

## LEARNING PATHS FROM SIGNATURE TENSORS\*

MAX PFEFFER<sup>†</sup>, ANNA SEIGAL<sup>‡</sup>, AND BERND STURMFELS<sup>§</sup>

**Abstract.** Matrix congruence extends naturally to the setting of tensors. We apply methods from tensor decomposition, algebraic geometry, and numerical optimization to this group action. Given a tensor in the orbit of another tensor, we compute a matrix which transforms one to the other. Our primary application is an inverse problem from stochastic analysis: the recovery of paths from their third order signature tensors. We establish identifiability results, both exact and numerical, for piecewise linear paths, polynomial paths, and generic dictionaries. Numerical optimization is applied for recovery from inexact data. We also compute the shortest path with a given signature tensor.

**Key words.** signature tensors, congruence action, tensor decomposition, identifiability, inverse problems, optimization

**AMS subject classifications.** 14Q15, 15A72, 65K10

**DOI.** 10.1137/18M1212331

**1. Introduction.** In many areas of applied mathematics, tensors are used to encode features of geometric data. The tensors then serve as the input to algorithms aimed at classifying and understanding the original data. This set-up comes with a natural inverse problem, namely, to recover the geometric objects from the tensors that represent them. The aim of this article is to solve this inverse problem.

Our motivation comes from the *signature method in machine learning* [8]. In this setting, the geometric object is a path  $[0, 1] \rightarrow \mathbb{R}^d$ . The path is encoded by its signature, an infinite sequence of tensors that are interrelated through a Lie algebra structure. Signature tensors were introduced by Chen [7], and they play an important role in stochastic analysis [12, 22]. We refer to [23, 24] for the recovery problem and to [10, 13, 17, 18, 21] for algorithms and applications. Our point of departure is the approach to signature tensors via algebraic geometry that was proposed in [1].

The problem we address is path recovery from the *signature tensor of order three*. This tensor is the third term in the signature sequence. Higher order signature tensors encode finer representations of a path than lower order signatures: if two paths (which are not loops) agree at the signature tensor of some order, then they agree up to scale at all lower orders [1, section 6]. We focus on order three because, like in many similar contexts [16], tensors under the congruence action have useful uniqueness properties that do not hold for matrices. The full space of paths is too detailed for meaningful recovery from finitely many numbers. As is often done [23], we restrict to paths which lie in a particular family. We consider paths whose coordinates can be written as linear combinations of functions in a fixed *dictionary*. The dictionary determines a core tensor, which is transformed by the congruence action into signatures of paths in the family. The family of piecewise linear paths is a main example.

\*Received by the editors September 7, 2018; accepted for publication (in revised form) by L. De Lathauwer January 8, 2019; published electronically April 3, 2019.

<http://www.siam.org/journals/simax/40-2/M121233.html>

**Funding:** The work of the second and third authors was partially supported by the National Science Foundation (DMS-1419018).

<sup>†</sup>MPI-MiS Leipzig, 04103 Leipzig, Germany (max.pfeffer@mis.mpg.de).

<sup>‡</sup>Department of Mathematics, University of California, Berkeley, CA 94720 (seigal@berkeley.edu).

<sup>§</sup>MPI-MiS Leipzig, 04103 Leipzig, Germany and Department of Mathematics, University of California, Berkeley, CA, 94720 (bernd@math.berkeley.edu).

This article is organized as follows. In section 2 we describe our set-up which emphasizes the notion of a dictionary to describe a family of paths. Our main contributions begin in section 3, where we investigate tensors under the congruence action by matrices. We obtain necessary and sufficient conditions for the size of the stabilizer under this group action. This leads to conditions under which a path can be recovered uniquely, or up to a finite list of choices, from the third order signature tensor, in section 4. We then apply these conditions to give identifiability results for generic dictionaries, in section 5, and piecewise linear paths, in section 6, where our results prove part of [1, Conjecture 6.10]. In section 7 we study numerical identifiability for the recovery of paths from signature data. Both upper bounds and lower bounds are given for the numerical nonidentifiability, the scaled inverse distance to the set of instances where the recovery problem is ill-posed. In section 8 we turn to numerical optimization and we present experimental results on unique recovery of low-complexity paths. Section 9 addresses the problem of finding the shortest path with given third signature tensor.

**2. Dictionaries and their core tensors.** We fix a *dictionary*  $\psi = (\psi_1, \psi_2, \dots, \psi_m)$  of piecewise differentiable functions  $\psi_i : [0, 1] \rightarrow \mathbb{R}$ . The dictionary corresponds to a path in  $\mathbb{R}^m$ , also denoted  $\psi$ , whose  $i$ th coordinate is  $\psi_i$ . The path  $\psi$  is regarded as a fixed reference path in  $\mathbb{R}^m$ . Its signature is a formal series of tensors

$$\sigma(\psi) = \sum_{k=1}^{\infty} \sigma^{(k)}(\psi),$$

whose  $k$ th term is a tensor in  $(\mathbb{R}^m)^{\otimes k}$  with entries that are iterated integrals of  $\psi$ :

$$(2.1) \quad (\sigma^{(k)}(\psi))_{i_1 i_2 \dots i_k} = \int_0^1 \cdots \left( \int_0^{t_3} \left( \int_0^{t_2} d\psi_{i_1}(t_1) \right) d\psi_{i_2}(t_2) \right) \cdots d\psi_{i_k}(t_k).$$

Evaluating (2.1) for  $k = 1$  shows that the first signature  $\sigma^{(1)}(\psi)$  is the vector  $\psi(1) - \psi(0)$ . The second signature  $\sigma^{(2)}(\psi)$  is the matrix  $\frac{1}{2}(\psi(1) - \psi(0))^{\otimes 2} + Q$ , where  $Q$  is skew-symmetric. Its entry  $q_{ij}$  is the *Lévy area* of the projection of  $\psi$  onto the plane indexed by  $i$  and  $j$ , the signed area between the planar path and the segment connecting its endpoints. For background on signature tensors of paths and their applications see [1, 8, 10, 12, 22, 23, 24].

This article is based on the following two premises:

- (a) We study the images of a fixed reference path  $\psi$  under linear maps.
- (b) We focus on the third order signature  $\sigma^{(3)}(\psi)$ .

We first discuss premise (a). Consider a linear map  $\mathbb{R}^m \rightarrow \mathbb{R}^d$  given by a  $d \times m$  matrix  $X = (x_{ij})$ . The image of the path  $\psi$  under  $X$  is the path  $X\psi : [0, 1] \rightarrow \mathbb{R}^d$  given by

$$t \mapsto \left( \sum_{j=1}^m x_{1j} \psi_j(t), \sum_{j=1}^m x_{2j} \psi_j(t), \dots, \sum_{j=1}^m x_{dj} \psi_j(t) \right).$$

The following key lemma relates the linear transformation of a path to the induced linear transformation of its signature tensor. The proof follows directly from the iterated integrals in (2.1), bearing in mind that integration is a linear operation.

