T} = \mathcal{T}_2^{(2)}$, $G = \{\{00, 01, 10\}, \{11, 31\}, \{20, 21, 30\}\}$ and $\mathcal{T}_G = (\mathcal{B}_1, \{s_8, s_9\}) \sqcup (\mathcal{B}_2, \{s_{12}, s_{13}\}) \sqcup (\mathcal{B}_3, \{s_{10}, s_{11}\})$. Then $\mathcal{T}_2 = [\mathcal{T}, \mathcal{T}_G]$. The set G defines a multigrading on $\mathbb{R}[p]_{\mathcal{T}}$ with $\mathcal{A} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\} \subset \mathbb{Z}^3$. The ideal $\ker(\varphi_{\mathcal{T}}) = \langle p_{00}^1 p_{11}^2 - p_{10}^1 p_{01}^1, p_{20}^3 p_{31}^2 - p_{30}^3 p_{21}^3 \rangle$ is not $\mathcal{A}$-graded. Thus in this case $\ker(\varphi_{\mathcal{T}_2}) \neq \ker(\varphi_{\mathcal{T}}) \times_{\mathcal{A}} \langle 0 \rangle$.

Example 3.9. Let $\mathcal{T}_3$ be the staged tree from Figure 1 and $\mathcal{T} = \mathcal{T}_3^{(2)}$. As in the previous examples, fix $G = \{\{00, 01, 10, 31\}, \{20, 21, 30, 11\}\}$ and $\mathcal{T}_G = (\mathcal{B}_1, \{s_8, s_9\}) \sqcup (\mathcal{B}_2, \{s_{10}, s_{11}\})$ so $\mathcal{T}_3 = [\mathcal{T}, \mathcal{T}_G]$. The set G defines a multigrading on $\mathbb{R}[p]_{\mathcal{T}}$ with $\mathcal{A} = \{\boldsymbol{e}_1, \boldsymbol{e}_2\} \subset \mathbb{Z}^2$. The ideal $\ker(\varphi_{\mathcal{T}}) = \langle p_{00}^1 p_{11}^2 - p_{10}^1 p_{01}^1, p_{20}^2 p_{31}^1 - p_{30}^2 p_{21}^2 \rangle$ is not $\mathcal{A}$-graded. However there is a nonempty principal subideal of $\ker(\varphi_{\mathcal{T}})$ that is $\mathcal{A}$-homogeneous. This principal ideal Q is generated by the quartic $p_{00}^1 p_{11}^2 p_{20}^2 p_{31}^1 - p_{01}^1 p_{10}^1 p_{21}^2 p_{30}^2$. In this case $\ker(\varphi_{\mathcal{T}_3}) = Q \times_{\mathcal{A}} \langle 0 \rangle$. This does not fall in the context of Proposition 3.5 since $\ker(\varphi_{\mathcal{T}_3}) \neq \ker(\varphi_{\mathcal{T}}) \times_{\mathcal{A}} \langle 0 \rangle$.

4. Proof of main theorem

The main ingredient in the proof of Theorem 2.5 is the toric fiber product. The idea of the proof is that when $(\mathcal{T}, \theta)$ is balanced and stratified, we can construct $\ker(\varphi_{\mathcal{T}})$ in a finite number of steps using Proposition 3.5.

Start with a one-level probability tree $\mathcal{T}_1$. Let G^1 be a partition of $\boldsymbol{i}_{\mathcal{T}_1}$, $\mathcal{T}_{G^1}$ be a gluing component and set $\mathcal{T}_2 = [\mathcal{T}_1, \mathcal{T}_{G^1}]$. In the inductive step, $\mathcal{T}_{j+1} = [\mathcal{T}_j, \mathcal{T}_{G^j}]$ with $r_j := |G^j|$. At each step j we also require that the label set of $\mathcal{T}_{G^j}$ is disjoint from the label set of $\mathcal{T}_j$. After n iterations, we obtain a stratified staged tree $\mathcal{T}_n$ whose set of stages is exactly $\cup_{j=1}^{n-1} G^j$. Whenever a staged tree $(\mathcal{T}, \theta)$ is constructed in this way so $\mathcal{T} = \mathcal{T}_n$ for some n we say $\mathcal{T}$ is an *inductively constructed* staged tree.

We recall a theorem from [17] concerning the defining equations of a toric fiber product. To state this theorem we use the notation for toric fiber products from Section 3.

Theorem 4.1 ([17, Theorem 2.9])**.** *Suppose that $\mathcal{A}$ is linearly independent. Let $F \subset I$ be a homogeneous Gröbner basis for I with respect to the weight vector ω_1 and let $H \subset J$ be a homogeneous Gröbner basis for J with respect to the weight vector ω_2. Let ω be a weight vector such that* Quad_B *is a Gröbner basis for I_B. Then*

$$\mathrm{Lift}(F) \cup \mathrm{Lift}(H) \cup \mathrm{Quad}_B$$

is a Gröbner basis for $I \times_{\mathcal{A}} J$ with respect to the weight order $\phi_B^*(\omega_1, \omega_2) + \epsilon\omega$ for sufficiently small $\epsilon > 0$.

We use this result to obtain a Gröbner basis of any inductively constructed staged tree. Theorem 4.1 has two important ingredients. The first is the set of equations denoted by Quad_B, these are quadratic equations that emerge from the construction of the toric fiber product. The second ingredient is the set $\mathrm{Lift}(F) \cup \mathrm{Lift}(H)$, these are the lifts of generators of the ideals I and J respectively.

We start by explaining the construction of the elements in Quad_B. Let $\mathcal{T}_j$ be an inductively constructed staged tree and $\mathcal{T}_{j+1} = [\mathcal{T}_j, \mathcal{T}_{G^j}]$ with $r_j = |G^j|$. Consider the monomial map

$$\begin{aligned} \phi_{B_j} : \mathbb{R}[p]_{\mathcal{T}_{j+1}} &\to \mathbb{R}[p^i_{\boldsymbol{a}}, p^i_k \mid \boldsymbol{a} \in G^j_i, k \in \boldsymbol{i}_{\mathcal{B}^j_i},\ i \in [r_j]] \\ p^i_{\boldsymbol{a}k} &\mapsto p^i_{\boldsymbol{a}} p^i_k \end{aligned} \tag{4}$$

where B_j denotes the exponent matrix of the monomial map ϕ_{B_j}. Set $I_{B_j} = \ker(\phi_{B_j})$ and

$$\mathrm{Quad}_{B_j} = \{\underline{p^i_{\boldsymbol{a}k_1} p^i_{\boldsymbol{b}k_2}} - p^i_{\boldsymbol{b}k_1} p^i_{\boldsymbol{a}k_2} \mid \boldsymbol{a}, \boldsymbol{b} \in G^j_i, k_1 \neq k_2 \in \boldsymbol{i}_{\mathcal{B}^j_i}, i \in [r_j]\}.$$

By Proposition 10 in [17] Quad_{B_j} is a Gröbner basis for I_{B_j} with respect to any term order that selects the underlined terms as leading terms.

