Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig

Toric geometry of path signature varieties

by

Laura Colmenarejo, Francesco Galuppi, and Mateusz Michałek

Preprint no.: 28 2019

Toric geometry of path signature varieties

Laura Colmenarejo[*], Francesco Galuppi[†], Mateusz Michałek[‡§]

March 12, 2019

Abstract

In stochastic analysis, a standard method to study a path is to work with its signature. This is a sequence of tensors of different order that encode information of the path in a compact form. When the path varies, such signatures parametrize an algebraic variety in the tensor space. The study of these signature varieties builds a bridge between algebraic geometry and stochastics, and allows a fruitful exchange of techniques, ideas, conjectures and solutions.

In this paper we study the signature varieties of two very different classes of paths. The class of rough paths is a natural extension of the class of piecewise smooth paths. It plays a central role in stochastics, and its signature variety is toric. The class of axis-parallel paths has a peculiar combinatoric flavour, and we prove that it is toric in many cases.

1 Introduction

A *path* is a continuous map $X\colon [0,1] \to \mathbb{R}^d$. This very simple mathematical object can be used to interpret a wide range of situations. From a physical transformation to a meteorological model, from a medical experiment to the stock market, everything that involves parameters changing with time can be described by a path. This makes paths priceless tools in many branches of mathematics, as well as in a number of applied sciences. The downside is that, being a continuous object, explicit computations on a path are not easy to handle. A typical way to overcome this problem is to find invariants that are simpler to understand and can provide us enough information.

For paths, this problem was faced in [Che54]. Assume that X is piecewise smooth, and fix $k \in \mathbb{N}$. Let X_i be the composition of X with the projection to the i-th coordinate. Chen defined the *k-th signature* of X to be the order k tensor $\sigma^{(k)}(X)$ whose $(i_1 \dots i_k)$-th entry is

$$\int_0^1 \int_0^{t_k} \cdots \int_0^{t_3} \int_0^{t_2} \dot{X}_{i_1}(t_1) \cdot \ldots \cdot \dot{X}_{i_k}(t_k) dt_1 \dots dt_k.$$

By convention, $\sigma^{(0)}(X) = 1$. The sequence

$$\sigma(X) := (\sigma^{(k)}(X) \mid k \in \mathbb{N})$$

[*]Laura Colmenarejo, University of Massachusetts at Amherst (Amherst, US), and Max Planck Institute for Mathematics in the Sciences (Leipzig, Germany)

[†]Francesco Galuppi, Max Planck Institute for Mathematics in the Sciences (Leipzig, Germany), galuppi@mis.mpg.de

[‡]Mateusz Michałek, Max Planck Institute for Mathematics in the Sciences (Leipzig, Germany), michalek@mis.mpg.de, and Polish Academy of Sciences, Institute of Mathematics (Warsaw, Poland), and Aalto University (Aalto, Finland). MM was supported by the Polish National Science Center Project 2013/08/A/ST1/00804 affiliated at the University of Warsaw.

[§]The authors would like to thank Lara Bossinger for the discussion in the early stage of this project, Avinash Kulkarni for the helpful suggestions and Joscha Diehl for his feedback.

is called the *signature* of X. Sometimes it is also convenient to consider a truncated signature $\sigma^{\leq m}(X) := (\sigma^{(k)}(X) \mid k \in \{0, \dots, m\})$. In this context, it is natural to ask how much these tensors tell us about X. In [Che58, Theorem 4.1] Chen proved that, up to a mild equivalence relation, the signature allows to uniquely recover a piecewise smooth path.

Signatures are appreciated not only in stochastics, but also in topological data analysis, financial mathematics or machine learning among others (see [GLKF13, CK16, CNO18, CO18]). Chen's iterated integrals have deep connections with de Rham homotopy theory, as shown in [Hai02]. In this paper we are interested in the geometric side, illustrated in Section 1.2. We study the so-called *signature varieties* – roughly speaking the geometric locus of all possible signatures. In Section 2 we consider the variety $\mathcal{R}_{d,k,m}$ of signatures of rough paths. Our results often rely on toric geometry, which allows us to answer several questions about their equations and singularities:

- Answering a question raised in [AFS18] and [Gal18], we show that the ideal of $\mathcal{R}_{d,k,m}$ may be generated in arbitrary high degree - Proposition 2.9. However, we prove that quadratic polynomials define a (possibly reducible) variety, of which $\mathcal{R}_{d,k,m}$ is an irreducible component - Proposition 2.11.
- We characterize the cases in which $\mathcal{R}_{d,k,m}$ is an embedding of the weighted projective space. We find conditions that make it (projectively) normal and examples when it is not - Subsection 2.3.
- We study the asymptotic behavior of the degree of $\mathcal{R}_{d,k,m}$ and we give explicit formulas for some special cases - Subsection 2.4.

In Section 3 we study the geometry of the signature variety $\mathcal{A}_{\nu,k}$ of axis parallel paths:

- We provide an easy, combinatorial parametrization of the variety - Lemma 3.3.
- We apply the above description to prove that $\mathcal{A}_{\nu,k}$ is toric in several cases - Section 3.1. We study the dimension of $\mathcal{A}_{\nu,k}$ exhibiting defective cases - Section 3.2.

Finally, we use our knowledge of axis paths to prove a general formula for the determinant of the signature matrix of any path, Theorem 3.19 and Corollary 3.23.

In order to present those results we need to recall the algebraic background.

1.1 The tensor algebra

The k-th signature of a path X belongs to $(\mathbb{R}^d)^{\otimes k}$. We now introduce an ambient space for the whole signature $\sigma(X)$.

Definition 1.1. *The* tensor algebra *over $\mathbb{R}^d$ is the graded $\mathbb{R}$-vector space*

$$T((\mathbb{R}^d)) := \mathbb{R} \times \mathbb{R}^d \times (\mathbb{R}^d)^{\otimes 2} \times \dots$$

of formal power series in the non-commuting variables $x_1, \dots, x_d$. It is an $\mathbb{R}$-algebra with respect to the tensor product. The algebraic dual of $T((\mathbb{R}^d))$ is the graded free $\mathbb{R}$-algebra

$$T(\mathbb{R}^d) = \mathbb{R}\langle x_1, \dots, x_d\rangle$$

of polynomials in the non-commuting variables $x_1, \dots, x_d$.

These spaces and their rich algebraic structures are well studied. In this section we recall the features we need, and we refer to [Reu93] for a detailed treatment. Next result, proven in [Che57, Section 2], gives a first taste of the correlation between the tensor algebra and signatures of paths. Recall that the *concatenation* of two paths X and Y is the path $X \sqcup Y\colon [0,1] \longrightarrow \mathbb{R}^d$ given by

$$(X \sqcup Y)(t) = \begin{cases} X(2t) & \text{if } t \in \left[0, \frac{1}{2}\right], \\ X(1) + Y(2t-1) & \text{if } t \in \left[\frac{1}{2}, 1\right]. \end{cases}$$

Lemma 1.2 (Chen's identity). *If $X, Y\colon [0,1] \to \mathbb{R}^d$ are piecewise smooth paths, then $\sigma(X \sqcup Y) = \sigma(X) \otimes \sigma(Y)$ as formal power series in $T((\mathbb{R}^d))$.*

We now introduce a useful notation.

Notation 1.3. *For $T \in T((\mathbb{R}^d))$, denote by $T_{i_1 \dots i_k}$ the $(i_1 \dots i_k)$-th entry of the order k element of T. For $y \in \mathbb{R}$, we set $T_y((\mathbb{R}^d)) := \{T \in T((\mathbb{R}^d)) \mid T_0 = y\}$ to be the space of tensor sequence with constant entry 0. Moreover, we will identify a degree k monomial $x_{i_1} \cdot \dots \cdot x_{i_k}$ with the word $w = i_1 \dots i_k$ in the alphabet $\{1, \dots, d\}$. The number k is called the* length *of w and it is denoted by $\ell(w)$. The degree 0 monomial corresponds to the empty word e. In this way, the tensor product of the two monomials corresponding to the words v and w is the word obtained by writing v followed by w, and it is called the* concatenation product*. The natural duality pairing*

$$\langle -, - \rangle \colon T((\mathbb{R}^d)) \times T(\mathbb{R}^d) \to \mathbb{R}$$

is given by $\langle T, i_1 \dots i_k \rangle = T_{i_1 \dots i_k}$, and extended by linearity.

Besides the concatenation of words, there is another product on $T(\mathbb{R}^d)$. It has a strong combinatoric flavour, but it will also allow us to define very interesting algebraic objects.

Definition 1.4. *We define the* shuffle product *of two words recursively by*

$$\begin{aligned} & e \sqcup\!\sqcup\, w = w \sqcup\!\sqcup\, e = w, \text{ and} \\ & (w_1 \otimes a) \sqcup\!\sqcup\, (w_2 \otimes b) = (w_1 \sqcup\!\sqcup\, (w_2 \otimes b)) \otimes a + ((w_1 \otimes a) \sqcup\!\sqcup\, w_2) \otimes b. \end{aligned}$$

Less formally, $v \sqcup\!\sqcup\, w$ is the sum of all order-preserving interleavings of v and w.

For instance, $1 \sqcup\!\sqcup\, 23 = 123 + 213 + 231$. Despite its apparently complicated definition, the shuffle product enjoys good properties. For instance, $(T(\mathbb{R}^d), \sqcup\!\sqcup, e)$ is a commutative algebra. Moreover next Lemma, proven in [Reu93, Proof of Corollary 3.5], shows that the shuffle product behaves nicely with respect to the signatures.

Lemma 1.5 (Shuffle identity). *If X is a piecewise smooth path, then*

$$\langle \sigma(X), v \rangle \cdot \langle \sigma(X), w \rangle = \langle \sigma(X), v \sqcup\!\sqcup\, w \rangle$$

for all words $v, w \in T(\mathbb{R}^d)$.

It follows that signatures do not fill the tensor space $T((\mathbb{R}^d))$, but rather they live in a subset. In order to make this observation precise, we introduce an important subspace of the tensor algebra.

Definition 1.6. *Consider the Lie bracketing $[T, S] = TS - ST$ on $T((\mathbb{R}^d))$. We define* $\mathrm{Lie}(\mathbb{R}^d) \subset T_0((\mathbb{R}^d))$ *to be the free Lie algebra generated by $x_1, \dots, x_d$, that is, the smallest vector subspace of $T((\mathbb{R}^d))$ that contains $x_1, \dots, x_d$ and is closed with respect to the bracketing.*

The Lie group associated to $\mathrm{Lie}(\mathbb{R}^d)$ is an important object. One way to define it is as the image of $\mathrm{Lie}(\mathbb{R}^d)$ under the exponential map.

Definition 1.7. *Define* $\exp : T_0((\mathbb{R}^d)) \to T_1((\mathbb{R}^d))$ *by the formal power series*

$$\exp(T) := \sum_{n=0}^{\infty} \frac{T^{\otimes n}}{n!}.$$

We denote $\mathcal{G}(\mathbb{R}^d) := \exp(\mathrm{Lie}(\mathbb{R}^d))$.

By construction, $(\mathcal{G}(\mathbb{R}^d), \otimes, e)$ is a group. However, there is another way to characterize it in terms of shuffle product. By [Reu93, Theorem 3.2],

$$\mathcal{G}(\mathbb{R}^d) = \{T \in T_1((\mathbb{R}^d)) \mid \langle T, v\rangle \cdot \langle T, w\rangle = \langle T, v \sqcup\!\sqcup\, w\rangle \text{ for all words } v, w \in T(\mathbb{R}^d)\}.$$

Now we see the first clear connection with the signatures. By the shuffle identity, $\mathcal{G}(\mathbb{R}^d)$ contains the signatures of all piecewise smooth paths.

For our purposes, we need to point out that every definition we recalled has a truncated version. Namely, one can fix $m \in \mathbb{N}$ and consider

$$T^m(\mathbb{R}^d) := \bigoplus_{k=0}^{m} (\mathbb{R}^d)^{\otimes k},$$

where tensors of order greater than m are set to zero. Inside $T^m(\mathbb{R}^d)$ there are $\mathcal{G}^m(\mathbb{R}^d)$ and $\mathrm{Lie}^m(\mathbb{R}^d)$. The restriction of the map exp is defined in the same way, and we will write $\exp^{(m)}$ to denote the map $T_0^m((\mathbb{R}^d)) \to T_1^m((\mathbb{R}^d))$.

Another feature of the group $\mathcal{G}(\mathbb{R}^d)$ is that, for every $m \in \mathbb{N}$, $\mathcal{G}^m(\mathbb{R}^d)$ not only contains, but actually coincides with the set of all truncated signatures of smooth paths (see [AFS18, Theorem 4.4]) and it is a Lie group (see [FV10, Theorem 7.30]). Moreover, it is defined by finitely many polynomials - namely, the shuffle relations with words of length at most m - hence it is an algebraic variety. It is irreducible by [AFS18, Theorem 4.10]. On the other hand, $\mathrm{Lie}^m(\mathbb{R}^d)$ is a finite dimensional vector space. In order to provide it with a basis, it is time to introduce another powerful combinatoric concept.

Definition 1.8. *A non-empty word* w *is* Lyndon *if, whenever we write* $w = pq$ *as the concatenation of two nonempty words, we have* $w < q$ *in the lexicographic order. In this case there is a unique pair* (p, q) *of nonempty words such that* $w = pq$ *and* q *is minimal with respect to lexicographic order. The* bracketing *of* w *is* $[p, q] = pq - qp$.