LEMMA 2.1. *The signature map is equivariant under linear transformations, i.e.,*

$$(2.2) \quad \sigma(X\psi) = X(\sigma(\psi)).$$

The action of the linear map  $X$  on the signature  $\sigma(\psi)$  is as follows. The  $k$ th order signature of  $\psi$  is a tensor in  $(\mathbb{R}^m)^{\otimes k}$ . We multiply this tensor on all  $k$  sides by the  $d \times m$  matrix  $X$ . The result of this tensor-matrix product, which is known as multilinear multiplication, is a tensor in  $(\mathbb{R}^d)^{\otimes k}$ . Using notation from the theory of tensor decomposition [16], the identity (2.2) can be written as

$$(2.3) \quad \sigma^{(k)}(X\psi) = \llbracket \sigma^{(k)}(\psi); X, X, \dots, X \rrbracket \quad \text{for } k = 1, 2, 3, \dots$$

For  $k = 1$  this is the matrix-vector product  $\sigma^{(1)}(X\psi) = X \cdot \sigma^{(1)}(\psi)$ . For  $k = 2$  the rectangular matrix  $X$  acts on the square signature matrix via the congruence action:

$$\sigma^{(2)}(X\psi) = X \cdot \sigma^{(2)}(\psi) \cdot X^T.$$

Lemma 2.1 means that, once the signature of the dictionary  $\psi$  is known, integrals no longer need to be computed. Signature tensors of a path arising from  $\psi$  by a linear transformation are obtained by tensor-matrix multiplication. This works for many useful families of paths.

We next justify our premise (b). For a path in  $\mathbb{R}^d$ , the number of entries in the  $k$ th signature tensor is  $d^k$ . For  $k \geq 4$ , this quickly becomes prohibitive. But  $k = 2$  is too small: signature matrices do not contain enough information to recover paths in a meaningful way. Even after fixing all  $\binom{d}{2}$  Lévy areas, there are too many paths between two points in  $\mathbb{R}^d$ . The third signature is the right compromise. The number  $d^3$  of entries is reasonable, paths are identifiable from third signatures under certain conditions, and we propose practical algorithms for path recovery. This paper establishes the last two points, assuming premise (a).

With our two premises in mind, we give more detail for the case of third signature tensors,  $k = 3$ . We fix a dictionary  $\psi$  consisting of  $m$  functions. We refer to its third signature  $C_\psi = \sigma^{(3)}(\psi) \in \mathbb{R}^{m \times m \times m}$  as the *core tensor* of  $\psi$ . This tensor has entries

$$(2.4) \quad c_{ijk} = \int_0^1 \left( \int_0^{t_3} \left( \int_0^{t_2} d\psi_i(t_1) \right) d\psi_j(t_2) \right) d\psi_k(t_3) \quad \text{for all } 1 \leq i, j, k \leq m.$$

The first and second signatures of a real path are determined by the third signature, provided the path is not a loop, just as any lower order signature tensor can be recovered up to scale from higher order signatures. This follows from the *shuffle relations* [1, Lemma 4.2]. Writing  $c_i$  and  $c_{ij}$  for the entries of the first and second signatures, respectively, we have the identities

$$(2.5) \quad c_i c_j = c_{ij} + c_{ji} \quad \text{and} \quad c_i c_{jk} = c_{ijk} + c_{jik} + c_{jki}.$$

Given a  $d \times m$  matrix  $X = (x_{ij})$ , the third signature of the image path  $X\psi$  in  $\mathbb{R}^d$  is denoted by  $\sigma^{(3)}(X)$ , as shorthand for  $\sigma^{(3)}(X\psi)$ . Following (2.3), this  $d \times d \times d$  tensor is obtained from  $C_\psi = (c_{ijk})$  by multiplying by  $X$  on each side. The entry of  $\sigma^{(3)}(X)$  in position  $(\alpha, \beta, \gamma)$  is

$$(2.6) \quad \llbracket C_\psi; X, X, X \rrbracket_{\alpha\beta\gamma} = \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m c_{ijk} x_{\alpha i} x_{\beta j} x_{\gamma k}.$$

The expression (2.6) for the signature tensor in terms of the core tensor is closely related to the *Tucker decomposition* [16, 29] that arises frequently in tensor compression. Our application differs from the usual setting in that the core tensor has fixed size and fixed entries. Furthermore, we multiply each side by the same matrix.

We next discuss some specific dictionaries and the paths they encode, starting with two dictionaries studied in [1]. The first dictionary is  $\psi(t) = (t, t^2, \dots, t^m)$ . Multiplying this dictionary by matrices  $X$  of size  $d \times m$  gives all *polynomial paths* of degree at most  $m$  that start at the origin in  $\mathbb{R}^d$ . The core tensor of  $\psi$  is denoted by  $C_{\text{mono}}$  to indicate the monomials  $t^i$ . By [1, Example 2.2], its entries are

$$(2.7) \quad c_{ijk} = \frac{j}{i+j} \cdot \frac{k}{i+j+k}.$$

Our second dictionary comes from an axis path in  $\mathbb{R}^m$ . It encodes all *piecewise linear paths* with  $\leq m$  steps. The  $i$ th entry in the dictionary is the piecewise linear basis function

$$(2.8) \quad \psi_i(t) = \begin{cases} 0 & \text{if } t \leq \frac{i-1}{m}, \\ mt - (i-1) & \text{if } \frac{i-1}{m} < t < \frac{i}{m}, \\ 1 & \text{if } t \geq \frac{i}{m}. \end{cases}$$

By [1, Example 2.1], the associated core tensor  $C_{\text{axis}}$  is “upper-triangular,” namely,

$$(2.9) \quad c_{ijk} = \begin{cases} 1 & \text{if } i < j < k, \\ \frac{1}{2} & \text{if } i < j = k \text{ or } i = j < k, \\ \frac{1}{6} & \text{if } i = j = k, \\ 0 & \text{otherwise.} \end{cases}$$

The tensors  $C_{\text{mono}}$  and  $C_{\text{axis}}$  are real points in the *universal variety*  $\mathcal{U}_{m,3} \subset (\mathbb{C}^m)^{\otimes 3}$ . This consists of all third order signatures of paths in  $\mathbb{C}^m$ , or equivalently all core tensors of dictionaries of size  $m$ . At present, we do not know whether all real points in  $\mathcal{U}_{m,3}$  are in the topological closure of the signature tensors of real paths.

**PROPOSITION 2.2.** *The variety  $\mathcal{U}_{m,3}$  is irreducible of dimension  $\frac{1}{3}m^3 + \frac{1}{2}m^2 + \frac{1}{6}m$ .*

*Proof.* This follows from [1, Theorem 6.1]. Note that, in the present paper,  $\mathcal{U}_{m,3}$  denotes the affine variety, whereas [1] refers to projective varieties. The dimension of  $\mathcal{U}_{m,3}$  is the number of Lyndon words (see [28]) on  $m$  letters of length 1, 2, or 3. These three numbers are  $m$ ,  $\binom{m}{2}$ , and  $\frac{1}{3}(m^3 - m)$ . Their sum equals  $\frac{1}{3}m^3 + \frac{1}{2}m^2 + \frac{1}{6}m$ .  $\square$

The polynomials that define  $\mathcal{U}_{m,3}$  are obtained by eliminating the unknowns  $c_i$  and  $c_{ij}$  from the equations (2.5). We provide more details at the end of section 5.