We now explain the construction of the elements in $\mathrm{Lift}(F) \cup \mathrm{Lift}(H)$. Fix $\mathcal{T}, G, \mathcal{T}_G$ and $[\mathcal{T}, \mathcal{T}_G]$ as in Definition 3.3 and denote by $\mathcal{A}$ the multigrading of the rings $\mathbb{R}[p]_{\mathcal{T}}, \mathbb{R}[p]_{[\mathcal{T},\mathcal{T}_G]}$ defined by G. We recall the definition of a lift of an $\mathcal{A}$-graded polynomial in $\mathbb{R}[p]_{\mathcal{T}}$ to the polynomial ring $\mathbb{R}[p]_{[\mathcal{T},\mathcal{T}_G]}$ of the toric fiber product. Since we only consider pure quadratic binomials, we restrict the definition from [17] of lift to this particular case. Consider the $\mathcal{A}$-graded polynomial

$$f = p^{i_1}_{\boldsymbol{a}_1} p^{i_2}_{\boldsymbol{a}_2} - p^{i_1}_{\boldsymbol{a}_3} p^{i_2}_{\boldsymbol{a}_4} \in \mathbb{R}[p]_{\mathcal{T}},$$

where $\boldsymbol{a}_1, \boldsymbol{a}_3 \in G_{i_1}$, $\boldsymbol{a}_2, \boldsymbol{a}_4 \in G_{i_2}$ and $i_1, i_2 \in [r]$. Set $k = (k_1, k_2)$ with $k_1 \in \boldsymbol{i}_{\mathcal{B}_{i_1}}, k_2 \in \boldsymbol{i}_{\mathcal{B}_{i_2}}$ and consider $f_k \in \mathbb{R}[p]_{[\mathcal{T},\mathcal{T}_G]}$ defined by

$$f_k = p^{i_1}_{\boldsymbol{a}_1 k_1} p^{i_2}_{\boldsymbol{a}_2 k_2} - p^{i_1}_{\boldsymbol{a}_3 k_1} p^{i_2}_{\boldsymbol{a}_4 k_2}.$$

Definition 4.2. Let $\mathcal{A}$ be the multigrading of the rings $\mathbb{R}[p]_{\mathcal{T}}, \mathbb{R}[p]_{[\mathcal{T},\mathcal{T}_G]}$ defined by G and let $F \in \ker(\varphi_{\mathcal{T}})$ be a collection of pure $\mathcal{A}$-graded binomials. We associate to each $f \in F$ the set $T_f = \boldsymbol{i}_{\mathcal{B}_{i_1}} \times \boldsymbol{i}_{\mathcal{B}_{i_2}}$ of indices and define

$$\mathrm{Lift}(F) = \{f_k : f \in F, k \in T_f\}.$$

The set $\mathrm{Lift}(F)$ is called the lifting of F to $\ker(\varphi_{\mathcal{T}}) \times_{\mathcal{A}} \langle 0 \rangle$ (see [17]).

Definition 4.3. Let $\mathcal{T}_n$ be an inductively constructed staged tree with $\mathcal{T}_i = [\mathcal{T}_{i-1}, \mathcal{T}_{G^i}]$ for $1 \leq i \leq n$ and let $\mathcal{A}_i$ be the grading in $\mathbb{R}[p]_{\mathcal{T}_i}$ determined by G^i. Fix two nonnegative integers i, q with $0 \leq i + q \leq n - 1$. We define

$$\mathrm{Lift}^q(\mathrm{Quad}_{B_i}) := \mathrm{Lift}_{\mathcal{A}_{i+q}}(\cdots(\mathrm{Lift}_{\mathcal{A}_{i+2}}(\mathrm{Lift}_{\mathcal{A}_{i+1}}(\mathrm{Quad}_{B_i})))\cdots)$$

where the subscript in $\mathrm{Lift}_{\mathcal{A}}(\,\cdot\,)$ indicates the grading of the argument is with respect to $\mathcal{A}$.

We formulate a lemma that says that certain staged subtrees of balanced and stratified staged trees are also balanced and stratified.

Lemma 4.4. *Let $(\mathcal{T},\theta)$ be a staged tree of level m and let q be a positive integer with $1 \leq q \leq m-1$. If $(\mathcal{T},\theta)$ is balanced and stratified then the staged subtree $\mathcal{T}^{(q)}$ of $\mathcal{T}$ as defined in Definition 3.6 is also balanced and stratified.*

Proof. We must prove that if $\boldsymbol{a},\boldsymbol{b}$ are two vertices in $\mathcal{T}^{(q)}$ that are in the same stage, then they satisfy condition $(\star)$ in $\mathbb{R}[\Theta]_{\mathcal{T}^{(q)}}$. Since $\mathcal{T}$ is balanced, then $\boldsymbol{a},\boldsymbol{b}$ satisfy condition $(\star)$ in $\mathbb{R}[\Theta]_{\mathcal{T}}$. Namely,

$$t(\boldsymbol{a}k_1)t(\boldsymbol{b}k_2) = t(\boldsymbol{a}k_2)t(\boldsymbol{b}k_1) \text{ in } \mathbb{R}[\Theta]_{\mathcal{T}} \quad \text{for all } k_1,k_2 \in \{0,\dots,|\mathrm{ch}(\boldsymbol{a})|\}. \tag{5}$$

For a vertex v in $\mathcal{T}^{(q)}$ define $[v] = \{\boldsymbol{u} \in \boldsymbol{i}_{\mathcal{T}^{(q)}} \,|\, \text{the root-to-}\boldsymbol{u}\text{ path in } \mathcal{T}^{(q)} \text{ goes through } v\}$. Then by a repeated use of Lemma 2.8, for $\boldsymbol{c} \in \{\boldsymbol{a}k_1, \boldsymbol{b}k_2, \boldsymbol{a}k_2, \boldsymbol{b}k_1\}$,

$$t(\boldsymbol{c}) = \sum_{\boldsymbol{u}\in[\boldsymbol{c}]} \prod_{e \in E(\lambda_{\boldsymbol{u}})} \theta(e)t(\boldsymbol{u})$$

where $\lambda_{\boldsymbol{u}}$ is the $\boldsymbol{c}$ to $\boldsymbol{u}$ path in $\mathcal{T}^{(q)}$. Here $t(\boldsymbol{c})$ is an element of $\mathbb{R}[\Theta]_{\mathcal{T}}$. Denote by $t(\boldsymbol{c})|_{\mathcal{T}^{(q)}}$ the polynomial obtained from $t(\boldsymbol{c})$ by specializing $t(\boldsymbol{u}) = 1$ for all $\boldsymbol{u} \in [\boldsymbol{c}]$. Since $\mathcal{T}$ is stratified, $t(\boldsymbol{c})|_{\mathcal{T}^{(q)}}$ is the interpolating polynomial of the subtree $\mathcal{T}^{(q)}$ rooted at $\boldsymbol{c}$. Applying this specialization to (5) yields condition $(\star)$ for the vertices $\boldsymbol{a},\boldsymbol{b}$ in $\mathbb{R}[\Theta]_{\mathcal{T}^{(q)}}$.

□

Proposition 4.5. *Let $\mathcal{T}_i$ be a balanced and inductively constructed staged tree. Suppose $\mathcal{T}_{i+1} = [\mathcal{T}_i, \mathcal{T}_{G^i}]$ and $\mathcal{T}_{i+1}$ is balanced. Then the elements in*

$$\mathrm{Lift}^{i-2}(\mathrm{Quad}_{B_1}), \mathrm{Lift}^{i-3}(\mathrm{Quad}_{B_2}), \dots, \mathrm{Lift}(\mathrm{Quad}_{B_{i-2}}), \mathrm{Quad}_{B_{i-1}}$$

are $\mathcal{A}_i$-graded.