As shown in [Reu93, Corolllary 4.14] and [AFS18, Proposition 4.7], the bracketings of all Lyndon words of length at most m form a basis for $\mathrm{Lie}^m(\mathbb{R}^d)$. Therefore, in order to compute its dimension - and thus the dimension of the variety $\mathcal{G}^m(\mathbb{R}^d)$ - it is enough to count Lyndon words. First, recall that the Möbius function $\mu : \mathbb{N} \to \mathbb{N}$ sends a natural number t to

$$\mu(t) = \begin{cases} 0 & \text{if } t \text{ is divisible by the square of a prime,} \\ 1 & \text{if } t \text{ is the product of an even number of distinct primes,} \\ -1 & \text{if } t \text{ is the product of an odd number of distinct primes.} \end{cases}$$

Then the number of Lyndon words of length l in the alphabet $\{1, \ldots, d\}$ is $\sum_{t|l} \frac{\mu(t)}{l} d^{\frac{l}{t}}$, and therefore

$$\dim \mathrm{Lie}^m(\mathbb{R}^d) = \sum_{l=1}^{m} \sum_{t|l} \frac{\mu(t)}{l} d^{\frac{l}{t}}.$$

1.2 Signatures from a geometric viewpoint

Chen's theorem assumes the knowledge of the k-th signature for every $k \in \mathbb{N}$. What happens when we know only one of them? In [AFS18], the authors consider the problem from an algebraic geometry perspective. If we fix a certain class of paths and the order k of the tensors, then the k-th signature $\sigma^{(k)}$ is an algebraic map into $(\mathbb{R}^d)^{\otimes k}$. The closure of the image of $\sigma^{(k)}$ is called the *k-signature variety*. We get a first example by considering the class of all smooth paths in $\mathbb{R}^d$. If X is smooth, then the k-th signature of X is just the k-th entry of $\sigma(X) \in \mathcal{G}(\mathbb{R}^d)$. This leads to the following definition.

Definition 1.9. *The* universal variety $\mathcal{U}_{d,k} \subset (\mathbb{R}^d)^{\otimes k}$ *is the projection of* $\mathcal{G}(\mathbb{R}^d)$ *onto the k-th factor. It is the closure of the set of all k-th signatures of smooth paths.*

Since every $\mathcal{G}^m(\mathbb{R}^d)$ is an irreducible variety, $\mathcal{U}_{d,k}$ is irreducible as well. By [AFS18, Theorem 6.1], its dimension coincides with $\dim \mathcal{G}^k(\mathbb{R}^d)$. Equivalently, we can compute it as the dimension of the vector space $\mathrm{Lie}^k(\mathbb{R}^d)$, that is the number of Lyndon words of length at most k in the alphabet $\{1, \ldots, d\}$.

This geometric approach provides a way to translate questions about paths into questions about the signature variety. For instance, the crucial problem of reconstructing uniquely the path from its k-signature translates into the injectivity of $\sigma^{(k)}$. This leads to study the fibers of $\sigma^{(k)}$ and therefore the dimension of the signature variety. Provided that a preimage can be reconstructed, such computation consists in solving a system of polynomial equations, and we can get an idea of how difficult that is by looking at the degree of the variety or at the degrees of the generators of its ideal. Taking a step further, the singular locus can point out that some signatures are different from the others, so we have a meaningful way to distinguish special paths in our class.

There are many interesting questions when it comes to study a signature variety, and we do not hope to give full answers to all of them. Path recovery from the third signature is the main goal of [PSS18]. Further interplay between the tensor algebra and the signatures is explored in [CP18]. Our contribution to this topic is a detailed study of the rough paths signature variety, presented in Section 2, as well as the analysis of the axis parallel signature variety, which we describe in Section 3. Both varieties present several geometric subtleties. When dealing with them, we could take advantage of a surprisingly rich combinatorial structure, that allows us to integrate a projective geometric approach with more computational techniques.

2 The rough Veronese variety

Researchers in stochastic analysis usually work with paths that are not piecewise smooth, and therefore do not have a signature in the sense of Chen's definition. One of the most important examples is the class of rough paths. The interest in rough paths is quickly growing. Applications include the study of controlled ODEs and stochastic PDEs (see [LCL07]) as well as sound compression (see [LS05]), not to mention mathematical physics (see [FGL15]). The main references on rough paths are [FV10] and [FH14]. In this section we recall what the signature of a rough path is, and study the corresponding signature variety.

The signature variety of all rough paths is the whole universal variety. In the same way as polynomial paths are an interesting subclass of smooth paths, in [AFS18, Section 5.4] the authors consider a nice subclass of rough paths, parametrized by $\mathrm{Lie}^m(\mathbb{R}^d)$. Their signature variety $\mathcal{R}_{d,k,m}$ exhibits similarities with the classical Veronese variety and it was therefore named the *rough Veronese variety* - see Definition 2.3. Another reason to study

$\mathcal{R}_{d,k,m}$ is that we can deduce many properties of the universal varieties from it. Indeed, $\mathcal{U}_{d,k} = \mathcal{R}_{d,k,m}$ for every $m \geq k$ and moreover, $\mathcal{U}_{d,k} = \mathcal{R}_{d,k,k}$ is a cone over $\mathcal{R}_{d,k,k-1}$ (see [Gal18, Proposition 24]).

2.1 Preliminaries

One purpose of the definition of a rough path is to generalize the concept of smooth path. We consider then a smooth path X, and without loss of generality we assume $X(0) = 0$. Fix $t \in [0,1]$. In the definition of k-th signature we can replace the integral on $[0,1]$ with an integral on $[0,t]$. This is the same as restricting X to $[0,t]$, hence we will denote such integral by $\sigma^{(k)}(X_{|[0,t]})$. For every k, we notice that $\sigma^{(k)}(X_{|[0,t]})$, as a function of t, is a path $[0,1] \to (\mathbb{R}^d)^{\otimes k}$. If we look at the full signature $\sigma(X_{|[0,t]})$, we get a path $[0,1] \to \mathcal{G}(\mathbb{R}^d)$. Notice that the signature of X is the endpoint of such path. By [Gal18, Lemma 9], this $\mathcal{G}(\mathbb{R}^d)$-valued path satisfies the Hölder-like inequality

$$\left|\sigma^{(k)}(X_{|[s,t]})\right| \lesssim |t-s|^k, \tag{1}$$

where $f(x) \lesssim g(x)$ means that there is a constant c such that $f(x) \leq c \cdot g(x)$ for every x. Summing up, a smooth path $X\colon [0,1] \to \mathbb{R}^d$ induces a path $\sigma(X_{|[0,\cdot]})\colon [0,1] \to \mathcal{G}(\mathbb{R}^d)$ satisfying inequality (1). If we want a rough path to be a generalization of a smooth path, we can use this property, also allowing different exponents. Let $p_k : T((\mathbb{R}^d)) \to (\mathbb{R}^d)^{\otimes k}$ be the projection.

Definition 2.1. *A* rough path *of order m is a path $\mathbf{X}\colon [0,1] \to \mathcal{G}^m(\mathbb{R}^d)$ such that $|p_k(\mathbf{X}(s)^{-1} \otimes \mathbf{X}(t))| \lesssim |t-s|^{\frac{k}{m}}$ for every $k \in \{1,\dots,m\}$ and every $s,t \in [0,1]$. The inverse is taken in the group $\mathcal{G}^m(\mathbb{R}^d)$.*

As we anticipated, we will focus on a special subclass of rough paths of order m, indexed by elements $L \in \mathrm{Lie}^m(\mathbb{R}^d)$.

Definition 2.2. *For $L \in \mathrm{Lie}^m(\mathbb{R}^d)$, consider the path $\mathbf{X}_L\colon [0,1] \to \mathcal{G}^m(\mathbb{R}^d)$ sending t to $\exp^{(m)}(tL)$. By [FV10, Exercise 9.17], this is indeed an order m rough path, and we define its* signature *to be its endpoint $\sigma(\mathbf{X}_L) := \exp(L) \in \mathcal{G}(\mathbb{R}^d)$.*

We want to parametrize the variety containing all $\sigma^{(k)}(\mathbf{X}_L)$ when L ranges in $\mathrm{Lie}^m(\mathbb{R}^d)$, so we are interested in the image of $p_k \circ \exp : \mathrm{Lie}^m(\mathbb{R}^d) \to (\mathbb{R}^d)^{\otimes k}$. While exp is not an algebraic map, $p_k \circ \exp$ is. In general, the image of an algebraic map is only a semialgebraic subset of the real affine space $(\mathbb{R}^d)^{\otimes k}$. However, it is simpler to work with complex projective varieties, so we will follow a common approach in applied algebraic geometry and consider the Zariski closure of the image, take the complexification and pass to the projectivization.

Definition 2.3. *The* rough Veronese variety *$\mathcal{R}_{d,k,m}$ is the closure of the image of the composition*

$$f_{d,k,m}\colon \mathrm{Lie}^m(\mathbb{R}^d) \xrightarrow{\exp} \mathcal{G}(\mathbb{R}^d) \xrightarrow{p_k} (\mathbb{R}^d)^{\otimes k} \to (\mathbb{R}^d)^{\otimes k} \otimes \mathbb{C} = (\mathbb{C}^d)^{\otimes k} \dashrightarrow \mathbb{P}^{d^k-1}.$$

In general it is not easy to work out all the invariants of a given variety. Luckily, we will shortly see that the rough Veronese variety is toric. In other words it is the closure of the image of a monomial map. There are a lot of tools and techniques that make toric varieties accessible and easy to work with. Classical references on toric varieties are [CLS11, Stu96]. First, we fix some notation.

Notation 2.4. *For every $d \geq 2$, let $W_{d,m}$ be the set of Lyndon words of length at most m in the alphabet with d letters. Let $S_{d,m} := \mathbb{C}[x_w \mid w \in W_{d,m}]$ be a polynomial ring with as many variables as Lyndon words. On this algebra we define a grading by setting the weight of x_w to be the length of w.*

Let A be the set of all monomials in $\mathbb{C}[x_w \mid w \in W_{d,m}]$ of weighted degree k. Define another polynomial ring $R_{d,k} := \mathbb{C}[y_\alpha \mid \alpha \in A]$ with as many variables as elements of A, and the usual polynomial grading.

Our starting point will be the following result, proven in [Gal18, Proposition 18]. Not only we know that $\mathcal{R}_{d,k,m}$ is toric, but we also know what are the monomials parametrizing it.

Theorem 2.5. *Up to projectivization, $\mathcal{R}_{d,k,m}$ is isomorphic to the closure of the image of the map $\operatorname{Spec} S_{d,m} \to \operatorname{Spec} R_{d,k}$ given by all monomials of weighted degree k. That is, the kernel of the map $R_{d,k} \to S_{d,m}$ is the homogeneous prime ideal defining $\mathcal{R}_{d,k,m}$.*

More precisely, it is the image of a weighted projective space by the map given by all sections of $\mathcal{O}(k)$, i.e. all monomials of (weighted) degree k. Such a map does not have to be an embedding. In fact, $\mathcal{O}(k)$ does not need to be a Cartier divisor.

Remark 2.6. *The classical k-Veronese variety is the image of the map given by all degree k monomials in the usual grading. Theorem 2.5 shows that $\mathcal{R}_{d,k,m}$ can be seen as a weighted version of the Veronese variety, thereby justifying its name.*

Theorem 2.5 allows us to investigate $\mathcal{R}_{d,k,m}$ with the tools of toric geometry. In some of our combinatoric arguments, we will find it convenient to use the following notation.

Notation 2.7. *For any integers a, b we denote by (a^b) a sequence or multiset of b elements equal to a. For example $(1^3, 2^2) = (1, 1, 1, 2, 2)$.*

We start by describing the ideal of this variety, i.e. the polynomials that vanish on it.

2.2 Polynomials defining the rough Veronese variety

The classical Veronese variety is defined by quadrics, so it is natural to ask whether $\mathcal{R}_{d,k,m}$ enjoys the same property. Despite holding in many examples (see [AFS18, Section 4.3] and the computations in [Gal18]), the conjecture that $\mathcal{R}_{d,k,m}$ is always defined by quadrics is false. In [Gal18, Proposition 26] we see that the ideal $\mathcal{R}_{2,20,14}$ has a cubic generator. One could still hope to bound the degree of the generators. Surprisingly no such bound exists, as we will shortly prove. First we need the following lemma from linear algebra.

Lemma 2.8. *For every $n \geq 3$, there exist $k_n \in \mathbb{N}$ and an $n \times n$ matrix M_n with positive integer entries satisfying the following properties:*

1. *the sum of the entries of each row and each column of M_n equals k_n,*
2. *the only way to decompose the multiset of entries of M_n into n-tuples summing to k_n is by rows or by columns,*
3. *each column, as a multiset, is distinct from each row.*

We refer to the matrix M_n as a rigid k_n-square matrix.

Proof. We proceed by induction on n. If $n = 3$, we take $k_3 = 12$ and

$$M_3 = \begin{pmatrix} 4 & 5 & 3 \\ 6 & 5 & 1 \\ 2 & 2 & 8 \end{pmatrix}.$$

Assume the statement holds for n and let us prove it for $n+1$. By induction hypothesis, there exist $k_n \in \mathbb{N}$ and an $n \times n$ matrix M_n with the required properties. Let $\lambda > n^2 + 1$ be a natural number and define $k_{n+1} = \lambda k_n + n + 1$. Moreover, define the $(n+1) \times (n+1)$ matrices

$$A = \left(\begin{array}{cccc|c} & & & & 0 \\ & \lambda M_n & & & 0 \\ & & & & \vdots \\ & & & & 0 \\ \hline 0 & \dots & 0 & 0 & \lambda k_n \end{array}\right) \text{ and } B = \left(\begin{array}{cccc|c} 0 & 0 & \dots & 0 & n+1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 0 & n+1 \\ 1 & 1 & \dots & 1 & 1 \\ \hline n & n & \dots & n & -n^2+n+1 \end{array}\right).$$

Then we take $M_{n+1} := A + B$. Since M_n has positive entries, the same holds for M_{n+1}. Moreover, the sum of the entries of each row and each column of M_{n+1} equals k_{n+1}. Fix a partition of the entries of M_{n+1} into $n+1$ multisets $\{r_{i,1}, \dots, r_{i,n+1}\}$, for $1 \le i \le n+1$, each summing up to k_{n+1}. Without loss of generality, we assume that $r_{1,1} = M_{n+1,n+1} = \lambda k_n - n^2 + n + 1$. None of the other $r_{1,j}$'s can be larger than or equal to λ, otherwise their sum would exceed $M_{n+1,n+1} + \lambda > \lambda k_n + n + 1 = k_{n+1}$, which is a contradiction. Therefore the only possible choices for the other $r_{1,j}$'s are either from the $(n+1)$-st row or the $(n+1)$-st column of M_{n+1}. If $r_{1,j_0} = 1$ for some j_0, then $r_{1,j} = n+1$ for $j \in \{2, \dots, \hat{j_0}, \dots, n\}$, hence $\{r_{1,j}\}_j$ must be the last column of M_{n+1}. Otherwise, we have $r_{1,j} \ge n$ for all $2 \le j \le n+1$, and as $\sum_{j=1}^{n+1} r_{1,j} = k_{n+1}$ we must have $r_{1,j} = n$ for $2 \le j \le n+1$. In this case, $\{r_{1,j}\}_j$ must be the last row of M_{n+1}. We now consider the other multisets $r_{i,j}$, for $i > 1$. By the previous argument, they induce a partition of either the first n columns or rows of M_{n+1}. We first consider the induced partition on A, i.e. the partition of λM_n. We claim that, restricted to that matrix, each part must sum up to λk_n. Otherwise, one part would need to have a smaller sum, and by divisibility by λ, it would be at most $\lambda k_n - \lambda$. However, none of the $n+1$ entries in the first n columns or first n rows in B sums up to $\lambda + n + 1$. This would contradict the fact that the multiset sums up to k_{n+1}.