**Example 2.3** (generic dictionaries). We describe a method for sampling real points in the universal variety  $\mathcal{U}_{m,3}$ , assuming [1, Conjecture 6.10]. Pick  $M$  random vectors  $Y_1, Y_2, \dots, Y_M$  in  $\mathbb{R}^m$ , where  $M$  exceeds  $\frac{1}{3}m^3 + \frac{1}{2}m^2 + \frac{1}{6}m = \dim(\mathcal{U}_{m,3})$ , and take the piecewise linear path with steps  $Y_1, Y_2, \dots, Y_M$ . By [1, Example 5.4], the resulting *generic core tensor* equals

$$(2.10) \quad C_{\text{gen}} = \frac{1}{6} \cdot \sum_{i=1}^M Y_i^{\otimes 3} + \frac{1}{2} \cdot \sum_{1 \leq i < j \leq M} (Y_i^{\otimes 2} \otimes Y_j + Y_i \otimes Y_j^{\otimes 2}) + \sum_{1 \leq i < j < k \leq M} Y_i \otimes Y_j \otimes Y_k.$$

The coefficients in (2.10) match the tensor entries in (2.9). By Chen’s formula [1, equation (38)], the signature tensor  $C_{\text{gen}}$  is the degree 3 component in the tensor series  $\sigma(\psi) = \exp(Y_1) \otimes \exp(Y_2) \otimes \dots \otimes \exp(Y_M)$ , where  $\exp(Y_i) = \sum_{k=0}^{\infty} \frac{1}{k!} Y_i^{\otimes k}$ .

An alternative method for sampling from  $\mathcal{U}_{m,3}$  uses the Gröbner basis in [1, Theorem 4.10]. We write  $\sigma_{\text{lyndon}}$  for the vector of all signatures  $\sigma_i$ ,  $\sigma_{ij}$ , and  $\sigma_{ijk}$  whose indices are Lyndon words. This includes all  $m$  first order signatures  $\sigma_i$ , all

$\binom{m}{2}$  second order signatures  $\sigma_{ij}$  with  $i < j$ , and all  $\frac{1}{3}(m^3 - m)$  third order signatures  $\sigma_{ijk}$  satisfying  $i < \min(j, k)$  or  $i = j < k$ . We pick these  $m + \binom{m}{2} + \frac{1}{3}(m^3 - m)$  signature values to be random real numbers and substitute these numbers into the vector  $\sigma_{\text{lyndon}}$ . The non-Lyndon signatures  $\sigma_{ijk}$  are then computed by evaluating  $\phi_{ijk}(\sigma_{\text{lyndon}})$ , where  $\phi_{ijk}$  is the normal form polynomial in [1, Theorem 4.10].

We now define what we mean by “learning paths” in the title of this paper. Let  $C$  be a fixed core tensor of format  $m \times m \times m$ , such as  $C_{\text{axis}}$ ,  $C_{\text{mono}}$  or  $C_{\text{gen}}$ . Our data is a  $d \times d \times d$  tensor  $S = (s_{ijk})$  that is the third signature of some path in  $\mathbb{R}^d$ . Our hypothesis is that the path can be represented by the dictionary  $\psi$ , i.e., it is the image of  $\psi$  under a linear map. We seek a  $d \times m$  matrix  $X = (x_{ij})$  that satisfies  $S = \sigma^{(3)}(X)$ . In other words, given  $C$  and  $S$ , we wish to solve the tensor equation

$$\llbracket C; X, X, X \rrbracket = S.$$

This is the system of  $d^3$  cubic equations in  $md$  unknowns  $x_{ij}$

$$(2.11) \quad \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m c_{ijk} x_{\alpha i} x_{\beta j} x_{\gamma k} = s_{\alpha\beta\gamma} \quad \text{for } 1 \leq \alpha, \beta, \gamma \leq d.$$

The system (2.11) has a solution  $X$  if and only if the dictionary with core tensor  $C$  admits a path with signature tensor  $S$ . For the dictionaries we consider, the solution  $X$  is conjectured to be unique among real matrices  $X$  provided  $m < \frac{1}{3}d^2 + \frac{1}{2}d + \frac{1}{6}$  and unique up to scaling by a third root of unity if we allow complex matrices  $X$ . The inequality means that the dimension of the universal variety  $\mathcal{U}_{d,3}$  exceeds the number  $md$  of unknowns, which is necessary for identifiability. For piecewise linear and polynomial paths, this is presented in Conjecture 6.12 and Lemma 6.16 of [1].

We experimented with *Gröbner bases* for solving the equations  $\llbracket C; X, X, X \rrbracket = S$ , thereby extending the computations in [1, section 6]. Table 1 summarizes our findings for various values of  $d$  and  $m$ . The number  $md$  is shown with lower index  $d^3$ . After the hyphen we report timings (in seconds) for the following computation. We pick a  $d \times m$  matrix  $X_0$  with random integer entries, sampled uniformly between  $-15$  and  $15$ , and we consider the system of equations  $\llbracket C; X, X, X \rrbracket = \llbracket C; X_0, X_0, X_0 \rrbracket$ . We compute two Gröbner bases, the first for  $C = C_{\text{axis}}$  and the second for  $C = C_{\text{mono}}$ . We did this experiment in Maple 16, using the command **Basis** with the default order (degree reverse lexicographic) in the **Groebner** package. The results are shown in Table 1. In all cases for which the computation succeeded, there were three standard monomials, corresponding to the matrices  $\eta X_0$ , where  $\eta^3 = 1$ . This confirms rational identifiability for the core tensor.

TABLE 1

*Using Gröbner bases to recover a path from its third signature. The first two numbers in each box are the size of the problem: the first number counts the unknowns; its index counts the number of equations. The next two numbers are timings for Gröbner basis computations in Maple 16. The first entry is for  $C_{\text{axis}}$ , the second for  $C_{\text{mono}}$ . The units are seconds, rounded down. An entry NI means that the model is not identifiable, while F means that the computation failed to terminate (after allowing 5 hours, on a computer with a CPU clock speed of 2.3GHz).*

$m \backslash d$	2	3	4	5	6	7
2	4 <sub>8</sub> – 0, 0	6 <sub>27</sub> – 0, 0	8 <sub>64</sub> – 0, 0	10 <sub>125</sub> – 0, 0	12 <sub>216</sub> – 0, 1	14 <sub>343</sub> – 0, 1
3	6 <sub>8</sub> – NI	9 <sub>27</sub> – 0, 0	12 <sub>64</sub> – 0, 0	15 <sub>125</sub> – 0, 1	18 <sub>216</sub> – 0, 1	21 <sub>343</sub> – 1, 3
4	8 <sub>8</sub> – NI	12 <sub>27</sub> – 0, 0	16 <sub>64</sub> – 0, 1	20 <sub>125</sub> – 1, 1	24 <sub>216</sub> – 2, 4	28 <sub>343</sub> – 5, 9
5	10 <sub>8</sub> – NI	15 <sub>27</sub> – NI	20 <sub>64</sub> – 4, 22	25 <sub>125</sub> – 16, 43	30 <sub>216</sub> – 72, 188	35 <sub>343</sub> – 601, F