Proof. Note that any inductively constructed staged tree is stratified, therefore $\mathcal{T}_i$ and $\mathcal{T}_{i+1}$ are stratified. By assumption $\mathcal{T}_i$ is inductively constructed, therefore there is a sequence of trees and gluing components $(\mathcal{T}_1, \mathcal{T}_{G^1}), \dots, (\mathcal{T}_{i-1}, \mathcal{T}_{G^{i-1}})$ from which $\mathcal{T}_i$ is constructed. Fix $q \in \{0,1,\dots,i-2\}$ and $j = i-q-1$, we show that the binomials in $\mathrm{Lift}^q(\mathrm{Quad}_{B_j})$ are $\mathcal{A}_i$-graded. To this end we prove that for m such that $0 \leq m \leq q$, the equations in $\mathrm{Lift}^m(\mathrm{Quad}_{B_j})$ are $\mathcal{A}_{j+m+1}$-graded. The proof is by induction on m.

Fix $m = 0$, we show that the elements in Quad_{B_j} are $\mathcal{A}_{j+1}$-graded. The equations

$$\mathrm{Quad}_{B_j} = \bigcup_{\alpha=1}^{r_j} \{p_{\boldsymbol{a}k_1}p_{\boldsymbol{b}k_2} - p_{\boldsymbol{b}k_1}p_{\boldsymbol{a}k_2} \mid \boldsymbol{a},\boldsymbol{b} \in G^j_\alpha, k_1,k_2 \in \boldsymbol{i}_{\mathcal{B}^j_\alpha}\}$$

are the generators of I_{B_j}. The multidegrees $\mathcal{A}_{j+1}$ are defined according to the partition G^{j+1} of the leaves of $\mathcal{T}_{j+1}$. If two leaves in $\mathcal{T}_{j+1}$ are in the same set of the partition G^{j+1} then they have the same degree. Since $\mathcal{T}_i$ is balanced and $\mathcal{T}_{j+2} = \mathcal{T}_i^{(j+2)}$ then by Lemma 4.4 the staged tree $\mathcal{T}_{j+2}$ is balanced. Therefore all the vertices in the stages $G^j = \{G^j_1,\dots,G^j_{r_j}\}$ satisfy condition $(\star)$ in $\mathbb{R}[\Theta]_{\mathcal{T}_{j+2}}$. This means that for all $\alpha \in \{1,\dots,r_j\}$ and $\boldsymbol{a},\boldsymbol{b} \in G^j_\alpha$

$$t(\boldsymbol{a}k_1)t(\boldsymbol{b}k_2) = t(\boldsymbol{b}k_1)t(\boldsymbol{a}k_2) \text{ for } k_1,k_2 \in \boldsymbol{i}_{\mathcal{B}^j_\alpha} \tag{6}$$

where $\mathrm{ch}(\boldsymbol{a}) = \{\boldsymbol{a}k \,|\, k \in \boldsymbol{i}_{\mathcal{B}^j_\alpha}\}$ and $\mathrm{ch}(\boldsymbol{b}) = \{\boldsymbol{b}k \,|\, k \in \boldsymbol{i}_{\mathcal{B}^j_\alpha}\}$.

Using the construction of $\mathcal{T}_{j+2}$ from $\mathcal{T}_{j+1}$ and $\mathcal{T}_{G^{j+1}}$, we see that for any index $\boldsymbol{c} \in \{\boldsymbol{a}k, \boldsymbol{b}k \mid k \in \boldsymbol{i}_{\mathcal{B}^j_\alpha}\}$ of the leaves of $\mathcal{T}_{j+1}$, $t(\boldsymbol{c})$ is equal to the interpolating polynomial $t_{\mathcal{B}^{j+1}}$ of some one-level probability tree $\mathcal{B}^{j+1}$ in $\mathcal{T}_{G^{j+1}}$. Also, the assignment $t(\boldsymbol{c}) = t_{\mathcal{B}^{j+1}}$ determines the degree of $p_{\boldsymbol{c}}$ in $\mathbb{R}[p]_{\mathcal{T}_{j+1}}$. To be precise, if $t(\boldsymbol{c}) = t(\boldsymbol{c}')$ for $\boldsymbol{c}, \boldsymbol{c}' \in \{\boldsymbol{a}k, \boldsymbol{b}k \mid k \in \boldsymbol{i}_{\mathcal{B}^j_\alpha}\}$ then $\boldsymbol{c}$ and $\boldsymbol{c}'$ are in the same set of the partition G^{j+1} hence $\deg(p_{\boldsymbol{c}}) = \deg(p_{\boldsymbol{c}'})$. The fact that Equation (6) holds for $\mathcal{T}_{j+2}$ means that in $\mathbb{R}[\Theta]_{\mathcal{T}_{j+2}}$ we have

$$
\begin{aligned}
t(\boldsymbol{a}k_1)t(\boldsymbol{b}k_2) &= t_{\mathcal{B}_1^{j+1}} t_{\mathcal{B}_2^{j+1}} \\
&= t(\boldsymbol{b}k_1)t(\boldsymbol{a}k_2)
\end{aligned} \tag{7}
$$

for some $\mathcal{B}_1^{j+1}, \mathcal{B}_2^{j+1} \in \mathcal{T}_{G^{j+1}}$. Therefore $\deg(p_{\boldsymbol{a}k_1} p_{\boldsymbol{b}k_2}) = \deg(p_{\boldsymbol{b}k_1} p_{\boldsymbol{a}k_2})$ with respect to $\mathcal{A}_{j+1}$. This completes the proof for $m = 0$.

As a result, all the equations in Quad_{B_j} can be lifted to elements in $\ker(\varphi_{\mathcal{T}_{j+2}})$. Using the notation in Equation (7), the lift of an element $f = p_{\boldsymbol{a}k_1} p_{\boldsymbol{b}k_2} - p_{\boldsymbol{b}k_1} p_{\boldsymbol{a}k_2} \in \mathrm{Quad}_{B_j}$ depends on the assignment $\{\boldsymbol{a}k_1, \boldsymbol{a}k_2, \boldsymbol{b}k_1, \boldsymbol{b}k_2\} \to \{\mathcal{B}_1^{j+1}, \mathcal{B}_2^{j+1}\}$. For instance, if $t(\boldsymbol{a}k_1) = t(\boldsymbol{b}k_1) = t_{\mathcal{B}_1^{j+1}}$ and $t(\boldsymbol{b}k_2) = t(\boldsymbol{a}k_2) = t_{\mathcal{B}_2^{j+1}}$ then $T_f = \boldsymbol{i}_{\mathcal{B}_1^{j+1}} \times \boldsymbol{i}_{\mathcal{B}_2^{j+1}}$ so

$$
\mathrm{Lift}(f) = \{f_\beta \mid \beta \in T_f\} = \{p^1_{\boldsymbol{a}k_1\beta_1} p^2_{\boldsymbol{b}k_2\beta_2} - p^1_{\boldsymbol{b}k_1\beta_1} p^2_{\boldsymbol{a}k_2\beta_2} \mid \beta_1 \in \boldsymbol{i}_{\mathcal{B}_1^{j+1}}, \beta_2 \in \boldsymbol{i}_{\mathcal{B}_2^{j+1}}\}.
$$

Suppose we have constructed $\mathrm{Lift}^{m-1}(\mathrm{Quad}_{B_j})$ inductively by lifting the equations in Quad_{B_j} and at each step all equations lift. An element in $\mathrm{Lift}^{m-1}(\mathrm{Quad}_{B_j})$ is a binomial of the form

$$
f = p_{\boldsymbol{a}k_1 s} p_{\boldsymbol{b}k_2 u'} - p_{\boldsymbol{b}k_1 u} p_{\boldsymbol{a}k_2 s'} \tag{8}
$$

where $\alpha \in \{1, \ldots, r_j\}$, $\boldsymbol{a}, \boldsymbol{b} \in G^j_\alpha$, $k_1, k_2 \in \boldsymbol{i}_{\mathcal{B}^j_\alpha}$ and s, s', u, u' are sequences of nonnegative integers of length $m-1$ that arise as subindices after lifting $m-1$ times. Note that $\boldsymbol{a}k_1 s, \boldsymbol{b}k_2 u', \boldsymbol{b}k_1 u, \boldsymbol{a}k_2 s' \in \boldsymbol{i}_{\mathcal{T}_{j+m}}$The claim is that (8) is $\mathcal{A}_{j+m}$-graded.