We notice that at this point we are not allowed to use the inductive assumption on M_n. Indeed, although we know that the induced partition of λM_n has correct sum, we do not know that each induced multiset has exactly n elements. However, we know that the multisets induced on B must sum up exactly to $n+1$. This implies that each of the first n elements of the last column (if $\{r_{1,j}\}_j$ was the last row) or last row (if $\{r_{1,j}\}_j$ was the last column) must belong to a different multiset $\{r_{i,j}\}_j$. Hence, there is precisely one such element in each such multiset. It follows that in each $\{r_{i,j}\}_j$ there are precisely n elements from the upper left $n \times n$ submatrix of M_{n+1}. Hence, by induction, the induced partition of λM_n is the same as a partition by columns or by rows. Suppose it is a partition by columns. It is only compatible with a partition of M_{n+1} by extending it by n. Further, looking at B, each partition must contain exactly one n and one 1. Thus, the considered partition is the same as the one given by columns. The argument for a row partition is similar. ☐

Thanks to Lemma 2.8, we are now able to produce instances of $\mathcal{R}_{d,k,m}$ generated in arbitrarily high degree.

Proposition 2.9. *For every $n \ge 3$ and every $d \ge 2$, there exist $k, m \in \mathbb{N}$ such that $k \ge m$ and the ideal of $\mathcal{R}_{d,k,m}$ is not generated in degree n.*

Proof. By Lemma 2.8, there exists $k \in \mathbb{N}$ and a rigid k-square matrix M of size $(n+1) \times (n+1)$. Let m be the largest entry of M. By Theorem 2.5, the ideal I of $\mathcal{R}_{d,k,m}$ is the kernel of the ring homomorphism

$$\varphi : \mathbb{C}[y_\alpha \mid \alpha \in A] \to \mathbb{C}[x_w \mid w \in W_{d,m}]$$

sending y_α to the weighted monomial α.

For every $i \in \{1, \ldots, m\}$, fix once and for all a variable of weight i. This means that we fix a length i Lyndon word w and we associate to i the variable x_w. For every $j \in \{1, \ldots, n+1\}$, consider the monomial α_j defined by the product of $n+1$ variables x_w whose weights are the entries of the j-th row of M. By construction, this monomial has weighted degree k and therefore $\alpha_1, \ldots, \alpha_{n+1} \in A$. In a similar way, we can consider the monomials $\beta_1, \ldots, \beta_{n+1}$ defined by the columns of M. We define

$$f := y_{\alpha_1} \cdot \ldots \cdot y_{\alpha_{n+1}} - y_{\beta_1} \cdot \ldots \cdot y_{\beta_{n+1}}.$$

This is a degree $n+1$ element of $\mathbb{C}[y_\alpha \mid \alpha \in A]$. Moreover, since the degrees appearing in the rows of M are the same as the degrees appearing in the columns, $f \in \ker(\varphi) = I$. Let us show that f is not generated by degree n elements. Let $q_1, \ldots, q_r$ be the polynomials spanning the degree n part of I. Since $\mathcal{R}_{d,k,m}$ is toric, we can assume these are binomials [CLS11, Proposition 1.1.9]. Hence, $q_i = g_i - h_i$, where each g_i and each h_i is a degree n monomial. Suppose by contradiction that f can be generated by $q_1, \ldots, q_r$. Then there exist variables $l_1, \ldots, l_r$ such that f is the sum

$$f = l_1(g_1 - h_1) + \cdots + l_r(g_r - h_r).$$

Without loss of generality, assume that $y_{\alpha_1} \cdot \ldots \cdot y_{\alpha_{n+1}} = l_1 g_1$. Since $l_1 g_1 - l_1 h_1 \in I$, we know that $\varphi(l_1 h_1) = \varphi(l_1 g_1)$. As a multiset, the variables in $\varphi(l_1 h_1)$ are the entries of the matrix M. Furthermore, the variables of $l_1 h_1$ provide a partition into $n+1$ sets of cardinality $n+1$, each with sum k. Since $h_1 \neq g_1$, by assumption on M, this partition must correspond to the column partition of M. Hence, $l_1(h_1 - g_1) = f$. This is a contradiction, because no variable divides f by construction. □

In applications, sometimes one has a signature coming from experiments and wants to know whether it belongs to a path of a certain class. In other words, we have a point in $\mathcal{U}_{d,k}$ and we want to understand if it belongs to a given signature subvariety. From this viewpoint Proposition 2.9 may seem disappointing. In order to check whether a given point belongs to $\mathcal{R}_{d,k,m}$, we might have to evaluate polynomials of very high degree. The good news is that in most cases it is enough to only evaluate quadrics. More precisely, $\mathcal{R}_{d,k,m}$ is always defined by quadrics outside of a coordinate linear subspace of large codimension. In order to prove that, we recall a classical lemma from toric geometry.

Lemma 2.10. *Let $s \in \mathbb{N}$ and let I be the homogeneous toric ideal associated to a set of lattice points $A \subset \mathbb{Z}^n$. Let g be a binomial which we write as $p_1 + \ldots + p_t - (p'_1 + \ldots + p'_t)$, where $p_i, p'_j \in A$. Then $g \in I(s)$ if and only if starting from the multiset $\{p_i\}$ we can reach the multiset $\{p'_j\}$ in finitely many steps of the following form. In each step we replace $p_{i_1}, \ldots, p_{i_s}$ with some other $r_{i_1}, \ldots, r_{i_s} \in A$ such that $p_{i_1} + \ldots + p_{i_s} = r_{i_1} + \ldots + r_{i_s}$.*

Proof. A toric ideal is known to be binomial. Thus, $f \in I(s)$ if and only if:

$$f = \sum_{i=1}^{k} m_i b_i,$$

where m_i are monomials and $b_i \in I$ are binomials of degree at most s. The proof is by induction on k. The monomial of f corresponding to $p_1+\cdots+p_t$ must appear on the right hand side, say in $m_1 b_1$. Then $f - m_1 b_1$ is obtained by applying one step of the procedure described in the lemma. We conclude by induction. □

Proposition 2.11. *Fix $R = R_{d,k}$ as in Notation 2.4. Let $I \subset R$ be the ideal of $\mathcal{R}_{d,k,m}$ and let $I(2) \subset I$ be the vector subspace of degree two forms. Let $V \subset \mathbb{P}^{d^k-1}$ be the variety defined by $I(2)$. Then $\mathcal{R}_{d,k,m}$ is an irreducible component of V.*

Proof. Since $I(2) \subset I$, it is clear that $\mathcal{R}_{d,k,m} \subset V$. We want to find a polynomial $f \in R \setminus I$ such that $(I(2) : f^\infty) = I$. This implies that I and $I(2)$ coincide in the localization R_f and therefore V and $\mathcal{R}_{d,k,m}$ coincide in the affine open subset $(f \neq 0)$. In particular, $\mathcal{R}_{d,k,m}$ is an irreducible component of V.

Let $\mu = \dim \operatorname{Lie}^m(\mathbb{R}^d)$ and let $A \subset \mathbb{N}^\mu$ be the set of lattice points associated to the monomials defining the toric variety $\mathcal{R}_{d,k,m}$. By Theorem 2.5, points of A are of the form

$$(w_{1,1}, \ldots, w_{1,a_1}, w_{2,1}, \ldots, w_{2,a_2}, \ldots, w_{m,1}, \ldots, w_{m,a_m}) \text{ where } \sum_{i=1}^{m} \sum_{j=1}^{a_i} i w_{i,j} = k.$$

By [CLS11, Proposition 1.1.9], there exists a set of binomial generators of I. Each of such binomials corresponds to an integral relation $p_1 + \ldots + p_t = p'_1 + \ldots + p'_t$, where all p_i and p'_i are in A (the sum of points corresponds to the product of variables). Our task is to find $f \notin I$ such that $f^n g \in I(2)$ for every generator $g = p_1 + \ldots + p_t - p'_1 - \ldots - p'_t$ of I for some n. We will define f to be a variable. In toric words, f corresponds to a lattice point $p \in A$. By Lemma 2.10, we want to find $p \in A$ such that $p + \ldots + p + p_1 + \ldots + p_t$ can be turned in $p + \ldots + p + p'_1 + \ldots + p'_t$ by repeatedly replacing *a pair* of summands with another pair having the same sum. Let

$$f = p := (k, 0, \ldots, 0)$$

and $p_1 = (w_{1,1}, \ldots, w_{1,a_1}, w_{2,1}, \ldots, w_{2,a_2}, \ldots, w_{m,1}, \ldots, w_{m,a_m})$. Assuming $w_{2,1} > 0$, we can replace $p + p_1$ by

$$(k-2, 0, \ldots, 0, 1, 0, \ldots, 0) + (w_{1,1}+2, w_{1,2}, \ldots, w_{1,a_1}, w_{2,1}-1, w_{2,2} \ldots, w_{2,a_2}, \ldots, w_{m,a_m}),$$

i.e. we replace two on the first coordinate with one on the coordinate corresponding to $w_{2,1}$. By iterating this process, we add more copies of p so that we can replace $p + \ldots + p + p_1$ by points for which *at most one coordinate* $w_{i,j}$ is nonzero for $(i,j) \neq (1,1)$, and that coordinate is equal to one. We do the same to $p_2, \ldots, p_t$, so that we can replace $p + \ldots + p + p_1 + \ldots + p_t$ by a sum of points for which at most one coordinate $w_{i,j}$ is nonzero for $(i,j) \neq (1,1)$, and that coordinate is equal to one. We apply the same process to $p'_1 + \ldots + p'_t$. Since now both sides are broken into this kind of simple pieces, and since the condition of having weighted sum k is always preserved, the only possibility is that the summands are pairwise the same. □

Remark 2.12. *In the proof of Proposition 2.11 we could have chosen the point p to be any point with coordinates $w_{i,j} = 0$ for $i > 1$. This proves that $I(2)$ and I define the same scheme (in particular, the quadrics define the correct set) outside of the locus where* all *these coordinates vanish. Therefore the components of the variety defined $I(2)$ (embedded or not) must be supported on a coordinate subspace of large codimension.*

2.3 Normality of the rough Veronese Variety

A classical approach to study geometry of a projective toric variety is to look at the associated lattice polytope [CLS11, Stu96, Ful93]. In our case the central object is presented in the following definition.

Definition 2.13. *For $k \in \mathbb{N}$ and $w = (w_1, \ldots, w_r) \in \mathbb{N}^r$, we define the lattice polytope $P(w, k)$ as the convex hull of*

$$\left\{ (t_1, \ldots, t_r) \in \mathbb{N}^r \;\middle|\; \sum_{i=1}^{r} w_i t_i = k \right\}.$$

In this notation, the polytope associated to $\mathcal{R}_{d,k,m}$ is $P((1^{a_1}, 2^{a_2}, \ldots, m^{a_m}), k)$, where a_i is the number of Lyndon words of length i in the alphabet $\{1, \ldots, d\}$. Many authors, like [CLS11, Ful93], require a toric variety to be *normal.* This is equivalent to the fact that the toric variety may be represented by a fan. Hence, the first important task is to decide when $\mathcal{R}_{d,k,m}$ is normal. One advantage is that normality can be checked on the polytope.

Lemma 2.14. *A polytope P is normal if and only if kP is normal for every $k \in \mathbb{N}$. Moreover, if X_P is the associated toric variety, then*

1. *X_P is normal if and only if P is very ample,*
2. *X_P is projectively normal (i.e. the affine cone over it is normal) if and only if P is normal.*

Proof. This results are proven in [CLS11, Chapter 2], together with many other features of normal polytopes. □

Lemma 2.14 will help us to determine when $\mathcal{R}_{d,k,m}$ is normal. We start with the following result, that simplifies the problem.

Lemma 2.15. *$P((w_1^{a_1}, w_2^{a_2}, \ldots, w_m^{a_m}), k)$ is normal if and only if $P((w_1, w_2, \ldots, w_m), k)$ is normal. In particular, the normality of $\mathcal{R}_{d,k,m}$ does not depend on d.*

Proof. By induction, it is enough to check what happens when we add or discard one of the entries. In order to simplify notation, set

$$P := P((w_1, \ldots, w_{i-1}, w_i, w_i', w_{i+1}, \ldots, w_m), k),$$

where $w_i = w_i'$ and we do not assume that the w_j's are distinct. Observe that one of the facets of P is precisely the polytope

$$Q := P((w_1, \ldots, w_i, w_{i+1}, \ldots, w_m), k).$$

Since every face of a normal polytope is normal, the first implication follows.