We conclude this section by mentioning two group actions, closely related to ours, which have been studied extensively. The first concerns homogeneous polynomials  $f(x_1, \dots, x_m)$ . Matrices  $Z$  in  $\text{GL}(m, \mathbb{C})$  act by linear change of variables  $f(x) \mapsto f(Z \cdot x)$ . This is precisely our congruence action  $C \mapsto \llbracket C; Z, Z, Z \rrbracket$  in the special case where  $C$  is a *symmetric* tensor that corresponds to the cubic polynomial

$$f(x_1, \dots, x_m) = \sum_{i,j,k=1}^m c_{ijk} x_i x_j x_k.$$

Another well-studied action (cf. [6, 20]) concerns tensors of any size  $m_1 \times \dots \times m_k$ . The group  $\text{GL}(m_1, \mathbb{C}) \times \dots \times \text{GL}(m_k, \mathbb{C})$  acts via  $C \mapsto \llbracket C; Z_1, \dots, Z_k \rrbracket$ , where  $Z_i \in \text{GL}(m_i, \mathbb{C})$ . For  $k = 3$  this is  $C \mapsto \llbracket C; Z_1, Z_2, Z_3 \rrbracket$ . Our action is the restriction to the diagonal  $Z := Z_1 = Z_2 = Z_3$ . There is a literature on the above two group actions, but much less on the *congruence action* of  $\text{GL}(m, \mathbb{C})$  on  $(\mathbb{C}^m)^{\otimes k}$  which is needed here.

**3. Stabilizers under congruence.** From now on, the letter  $\mathbb{K}$  denotes a field, usually either the real numbers  $\mathbb{R}$  or the complex numbers  $\mathbb{C}$ . We study the congruence action of invertible matrices  $X \in \mathbb{K}^{m \times m}$  on the space of tensors  $T \in (\mathbb{K}^m)^{\otimes k}$  via

$$T \mapsto \llbracket T; X, X, \dots, X \rrbracket.$$

Writing  $X = (x_{ij})$ ,  $T = (t_{\alpha_1 \dots \alpha_k})$ , the entries of the transformed tensor are

$$\llbracket T; X, X, \dots, X \rrbracket_{\beta_1 \dots \beta_k} = \sum_{\alpha_1, \dots, \alpha_k} t_{\alpha_1 \dots \alpha_k} x_{\beta_1 \alpha_1} x_{\beta_2 \alpha_2} \dots x_{\beta_k \alpha_k}.$$

This tensor is the image of  $T$  under the congruence action by  $X$ . The *stabilizer* of  $T$  under the group action is the subgroup of matrices  $X$  in  $\text{GL}(m, \mathbb{K})$  that satisfy  $\llbracket T; X, X, \dots, X \rrbracket = T$ . We denote it by  $\text{Stab}_{\mathbb{K}}(T)$ . The stabilizer is defined by a system of polynomial equations of degree  $k$  in the entries of  $X$ . Matrices  $\eta I$  with  $\eta^k = 1$  are always among the solutions.

Section 4 will relate the stabilizer of  $T$  under the congruence action to the identifiability of path recovery within the family of paths whose dictionary has signature tensor  $T$ . It is an open problem to characterize tensors  $T$  in  $(\mathbb{K}^m)^{\otimes k}$  whose stabilizer under congruence is *nontrivial*, i.e., for which  $\text{Stab}_{\mathbb{K}}(T)$  strictly contains  $\{\eta I : \eta^k = 1\}$ .

We introduce an important notion for stabilizers under congruence, which we call *symmetrically concise*. It means that, for  $T \in (\mathbb{K}^m)^{\otimes k}$ , there is no subspace  $W \subsetneq \mathbb{K}^m$  such that  $T \in W^{\otimes k}$ .

We can define symmetrically concise in terms of flattenings. The tensor  $T$  has  $m^k$  entries and  $k$  principal flattenings, matrices of size  $m \times m^{k-1}$ . The  $i$ th flattening  $T^{(i)}$  has rows labeled by the  $i$ th index of  $T$  and its columns labeled by a multi-index from all remaining indices [16, 20]. Flattenings are also known as matricizations. We recall from [30] that a tensor  $T \in (\mathbb{K}^m)^{\otimes k}$  is *concise* if it has flattening ranks  $(m, m, \dots, m)$ . We concatenate the  $k$  flattening matrices to form a single matrix of size  $m \times km^{k-1}$ . The tensor is symmetrically concise if this matrix has full rank  $m$ .

For symmetric tensors concise and symmetrically concise are equivalent, because the  $m \times km^{k-1}$  matrix consists of  $k$  identical blocks of size  $m \times m^{k-1}$ . However, symmetrically concise is weaker than concise for nonsymmetric tensors. For instance, the  $3 \times 3 \times 3$  basis tensor  $T = e_1 \otimes e_2 \otimes e_3$  is symmetrically concise but not concise: there exist subspaces  $W_i \subsetneq \mathbb{K}^3$  with  $T \in W_1 \otimes W_2 \otimes W_3$ , but we cannot find the same subspace  $W \subsetneq \mathbb{K}^3$  across all modes such that  $T \in W^{\otimes 3}$ .

We can also define symmetrically concise from a decomposition into rank one terms. A tensor  $T \in (\mathbb{K}^m)^{\otimes k}$  is *rank one* if  $T = v^{(1)} \otimes v^{(2)} \otimes \cdots \otimes v^{(k)}$  for some nonzero vectors  $v^{(j)} \in \mathbb{K}^m$ . The *rank* of  $T$  (over  $\mathbb{K}$ ) is the minimal number of terms in an expression for  $T$  as a sum of rank one tensors,  $T = \sum_{l=1}^r T_l$  for  $T_l = v_l^{(1)} \otimes v_l^{(2)} \otimes \cdots \otimes v_l^{(k)}$  with  $v_l^{(j)} \in \mathbb{K}^m$ . We call a decomposition of  $T$  of minimal length a *minimal decomposition*. A tensor is symmetrically concise if the  $kr$  vectors  $v_l^{(j)}$  in any minimal decomposition span the ambient space  $\mathbb{K}^m$  (by Lemma 4.1).

**PROPOSITION 3.1.** *Let  $T \in (\mathbb{K}^m)^{\otimes k}$  be a tensor that is not symmetrically concise. Then the stabilizer of  $T$  under the congruence action is nontrivial.*

*Proof.* Since  $T$  is not symmetrically concise, there exists a vector  $v \in \mathbb{K}^m$  of norm one such that  $v^\top T^{(i)} = 0$  for all  $i = 1, \dots, k$ . This condition implies  $\llbracket T; I + vv^\top, I + vv^\top, \dots, I + vv^\top \rrbracket = T$ . Hence the invertible matrix  $I + vv^\top$  is in the stabilizer of  $T$ .  $\square$

Tensors with trivial stabilizer are symmetrically concise but not always concise.

**Example 3.2.** Consider the rank one tensor  $T = e_1 \otimes e_2 \otimes (e_1 + e_2)$ . Each  $2 \times 4$  flattening matrix of  $T$  is rank-deficient. This means that  $T$  has flattening ranks  $(1, 1, 1)$ , so  $T$  is not concise. However, the  $2 \times 12$  matrix we obtain by concatenating the three flattening matrices has full rank, hence the tensor  $T$  is symmetrically concise. The stabilizer of  $T$  is directly computed to be trivial.