Following a similar argument as for $m = 0$, we know that two elements in the same set of the partition G^{j+m} have the same multidegree with respect to $\mathcal{A}_{j+m}$. As before, this condition can be verified for f by checking that

$$
t(\boldsymbol{a}k_1 s)t(\boldsymbol{b}k_2 u') = t(\boldsymbol{b}k_1 u)t(\boldsymbol{a}k_2 s') \text{ in } \mathbb{R}[\Theta]_{\mathcal{T}_{j+m+1}}. \tag{9}
$$

For $\boldsymbol{c} \in \{\boldsymbol{a}k_1, \boldsymbol{b}k_2, \boldsymbol{a}k_2, \boldsymbol{b}k_1\}$, $[\boldsymbol{c}] := \{\beta \in \boldsymbol{i}_{\mathcal{T}_{j+m}} \mid \text{the root-to-}\beta \text{ path in } \mathcal{T}_{j+m} \text{ goes trough } \boldsymbol{c}\}$. To check that Equation 9 holds, consider (2) from Definition 2.10 for the vertices $\boldsymbol{a}, \boldsymbol{b} \in G^j_\alpha$. This equation is $t(\boldsymbol{a}k_1)t(\boldsymbol{b}k_2) = t(\boldsymbol{b}k_1)t(\boldsymbol{a}k_2)$ where $k_1, k_2 \in \boldsymbol{i}_{\mathcal{B}^j_\alpha}$. We use Lemma 2.8 to rewrite this equation as

$$
\begin{aligned}
&\left(\sum_{\boldsymbol{a}k_1 s \in [\boldsymbol{a}k_1]} \left(\prod_{e \in E(\boldsymbol{a}k_1 \to \boldsymbol{a}k_1 s)} \theta(e)\right) t(\boldsymbol{a}k_1 s)\right) \cdot \left(\sum_{\boldsymbol{b}k_2 u' \in [\boldsymbol{b}k_2]} \left(\prod_{e \in E(\boldsymbol{b}k_2 \to \boldsymbol{b}k_2 u')} \theta(e)\right) t(\boldsymbol{b}k_2 u')\right) \\
&= \left(\sum_{\boldsymbol{b}k_1 u \in [\boldsymbol{b}k_1]} \left(\prod_{e \in E(\boldsymbol{b}k_1 \to \boldsymbol{b}k_1 u)} \theta(e)\right) t(\boldsymbol{b}k_1 u)\right) \cdot \left(\sum_{\boldsymbol{a}k_2 s' \in [\boldsymbol{a}k_2]} \left(\prod_{e \in E(\boldsymbol{a}k_2 \to \boldsymbol{a}k_2 s')} \theta(e)\right) t(\boldsymbol{a}k_2 s')\right).
\end{aligned} \tag{10}
$$

When we specialize $t(\boldsymbol{a}k_1 s) = t(\boldsymbol{b}k_2 u') = t(\boldsymbol{b}k_1 u) = t(\boldsymbol{a}k_2 s') = 1$ in each sum in Equation (10) we recover the interpolating polynomials for $t(\boldsymbol{a}k_1), t(\boldsymbol{b}k_2), t(\boldsymbol{b}k_1), t(\boldsymbol{a}k_2)$ in $\mathbb{R}[\Theta]_{\mathcal{T}_{j+m}}$. By Lemma 4.4,

$\mathcal{T}_{j+m}$ satisfies condition ($\star$) therefore

$$\left(\sum_{\boldsymbol{a}k_1 s\in[\boldsymbol{a}k_1]}\prod_{e\in E(\boldsymbol{a}k_1\to\boldsymbol{a}k_1 s)}\theta(e)\right)\cdot\left(\sum_{\boldsymbol{b}k_2 u'\in[\boldsymbol{b}k_2]}\prod_{e\in E(\boldsymbol{b}k_2\to\boldsymbol{b}k_2 u')}\theta(e)\right)= \tag{11}$$

$$\left(\sum_{\boldsymbol{b}k_1 u\in[\boldsymbol{b}k_1]}\prod_{e\in E(\boldsymbol{b}k_1\to\boldsymbol{b}k_1 u)}\theta(e)\right)\cdot\left(\sum_{\boldsymbol{a}k_2 s'\in[\boldsymbol{a}k_2]}\prod_{e\in E(\boldsymbol{a}k_2\to\boldsymbol{a}k_2 s')}\theta(e)\right). \tag{12}$$

The factors in the above equality are sums of monomials all with coefficients equal to one. Thus for every pair $\boldsymbol{a}k_1 s\in[\boldsymbol{a}k_1], \boldsymbol{b}k_2u'\in[\boldsymbol{b}k_2]$ in the product of the left hand side of the equation, there exists a pair $\boldsymbol{a}k_2s'\in[\boldsymbol{a}k_2], \boldsymbol{b}k_2u\in[\boldsymbol{b}k_1]$ in the product of the right hand side of the equation such that

$$\left(\prod_{e\in E(\boldsymbol{a}k_1\to\boldsymbol{a}k_1 s)}\theta(e)\right)\cdot\left(\prod_{e\in E(\boldsymbol{b}k_2\to\boldsymbol{b}k_2 u')}\theta(e)\right)=\left(\prod_{e\in E(\boldsymbol{b}k_1\to\boldsymbol{b}k_1 u)}\theta(e)\right)\cdot\left(\prod_{e\in E(\boldsymbol{a}k_2\to\boldsymbol{a}k_2 s')}\theta(e)\right). \tag{13}$$

Hence condition ($\star$) for the vertices $\boldsymbol{a},\boldsymbol{b}$ in $\mathcal{T}_{j+m+1}$ can be rewritten as

$$\sum_{\boldsymbol{a}k_1 s\in[\boldsymbol{a}k_1],\boldsymbol{b}k_2u'\in[\boldsymbol{b}k_2]}\left(\prod_{e\in E(\boldsymbol{a}k_1\to\boldsymbol{a}k_1 s)}\theta(e)\right)\left(\prod_{e\in E(\boldsymbol{b}k_2\to\boldsymbol{b}k_2 u')}\theta(e)\right)(t(\boldsymbol{a}k_1s)t(\boldsymbol{b}k_2u')-t(\boldsymbol{b}k_1u)t(\boldsymbol{a}k_2s'))=0.$$

Since $\mathcal{T}_{j+m+1}$ is stratified, the variables involved in the factored monomials above are disjoint from the variables involved in the factors of the form $t(\boldsymbol{a}k_1s)t(\boldsymbol{b}k_2u')-t(\boldsymbol{b}k_1u)t(\boldsymbol{a}k_2s')$, therefore the equation above holds if and only if $t(\boldsymbol{a}k_1s)t(\boldsymbol{b}k_2u')-t(\boldsymbol{b}k_1u)t(\boldsymbol{a}k_2s')=0$ for each summand. This proves that the elements in $\mathrm{Lift}^{m-1}(\mathrm{Quad}_{B_j})$ are $\mathcal{A}_{j+m}$-graded. □