Assume now that Q is normal. Consider the linear surjection $\sim: P \to Q$, defined by the sum of the two entries corresponding to w_i and w_i'. If $t = (t_1, \ldots, t_{i-1}, t_i, t_i', t_{i+1}, \ldots, t_m) \in P$, then $\tilde{t} = (t_1, \ldots, t_{i-1}, t_i + t_i', t_{i+1}, \ldots, t_m) \in Q$. Note that $\sim$ is also surjective on lattice points, that is $\sim: P \cap \mathbb{N}^r \to Q \cap \mathbb{N}^{r-1}$ is surjective. Let p be a lattice point in some multiple sP of P. We can write $p = \lambda_1 p_1 + \ldots + \lambda_s p_s$, where $p_1, \ldots, p_s \in P$ are lattice points and $\lambda_1 + \ldots + \lambda_s = s$. Since $\tilde{p} \in sQ$, Q is normal by hypothesis and $\sim$ is surjective on lattice points, there are $x_1, \ldots, x_s \in P \cap \mathbb{N}^r$ such that $\tilde{p} = \lambda_1 \tilde{p}_1 + \ldots + \lambda_s \tilde{p}_s = \tilde{x}_1 + \ldots + \tilde{x}_s$. Let $x_i \in P$ be the preimage of $\tilde{x}_i$ having 0 in the entry corresponding to w_i', and let $x = x_1 + \ldots + x_s$. Since $\tilde{p} = \tilde{x}$, p and x coincide on every coordinate except the ones corresponding to w_i and w_i'. Moreover, these entries have the same sum, and such sum is an integer by construction of x. Then it is enough to increase the zero entry of some suitable x_i, keeping such sum untouched, to obtain from $x_1, \dots, x_s$ another set of s lattice points of P whose sum is p. Therefore P is normal. □

In general, the rough Veronese variety does not need to be normal. As we show using the code available online [CGM], $P((1, 2, \dots, 9), 18)$ is not very ample, so $\mathcal{R}_{d,18,9}$ is not normal for any $d \geq 2$.

In algebraic geometry, given a map $f = (f_1, \dots, f_s)$ defined by homogeneous polynomials of the same degree, it is natural to study the induced *rational map* $\tilde{f}$ between projective spaces. The map $\tilde{f}$ may be not defined everywhere - the locus where all of the polynomials f_i vanish is called the *base locus* or *indeterminacy locus* of $\tilde{f}$. In our setting, the monomials defining $\mathcal{R}_{d,k,m}$ are of degree k, however in variables that are of different degrees. This implies that instead of considering the classical projective space, we need the *weighted projective space*. Theorem 2.5 tells us that $\mathcal{R}_{d,k,m}$ is the closure of the image of a rational map

$$\mathbb{P}(1^{a_1}, \dots, m^{a_m}) \dashrightarrow \mathbb{P}^{d^k-1}.$$

The codomain is the usual projective space and the domain is the weighted projective space with $\sum_i a_i$ variables of weights as given in the brackets. The map is given by all monomials of degree k.

Before we proceed further, let us recall that maps from an algebraic variety X to a projective space are studied through line bundles (or equivalently Cartier divisors) on X [Laz04]. The theory of line bundles on weighted projective spaces is very well understood. The Picard group equals $\mathbb{Z}$ where $\mathbf{1}$ is an ample generator. The global sections of this generator may be identified with all monomials of degree l equal to the least common multiple of all the degrees appearing in the weighted projective space. Thus, it is customary to denote it by $\mathcal{O}(l)$. In particular, the elements of the Picard group will be denoted by $\mathcal{O}(s)$, for s divisible by l. Although the class group is abstractly also equal to $\mathbb{Z}$, it is larger. The inclusion of the Picard group in the class group $\mathbb{Z} \hookrightarrow \mathbb{Z}$ is given by multiplication by l. In particular, the elements of the class group will be denoted by $\mathcal{O}(s)$, for arbitrary s.

Lemma 2.16. *Let $\mathbb{P}(w_1, \dots, w_s)$ be any weighted projective space and let $l = \operatorname{lcm}(w_1, \dots, w_s)$. The map from $\mathbb{P}(w_1, \dots, w_s)$ to a projective space given by all monomials of degree k does not have a base locus if and only if $l|k$.*

Proof. If $w_1 \nmid l$, then $[1 : 0 : \dots : 0]$ is a base point. If all $w_i | l$ then the monomials $x_i^{\frac{l}{w_i}}$ do not have a base locus. □

As there is some confusion in the literature, we stress that even if $l|k$ the map from the previous lemma does not have to be an embedding. In other words $\mathcal{O}(l)$, although always ample, may be not very ample.

Example 2.17. *The polytope $P((1, 6, 10, 15), 30)$ is not very ample. Equivalently, the map $\mathbb{P}(1, 6, 10, 15) \to \mathbb{P}^{17}$ given by all monomials of weighted degree 30 is not an embedding. This can be checked with the software* Normaliz *[BIR+], or by studying the polytope.*

This raises the following question: when is $\mathcal{R}_{d,k,m}$ an embedding of the weighted projective space $\mathbb{P}(1^{a_1}, \dots, m^{a_m})$? We will prove that this happens if and only if k is a multiple of all natural numbers between 1 and m. Equivalently, $\mathcal{O}(k)$ is a very ample line bundle if and only if $\operatorname{lcm}(2, \dots, m) \mid k$. We will need the following technical lemma.

Lemma 2.18. *If $m \geq 7$ and $m \neq 10$, then there exist two distinct prime numbers strictly larger than $\frac{m}{2}$ and at most equal to m such that their sum is not a power of two.*

Proof. If $m \leq 56$, we can easily find the required primes.

$32 \leq m \leq 56$	29 and 31
$20 \leq m \leq 31$	17 and 19
$14 \leq m \leq 19$	11 and 13
$11 \leq m \leq 13$	7 and 11
$m \in \{7, 8, 9\}$	5 and 7

Assume then $m \geq 57$. For $x \in \mathbb{N}$, let $\pi(x)$ be the number of prime numbers smaller or equal than x. In [RS62], the authors show that

$$\frac{x}{\log x} < \pi(x) < 1.3 \frac{x}{\log x}$$

for every $x \geq 17$. Now,

$$\pi(m) - \pi\left(\frac{m}{2}\right) \geq \frac{m}{\log m} - 1.3 \frac{m}{2 \log\left(\frac{m}{2}\right)},$$

and the latter is at least 3 for $m \geq 57$, so there are at least three prime numbers $a < b < c$ between $\frac{m}{2}$ and m. If $b+c$ is not a power of 2, we are done. Otherwise we have $b + c = 2^n$. Then

$$2^{n-1} = \frac{b+c}{2} \leq m < a + b < b + c = 2^n,$$

so $a + b$ cannot be a power of 2. □

Theorem 2.19. *If $i \mid k$ for every $i \in \{1, \ldots, m\}$, then $\mathcal{R}_{d,k,m}$ is projectively normal.*

Proof. For $m \leq 6$, normality can be explicitly checked with the software *Polymake* [AGH+17], so we assume $m \geq 7$. By Lemmas 2.14 and 2.15, we may assume that $k = \operatorname{lcm}(1, \ldots, m)$ and study the map

$$\mathbb{P}(1, 2, \ldots, m) \to \mathbb{P}^N$$

defined by all monomials of weighted degree k. Let $P := P((1, 2, \ldots, m), k)$ be the associated polytope. For $s \in \mathbb{N}$, consider the dilation sP of P. A lattice point of sP is of the form $(a_1, \ldots, a_m) \in \mathbb{N}^m$ with $a_1 + 2a_2 + \ldots + ma_m = sk$. We can write it as a multiset

$$M = \{1^{a_1}, \ldots, m^{a_m}\}.$$

By induction on s, we only need to prove that there exists a submultiset S of M whose entries sum up to k. Let us modify M to M' in the following way. If there are $i, j \in M$ such that $i, j \leq \frac{m}{2}$, then discard i and j and add $i + j$. Notice that $i + j \leq m$. If S' is a submultiset of M' whose entries sum up to k, then either S does not contain $i + j$, and therefore $S = S'$ is a submultiset of M as well, or S' contains $i + j$, and we define S by replacing back $i + j$ with i, j. By iterating this argument, we can assume that there is at most one $i \in \{1, \ldots, m\}$ such that $i \leq \frac{m}{2}$.

The resulting multiset M' is of the form $\left\{i_1^{b_{i_1}}, \ldots, i_t^{b_{i_t}}\right\}$. Let $k' = \operatorname{lcm}(i_1, \ldots, i_t)$ and consider the map

$$\mathbb{P}(i_1, i_2, \ldots, i_t) \to \mathbb{P}^T$$

given by all the monomials of weighted degree k'. Its image is associated to the polytope $P' := P((i_1, i_2, \ldots, i_t), k')$. By construction, k' divides k and $\dim P' \leq t - 1 \leq \frac{m}{2}$. Observe that any point of P' has entries summing up to k'. Since the entries of M' sum up to sk, M' is a point in rP' for $r = \frac{sk}{k'}$. In fact, $M' \in s\left(\frac{k}{k'}P'\right)$ and so, it is enough to prove that

$\frac{k}{k'}P'$ is normal. This will imply that M' is a sum of lattice points of $\frac{k}{k'}P'$, and therefore it admits the required submultiset S'.

If there exists a prime number $\frac{m}{2} \leq p \leq m$ such that $p \nmid k'$, then $p \mid \frac{k}{k'}$ and so $\frac{k}{k'} \geq p \geq \frac{m}{2} \geq \dim P'$. In this case, $\frac{k}{k'}P'$ is normal by [CLS11, Theorem 2.2.12]. This guarantees the existence of the desired S'. Otherwise, assume $m \neq 10$. By Lemma 2.18, there exist two prime numbers p_1, p_2 between $\frac{m}{2}$ and m such that $p_1 + p_2$ is not a power of 2. Up to relabelling, we can assume that $b_{p_1} \leq b_{p_2}$.

Once more, we have to modify M' by making some replacements. For b_{p_1} times, discard an entry p_1 and an entry p_2 and add an entry p_1+p_2. Denote by M'' the resulting multiset and by P'' the associated polytope, that satisfies $\dim P'' \leq \frac{m}{2}$. As before, it is enough to find a submultiset S'' of M'' such that the entries of S'' sum up to k. By construction, the entries of M'' still sum up to sk and their least common multiple k'' divides k. Indeed, the sum $p_1 + p_2$ is even and not a power of two, thus all powers of prime numbers that divide it are smaller or equal to m. Notice that neither p_1, nor any of its multiples appear in M''. Therefore, $p_1 \mid k$ and $p_1 \nmid k''$, which implies that $p_1 \mid \frac{k}{k''}$ and so $\frac{k}{k''} \geq p_1 \geq \frac{m}{2} \geq \dim P''$. By [CLS11, Theorem 2.2.12], $\frac{k}{k''}P''$ is normal. This guarantees the existence of the desired S''.

The last case $m = 10$ can be explicitly checked with the software Normaliz. □

Corollary 2.20. *The map $\mathbb{P}(1^{a_1}, \ldots, m^{a_m}) \to \mathbb{P}^N$ given by all monomials of weighted degree k is an embedding if and only if $i \mid k$ for every $i \in \{1, \ldots, m\}$. If so, $\mathcal{R}_{d,k,m} \simeq \mathbb{P}(1^{a_1}, \ldots, m^{a_m})$ is embedded as a projectively normal variety.*

Proof. By Theroem 2.19, the polytope associated to $\mathcal{R}_{d,k,m}$ is normal and thus very ample. This implies that the map is an embedding and the image is projective normal. □

2.4 Dimension and Degree of the rough Veronese variety

In this section we provide formulas for the dimension and degree of $\mathcal{R}_{d,k,m}$. We work under the assumption that $k \geq m$, as otherwise some variables do not appear at all in the parametrization and hence we can easily reduce to this case. Recall that the *normalized volume* of a polytope $P \subset \mathbb{R}^n$, denoted by $\operatorname{vol} P$, is $n!$ times its Lebesgue measure.

Proposition 2.21. *Let $d, k, m \in \mathbb{N}$ with $d \geq 2$ and $k \geq m$. Let $W_{d,m}$ be the set of Lyndon words of length at most m in the alphabet with d letters.*

1. *The dimension of $\mathcal{R}_{d,k,m}$ is $\#(W_{d,m}) - 1$.*
2. *Let $l = \operatorname{lcm}(1, 2, \ldots, m)$ and let Δ be the convex hull of the integral points*

$$\left(\frac{l}{1}, 0, \ldots, 0\right), \left(0, \frac{l}{1}, 0, \ldots, 0\right), \ldots, \left(0, \ldots, 0, \frac{l}{m}\right),$$

where the number of occurrences of $\frac{l}{i}$ is the number of Lyndon words of length i. Then

$$\deg \mathcal{R}_{d,k,m} \leq \operatorname{vol}\left(\frac{k}{l}\Delta\right) \tag{2}$$

and

$$\lim_{k\to\infty} \frac{\deg \mathcal{R}_{d,k,m}}{\operatorname{vol}\left(\frac{k}{l}\Delta\right)} = 1.$$

Moreover, equality holds in (2) if and only if $l|k$.