We next derive a Jacobian criterion which gives a sufficient condition for the stabilizer of a tensor under the congruence action to be finite. For notational simplicity we state the criterion only for order three tensors. The Jacobian  $\nabla f(X) \in \mathbb{K}^{m^3 \times m^2}$  of the function  $f(X) = \llbracket T; X, X, X \rrbracket$  has entries

$$\nabla f(X)_{(i,j,k),(u,v)} = \frac{\partial f_{ijk}}{\partial x_{uv}} = \sum_{\alpha, \beta} (t_{v\alpha\beta} \delta_{ui} x_{j\alpha} x_{k\beta} + t_{\alpha v\beta} \delta_{uj} x_{i\alpha} x_{k\beta} + t_{\alpha\beta v} \delta_{uk} x_{i\alpha} x_{j\beta}),$$

where  $\delta_{ij}$  is the Kronecker delta. The entries of the Jacobian at  $X = I$  are

$$(3.1) \quad \nabla f(I)_{(i,j,k),(u,v)} = t_{vj k} \delta_{ui} + t_{iv k} \delta_{uj} + t_{ij v} \delta_{uk}.$$

Consider the  $m^2 \times m^2$  submatrix  $J_1$  of the Jacobian obtained by setting  $k = 1$  in (3.1). The entry of  $J_1$  in row  $(i, j)$  and column  $(u, v)$  is the linear form

$$J_1((i, j), (u, v)) = \delta_{ui} t_{vj1} + \delta_{uj} t_{iv1} + \delta_{u1} t_{ijv}.$$

**PROPOSITION 3.3.** *Let  $T$  be a tensor whose  $m^2 \times m^2$  matrix  $J_1$  as above is invertible. Then the stabilizer of  $T$  under the congruence action by  $\mathrm{GL}(m, \mathbb{K})$  is finite.*

*Proof.* The stabilizer of  $T$  under congruence is infinite when the map  $f : Z \mapsto \llbracket T; Z, Z, Z \rrbracket$  has positive-dimensional fibers. If the matrix  $J_1$  is invertible, then the Jacobian  $\nabla f$  has full rank at  $Z = I$ . This implies that a connected component of the stabilizer consists of the single matrix  $I$ . Consider another connected component of the stabilizer, containing a matrix  $X$ . Applying  $X^{-1}$  to the component gives a connected component of the stabilizer containing  $I$ , which therefore must be the single matrix  $I$ . Hence all connected components are zero-dimensional, and the stabilizer is finite.  $\square$

The same conclusion holds if any of the maximal minors of the Jacobian in  $\mathbb{K}^{m^k \times m^2}$  is nonzero. For a fixed tensor  $T$ , of small enough size,  $\det(J_1)$  can be computed. Consider the tensor  $C_{\text{mono}}$  with entries (2.7) for  $m = 10$ . We compute in Maple  $\det(J_1)^{-1} = 2^{288} 3^{160} 5^{81} 7^{75} 11^{96} 13^{86} 17^{52} 19^{35}$ , and Proposition 3.3 allows us to

conclude that the stabilizer under congruence is finite. Repeating this computation for all  $m \leq 30$ , we observe that polynomial paths are algebraically identifiable from their third signature tensors when  $m \leq 30$  and  $m \leq d$ . This doubles the bound  $m = 15$  from [1, Lemma 6.16], contributing progress toward the proof of [1, Conjecture 6.10]. We will study the numerical identifiability of such paths in section 7.

**4. Criteria for identifiability.** In this section, we relate the stabilizer of a core tensor  $C$  under congruence to the identifiability of path recovery from the signature tensor  $\llbracket C; X, X, \dots, X \rrbracket$ , where  $X$  is a rectangular matrix in  $\mathbb{K}^{d \times m}$ . We focus on the case  $m \leq d$ , when the dictionary size is smaller than the dimension of the ambient space. We shall argue that to study the identifiability of paths from their signatures, it suffices to study identifiability of  $C$  under congruence. The signature tensor is *identifiable* if the matrix  $X$  can be recovered up to scale. The signature tensor is *algebraically identifiable* if  $X$  can be recovered up to a finite list of choices. Finally, it is *rationally identifiable* if, up to scale,  $X$  can be written as a rational function in  $C$ .

The following result compares minimal decompositions of smaller tensors with those of larger tensors in which they appear as a block. Any decomposition of the larger tensor is obtained from a decomposition of the smaller tensor by adding zeros.

**LEMMA 4.1.** *Let  $T \in (\mathbb{K}^d)^{\otimes k}$  be a tensor with all entries zero outside a block of size  $m \times m \times \dots \times m$  for some  $m \leq d$ . Then any rank one term in a minimal decomposition of  $T$  is also zero outside of the block.*

*Proof.* Let  $T = \sum_{l=1}^r v_l^{(1)} \otimes v_l^{(2)} \otimes \dots \otimes v_l^{(k)}$  be a minimal decomposition. Assume that a rank one term is nonzero outside of the block, i.e., the coordinate  $v_l^{(j)}(\alpha)$  is nonzero for some  $l$ , some  $j$ , and some index  $\alpha$  not contained in the block. The terms  $v_l^{(1)} \otimes \dots \otimes v_l^{(j)}(\alpha) \otimes \dots \otimes v_l^{(k)}$  sum to zero. However, the order  $k-1$  tensors in a minimal decomposition, resulting from removing the  $j$ th vector from each rank one term, are linearly independent. This implies  $v_l^{(j)}(\alpha) = 0$  for all  $l$ , a contradiction.  $\square$

We note that Lemma 4.1 also appears as [20, Proposition 3.1.3.1]. We use this lemma to relate the stabilizer of the core tensor  $C$  under the congruence action to the set of paths with the same signature tensor  $\llbracket C; X, X, \dots, X \rrbracket$ .

**THEOREM 4.2.** *Fix a symmetrically concise tensor  $C$  in  $(\mathbb{K}^m)^{\otimes k}$ . Let  $\text{Stab}_{\mathbb{K}}(C)$  be its stabilizer under the congruence action by  $\text{GL}(m, \mathbb{K})$ . For any matrix  $X \in \mathbb{K}^{d \times m}$  of rank  $m \leq d$ , we have*

$$(4.1) \quad \{Y \in \mathbb{K}^{d \times m} : \llbracket C; X, X, \dots, X \rrbracket = \llbracket C; Y, Y, \dots, Y \rrbracket\} = \{XZ : Z \in \text{Stab}_{\mathbb{K}}(C)\}.$$

*Proof.* Suppose  $\llbracket C; X, \dots, X \rrbracket = \llbracket C; Y, \dots, Y \rrbracket$ . Let  $\tilde{C}$  be the  $d \times d \times \dots \times d$  tensor with entries

$$\tilde{c}_{i_1 \dots i_k} = \begin{cases} c_{i_1 \dots i_k} & \text{if } 1 \leq i_1, \dots, i_k \leq m, \\ 0 & \text{otherwise.} \end{cases}$$

Let  $\tilde{X}$  be an invertible  $d \times d$  matrix whose first  $m$  columns are  $X$ , and likewise construct  $\tilde{Y}$ . Then  $\llbracket \tilde{C}; \tilde{X}, \dots, \tilde{X} \rrbracket = \llbracket \tilde{C}; \tilde{Y}, \dots, \tilde{Y} \rrbracket$ . We multiply by  $\tilde{X}^{-1}$  to get  $\tilde{C} = \llbracket \tilde{C}; \tilde{Z}, \tilde{Z}, \dots, \tilde{Z} \rrbracket$  where  $\tilde{Z} = \tilde{X}^{-1}\tilde{Y}$  and the top-left  $m \times m$  block of  $\tilde{Z}$ , denoted  $Z$ , satisfies  $\llbracket C; Z, Z, \dots, Z \rrbracket = C$ .