Proof of Theorem 2.5. If $\mathcal{T}$ is stratified, then $\mathcal{T}$ is an iteratively constructed staged tree and $\mathcal{T}=\mathcal{T}_n$ for some n. Set $F_n=\mathrm{Lift}^{n-2}_{\mathcal{A}_n}(\mathrm{Quad}_{B_1})\cup\mathrm{Lift}^{n-3}_{\mathcal{A}_n}(\mathrm{Quad}_{B_1})\cup\cdots\cup\mathrm{Quad}_{B_{n-1}}$. We prove by induction on n that $\ker(\varphi_{\mathcal{T}_n})$ is generated by F_n and that F_n is a Gröbner basis with squarefree intial ideal. The first non-trivial case is $n=2$. We have $F_2=\mathrm{Quad}_{B_1}$ and from Proposition 10 in [17], F_2 is a Gröbner basis for the ideal $\ker(\varphi_{\mathcal{T}_2})=\ker(\varphi_{\mathcal{T}_1})\times_{\mathcal{A}_1}\langle 0\rangle$. Suppose the statement is true for i, so the elements in F_i are a Gröbner basis for $\ker(\varphi_{\mathcal{T}_i})$. Since $\mathcal{T}_n$ is balanced, by Lemma 4.4 the trees $\mathcal{T}_i$ and $\mathcal{T}_{i+1}$ are also balanced. Then from Proposition 4.5 the elements in F_i are $\mathcal{A}_i$ homogeneous, so by [17, Proposition 10] the set F_{i+1} is a Gröbner basis for $\ker(\varphi_{\mathcal{T}_{i+1}})$. Since the elements in F_n are all extensions of elements in Quad_{B_j} for j with $1\le j\le n-1$ we see that the leading terms of these binomials are squarefree. Hence the initial ideal of $\langle F_n\rangle$ is squarefree. □

Corollary 4.6. *Let $(\mathcal{T},\theta)$ be a balanced and stratified staged tree. Fix Δ to be the polytope defined by the convex hull of the lattice points in the exponent matrix of the map $\varphi_{\mathcal{T}}$. Then Δ has a regular unimodular triangulation. In particular the toric variety defined by* $\ker(\varphi_{\mathcal{T}})$ *is Cohen-Macaulay.*

Proof. The ideal $\ker(\varphi_{\mathcal{T}})$ has a square free quadratic Gröbner basis with respect to a term order $\prec$. From [16, Corollary 8.9], this induces a regular unimodular triangulation of Δ. □

5. Connections to discrete statistical models

Staged tree models are a class of graphical discrete statistical models introduced by Anderson and Smith in [14]. While Bayesian networks and decomposable models are defined via conditional independence statements on random variables corresponding to the vertices of a graph, staged tree models encode independence relations on the events of an outcome space represented by a tree. In the statistical literature these models are also referred to as chain event graphs. We refer the reader to the book [15] and to [19] to find out more about their statistical properties, practical implementation and causal interpretation. In this section we give a formal definition of these models and recall results from [5] and [8] about their defining equations.

Given a discrete random variable X with state space $\{0, \dots, n\}$, a probability distribution on X is a vector $(p_0, \dots, p_n) \in \mathbb{R}^{n+1}$ where $p_i = P(X = i)$, $i \in \{0, \dots, n\}$, $p_i \geq 0$ and $\sum_{i=0}^{n} p_i = 1$. The open probability simplex

$$\Delta_n^\circ = \{(p_0, \dots, p_n) \in \mathbb{R}^{n+1} \mid p_i > 0, p_0 + \dots + p_n = 1\}$$

consists of all the possible positive probability distributions for a discrete random variable with state space $\{0, \dots, n\}$. A *discrete statistical model* is a subset of Δ_n°. In the next definition we associate a discrete statistical model to a given staged tree.

Definition 5.1. Let $(\mathcal{T}, \theta)$ be a staged tree and let $\boldsymbol{\theta} = (\theta(e) \mid \theta(e) \in \mathrm{im}(\theta))$ be a vector of parameters where each entry is a label in $\mathcal{L}$. We define the parameter space $\Theta_{\mathcal{T}} := \{\boldsymbol{\theta} \mid \theta(e) \in (0,1)$ and for all $\boldsymbol{a} \in V$, $\sum_{e \in E(\boldsymbol{a})} \theta(e) = 1\}$. Note that $\Theta_{\mathcal{T}}$ is a product of simplices. A *staged tree model* $\mathcal{M}_{(\mathcal{T},\theta)}$ is the image of the map $\Psi_{\mathcal{T}} : \Theta_{\mathcal{T}} \to \Delta_{|\boldsymbol{i}_{\mathcal{T}}|-1}^\circ$ defined by

$$\boldsymbol{\theta} \mapsto p_{\boldsymbol{\theta}} = \left(\prod_{e \in E(\lambda_{\boldsymbol{j}})} \theta(e) \right)_{\boldsymbol{j} \in \boldsymbol{i}_{\mathcal{T}}}.$$

We can check that for every $\boldsymbol{\theta} \in \Theta_{\mathcal{T}}$, $p_{\boldsymbol{\theta}}$ is a probability distribution and therefore $\Psi(\Theta_{\mathcal{T}}) \subset \Delta_{|\boldsymbol{i}_{\mathcal{T}}|-1}^\circ$. Two staged trees $(\mathcal{T}, \theta)$ and $(\mathcal{T}', \theta')$ are said to be *statistically equivalent* if there exists a bijection between the sets $\Lambda_{\mathcal{T}}$ and $\Lambda_{\mathcal{T}'}$ in such a way that the image of $\Psi_{\mathcal{T}}$ is equal to the image of $\Psi_{\mathcal{T}'}$ under this bijection.

Example 5.2. The staged tree $\mathcal{T}_1$ in Figure 1 is the staged tree representation of the decomposable model associated to the undirected graph $G = [12][23][34]$ on four nodes.

Remark 5.3. For staged tree models, the root-to-leaf paths in the tree represent the possible unfoldings of a sequence of events. Given an edge (v, w) in $\mathcal{T}$, the label $\theta(v, w)$ is the transition probability from v to w given arrival at v.

Remark 5.4. A staged tree model $\mathcal{M}_{(\mathcal{T},\theta)}$ is a discrete statistical model parameterized by polynomials. The domain of this model is a semialgebraic set given by a product of simplices. As a consequence the image of $\Psi_{\mathcal{T}}$ is also a semialgebraic set. An important property of these models as noted in [8] is that the only inequality constraints of the image of $\Psi_{\mathcal{T}}$ are the ones imposed by the probability simplex, namely $0 \leq p_{\boldsymbol{j}} \leq 1$ for $\boldsymbol{j} \in \boldsymbol{i}_{\mathcal{T}}$ and $\sum_{\boldsymbol{j} \in \boldsymbol{i}_{\mathcal{T}}} p_{\boldsymbol{j}} = 1$.

In Definition 2.2 we defined the toric ideal associated to a staged tree $(\mathcal{T}, \theta)$. Now we define the ideal associated to a staged tree model $\mathcal{M}_{(\mathcal{T},\theta)}$. For this we use the rings $\mathbb{R}[p]_{\mathcal{T}}$ and $\mathbb{R}[\Theta]_{\mathcal{T}}$ from Definition 2.2. Consider the ideal $\mathfrak{q}$ of $\mathbb{R}[\Theta]_{\mathcal{T}}$ generated by all sum-to-one conditions $1 - \sum_{e \in E(\boldsymbol{a})} \theta(e)$ for $\boldsymbol{a} \in V$ and let $\mathbb{R}[\Theta]_{\mathcal{M}_{\mathcal{T}}} := \mathbb{R}[\Theta]_{\mathcal{T}}/\mathfrak{q}$. Denote by π the canonical projection from $\mathbb{R}[\Theta]_{\mathcal{T}}$ to the quotient ring $\mathbb{R}[\Theta]_{\mathcal{M}_{\mathcal{T}}}$.