Proof. Let a_i be the number of Lyndon words of length i. We know that $\mathcal{R}_{d,k,m}$ is the toric variety associated to the polytope

$$P := P((1^{a_1}, \dots, m^{a_m}), k).$$

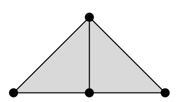
1. The dimension of $\mathcal{R}_{d,k,m}$ was already computed in [AFS18, Remark 6.5] by proving that the fibers of the map $f_{d,k,m}$ have dimension 0. However, here we show a different proof based on toric techniques. Note that $\dim \mathcal{R}_{d,k,m} \leq \dim \mathrm{Lie}^m(\mathbb{R}^d) - 1 = \#(W_{d,m}) - 1$, as this is the dimension of the parameterizing weighted projective space. Thus, in order to prove the statement we only have to show that P has maximal possible dimension. In fact, we prove that the lattice points of P generate the lattice
$$L = \left\{ (x_{1,1}, \dots, x_{1,a_1}, x_{2,1}, \dots, x_{m,a_m}) \in \mathbb{Z}^{\# W_{d,m}} \mid k \text{ divides } \sum_{i=1}^{m} \sum_{j=1}^{a_i} i x_{i,j} \right\}.$$
If $i \in \{1, \dots, m\}$, $j \in \{1, \dots, a_i\}$ and $(i, j) \neq (1, 1)$, then there exists a lattice point $p \in P$ with $x_{i,j} = 1$ and $x_{i',j'} = 0$, unless $(i', j') = (i, j)$ or $(i', j') = (1, 1)$. Take any $z \in L$. By adding and subtracting p of the above type, we can reduce to the situation when z has only the first coordinate nonzero. In such a case, the claim is obvious.

2. We note that P is the convex hull of lattice points in $\frac{k}{l}\Delta$. Thus,
$$\deg \mathcal{R}_{d,k,m} = \mathrm{vol}\, P \leq \mathrm{vol}\, \frac{k}{l}\Delta.$$
Now, equality holds if and only if $P = \frac{k}{l}\Delta$. However, the latter is a lattice polytope if and only if $l|k$. Finally, note that $P((1^{a_1}, \dots, m^{a_m}), k)$ is injected into $P((1^{a_1}, \dots, m^{a_m}), k + 1)$ simply by adding $(1, 0, \dots, 0)$. Hence
$$\mathrm{vol} \left\lfloor \frac{k}{l} \right\rfloor \Delta \leq \mathrm{vol}\, P = \deg \mathcal{R}_{d,k,m} \leq \mathrm{vol} \left\lceil \frac{k}{l} \right\rceil \Delta,$$
which finishes the proof. □

Many software-aided computations on the dimension, the degree and even the generators of the ideal of $\mathcal{R}_{d,k,m}$ are presented in [Gal18]. We now want to use Proposition 2.21 to deal with this numbers by using toric geometry.

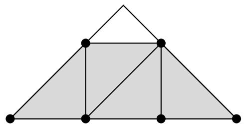
Example 2.22. *Fix $d = m = 2$. Since $W_{2,2} = \{1, 2, 12\}$, $\dim \mathcal{R}_{2,k,2} = 2$. We now compute the degree for small values of k. Since $\mathcal{R}_{2,k,2}$ is a surface, we will deal with 2-dimensional polytopes. Here, $l = \mathrm{lcm}(1, 2) = 2$ and we denote $P := P((1^2, 2), k)$.*

($k = 2$) *As in Notation 2.4, $R_{2,2} = \mathbb{C}[x^2, xy, y^2, a]$ and $P = \Delta$ is the convex hull of the points $(2, 0, 0), (0, 2, 0), (0, 0, 1)$. It is a triangle*



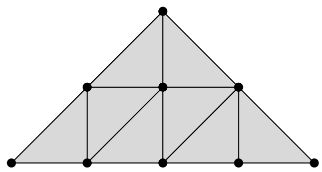
of normalized area 2, hence $\deg \mathcal{R}_{2,2,2} = 2$. As [Gal18, Example 14] shows, $\mathcal{R}_{2,2,2}$ is the cone over a smooth conic in $\mathbb{P}^3$.

($k = 3$) *Since* $R_{2,3} = \mathbb{C}[x^3, x^2y, xy^2, y^3, xa, ya]$, $\frac{3}{2}\Delta = \operatorname{Conv}\left\{(3,0,0),(0,3,0),(0,0,\frac{3}{2})\right\}$, *while* $P = \operatorname{Conv}\{(3,0,0),(0,3,0),(1,0,1),(0,1,1)\}$.

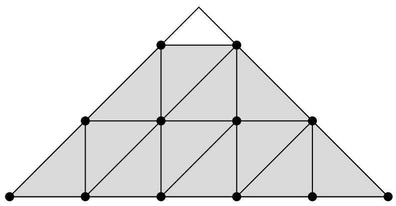


Therefore, $\deg \mathcal{R}_{2,3,2} = \operatorname{vol} P = 4 < \frac{9}{2} = \operatorname{vol} \frac{3}{2}\Delta$.

($k = 4$) *In this case,* $R_{2,4} = \mathbb{C}[x^4, x^3y, x^2y^2, xy^3, y^4, x^2a, xya, y^2a, a^2]$, *and so,* $P = 2\Delta = \operatorname{Conv}\{(4,0,0),(0,4,0),(0,0,2)\}$ *and* $\deg \mathcal{R}_{2,4,2} = \operatorname{vol} P = 8$.



($k = 5$) *As in the case* $k = 3$, $P \subsetneq \frac{5}{2}\Delta$. *More precisely,* $\frac{5}{2}\Delta = \operatorname{Conv}\left\{(5,0,0),(0,5,0),(0,0,\frac{5}{2})\right\}$, *and* $P = \operatorname{Conv}\{(5,0,0),(0,5,0),(1,0,2),(0,1,2)\}$.



Hence, $\deg \mathcal{R}_{2,5,2} = \operatorname{vol} P = 12 < \frac{25}{2} = \operatorname{vol} \frac{5}{2}\Delta$.

In this way, it is straightforward to see that, up to a linear isometry, the simplex $\frac{k}{2}\Delta$ *is the triangle with vertices* $(k,0,0),(0,k,0)$ *and* $\left(0,0,\frac{k}{2}\right)$. *If* k *is even, then they are lattice points,* $P = \frac{k}{2}\Delta$, *and*

$$\deg \mathcal{R}_{2,k,2} = \operatorname{vol} P = \operatorname{vol} \frac{k}{2}\Delta = \frac{k^2}{2}.$$

If k *is odd, then* P *is the trapezium with vertices* $(k,0,0),(0,k,0),\left(1,0,\frac{k-1}{2}\right)$ *and* $\left(0,1,\frac{k-1}{2}\right)$, *implying that*

$$\deg \mathcal{R}_{2,k,2} = \operatorname{vol} P = \frac{k^2-1}{2} < \frac{k^2}{2} = \operatorname{vol} \frac{k}{2}\Delta.$$

3 Axis-parallel paths

Besides $\mathcal{R}_{d,k,m}$, the universal variety contains other interesting subvarieties. One of them is the signature variety $\mathcal{L}_{d,k,m}$ of piecewise linear paths. Up to translation, a piecewise linear path X with m steps is the concatenation of m linear paths, each represented by a vector v_i. This decomposition is unique, provided that v_{i+1} is not a multiple of v_i for any i. In this section, we study the subfamily of *axis-parallel paths*.

Definition 3.1. *Let* $\{e_1, \ldots, e_d\}$ *be the standard basis for* $\mathbb{R}^d$. *A piecewise linear path* $X = v_1 \sqcup \ldots \sqcup v_m$ *is an axis-parallel path (or simply an axis path) if there are* $a_1, \ldots, a_m \in \mathbb{R}$ *such that* $v_i = a_i e_{\nu_i}$ *for every* i, *where* $\nu_i \in \{1, \ldots, d\}$.

In other words, each step is a multiple of a basis vector. Therefore, an axis-parallel path is characterized by two sequences. One of them is the sequence $\nu = (\nu_1, \ldots, \nu_m)$, called *shape* of X, that stores in each $\nu_i \in \{1, 2, \ldots, d\}$ the direction of the i-th step. The other sequence, $a = (a_1, \ldots, a_m) \in \mathbb{R}^m$, stores the length of each step and we call it the *sequence of lengths*. Notice that $\ell(\nu) = \ell(a) = m$. When we study an axis-parallel path $X: [0,1] \to \mathbb{R}^d$, we may assume that the image is not contained in any hyperplane, that is, it is nondegenerate. This means that $\{\nu_1, \ldots, \nu_m\} = \{1, \ldots, d\}$.

The k-th signature of an axis-parallel path X can be computed combinatorially in a very nice way. Recall that a *partition* of a set S is a collection of subsets (called blocks) such that their union is S. Now each sequence ν induces a partition $\pi_\nu = \{\pi_1|\pi_2|\ldots|\pi_d\}$ of the set $\{1, \ldots, m\}$, defined by $(\pi_\nu)_i = \{j \in \{1, \ldots, m\} \mid \nu_j = i\}$. For instance, if $\nu = (1, 2, 1, 3, 3, 1)$ then $\pi_\nu = \{1, 3, 6|2|4, 5\}$. We will write π instead of π_ν when there are no ambiguities. We can now introduce the main character of this section.

Definition 3.2. *Fix $\nu = (\nu_1, \ldots, \nu_m)$ and $k \in \mathbb{N}$. Let $d = \max(\nu_1, \ldots, \nu_m)$. For an axis parallel path $X = a_1 e_{\nu_1} \sqcup \ldots \sqcup a_m e_{\nu_m}$, let $g_{\nu,k}(X) := \sigma^{(k)}(X)$ be its k-th signature. As we did in Definition 2.3, we pass to the complex projective space and we define the* axis paths variety *$\mathcal{A}_{\nu,k}$ to be closure of the image of the composition*

$$\mathbb{R}^m \xrightarrow{g_{\nu,k}} (\mathbb{R}^d)^{\otimes k} \to (\mathbb{R}^d)^{\otimes k} \otimes \mathbb{C} = (\mathbb{C}^d)^{\otimes k} \dashrightarrow \mathbb{P}^{d^k-1}.$$

Instead of $g_{\nu,k}$, we can also consider

$$G_{\nu,k} : \mathbb{R}^m \to \mathbb{R}^d \times (\mathbb{R}^d)^{\otimes 2} \times \ldots \times (\mathbb{R}^d)^{\otimes k}$$

by sending $a \mapsto (g_{\nu,1}(a), \ldots, g_{\nu,k}(a))$. In this case we denote by $\mathcal{A}_{\nu,\le k} \subset \mathbb{C}^d \times (\mathbb{C}^d)^{\otimes 2} \times \ldots \times (\mathbb{C}^d)^{\otimes k}$ the closure of the complexification of the image of $G_{\nu,k}$.

There is a nice way to write down the polynomials defining the map $g_{\nu,k}$. The following result is a consequence of [AFS18, Corollary 5.3].

Lemma 3.3. *Let X be the axis-parallel path of shape ν and sequence of lengths a. Then the $(i_1 \ldots i_k)$-th entry of the k-th signature is*

$$\sigma(X)_{i_1 \ldots i_k} = \sum_{(j_1, \ldots, j_k)} \frac{1}{s_1! s_2! \cdots s_m!} a_{j_1} a_{j_2} \cdots a_{j_k},$$

where we sum over all the non-decreasing sequences $(j_1, j_2, \ldots, j_k)$ such that $j_l \in \pi_{i_l}$ for $l \in \{1, \ldots, k\}$, and s_l counts the number of times that l appears in $(j_1, j_2, \ldots, j_k)$.

Lemma 3.3 allows us to find the k-th signature of an axis-parallel path and to explicitly write the map $g_{\nu,k}$. We implemented a Macaulay2 code to compute the ideal of $\mathcal{A}_{\nu,k}$. This code can be found in [CGM].

Example 3.4. *Let σ be the 4-th signature of an axis path of shape $\nu = (1, 2, 1, 3, 2, 3, 1, 4)$ and sequence of lengths $a = (a_1, \ldots, a_8)$. Then $\pi_\nu = \{1, 3, 7|2, 5|4, 6|8\}$. By Lemma 3.3,*

$$\begin{aligned}
\sigma_{1234} &= a_1 a_2 a_4 a_8 + a_1 a_2 a_6 a_8 + a_1 a_5 a_6 a_8 + a_3 a_5 a_6 a_8, \\
\sigma_{2314} &= a_2 a_4 a_7 a_8 + a_2 a_6 a_7 a_8 + a_5 a_6 a_7 a_8, \\
\sigma_{4123} &= 0, \\
\sigma_{1124} &= \frac{1}{2} a_1^2 a_2 a_8 + \frac{1}{2} a_1^2 a_5 a_8 + a_1 a_3 a_5 a_8 + \frac{1}{2} a_3^2 a_5 a_8.
\end{aligned}$$

Remark 3.5. *As any signature variety, the axis paths variety is contained in the universal variety. Namely, $\mathcal{A}_{\nu,k} \subset \mathcal{U}_{d,k}$ for every $d \geq \max(\nu_1, \ldots, \nu_m)$. It is known that $\mathcal{L}_{d,k,m}$ coincides with $\mathcal{U}_{d,k}$ for m large enough, because every piecewise smooth path can be approximated by using piecewise linear paths. Since every piecewise linear path can be approximated by an axis path, we deduce that for every d and k, there exists ν such that $\max(\nu_i) = d$ and $\mathcal{A}_{\nu,k} = \mathcal{U}_{d,k}$.*

It is worth to write down what Chen's identity means for the axis paths signature variety.

Proposition 3.6. *Consider two sequences ν_1 and ν_2. If X_1 and X_2 are axis-parallel paths of shape ν_1 and ν_2 respectively, then*

$$g_{\nu_1\nu_2,k}(X_1 \sqcup X_2) = \sum_{a+b=k} g_{\nu_1,a}(X_1) \otimes g_{\nu_2,b}(X_2).$$

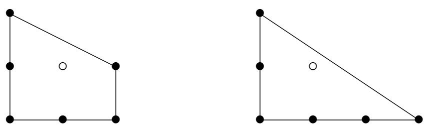
In particular, $\mathcal{A}_{\nu_1\nu_2,\leq k}$ is a projection of the Segre product $\mathcal{A}_{\nu_1,\leq k} \times \mathcal{A}_{\nu_2,\leq k}$.

In Section 2 we saw how important it is for a variety to be toric, so it is natural to ask whether $\mathcal{A}_{\nu,k}$ enjoys such property.