Let  $\tilde{C} = \sum_{l=1}^r \tilde{T}_l$  be a minimal decomposition, where  $\tilde{T}_l = \tilde{v}_l^{(1)} \otimes \tilde{v}_l^{(2)} \otimes \dots \otimes \tilde{v}_l^{(k)}$  with  $\tilde{v}_l^{(j)} \in \mathbb{K}^m \times \{0\}^{d-m} \subseteq \mathbb{K}^d$ . We obtain another minimal decomposition of  $\tilde{C}$ , by acting with  $\tilde{Z}$ , with rank one terms  $\llbracket \tilde{T}_l; \tilde{Z}, \tilde{Z}, \dots, \tilde{Z} \rrbracket = (\tilde{Z}\tilde{v}_l^{(1)}) \otimes (\tilde{Z}\tilde{v}_l^{(2)}) \otimes \dots \otimes$



$(\tilde{Z}\tilde{v}_l^{(k)})$ . By Lemma 4.1, all minimal decompositions of  $\tilde{C}$  come from those of  $C$  by adjoining zeros. This means that the  $d - m$  row vectors in the  $(d - m) \times m$  lower-left block of  $\tilde{Z}$  have dot product zero with every vector appearing in a minimal decomposition of  $C$ . Since  $C$  is symmetrically concise, these row vectors must be zero. The identity  $\tilde{Y} = \tilde{X}\tilde{Z}$  now implies  $Y = XZ$ . This concludes the proof.  $\square$

**COROLLARY 4.3.** *Let  $C \in (\mathbb{K}^m)^{\otimes k}$  be a symmetrically concise tensor whose stabilizer under congruence by  $\mathrm{GL}(m, \mathbb{K})$  has cardinality  $n$ . Then, for any matrix  $X \in \mathbb{K}^{d \times m}$  of rank  $m$ , there are  $n$  matrices in  $\mathbb{K}^{d \times m}$  with  $k$ th signature  $\llbracket C; X, X, \dots, X \rrbracket$ . In particular, if the stabilizer of  $C$  under congruence by  $\mathrm{GL}(m, \mathbb{K})$  is finite, then rank  $m$  matrices  $X$  are algebraically identifiable from the signature tensor  $\llbracket C; X, X, \dots, X \rrbracket$ .*

*Proof.* Let  $Y$  be a matrix in  $\mathbb{K}^{d \times m}$  that satisfies  $\llbracket C; X, X, \dots, X \rrbracket = \llbracket C; Y, Y, \dots, Y \rrbracket$ . By Theorem 4.2, we have  $Y = XZ$ , where  $Z$  is in the stabilizer under the congruence action. If there are  $n$  choices for  $Z$ , then there are  $n$  choices for  $Y$ .  $\square$

If  $C \in (\mathbb{K}^m)^{\otimes k}$  has trivial stabilizer under congruence, then it is already symmetrically concise by Proposition 3.1. We can thus simplify Corollary 4.3 as follows.

**COROLLARY 4.4.** *If  $C \in (\mathbb{K}^m)^{\otimes k}$  has trivial stabilizer under congruence, then rank  $m$  matrices  $X \in \mathbb{K}^{d \times m}$  are identifiable from the signature tensor  $\llbracket C; X, X, \dots, X \rrbracket$ .*

In the rest of this section we assume that  $k = 3$ . The following example illustrates why Theorem 4.2 and Corollary 4.3 fail when  $C$  is not symmetrically concise.

*Example 4.5.* Fix the  $2 \times 2 \times 2$  tensor  $C = e_1 \otimes e_1 \otimes e_1$ . Its stabilizer in  $\mathrm{GL}(2, \mathbb{R})$  is

$$Z = \begin{bmatrix} 1 & * \\ 0 & * \end{bmatrix},$$

where the  $*$  entries can take any value in  $\mathbb{R}$ . Setting  $m = 2, d = 3$ , we also introduce

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} 1 & * \\ 0 & * \\ 0 & * \end{bmatrix}.$$

The left-hand side of (4.1) is the set of all matrices of the form  $Y$ . This set strictly contains the right-hand side of (4.1), because not all matrices  $Y$  are expressible as  $XZ$  for some  $Z$ . This happens because the last row of  $Y$  has dot product zero with all vectors in the minimal decomposition of  $C$ , without being zero itself, i.e., the tensor  $C$  is not symmetrically concise.

Identifiability for tensors is usually studied in the context of minimal decompositions; see, e.g., [19]. The following result gives conditions under which algebraic identifiability of a minimal decomposition implies algebraic identifiability under congruence. We consider two decompositions of a tensor to be the same if they differ by a reordering of the rank one terms.

**THEOREM 4.6.** *Let  $\psi$  be a dictionary that is not a loop. Suppose that its core tensor  $C = C_\psi \in (\mathbb{R}^m)^{\otimes 3}$  is symmetrically concise and has  $\mathrm{rank}(C) = r$ , and the number  $\delta$  of minimal decompositions of  $C$  is finite. Given a generic matrix  $X \in \mathbb{R}^{d \times m}$  with  $m \leq d$ , there are at most  $\delta \cdot \frac{r!}{(r-m)!}$  matrices  $Y \in \mathbb{R}^{d \times m}$  that have the same third order signature tensor as  $X$ .*

*Proof.* We determine the number of solutions  $Y$  to the tensor equation

$$(4.2) \quad \sigma^{(3)}(X) = \llbracket C; X, X, X \rrbracket = \llbracket C; Y, Y, Y \rrbracket = \sigma^{(3)}(Y).$$

Let  $C = \sum_{l=1}^r T_l$  be a minimal decomposition. Consider a change of basis of  $\mathbb{R}^m$  such that all standard basis vectors  $e_1, \dots, e_m$  occur in the minimal decomposition. This exists because  $C$  is symmetrically concise. Let  $W$  be the change of basis matrix. By Theorem 4.2, it suffices to count  $m \times m$  matrices  $Z$  which stabilize  $C' = \llbracket C; W, W, W \rrbracket$ . This is the third signature of the path  $W\psi$ , and it also has  $\delta$  minimal decompositions.

Let  $C' = \sum_{l=1}^r S_l$  be one of the minimal decompositions. We have at most  $r$  choices for the image of  $e_1$  (up to scale) in this decomposition. Then, we have at most  $r-1$  choices for  $e_2$  up to scale, etc. This gives at most  $\frac{r!}{(r-m)!}$  choices of  $m \times m$  matrices  $N$  with  $Z = N\Lambda$ , where  $\Lambda$  is diagonal and invertible. Since  $\psi$  is not a loop, the first order signature  $v = \psi(1) - \psi(0)$  is recoverable from the diagonal entries of the third order signature and  $v \neq 0$  is also fixed by  $Z$ . Hence  $Zv = v$ , so  $\Lambda v = N^{-1}v$ . Evaluating the right-hand side allows us to find  $\Lambda$ .  $\square$

The following example attains the bound in Theorem 4.6 nontrivially.