Definition 5.5. Let $\mathcal{M}_{(\mathcal{T},\theta)}$ be a staged tree model and set $\overline{\varphi}_{\mathcal{T}} := \pi \circ \varphi_{\mathcal{T}}$. The ideal $\ker(\overline{\varphi}_{\mathcal{T}})$ is the *staged tree model ideal* associated to the model $\mathcal{M}_{(\mathcal{T},\theta)}$.

From the definition it follows that for every staged tree $(\mathcal{T}, \theta)$, the toric staged tree ideal is contained in the staged tree model ideal, i.e. $\ker(\varphi_{\mathcal{T}}) \subset \ker(\overline{\varphi}_{\mathcal{T}})$. It is not true in general that these two ideals are equal [5]. However, Theorem 10 in [5] states that if a staged tree $(\mathcal{T}, \theta)$ is balanced, then $\ker(\varphi_{\mathcal{T}}) = \ker(\overline{\varphi}_{\mathcal{T}})$.

Corollary 5.6. *If $(\mathcal{T}, \theta)$ is a balanced and stratified staged tree, then the ideal* $\ker(\overline{\varphi}_{\mathcal{T}})$ *has a quadratic Gröbner basis with squarefree initial ideal.*

Example 5.7. Consider the staged tree model defined by the tree in Figure 1 as in Example 5.2. Since this staged tree model is equal to the decomposable model given by $G = [12][23][34]$, from [7] we know it has a quadratic Gröbner basis. We recover the same result from the perspective of staged trees by using Corollary 5.6.

Corollary 5.6 is relevant in statistics because of the connection of Gröbner bases to sampling [1]. We presented Example 5.7 where a balanced and stratified staged tree represents an instance of a decomposable graphical model. We now provide more examples of staged tree models for which Corollary 5.6 holds. The first one is an explanation of the contraction axiom for conditional independence statements through the lens of staged trees. Before we present our examples we do a quick overview of discrete conditional independence models. Our exposition follows that in [18, Chapter 4], for more details we refer the reader to [13, Chapters 1,2,3] and [4].

Let $X = (X_1, \ldots, X_n)$ be a vector of discrete random variables, where X_i has state spaces $[d_i]$ for $i \in [n]$. The vector X has state space $\mathcal{X} = [d_1] \times \cdots \times [d_n]$ and we write $p_{u_1\cdots u_n}$ for the probability $P(X_1 = u_1, \ldots, X_n = u_n)$. For each subset $A \subset [n]$, X_A is the subvector of X indexed by the elements in A. Similarly, $\mathcal{X}_A = \prod_{i\in A}[d_i]$ and for a vector $x \in \mathcal{X}$, x_A denotes the restriction of x to the indexes in A.

Definition 5.8. Let A, B, C be pairwise disjoint subsets of $[n]$. The random vector X_A is conditionally indpendent of X_B given X_C if for every $a \in X_A, b \in X_B$ and $c \in X_C$

$$P(X_A = a, X_B = b | X_C = c) = P(X_A = a | X_C = c) \cdot P(X_B = b | X_C = c)$$

The notation $X_A \perp\!\!\!\perp X_B \,|\, X_C$ is used to denote that the random vector X satisfies the *conditional independence statement* that X_A is conditionally independent on X_B given X_C. When C is the empty set this reduces to marginal independence between X_A and X_B.

If C is a list of conditional independence statements among variables in a vector X, the *conditional independence model* $\mathcal{M}_C$ is the set of all probability distributions on $\mathcal{X}$ that satisfy the conditional independence statements in C. A conditional independence statement $X_A \perp\!\!\!\perp X_B \,|\, X_C$ translates into the condition that the joint probability distribution of the variables in X satisfies a set of quadratic equations. For elements $a \in X_A, b \in X_B$ and $c \in X_C$ we set $p_{a,b,c,+} = P(X_A = a, X_B = b, X_C = c)$.

Proposition 5.9 ([18])**.** *If X is a discrete random vector then the independence statement $X_A \perp\!\!\!\perp X_B \,|\, X_C$ holds for X if and only if*

$$p_{a_1,b_1,c,+}p_{a_2,b_2,c,+} - p_{a_1,b_2,c,+}p_{a_2,b_1,c,+} = 0$$

for all $a_1, a_2 \in \mathcal{X}_A$, $b_1, b_2 \in \mathcal{X}_B$ and $c \in \mathcal{X}_C$.

The conditional independence ideal $I_{A\perp\!\!\!\perp B|C}$ is the ideal generated by all quadrics in Proposition 5.9. If C is a list of conditional independence statements then we define I_C as the sum of all conditional independence ideals associated to statements in C.

Example 5.10. We consider the contraction axiom for positive distributions using staged tree models. Fix three discrete random variables X_1, X_2, X_3 with state spaces $[d_1 + 1], [d_2 + 1], [d_3 + 1]$ respectively. The contraction axiom states that the set of conditional independence statements $C = \{X_1 \perp\!\!\!\perp X_2 \,|\, X_3, X_2 \perp\!\!\!\perp X_3\}$ implies the statement $X_2 \perp\!\!\!\perp (X_1, X_3)$. A primary decomposition of the ideal I_C was obtained in [6, Theorem 1]. Here we provide a proof using staged trees, that one of the

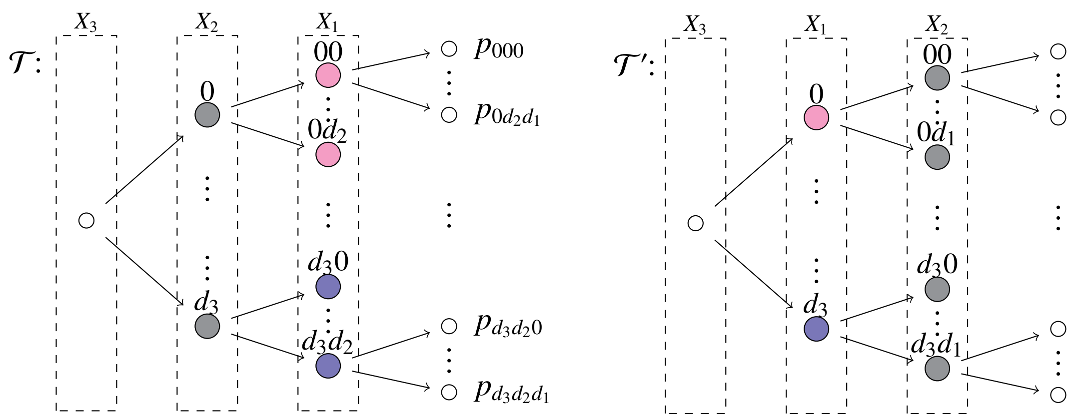

FIGURE 2. The staged trees $\mathcal{T}$ and $\mathcal{T}'$ are statistically equivalent, they represent the contraction axiom for three discrete random variables X_1, X_2 and X_3.