3.1 Toricness of $\mathcal{A}_{\nu,k}$

It is not difficult to find instances when the axis paths variety is toric. As an example we can take $\mathcal{A}_{\nu,1} = \mathbb{R}^d$, or consider $\nu = (1, 2, \ldots, d)$ and obtain the Veronese variety $\mathcal{V}_{d,k} = \mathcal{A}_{\nu,k}$. A less trivial example is given by Remark 3.5. Thanks to Lemma 3.3, we are able to perform efficient computations and check that many other occurrences of $\mathcal{A}_{\nu,k}$ are toric, and even find their polytopes.

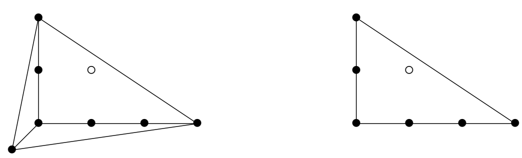
Example 3.7. *In [CGM] we compute the ideal of $\mathcal{A}_{(1,2,1),3}$ and we check that it is toric. Further computations on the ideal allow us to determine its polytope P. Indeed, $\mathcal{A}_{(1,2,1),3}$ is a degree 6 surface in $\mathbb{P}^6$, so P is a two-dimensional lattice polytope of normalized area 6 that contains exactly 7 lattice points. By checking the Hilbert and Erhart polynomials and the Betti table, we see that there are only two possible polytopes.*



By looking at the intersections of the tangent spaces at singular points with the variety, we see that the correct one is the one on the right.

In [Gal18] the second author presents a change of coordinates, based on the exponential description of the tensor signature, for which the universal variety $\mathcal{U}_{d,k}$ is the image of a monomial map, and therefore it is toric. However, such change of coordinates fails to make every $\mathcal{A}_{\nu,k}$ toric as well. For instance, it does not turn $\mathcal{A}_{(1,2,1),3}$ into a variety defined by binomials. One could still hope to find another change of coordinates that makes both ideals of $\mathcal{A}_{\nu,k}$ and $\mathcal{U}_{d,k}$ binomial. Unfortunately, this is not true.

Example 3.8. *No change of coordinates in $\mathbb{P}^7$ can make the ideals of both $\mathcal{A}_{(1,2,1),3}$ and $\mathcal{U}_{3,3}$ toric. The polytopes P and Q of $\mathcal{U}_{3,3}$ and $\mathcal{A}_{(1,2,1),3}$ are*



There are two cases how $\mathcal{A}_{(1,2,1),3}$ could be a toric subvariety of $\mathcal{U}_{3,3}$. Either it is a toric divisor or the inclusion is a toric morphism. A careful technical analysis shows that neither is possible.

Despite previous Example, the axis paths variety turns out to be toric in many cases.

Proposition 3.9. *Let $d = \max(\nu_1, \dots, \nu_m)$. Then*

1. *$\mathcal{A}_{\nu,3}$ is toric for every $d \le 2$.*
2. *$\mathcal{A}_{\nu,2}$ is toric for every $d \le 3$.*
3. *$\mathcal{A}_{(1,2,\dots,d,j),2}$, with $1 \le j < d$, is toric.*

Proof. It is not difficult to check all possible cases for the first two items. The computation can be found in [CGM]. For the third item, we have $m = d+1$. Let us apply the change of coordinates

$$\begin{cases} \tilde{a}_j = a_j + a_{d+1} \\ \tilde{a}_i = a_i \text{ for } i \neq j \end{cases}$$

on the domain $\mathbb{R}^{d+1}$. By Lemma 3.3, we know that the entries of $\sigma^{(2)}(X)$ are

$$\sigma^{(2)}(X)_{il} = \begin{cases} a_i a_l = \tilde{a}_i \tilde{a}_l & \text{for } i, l \neq j \\ a_i(a_j + a_{d+1}) = \tilde{a}_i \tilde{a}_j & \text{for } i < l = j \\ (a_j + a_{d+1})^2 = (\tilde{a}_j)^2 & \text{for } i = l = j \\ a_i a_{d+1} = \tilde{a}_i \tilde{a}_{d+1} & \text{for } i > l = j \\ 0 & \text{for } i = j < l \\ a_j a_i & \text{for } i = j > l. \end{cases}$$

Finally, if we perform the change of coordinates $\sigma^{(2)}(X)_{jd} \mapsto \sigma^{(2)}(X)_{jd} + \sigma^{(2)}(X)_{j,d+1}$, the map becomes monomial. □

There are further cases in which we can prove that the axis paths variety is toric.

Lemma 3.10. *If the ideal of $\mathcal{A}_{\nu,\le k}$ is binomial after a linear change of coordinates, then the same holds for the ideal of $\mathcal{A}_{\nu(d+1),k}$.*

Proof. Let X be an axis path of shape ν and sequence of length $(a_1, \dots, a_m)$. Since $d+1$ does not appear in ν, in the set partition $\pi_{\nu(d+1)}$ the $(d+1)$-st block only contains the entry a_{m+1}. Therefore, by Lemma 3.3,

$$\sigma(X)_{i_1 \dots i_{\tilde{k}} (d+1)^{k-\tilde{k}}} = \begin{cases} \frac{1}{(k-\tilde{k})!}(a_{m+1})^{k-\tilde{k}} \sigma(X)_{i_1 \dots i_{\tilde{k}}} & \text{if } d+1 \text{ does not appear in } (i_1, \dots, i_{\tilde{k}}) \\ 0 & \text{otherwise.} \end{cases}$$

Hence, $\mathcal{A}_{\nu(d+1),k}$, up to diagonal change of coordinates, is equal to the projectivization of $\mathcal{A}_{\nu,\le k}$. □

We finish this section with the following natural question.

Question. *Is every variety $\mathcal{A}_{\nu,k}$ toric?*

Our results provide a positive answer in all relatively small cases. However, since we found no general technique, we expect that the answer may be negative. The main obstacle to provide a nontoric example is that a first potential candidate is already too large to be dealt with, either with ad hoc geometric arguments or general computational methods [KMM17].

3.2 Dimension of $\mathcal{A}_{\nu,k}$

As we stressed in the introduction, a fundamental problem when we deal with signatures is to determine whether it is possible to recover the path from the signature. The best case scenario is injectivity, when we can uniquely reconstruct the path. In general this is too much to hope, so we focus on a weaker property, and we want to know when $g_{\nu,k}$ is generically finite. Therefore we are interested in the dimension of the fibers, or equivalently in $\dim \mathcal{A}_{\nu,k}$. Since $\mathcal{A}_{\nu,k} \subset \mathcal{U}_{d,k}$ is the image of $\mathbb{R}^{\ell(\nu)}$ under a polynomial map, we have

$$\dim \mathcal{A}_{\nu,k} \leq \min\{\ell(\nu), \dim(\mathcal{U}_{d,k})\}.$$

In general, the inequality can be strict. For instance, the dimension of $\mathcal{A}_{(1,2,1,2,3),3}$ is 4, while $\ell(\nu) = 5$ and $\dim(\mathcal{U}_{3,3}) = 7$. In order to better study the dimension of our axis-parallel signature variety, we introduce the following definition.

Definition 3.11. *For $d, k \in \mathbb{N}$, we define the set*

$$N_{d,k} := \{\nu \mid \max(\nu) \leq d \textit{ and } G_{\nu,k} \textit{ is not generically finite}\}.$$

We say that $\mathcal{A}_{\nu,k}$ is defective *if $\nu \in N_{d,k}$ for $d \geq \max(\nu)$. Otherwise, we say that $\mathcal{A}_{\nu,k}$ has the expected dimension.*

It may seem that we made a choice in using $G_{\nu,k}$ instead of $g_{\nu,k}$ in the previous definition. However, the k-th signature of a generic piecewise smooth path X defines all the previous signatures up to finitely many choices (see [AFS18, Section 6]). Therefore,

$$N_{d,k} = \{\nu \mid \max(\nu) \leq d \text{ and } g_{\nu,k} \text{ is not generically finite}\}.$$

For the same reason, $N_{d,k+1} \subset N_{d,k}$ for every k. Each $N_{d,k}$ is a language, i.e. a set of words. It would be very interesting to completely characterise it - cf. Conjecture 3.15. First, we prove that $N_{d,k}$ is absorbing with respect to concatenation.

Lemma 3.12. *Let ν_1, ν_2 be two shapes, let $\nu_1\nu_2$ be their concatenation and let $d = \max(\nu_1\nu_2)$. If $\nu_1 \in N_{d,k}$, then both $\nu_1\nu_2$ and $\nu_2\nu_1$ belong to $N_{d,k}$.*

Proof. Take a general point of $\mathcal{A}_{\nu_1\nu_2,\leq k}$. It is of the form $\sigma^{\leq k}(X)$ for some general axis-parallel path X. Since X has shape $\nu_1\nu_2$, we can write it as a concatenation X_1X_2, for general paths X_1 of shape ν_1 and X_2 of shape ν_2. By hypothesis, $\nu_1 \in N_{d,k}$, and so there exist infinitely many paths Y of shape ν_1 such that $g_{\nu_1,a}(Y) = g_{\nu_1,a}(X_1)$ for every $a \leq k$. By Proposition 3.6,

$$g_{\nu_1\nu_2,k}(YX_2) = \sum_{a+b=k} g_{\nu_1,a}(Y) \otimes g_{\nu_2,b}(X_2) = \sum_{a+b=k} g_{\nu_1,a}(X_1) \otimes g_{\nu_2,b}(X_2) = g_{\nu_1\nu_2,k}(X),$$

hence the fiber containing X is not finite and therefore $\nu_1\nu_2 \in N_{d,k}$. In the same way it is possible to prove that $\nu_2\nu_1 \in N_{d,k}$. □

On the other hand, if $\mathcal{A}_{\nu,k}$ is not defective, then we can add a new letter at any point in ν and the resulting variety is still of expected dimension.

Lemma 3.13. *Let $\nu \notin N_{d,k}$ and let $d = \max(\nu)$. If we write $\nu = \nu_1\nu_2$ and we take a new letter $l > d$, then $\nu_1 \cdot l \cdot \nu_2 \notin N_{l,k}$.*

Proof. Let us pick a general element in the affine cone over $\mathcal{A}_{\nu_1 \cdot l \cdot \nu_2,k}$. First we know that, up to finite number of choices, we can identify the parameter a associated to l, as $\sigma_{(l^k)} = \frac{a^k}{k!}$. We may also identify the other parameters as the signatures indexed by numbers from one to d are the same for ν and $\nu_1 \cdot l \cdot \nu_2$. □

We noticed several times that the universal variety can be obtained as $\mathcal{A}_{\nu,k}$ for sufficiently long ν. Now we want to prove that this can be done in an efficient way, namely by using a sequence with as many entries as $\dim \mathcal{U}_{d,k}$.

Lemma 3.14. *For every d and k, there is a shape ν such that $\max(\nu) = d$, $\mathcal{A}_{\nu,k} = \mathcal{U}_{d,k}$ and $\ell(\nu) = \dim \mathcal{U}_{d,k}$.*

Proof. By Remark 3.5, there is a ν' satisfying the first two properties. We build ν as a subsequence of ν' by deleting those entries that do not increase the dimension of the variety in the following way. If ν' satisfies the last requirement too, we are done. Suppose then that $\nu' = (\nu_1, \ldots, \nu_s)$ for $s > \dim \mathcal{U}_{d,k}$. By construction, there is an index $i \in \{1, \ldots, s-1\}$ such that $\mathcal{A}_{(\nu_1,\ldots,\nu_i),k} = \mathcal{A}_{(\nu_1,\ldots,\nu_{i+1}),k}$. If $i = s-1$, take $\nu' := (\nu_1, \ldots, \nu_{s-1})$ and proceed by induction on the length ν'. Otherwise, let $\nu := (\nu_1, \ldots, \widehat{\nu_{i+1}}, \ldots, \nu_s)$, where the entry with the hat is omitted. Consider the sequences

$$\alpha := (\nu_1, \ldots, \nu_i), \quad \alpha' := (\nu_1, \ldots, \nu_{i+1}), \quad \beta := (\nu_{i+2}, \ldots, \nu_s),$$

so that $\nu' = \alpha'\beta$ and $\nu = \alpha\beta$. By Proposition 3.6, the map

$$\sigma^{\leq k} \colon \mathbb{R}^{\ell(\nu')} \to \mathcal{A}_{\nu',\leq k}$$

factors as

$$\mathbb{R}^{\ell(\nu')} \to \mathcal{A}_{\alpha',\leq k} \times \mathcal{A}_{\beta,\leq k} \xrightarrow{\text{Segre}} \mathcal{A}_{\nu',\leq k}.$$

We also have another map

$$\mathbb{R}^{\ell(\nu)} \to \mathcal{A}_{\alpha,\leq k} \times \mathcal{A}_{\beta,\leq k} \xrightarrow{\text{Segre}} \mathcal{A}_{\nu,\leq k}.$$

Since $\mathcal{A}_{\alpha,\leq k} = \mathcal{A}_{\alpha',\leq k}$, we conclude that $\mathcal{A}_{\nu,\leq k} = \mathcal{A}_{\nu',\leq k}$. We continue the process until there is no such index i. The resulting subsequence of ν' satisfies the requirements of the statement. □

As an easy consequence, we notice that if $d = 2$ there is only one possible shape $(1, 2, 1, 2, \ldots)$. Since it is unique, it satisfies the properties of Lemma 3.14 and therefore $\mathcal{A}_{\nu,k}$ has always the expected dimension.

Now, we look back to our first defective example. The reason why $\mathcal{A}_{(1,2,1,2,3),3}$ is defective is that there is a subsequence $(1, 2, 1, 2)$ such that $\mathcal{A}_{(1,2,1,2),3} = \mathcal{U}_{2,3}$ but $\dim(\mathcal{U}_{2,3}) = 3$. This means that we are filling a 3-dimensional universal variety with a sequence of length four. We conjecture that this behavior is the only obstruction to having the expected dimension.