*Example 4.7* ( $m = d = 2$ ). Fix the dictionary  $\psi = (\psi_1, \psi_2)$  with basis functions  $\psi_1(t) = t - 10t^2 + 10t^3$  and  $\psi_2(t) = 11t - 20t^2 + 10t^3$ . By (2.4), the core tensor equals

$$C_\psi = \frac{1}{42} \left[ \begin{array}{cc|cc} 7 & -8 & 37 & -8 \\ -8 & 37 & -8 & 7 \end{array} \right].$$

Using Macaulay2 [15], we find that this tensor is symmetrically concise and has rank two and a unique rank two decomposition. The stabilizer consists of two matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Our upper bound of  $\delta \cdot \frac{r!}{(r-m)!} = 1 \cdot 2 = 2$  on the size of the stabilizer is attained. The stabilizer shows that  $C_\psi$  is unchanged under swapping the coordinates  $\psi_1$  and  $\psi_2$ .

**5. Identifiability for generic dictionaries.** Our main motivation for working with third order signatures is that, under reasonable hypotheses on the dictionary  $\psi$  and the ambient dimension  $d$ , the path  $X\psi$  can be recovered uniquely from  $\sigma^{(3)}(X)$ . We begin by placing this in the context of lower order signature tensors via a description of the set of paths with the same first and second order signatures. Thereafter, we turn to the universal variety  $\mathcal{U}_{m,3}$  of all third order signature tensors, and we apply results from invariant theory to show that generic dictionaries are identifiable.

We fix a dictionary  $\psi$ . The paths represented by this dictionary are paths  $X\psi$  as  $X$  varies over  $d \times m$  matrices. In this  $md$ -dimensional space, there is an  $m(d-1)$ -dimensional linear space of paths with the same first signature  $\sigma^{(1)}(X) = X\psi(1) - X\psi(0)$ . For second order signatures, the nonuniqueness of path recovery is quantified by the stabilizer of the  $m \times m$  core matrix  $C = C_\psi$  under the matrix congruence action:

$$\text{Stab}_{\mathbb{R}}(C) = \{ X \in \text{GL}(m, \mathbb{R}) : X C X^T = C \}.$$

**PROPOSITION 5.1.** *The stabilizer of a generic  $m \times m$  core matrix  $C$  is a variety of dimension  $\binom{m}{2}$ . Setting  $\det(X) > 0$ , it is conjugate to the symplectic group  $\text{Sp}(m, \mathbb{R})$  intersected with the codimension  $m$  group of matrices that fix a given vector in  $\mathbb{R}^m$ .*

*Proof.* The symplectic group  $\text{Sp}(m, \mathbb{R})$  is the set of all endomorphisms of  $\mathbb{R}^m$  that fix a skew-symmetric bilinear form of maximal rank. If  $m$  is even, then this is one of the classical semisimple Lie groups. If  $m$  is odd, then  $\text{Sp}(m, \mathbb{R})$  can be realized by extending the matrices in  $\text{Sp}(m-1, \mathbb{R})$  by a column of arbitrary entries. In both

cases, we have  $\dim(\mathrm{Sp}(m, \mathbb{R})) = \binom{m+1}{2}$ . The core matrix equals  $C = uu^\top + Q$ , where  $u$  is a general column vector and  $Q$  is a general skew-symmetric matrix. Our stabilizer consists of all  $m \times m$  matrices  $X$  that satisfy  $Xu = u$  and  $XQX^\top = Q$ . The second condition defines the symplectic group, up to change of coordinates, and the first condition specifies a general linear space of codimension  $m$ .  $\square$

The matrix  $C$  in Proposition 5.1 is a generic point in the universal variety  $\mathcal{U}_{m,2}$ . We now consider an  $m \times m \times m$  core tensor  $C$  that is generic in the universal variety  $\mathcal{U}_{m,3}$ . This is the third signature of a dictionary  $\psi$  which is generic in the sense of Example 2.3. In the following identifiability result, the field  $\mathbb{K}$  can be either  $\mathbb{R}$  or  $\mathbb{C}$ .

**THEOREM 5.2.** *Let  $C$  be an  $m \times m \times m$  tensor that is a generic point in the variety  $\mathcal{U}_{m,3}$ . The stabilizer of  $C$  under the congruence action by  $\mathrm{GL}(m, \mathbb{K})$  is trivial.*

*Proof.* We work over the complex numbers  $\mathbb{C}$  and show that the complex stabilizer of the real tensor  $C$  is trivial: it consists only of the scaled identity matrices  $\eta I$ , where  $\eta^3 = 1$ . For  $m \leq 3$ , this result is established by a direct Gröbner basis computation in Maple. For  $m \geq 4$  we use the following parametrization of the universal variety  $\mathcal{U}_{m,3}$ . Let  $P$  be a generic vector in  $\mathbb{C}^m$ , and let  $Q$  be a generic skew-symmetric  $m \times m$  matrix. Following the definitions in [1, section 4.1], we take  $L$  to be a generic element in the space  $\mathrm{Lie}^{[3]}(\mathbb{C}^m)$  of homogeneous Lie polynomials of degree 3. Then

$$(5.1) \quad C = \frac{1}{6}P^{\otimes 3} + \frac{1}{2}(P \otimes Q + Q \otimes P) + L.$$

Indeed,  $P + Q + L$  is a general Lie polynomial of degree  $\leq 3$ , and the right-hand side in (5.1) is the degree 3 component in the expansion of its exponential; see [1, Example 5.15]. The constituents  $P$ ,  $Q$ , and  $L$  are recovered from  $C$  by taking the logarithm of  $C$  in the tensor algebra and extracting the homogeneous components of degree 1, 2, and 3. In particular, since these computations are equivariant with respect to the congruence action by  $\mathrm{GL}(m, \mathbb{K})$ , the stabilizer of  $C$  is contained in the stabilizer of  $L$ .

By [1, Proposition 4.7], a basis for the vector space  $\mathrm{Lie}^{[3]}(\mathbb{C}^m)$  consists of the bracketings of all Lyndon words of length three on the alphabet  $\{1, 2, \dots, m\}$ . The number of these Lyndon triples is  $\frac{1}{3}(m^3 - m)$ . The group  $G = \mathrm{GL}(m, \mathbb{C})$  acts irreducibly on  $\mathrm{Lie}^{[3]}(\mathbb{C}^m)$ . By comparing dimensions, we see that

$$(5.2) \quad \mathrm{Lie}^{[3]}(\mathbb{C}^m) \simeq S_{(2,1)}(\mathbb{C}^m).$$

The right-hand side is the irreducible  $G$ -module associated with the partition  $(2, 1)$  of the integer 3; see [20]. By the hook length formula, the vector space dimension of (5.2) equals  $\frac{1}{3}(m^3 - m)$ . This number exceeds the dimension  $m^2$  of the group, since  $m \geq 4$ . The map  $C \mapsto L$  from the universal variety  $\mathcal{U}_{m,3}$  to the  $G$ -module in (5.2) is surjective, since the homogeneous Lie polynomial  $L$  in (5.1) can be chosen arbitrarily.