primary components of I_C is the prime binomial ideal $I_{X_2 \perp\!\!\!\perp (X_1,X_3)}$. As mentioned in [6] this is a well known fact. First we explain how to represent the two statements in C with a staged tree. Consider the tree $\mathcal{T}$ in Figure 2. This tree represents the state space of the vector (X_3, X_2, X_1) as a sequence of events where X_3 takes place first, X_2 second and X_1 third. The vertices of $\mathcal{T}$ are indexed recursively as defined at the beginning of Section 3. The statement $X_2 \perp\!\!\!\perp X_3$ is represented by the stage consisting of the vertices $\{0, \ldots, d_3\}$, these are colored gray in $\mathcal{T}$. The statement $X_1 \perp\!\!\!\perp X_2 \,|\, X_3$ is represented by the stages $S_0, \ldots, S_{d_3}$ where $S_i = \{ij \,|\, j \in \{0, \ldots, d_2\}\}$ and $i \in \{0, \ldots, d_3\}$. These stages mean that for a given outcome of X_3 the unfolding of the event X_2 followed by X_1 behaves as an independence model on two random variables. In Figure 2 the stage S_0 is colored in pink and the stage S_{d_3} is colored in purple. Although the gray vertices are not in the same position, we can easily check that $\mathcal{T}$ is balanced and stratified. Therefore $\ker(\varphi_{\mathcal{T}})$ has a quadratic Gröbner basis. Following the proof of Theorem 2.5 we can construct this basis explicitly. It consists of a set of quadratic equations given by the elements in Quad_{B_2} coming from the stages in $S_0, \ldots, S_{d_3}$ and the lifts of the equations Quad_{B_1} coming from the stage $\{0, \ldots, d_3\}$. If we swap the order of X_1 and X_2 in $\mathcal{T}$, we obtain the staged tree $\mathcal{T}'$ in Figure 2. This tree represents the same statistical model as $\mathcal{T}$ now with the unfolding of events X_3, X_2, X_1. The gray stages in $\mathcal{T}'$ represent the statement $X_2 \perp\!\!\!\perp (X_1, X_3)$. Hence, after establishing the evident bijection between the leaves of $\mathcal{T}$ and $\mathcal{T}'$ we see that $I_{X_2 \perp\!\!\!\perp (X_1,X_3)} = \ker(\varphi_{\mathcal{T}'}) = \ker(\varphi_{\mathcal{T}})$.

One of the main differences between staged tree models and discrete Bayesian networks is that the state space of a Bayesian network is equal to the product of the state spaces of the random variables in the vertices of the graph while the state space of a staged tree model does not necessarily have to equal a cartesian product. When $\mathcal{T}$ is not equal to the cartesian product of some finite sets we call the tree $\mathcal{T}$ *asymmetric*. The lemmas that follow are important to show that Theorem 2.5 also holds for the case when $\mathcal{T}$ is asymmetric. This implies that we can use Theorem 2.5 to construct quadratic Gröbner bases for staged tree models whose underlying tree does not necessarily represents the distribution of a vector of discrete random variables.

The definition of staged tree in [8] requires that each vertex in $\mathcal{T}$ has either no or at least two outgoing edges from v. We stepped away from making this requirement for the staged trees we consider in Section 2. In the next lemmas we explain how this mild extension of the definition behaves with respect to condition $(\star)$ and how trees defined according to [8] are recovered from the more general trees we consider. Throughout the next lemmas, we fix a staged tree $(\mathcal{T}, \theta)$ with

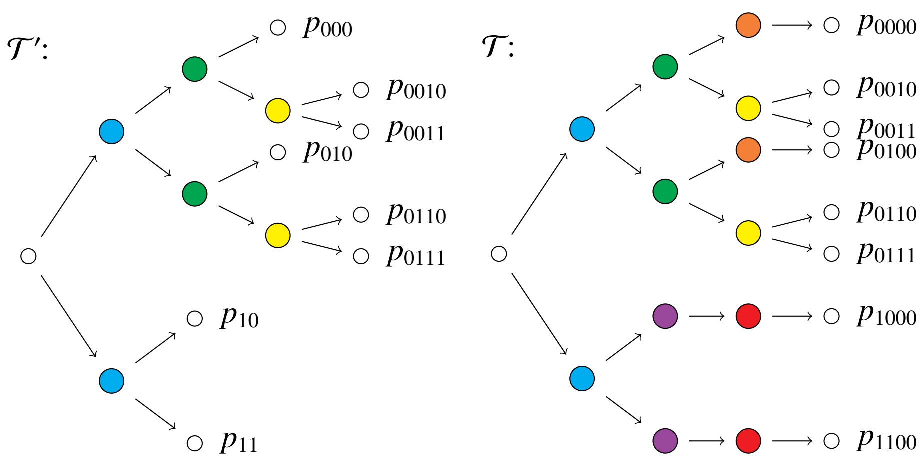

FIGURE 3. The staged trees $\mathcal{T}$ and $\mathcal{T}'$ are statistically equivalent.

edge set E and define $E_1 = \{e \in E \mid E(v) = \{e\}$ for some $v \in V\}$. For the trees in Figure 3, $\mathcal{T}$ has $|E_1| = 6$ while for $\mathcal{T}'$, $|E_1| = 0$.

Lemma 5.11. *Suppose $(\mathcal{T}, \theta)$ is a staged tree. Let $\mathcal{T}'$ be the staged tree obtained from $\mathcal{T}$ by contracting the edges in E_1. Then $\mathcal{M}_{(\mathcal{T},\theta)} = \mathcal{M}_{(\mathcal{T}',\theta)}$ and* $\ker(\overline{\varphi}_{\mathcal{T}}) = \ker(\overline{\varphi}_{\mathcal{T}'})$.

Proof. First, note that the number of root-to-leaf paths in $\mathcal{T}'$ is the same as in $\mathcal{T}$. Moreover, each root-to-leaf path λ' in $\mathcal{T}'$ is obtained from a unique root-to-leaf path λ in $\mathcal{T}$ by contracting the edges in E_1. Now let λ be a root-to-leaf path in $\mathcal{T}$. The λ-coordinate of the map $\Psi_{\mathcal{T}}$ applied to an element $\boldsymbol{\theta} \in \Theta_{\mathcal{T}}$ is

$$[\Psi_{\mathcal{T}}(\boldsymbol{\theta})]_{\lambda} = \prod_{e \in E(\lambda)} \theta(e) = \prod_{e \in E(\lambda')} \theta(e)$$
$$= [\Psi_{\mathcal{T}'}(\boldsymbol{\theta}\,|_{\mathcal{T}'})]_{\lambda'}$$

The second equality in the previous equation follows from taking a closer look at $\Theta_{\mathcal{T}}$. Indeed for all $e \in E_1$ we have $\theta(e) = 1$ because of the sum-to-one conditions imposed on $\Theta_{\mathcal{T}}$ in Definition 5.1. For the third equality, $\boldsymbol{\theta}\,|_{\mathcal{T}'}$ denotes the restriction of the vector $\boldsymbol{\theta}$ to the edge labels of $\mathcal{T}'$. It follows from the equalities above that the coordinates of $\Psi_{\mathcal{T}}$ and $\Psi_{\mathcal{T}'}$ are equal. Therefore $\mathcal{M}_{(\mathcal{T},\theta)} = \mathcal{M}_{(\mathcal{T}',\theta)}$. A similar argument applied to the maps $\overline{\varphi}_{\mathcal{T}}$ and $\overline{\varphi}_{\mathcal{T}'}$ shows that $\ker(\overline{\varphi}_{\mathcal{T}}) = \ker(\overline{\varphi}_{\mathcal{T}'})$. To carry out this argument we need to reindex the leaves of the trees, this can be done by dropping the index of the elements in E_1. □

We illustrate Lemma 5.11 in Figure 3 where $\mathcal{T}'$ is obtained from $\mathcal{T}$ by contracting the six edges in E_1. The two staged trees in this figure define the same statistical model.