Conjecture 3.15. *Let $\nu = (\nu_1, \ldots, \nu_m)$ be a sequence with $d = \max(\nu)$. Then $\mathcal{A}_{\nu,k}$ is defective if and only if there is a subsequence $\nu' = (\nu_i, \ldots, \nu_{i+r})$ of ν with $d' = \max(\nu')$, such that $\mathcal{A}_{\nu',k} = \mathcal{U}_{d',k}$ but $r + 1 > \dim(\mathcal{U}_{d',k})$.*

3.3 Determinant of axis-parallel signatures

Let us start with an observation about the entries of the k-th signature for a special family of axis-parallel paths.

Lemma 3.16. *Let $k \in \mathbb{N}$ and let X be an axis-parallel path with shape ν and sequence of length $(a_1, \ldots, a_m)$. If an entry ν_i of ν appears only once in the j-th block of π, then a_i divides all the entries of each j-th slice of $\sigma^{(k)}(X)$. If moreover $k = 2$, then $a_i^2 \mid \det(\sigma^{(2)}(X))$.*

Proof. If we look at one of the j-th slices, we are fixing one of the indices of $\sigma^{(k)}(X)$ to be j. By Lemma 3.3, every monomial of the sum is a multiple of a_i. When $k = 2$, we are dealing with a square matrix in which both the j-th row and the j-th column of $\sigma^{(2)}(X)$ are multiples of a_i. In order to conclude that $a_i^2 \mid \det(\sigma^{(2)}(X))$, it is enough to check that a_i^2 divides the diagonal entry $\sigma^{(2)}(X)_{jj}$, and this follows from Lemma 3.3. □

This is just a special case of a much more general phenomenon, corroborated by many experiments with Sage [The18] and the code included in [CGM]. The determinant of the second signature of an axis-parallel path is always the square of a polynomial in $a_1, \ldots, a_m$. This subsection is devoted to explaining this result and its consequences. In order to correctly state it, we need some definitions.

Definition 3.17. *For any shape $\nu = (\nu_1, \ldots, \nu_m)$ with $d = \max \nu$, we say that a subsequence $\mu = (\nu_{i_1}, \ldots, \nu_{i_d})$ is a* good subshape *if $i_1 < \cdots < i_d$ and $\{\nu_{i_1}, \ldots, \nu_{i_d}\} = \{1, \ldots, d\}$. Moreover, we define the* sign *of a good subshape, $\operatorname{sgn} \mu$, as the sign of the permutation $(\nu_{i_1}, \ldots, \nu_{i_d}) \in S_d$.*

Without loss of generality we assumed that $\{1, \ldots, d\} = \{\nu_1, \ldots, \nu_m\}$, so good subshapes always exist.

Definition 3.18. *Let X be an axis path of shape ν and let a_i be the parameter associated to ν_i. We define the degree d homogeneus polynomial $P(a) \in \mathbb{C}[a_1, \ldots, a_m]$ by*

$$P(a) := \sum_{\mu=(\nu_{i_1},\ldots,\nu_{i_d})} (\operatorname{sgn} \mu) \prod_{j=1}^{d} a_{i_j},$$

where the sum is taken over all good subshapes μ of ν. Moreover, we denote by $\det_2 \nu \in \mathbb{C}[a_1, \ldots, a_m]$ the determinant $\det\left(\sigma^{(2)}(X)\right)$. By Lemma 3.3, it is homogeneous of degree $2d$.

Theorem 3.19. *With the notation from Definition 3.18, we have*

$$2^d \det\left(\sigma^{(2)}(X)\right) = P(a)^2.$$

The proof we found is quite technical and we present it in Appendix A. However, Theorem 3.19 has several interesting consequences. Since we can write $\mathcal{U}_{d,2}$ as an axis paths variety, we now know that the determinant of the signature matrix of any path is a square. In particular, if X is any path, then $\det \sigma^{(2)}(X) \geq 0$, hence the real part of $\mathcal{U}_{d,2}$ lies in the semialgebraic set $\{M \in \mathbb{R}^{d \times d} \mid \det M \geq 0\}$. In other words, real points of $\mathcal{U}_{d,2}$ with negative determinant are not signature matrices of paths, but rather they come from taking the Zariski closure. It is even more interesting to think in terms of the shuffle identity.

Example 3.20. *Let $X\colon [0,1] \to \mathbb{R}^2$ be any path. By Lemma 1.5,*

$$\begin{aligned}
\det(\sigma^{(2)}(X)) &= \det \begin{pmatrix} \langle \sigma(X), 11 \rangle & \langle \sigma(X), 12 \rangle \\ \langle \sigma(X), 21 \rangle & \langle \sigma(X), 22 \rangle \end{pmatrix} \\
&= \langle \sigma(X), 11 \rangle \cdot \langle \sigma(X), 22 \rangle - \langle \sigma(X), 12 \rangle \cdot \langle \sigma(X), 21 \rangle \\
&= \langle \sigma(X), 11 \sqcup\!\sqcup\, 22 - 12 \sqcup\!\sqcup\, 21 \rangle \\
&= \frac{1}{4} \langle \sigma(X), (12 - 21)^{\sqcup\!\sqcup 2} \rangle.
\end{aligned}$$

In this case it was not too difficult to realize that $4(11 \sqcup\!\sqcup 22 - 12 \sqcup\!\sqcup 21) = (12-21)^{\sqcup\!\sqcup 2}$ and therefore that $\det(\sigma^{(2)}(X))$ is a square. Note that $12 - 21$ is twice the Lévy area defined by the path. This is a general behavior.

Definition 3.21. *Let S_d be the symmetric group in d elements. Define*

$$\mathrm{inv}_d = \sum_{\rho \in S_d} \mathrm{sgn}(\rho)\rho(1)\ldots\rho(d) \in T(\mathbb{R}^d).$$

It is the sum of $d!$ words of length d, so it is a degree d element of the shuffle algebra.

On one hand, this has a geometric meaning: as illustrated in [DR18, Section 3.2], inv_d has an interpretation in terms of the *signed volume* of the convex hull of the path. On the other hand, there is a relation with signatures.

Lemma 3.22. *Let $a = (a_1, \ldots, a_m) \in \mathbb{R}^m$ and let $X : [0,1] \to \mathbb{R}^d$ be the axis path corresponding to the sequence of lengths a and shape ν. Then $P(a) = \langle \sigma(X), \mathrm{inv}_d \rangle$.*

Proof. By Lemma 3.3, if $i_1, \ldots, i_d$ are pairwise disjoint then $\langle \sigma(X), i_1 \ldots i_d \rangle$ equals the sum of products $\prod_{l=1}^{d} a_{j_l}$ where $\nu_{j_l} = i_l$ and $j_1 < \cdots < j_d$. In particular, it is a subsum in the definition of $P(a)$ corresponding to the good subshapes associated to the permutation $(i_1, \ldots, i_d)$. Summing up over all possible permutations $(i_1, \ldots, i_d)$ with signs we obtain the statement of the lemma. □

It follows that Theorem 3.19 has an important consequence on the shuffle algebra.

Corollary 3.23. *Consider the matrix*

$$A = \begin{pmatrix} 11 & 12 & \ldots & 1d \\ 21 & 22 & \ldots & 2d \\ \vdots & \vdots & \ddots & \vdots \\ d1 & d2 & \ldots & dd \end{pmatrix}$$

with coefficients in the shuffle algebra $(T(\mathbb{R}^d), \sqcup\!\sqcup, e)$. Let $\det_{\sqcup\!\sqcup}(A)$ be its determinant, where the product is the shuffle. Then $2^d \det_{\sqcup\!\sqcup}(A) = (\mathrm{inv}_d)^{\sqcup\!\sqcup 2}$.

Proof. It is enough to prove that $\langle T, (\mathrm{inv}_d)^{\sqcup\!\sqcup 2} \rangle = 2^d \langle T, \det_{\sqcup\!\sqcup}(A) \rangle$ for every $T \in T((\mathbb{R}^d))$. Since the linear span of $\mathcal{G}(\mathbb{R}^d)$ is the whole $T((\mathbb{R}^d))$, by linearity we just have to show that such equality holds when $T \in \mathcal{G}(\mathbb{R}^d)$. Let us pick then $T \in \mathcal{G}(\mathbb{R}^d)$. Then there is an axis path X, corresponding to the sequence of lengths $a \in \mathbb{R}^m$, such that $\langle T, (\mathrm{inv}_d)^{\sqcup\!\sqcup 2} \rangle = \langle \sigma(X), (\mathrm{inv}_d)^{\sqcup\!\sqcup 2} \rangle$ and $\langle T, \det_{\sqcup\!\sqcup}(A) \rangle = \langle \sigma(X), \det_{\sqcup\!\sqcup}(A) \rangle$. By Theorem 3.19 and Lemma 3.22,

$$2^d \langle \sigma(X), \det_{\sqcup\!\sqcup}(A) \rangle = 2^d \det(\sigma^{(2)}(X)) = P(a)^2 = \langle \sigma(X), (\mathrm{inv}_d)^{\sqcup\!\sqcup 2} \rangle.$$ □

Even though the techniques we use are based on the combinatorics of $\mathcal{A}_{\nu,k}$, Corollary 3.23 has nothing to do with axis paths, nor with signatures at all. It would be very interesting to find a generalization for $k > 2$. In our opinion, this can give new insight on the properties of the shuffle algebra.

A Proof of Theorem 3.19

For this proof we fix an axis path X with shape $\nu \in \mathbb{R}^m$ and sequence of lengths $a \in \mathbb{R}^m$. The first step will be to reduce the problem to the case in which every entry of ν appears at most twice. Let us start with an observation.

Remark A.1. *The symmetric group S_d acts on $\mathbb{R}^d$ by permuting the basis elements. In the same way, it acts on the sequence ν, on the path X and thus on the polynomials P and $\det_2 \nu$. The determinant is invariant, while the sign of P changes according to the sign of the permutation. However, P^2 is invariant. This means that we are allowed to relabel the entries of ν. In other words, we can always pick $\sigma \in S_d$ and replace $(\nu_1, \ldots, \nu_m)$ with $(\sigma(\nu_1), \ldots, \sigma(\nu_m))$.*

In order to prove that two polynomials are equal, it is enough to prove that they have the same coefficient on each monomial.

Notation A.2. *For a polynomial Q and a monomial M let $Q_{|M}$ be the coefficient of Q corresponding to M.*

Since both P^2 and $\det_2 \nu$ are homogeneous of degree $2d$, we only have to take care of degree $2d$ monomials.

Lemma A.3. *Suppose that there are three (not necessarily distinct) indices $i, j, l \in \{1, \ldots, m\}$ such that $\nu_i = \nu_j = \nu_l$.*

1. *Let $M \in \mathbb{C}[a_1, \ldots, a_m]_{2d}$ be such a monomial that $a_i a_j a_l \mid M$. Then $P^2_{|M} = (\det_2 \nu)_{|M} = 0$.*
2. *Let ν' be the sequence obtained from ν by removing ν_l and let P' be the associated polynomial. If $M \in \mathbb{C}[a_1, \ldots, \hat{a}_l, \ldots, a_m]_{2d}$ is a monomial, then $P'_{|M} = P_{|M}$ and $\det_2 \nu_{|M} = \det_2 \nu'_{|M}$.*

Proof. Thanks to remark A.1, we may assume $\nu_i = \nu_j = \nu_l = 1$. By Lemma 3.3, the variables a_i, a_j, a_l only appear in the first row and in the first column of the signature matrix. Furthermore, among the monomials of the diagonal entry $\sigma(X)_{11}$ there are degree at most 2 monomials in the variables a_i, a_j, a_l, while away from the diagonal all monomials contain only one of these variables, with exponent 1. Therefore $a_i a_j a_l$ does not divide any monomial appearing in $\det_2 \nu$. On the other hand, every good subshape μ of ν contains the entry 1 exactly once, so each monomial of P contains at most one among a_i, a_j, a_l, and with exponent 1. In P^2 there can be monomials containing the product of two among a_i, a_j, a_l, but not containing all three of them. The second statement follows from the definitions of P and $\det_2 \nu$. ☐

Thanks to Lemma A.3, from now on we can restrict our attention to sequences ν with no triple entries. Now we want to take care of the case in which an entry appears only once. We find it useful to focus on a particular monomial.

Definition A.4. *Assume that ν has no triple entries. If $\mu = (\nu_{i_1}, \ldots, \nu_{i_\ell})$ is a subsequence of ν, we set*

$$e(i_j) = \begin{cases} 1 & \text{if } \nu_{i_j} \text{ appears twice in } \mu \\ 2 & \text{if } \nu_{i_j} \text{ appears onece in } \mu \end{cases}$$

and we define the monomial

$$M_\mu = \prod_{j=1}^{\ell} a_{i_j}^{e(i_j)} \in \mathbb{C}[a_1, \ldots, a_m].$$

Since we are assuming that ν has no triple entries, $\deg(M_\nu) = 2d$.

Lemma A.5. *If a monomial M appears in either P^2 or $\det_2 \nu$ and it is divisible by all variables $a_1, \ldots, a_m$, then $M = M_\nu$.*

Proof. Let M be such a monomial. By Lemma A.3, $M \mid M_\nu$. Indeed, if ν_i appears in ν at most once, then $a_i^3 \nmid M$. If $\nu_i = \nu_j$ appears exactly twice then $a_i a_j \mid M$, but $a_i^2, a_j^2 \nmid M$. As the degrees are the same, equality follows. □

The next results will explain what happens to the coefficient of M_ν when ν has a non-repeated entry.

Lemma A.6. *Assume ν contains at least one non-repeated entry. Suppose that the last one occurs in ν_{m-l}. Let ν_0 be the sequence obtained from ν by discarding ν_{m-l} and let P_0 be the associated polynomial. Then*

$$(P(a)^2)_{|M_\nu} = (-1)^l (P_0(a)^2)_{|M_{\nu_0}}.$$

Proof. By Remark A.1, we can assume that $\nu_{m-l} = d$. Since M_ν is a monomial of P^2, to compute $P^2_{|M_\nu}$ we have to consider the contribution of all pairs (α, β) of good subshapes of ν. Let (α, β) be a pair contributing to M_ν. The variable of any non-repeated entry - including d - has to appear in both α and β. On the other hand, the variable of an entry that appears twice appears once in α and once in β. Therefore there is a bijection between pairs (α, β) of good subshapes of ν contributing to M_ν and pairs (α_0, β_0) of good subshapes of ν_0 contributing to M_{ν_0}, where α_0 is the sequence obtained from α by removing d, and β_0 is defined in a similar way from β. If we set now a (resp. b) to be the number of entries of α (resp β) after the discarded one, then

$$\begin{aligned}(P(a)^2)_{|M_\nu} &= \sum_{(\alpha,\beta)} \operatorname{sgn}(\alpha)\operatorname{sgn}(\beta) = \sum_{(\alpha_0,\beta_0)} (-1)^a \operatorname{sgn}(\alpha_0)(-1)^b \operatorname{sgn}(\beta_0) \\ &= (-1)^l \sum_{(\alpha_0,\beta_0)} \operatorname{sgn}(\alpha_0)\operatorname{sgn}(\beta_0) = (-1)^l (P_0(a)^2)_{|M_{\nu_0}}.\end{aligned}$$
□

Before we prove the analogous statement for the polynomial $\det_2 \nu$, we introduce a combinatorial interpretation of the coefficient $\det_2(\nu)_{|M_\nu}$. By Laplace expansion, the contributions to $\det_2(\nu)_{|M_\nu}$ come from directed graphs (with possible loops) with vertices corresponding to symbols in ν and:

- if a symbol i appears twice in ν then precisely one such vertex is outgoing and one is incoming;
- if a symbol i appears once in ν then it has degree two and has one incoming and one outgoing edge - this is the only case where a vertex can be a loop;
- all edges are from left to right.

Each such graph contributes $\pm\frac{1}{2^m}$ where m is the number of loops and the sign is the sign of the corresponding permutation.

Example A.7. *Let $\nu = (1, 2, 1)$. We have two possible graphs. One with edges $(1, 2)$ and $(2, 1)$. It encodes the transposition $(1, 2)$ and contributes with -1. The other one has an edge $(1, 1)$ and a loop over two. It encodes the identity permutation and contributes $\frac{1}{2}$.*

Lemma A.8. *With the same hypothesis as in Lemma A.6,*

$$\det_2(\nu)_{|M_\nu} = \frac{(-1)^l}{2} \cdot \det_2(\nu_0)_{|M_{\nu_0}}.$$

Proof. We look at the possible cases for the position of d.

If d is the last symbol then the d-th column of the second signature matrix has only one nonzero entry, which equals $\frac{1}{2}$ times the square of the variable associated to d. Applying Laplace expansion we obtain the formula in this case.

We now assume that d is the last but one entry. The contributing graphs are of two types.

If d is a loop then (by forgetting the loop) we obtain graphs on ν_0. Hence, the contribution of those graphs is $\frac{1}{2}(\det_2(\nu_0))_{|M_{\nu_0}}$.

If d is not a loop, it must have an incoming edge, say (a, d) and an outgoing edge (d, b), where b must be the last symbol in ν. We may replace this by an edge (a, b). In the cycle presentation of the permutation we removed d from one cycle, i.e. changed the sign of the permutation. As we did not change the number of loops the contribution equals $-(\det_2(\nu_0))_{|M_{\nu_0}}$. Summing the two contributions we obtain the result in this case.

We may now assume that d is the last letter in ν that appears once and further there are at least two symbols after it, i.e. $\nu = \ldots dab \ldots$. Let $\nu' = \ldots abd$ be ν with d moved two places forward. We will prove that $(\det_2(\nu))_{|M_\nu} = (\det_2(\nu'))_{|M_{\nu'}}$. By induction this will finish the proof of the lemma.

We may identify graphs contributing to $(\det_2(\nu))_{|M_\nu}$ with those contributing to $(\det_2(\nu'))_{|M_{\nu'}}$ with the exception of graphs for which there is an edge from d either to a or b. Further, if there is an edge (d, a) and the considered b is an outgoing vertex, the encoded permutation is $x \to d \to a$, $b \to c$ for some x and c. We may associate to it a graph on ν' with edges $(x, a), (b, d), (d, c)$. This does not change the sign of permutation. We proceed in the same way if there is an edge (d, b) and a is outgoing. The remaining graphs are those where there is an edge from d to a or b and the other vertex is incoming. We will prove that the sum of contributions of such graphs equals zero.

Consider a graph with edges $(d, a), (c, b)$ for some c. We associate to it a graph with edges $(c, a), (d, b)$. In the cycle decomposition of the permutation this operation either joins two cycles or decomposes one cycle into two, i.e. changes the sign of the permutation. In particular, the contribution of each pair equals zero.

We also see that we obtain all graphs on ν', apart from those for which there is an edge from a or b to d and the other vertex is outgoing. Just as above one can show that the contribution of such graphs equals zero. □

We can now reduce to the case in which every entry of ν appears exactly twice.

Corollary A.9. *Assume Theorem 3.19 holds for sequences in which every entry appears exactly twice. Then it holds for every sequence.*

Proof. We compare the coefficients of monomials in both polynomials. If a monomial is different from M_ν, then it does not contain one of the variables. In this case Lemma A.5 allows us to consider a subsequence of ν and conclude by induction on $\ell(\nu)$. Consider then M_ν. By repeatedly applying Lemmas A.6 and A.8, we can discard all non-repeated entries and conclude by hypothesis. □

From now on we may assume that each of the letters $1, \ldots, d$ appears exactly twice in ν. In particular, $d = 2m$. As in the proof of Corollary A.9, we can assume, by induction on $\ell(\nu)$, that $P^2_{|M} = \det_2 \nu_{|M}$ for every monomial $M \neq M_\nu$. In order to prove that the two polynomials coincide, it is enough to find a point $q \in \mathbb{R}^{2m}$ such that $M_\nu(q) \neq 0$ and $P^2(q) = \det_2 \nu(q)$. Since we suppose that every entry of ν appears exactly twice, we can define $q \in \mathbb{R}^{2m}$ by

$$q_i = \begin{cases} 1 & \text{if } \nu_i \text{ appears for the first time in the } i\text{-th entry of } \nu, \\ -1 & \text{if } \nu_i \text{ appears for the second time in the } i\text{-th entry of } \nu. \end{cases}$$

The following is a straightforward consequence of Lemma 3.3.

Lemma A.10. *The matrix $\sigma^{(2)}(q)$ is skew-symmetric. If $i \neq j$, then*

$$\sigma^{(2)}(q)_{ij} = \begin{cases} -1 & \textit{if the subsequence of } \nu \textit{ with the symbols } i \textit{ and } j \textit{ is } jiji, \\ 1 & \textit{if the subsequence of } \nu \textit{ with the symbols } i \textit{ and } j \textit{ is } ijij, \\ 0 & \textit{otherwise.} \end{cases}$$

To finish the proof it is enough to show that $P(q)^2 = 2^d \det_2(\nu)(q)$. As the second signature matrix is skew-symmetric, it is enough to prove that $P(q) = 2^{d/2}\operatorname{Pf}_2(\nu)(q)$, where Pf is the Pfaffian. We note that when $\nu = (1, 1, 2, 2, \ldots, d, d)$ the claim is easy as both sides equal zero. The following two lemmas allow to reduce any ν to this case, finishing the proof.

Let ν_i, ν_{i+1} be two consecutive entries in ν. Let ν' be the sequence obtained from ν by switching ν_i and ν_{i+1}, and let $q' \in \mathbb{R}^{2m}$ be the corresponding point, defined in the same way we defined q. Let ν'' be the sequence obtained from ν by removing both occurrences of the symbol ν_i and both occurrences of the symbol ν_{i+1}. In a similar fashion we define $q'' \in \mathbb{R}^{2m-4}$. By Remark A.1, we may assume without loss of generality that $\nu_i = d - 1$ and $\nu_{i+1} = d$.

Lemma A.11. *Let*

$$e(d) = \begin{cases} 1 & \textit{if } d \textit{ appears for the first time in the } (i+1)\textit{-st entry of } \nu \\ -1 & \textit{if } d \textit{ appears for the second time in the } (i+1)\textit{-st entry of } \nu. \end{cases}$$

In a similar way we define $e(d-1)$. Then

1. $\operatorname{Pf}_2(\nu)(q) = \operatorname{Pf}_2(\nu')(q) + e(d-1)e(d)\operatorname{Pf}_2(\nu'')(q'')$ *and*
2. $P_\nu(q) = P_{\nu'}(q) + 2e(d-1)e(d)P_{\nu''}(q')$.

Proof. 1. The change from ν to ν' changes the second signature matrix by replacing the lower right 2×2 submatrix either from $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ to $\pm \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ or the other way round. The formula follows from the standard Laplace expansion for Pfaffians.

2. The good underlinings that involve at most one of the exchanged $d-1$ and d provide the same contribution both to ν and ν'. It remains to investigate the contribution of good underlinings containing both $d-1$ and d. These contribute to ν and ν' with opposite signs and are exactly the contributions of $P_{\nu''}(q')$. The sign depends only on the property if the underlined variables we forget are taken with plus or minus. □

We conclude by induction on the number of permutations needed to transform ν to $1122\ldots dd$ and the length of ν.

References

[AFS18] Carlos Améndola, Peter Friz, and Bernd Sturmfels. Varieties of signature tensors, 2018.

[AGH+17] Benjamin Assarf, Ewgenij Gawrilow, Katrin Herr, Michael Joswig, Benjamin Lorenz, Andreas Paffenholz, and Thomas Rehn. Computing convex hulls and counting integer points with `polymake`. *Math. Program. Comput.*, 9(1):1–38, 2017. https://polymake.org/doku.php.

[BIR+] W. Bruns, B. Ichim, T. Römer, R. Sieg, and C. Söger. Normaliz. algorithms for rational cones and affine monoids. Available at https://www.normaliz.uni-osnabrueck.de.

[CGM] Laura Colmenarejo, Francesco Galuppi, and Mateusz Michałek. Code available at L. Colmenarejo's webpage. https://sites.google.com/view/l-colmenarejo/publications.

[Che54] Kuo-Tsai Chen. Iterated integrals and exponential homomorphism. *Proceedings of London Mathematical Society*, 4:502–512, 1954.

[Che57] Kuo-Tsai Chen. Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Ann. of Math. (2)*, 65:163–178, 1957.

[Che58] Kuo-Tsai Chen. Integration of paths - a faithful representation of paths by noncommutative formal power series. *Transactions Amer. Math. Soc.*, 89:95–407, 1958.

[CK16] Ilya Chevyrev and Andrey Kormilitzin. A primer on the signature method in machine learning, 2016.

[CLS11] David A. Cox, John B. Little, and Henry K. Schenck. *Toric varieties*, volume 124 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2011.

[CNO18] Ilya Chevyrev, Vidit Nanda, and Harald Oberhauser. Persistence paths and signature features in topological data analysis, 2018.

[CO18] Ilya Chevyrev and Harald Oberhauser. Signature moments to characterize laws of stochastic processes, 2018.

[CP18] Laura Colmenarejo and Rosa Preiß. Signatures of paths transformed by polynomial maps, 2018.

[DR18] Joscha Diehl and Jeremy Reizenstein. Invariants of multidimensional time series based on their iterated-integral signature. *CoRR*, abs/1801.06104, 2018.

[FGL15] Peter Friz, Paul Gassiat, and Terry Lyons. Physical Brownian motion in a magnetic field as a rough path. *Trans. Amer. Math. Soc.*, 367(11):7939–7955, 2015.

[FH14] Peter Friz and Martin Hairer. *A course on rough paths*. Springer, 2014.

[Ful93] William Fulton. *Introduction to toric varieties*, volume 131 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1993. The William H. Roever Lectures in Geometry.

[FV10] Peter Friz and Nicolas Victoir. *Multidimensional stochastic processes as Rough Paths. Theory and applications.* Cambridge University Press, 2010.

[Gal18] Francesco Galuppi. The rough veronese variety, 2018.

[GLKF13] Lajos Gergely Gyurkó, Terry Lyons, Mark Kontkowski, and Jonathan Field. Extracting information from the signature of a financial data stream, 2013.

[Hai02] Richard Hain. *Iterated Integrals and Algebraic Cycles: Examples and Prospects.*, chapter 4, pages 55–118. NJ: World Scientific Publishing, 2002.

[KMM17] Lukas Katthän, Mateusz Michałek, and Ezra Miller. When is a polynomial ideal binomial after an ambient automorphism? *Foundations of Computational Mathematics*, pages 1–23, 2017.

[Laz04] Robert Lazarsfeld. *Positivity in algebraic geometry. I and II.* Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]. Springer-Verlag, Berlin, 2004. Classical setting: line bundles and linear series.

[LCL07] Terry J. Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*, volume 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004, With an introduction concerning the Summer School by Jean Picard.

[LS05] Terry Lyons and Nadia Sidorova. Sound compression - a rough path approach. *Proceedings of the 4th International Symposium on Information and Communication Technologies, Cape Town*, pages 223–229, 2005.

[PSS18] M. Pfeffer, A. Seigal, and B. Sturmfels. Learning paths from signature tensors, 2018. arXiv:1809.01588, to appear in SIAM Journal on Matrix Analysis and Applications.

[Reu93] Christophe Reutenauer. *Free Lie algebras.* Oxford University Press, 1993.

[RS62] J. Barkley Rosser and Lowell Schoenfeld. Approximate formulas for some functions of prime numbers. *Illinois J. Math.*, 6:64–94, 1962.

[Stu96] Bernd Sturmfels. *Gröbner bases and convex polytopes*, volume 8 of *University Lecture Series.* American Mathematical Society, Providence, RI, 1996.

[The18] The Sage Developers. *SageMath, the Sage Mathematics Software System (Version x.y.z)*, 2018. https://www.sagemath.org.