Remark 5.12. To prove Corollary 5.6 we used [5, Theorem 10]. The proof of Theorem 10 in [5] is presented for trees such that $E_1 = 0$. However the result still holds when $|E_1| > 1$ because the ideal I_{Paths} (from [5]) is contained in $\ker(\varphi_{\mathcal{T}})$ in this case also, see [5] for more details.

Lemma 5.13. *Suppose $(\mathcal{T}, \theta)$ is a balanced and stratified staged tree. Let $\mathcal{T}'$ be the tree obtained from $\mathcal{T}$ by contracting the edges in E_1. Then $(\mathcal{T}', \theta)$ is also balanced.*

Proof. Suppose $\mathcal{T}$ is balanced and $\boldsymbol{a}, \boldsymbol{b}$ are in the same stage. Following the notation from Definition 2.10, we have $t(\boldsymbol{a}i)t(\boldsymbol{b}j) = t(\boldsymbol{b}j)t(\boldsymbol{a}j)$ in $\mathbb{R}[\Theta]_{\mathcal{T}}$, for all $i \neq j \in \{0, 1, \ldots, k\}$. If we specialize

$\theta(e) = 1$ in this equation for all $e \in E_1$ and since $\mathcal{T}'$ is stratified, then $t(\boldsymbol{a}i)t(\boldsymbol{b}j)\ |_{\theta(e)=1, e\in E_1} = t(\boldsymbol{b}j)t(\boldsymbol{a}i)\ |_{\theta(e)=1, e\in E_1}$ in $\mathbb{R}[\Theta]_{\mathcal{T}'}$. Therefore $\mathcal{T}'$ is also balanced. □

Corollary 5.14. *Suppose $\mathcal{T}$ is a balanced and stratified staged tree with $|E_1| > 1$. Let $\mathcal{T}'$ be the staged tree obtained from $\mathcal{T}$ by contracting the edges in E_1. Then $\ker(\overline{\varphi}_{\mathcal{T}'})$ is a toric ideal with a quadratic Gröbner basis whose intial ideal is squarefree.*

Proof. From Corollary 5.6 it follows that $\ker(\overline{\varphi}_{\mathcal{T}})$ is a toric ideal with a quadratic Gröbner basis and squarefree initial ideal. After an appropiate bijection, by Lemma 5.11, $\ker(\overline{\varphi}_{\mathcal{T}}) = \ker(\overline{\varphi}_{\mathcal{T}'})$. □

We illustrate the result in Corollary 5.14 with an example.

Example 5.15. Fix $\mathcal{T}$ and $\mathcal{T}'$ to be the staged trees in Figure 3. The staged tree $\mathcal{T}'$ is considered in [5, Section 6] as an example of the possible unfolding of events in a cell culture. A thorough discussion of this example and its difference with other graphical models is also contained in [5, Section 6]. Here we explain how to obtain a Gröbner basis for $\ker(\varphi_{\mathcal{T}'})$ using Corollary 5.14. The tree $\mathcal{T}'$ is balanced and statistically equivalent to $\mathcal{T}$. By Corollary 5.6, $\mathcal{T}$ has a quadratic Gröbner basis with square free initial ideal. Using the lemmas preceeding this example, there is a bijection between the root-to-leaf paths in $\mathcal{T}$ and $\mathcal{T}'$ thus $\mathbb{R}[p]_{\mathcal{T}}$ and $\mathbb{R}[p]_{\mathcal{T}'}$ are isomorphic. Under this isomorphism, the Gröbner basis for $\ker(\overline{\varphi}_{\mathcal{T}})$ is a Gröbner basis for $\ker(\overline{\varphi}_{\mathcal{T}'})$ its generators are

$$p_{0111}p_{10} - p_{0011}p_{110},\ p_{0011}p_{0110} - p_{0010}p_{0111},\ p_{0110}p_{10} - p_{0010}p_{110},$$
$$p_{0010}p_{010} - p_{000}p_{0110},\ p_{0011}p_{010} - p_{000}p_{0111},\ p_{010}p_{10} - p_{000}p_{110}.$$

Acknowledgements

Computations using the symbolic algebra software *Macaulay2* [11] were crucial for the development of this paper. We thank Thomas Kahle for his support in the completion of this project and Bernd Sturmfels for a careful reading of an earlier version of this manuscript. Both authors were supported by the Deutsche Forschungsgemeinschaft (314838170, GRK 2297 MathCoRe).

References

[1] S. Aoki, H. Hara, and A. Takemura. *Markov bases in algebraic statistics*. Springer Series in Statistics. Springer, New York, 2012.
[2] Y. Chen, I. H. Dinwoodie, and S. Sullivant. Sequential importance sampling for multiway tables. *Ann. Statist.*, 34(1):523–545, 2006.
[3] P. Diaconis and B. Sturmfels. Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.*, 26(1):363–397, 1998.
[4] M. Drton, B. Sturmfels, and S. Sullivant. *Lectures on algebraic statistics*, volume 39. Springer Science & Business Media, 2008.
[5] E. Duarte and C. Görgen. Equations defining probability tree models. *Journal of Symbolic Computation*, to appear, arXiv:1802.04511.
[6] L. D. Garcia, M. Stillman, and B. Sturmfels. Algebraic geometry of Bayesian networks. *J. Symbolic Comput.*, 39(3-4):331–355, 2005.
[7] D. Geiger, C. Meek, and B. Sturmfels. On the toric algebra of graphical models. *Ann. Statist.*, 34(3):1463–1492, 2006.
[8] C. Görgen. *An algebraic characterisation of staged trees: their geometry and causal implications*. PhD thesis, University of Warwick, Department of Statistics, 2017.
[9] C. Görgen, A. Bigatti, E. Riccomagno, and J. Q. Smith. Discovery of statistical equivalence classes using computer algebra. *International Journal of Approximate Reasoning*, 95:167–184, 2018.
[10] C. Görgen, J. Q. Smith, et al. Equivalence classes of staged trees. *Bernoulli*, 24(4A):2676–2692, 2018.
[11] D. R. Grayson and M. E. Stillman. *Macaulay2, a software system for research in algebraic geometry*. Available at `http://www.math.uiuc.edu/Macaulay2/`.

[12] J. Herzog, T. Hibi, and H. Ohsugi. *Binomial ideals*, volume 279 of *Graduate Texts in Mathematics*. Springer, Cham, 2018.
[13] M. Maathuis, M. Drton, S. Lauritzen, and M. Wainwright. *Handbook of Graphical Models*. CRC Press, 2018.
[14] J. Q. Smith and P. E. Anderson. Conditional independence and chain event graphs. *Artificial Intelligence*, 172(1):42 – 68, 2008.
[15] J. Q. Smith, C. Görgen, and R. A. Collazo. Chain event graphs. *preparation for Chapman & Hall*, 2017.
[16] B. Sturmfels. *Gröbner bases and convex polytopes*, volume 8 of *University Lecture Series*. American Mathematical Society, Providence, RI, 1996.
[17] S. Sullivant. Toric fiber products. *Journal of Algebra*, 316(2):560–577, 2007.
[18] S. Sullivant. *Algebraic statistics*, volume 194 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2018.
[19] P. Thwaites, J. Q. Smith, and E. Riccomagno. Causal analysis with chain event graphs. *Artificial Intelligence*, 174(12-13):889–909, 2010.

Authors' addresses:
Lamprini Ananiadi, Otto-von-Guericke Universität, Magdeburg, lamprini.ananiadi@ovgu.de
Eliana Duarte, MPI-MiS ,Leipzig and Otto-von-Guericke Universität, Magdeburg, eliana.duarte@ovgu.de