We now apply Popov's classification [2, 26] of irreducible  $G$ -modules with non-trivial generic stabilizer. A recent extension to arbitrary fields due to Garibaldi and Guralnick can be found in [14]. A very special case of these general results says that the stabilizer of a generic point  $L$  in the  $G$ -module  $S_{(2,1)}(\mathbb{C}^m)$  is trivial. This implies that the stabilizer of the core tensor  $C$  under the congruence action of  $G$  is trivial.  $\square$

We conclude from Theorem 5.2 and Corollary 4.4 that the paths which are representable in a generic dictionary are identifiable from their third order signature.

**COROLLARY 5.3.** *Let  $m \leq d$  and let  $C \in \mathcal{U}_{m,3}$  be a generic dictionary. Given  $X \in \mathbb{R}^{d \times m}$  of rank  $m$ , the only real solution to  $\llbracket C; X, X, X \rrbracket = \llbracket C; Y, Y, Y \rrbracket$  is  $Y = X$ .*

We close this section with a remark about equations defining  $\mathcal{U}_{m,3}$ . The entries  $p_k$  of the vector  $P$  and the entries  $q_{ij}$  of the skew-symmetric matrix  $Q$  are recovered from the entries  $c_{ijk}$  of the core tensor  $C$  by the identities

$$(5.3) \quad \begin{aligned} p_k q_{ij} &= \frac{1}{2}(c_{kij} + c_{ikj} + c_{ijk}) - \frac{1}{2}(c_{kji} + c_{jki} + c_{jik}), \\ p_i p_j p_k &= c_{kij} + c_{ikj} + c_{ijk} + c_{kji} + c_{jki} + c_{jik}. \end{aligned}$$

These formulas are a very special case of [13, Proposition 18]. Here  $(i, j, k)$  runs over triples in  $\{1, 2, \dots, m\}$ . The linear forms on the right-hand side (5.3) span the shuffle linear forms, i.e., linear equations that cut out (5.2) as a subspace of  $(\mathbb{C}^m)^{\otimes 3}$ .

We regard  $P$  as a column vector of length  $m$ . For any matrix  $A$  we write  $\text{vec}(A)$  for the row vector whose coordinates are the entries of  $A$ . We apply this to the symmetric matrix  $PP^T$  and the skew-symmetric matrix  $Q$ , and we concatenate the resulting row vectors. Then the following matrix has  $m$  rows and  $2m^2$  columns:

$$H = P \cdot [\text{vec}(PP^T) \text{vec}(Q)].$$

By construction, the matrix  $H$  has rank 1, and we obtain  $P$  and  $Q$  from its rank 1 factorization. Each entry of  $H$  is one of the monomials on the left-hand side in (5.3).

Let  $H[C]$  denote the  $m \times 2m^2$  matrix that is obtained from  $H$  by replacing each monomial by the corresponding linear form on the right-hand side of (5.3). Thus the entries of  $H[C]$  are linear forms in  $C$ . We have shown that the  $2 \times 2$  minors of  $H[C]$  cut out the variety  $\mathcal{U}_{m,3}$ . A vast generalization of this fact is due to Galuppi [13]. His results also imply that the  $2 \times 2$  minors of  $H[C]$  generate the prime ideal of  $\mathcal{U}_{m,3}$ .

**6. Piecewise linear paths are identifiable.** In this section we prove that piecewise linear paths in real  $d$ -space with  $m \leq d$  steps are uniquely recoverable from their third order signatures. As before, we take  $\mathbb{K}$  to be  $\mathbb{R}$  or  $\mathbb{C}$ . Let  $C_{\text{axis}} \in \mathbb{K}^{m \times m \times m}$  be the piecewise linear core tensor in (2.9) and  $S$  any tensor in the orbit

$$\{ \llbracket C_{\text{axis}}; X, X, X \rrbracket \in \mathbb{K}^{d \times d \times d} : X \in \mathbb{K}^{d \times m} \}.$$

We show that there is a unique matrix  $X$ , up to third root of unity, such that  $S = \llbracket C_{\text{axis}}; X, X, X \rrbracket$ . This proves Conjectures 6.10 and 6.12 in [1] for  $m \leq d$ . In particular, if the field  $\mathbb{K}$  is the real numbers  $\mathbb{R}$ , the matrix  $X$  can be uniquely recovered from  $S$ .

**LEMMA 6.1.** *Let  $X \in \mathbb{K}^{m \times m}$  be in the stabilizer of  $C_{\text{axis}}$  under congruence. If  $e_m = (0, \dots, 0, 1)^T$  is an eigenvector of  $X$ , then  $e_m$  is also an eigenvector of  $X^T$ .*

*Proof.* Any matrix  $X$  in the stabilizer of  $C = C_{\text{axis}}$  also stabilizes the first and second order signatures, up to scaling by third root of unity, by (2.5). The core tensor  $C$  represents a path from  $(0, \dots, 0)$  to  $(1, \dots, 1)$ , so the first order signature is  $b = (1, \dots, 1)^T$ . The matrix  $X$  satisfies  $Xb = \eta b$ , where  $\eta^3 = 1$ , so  $b$  is an eigenvector of  $X$ . By [1, Example 2.1], the signature matrix of the piecewise linear dictionary is

$$C_2 = \begin{bmatrix} \frac{1}{2} & 1 & \cdots & 1 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \frac{1}{2} & 1 \\ 0 & \cdots & 0 & \frac{1}{2} \end{bmatrix}.$$

Since  $C_2$  differs from  $XC_2X^T$  by a third root of unity, denoted  $\eta'$ , we have

$$(6.1) \quad XC_2X^T = \eta' C_2 \implies \eta' C_2^{-1} = X^T C_2^{-1} X \implies \eta' C_2^{-1} b = \eta X^T C_2^{-1} b.$$

Hence  $v := \frac{1}{2} C_2^{-1} b = ((-1)^{m-1}, (-1)^{m-2}, \dots, -1, 1)^T$  is an eigenvector of  $X^T$ .

Consider the matrix obtained from  $C$  by multiplying by  $v$  along the second index,  $D = \llbracket C; \cdot, v, \cdot \rrbracket$  or  $D_{ik} = \sum_j c_{ijk} v_j$ . The matrix  $D$  is diagonal, by the following direct computation. If  $i < k$ , we get

$$D_{ik} = \sum_{j=1}^m (-1)^{m+j} c_{ijk} = (-1)^m \left( \frac{1}{2} (-1)^i + \sum_{i < j < k} (-1)^j + \frac{1}{2} (-1)^k \right) = 0.$$

If  $i > k$  all entries  $c_{ijk}$  vanish, hence the sum is zero. If  $i = k$ , the only nonzero entry of  $C$  that appears in the sum is  $c_{iii}$  and we obtain  $D_{ii} = \frac{1}{6}(-1)^{m+i}$ . Also, by definition of  $D$ , we find that the matrix  $X$  is in its stabilizer under congruence, up to scaling by third root of unity,  $D = \frac{\eta}{\eta'} XDX^\top$ .

Suppose that  $e_m$  is an eigenvector of  $X$ . By the same argument as in (6.1) we find that  $D^{-1}e_m = 6e_m$  is an eigenvector of  $X^\top$ . Hence,  $e_m$  is an eigenvector of  $X^\top$ .  $\square$

**THEOREM 6.2.** *The stabilizer of the piecewise linear core tensor  $C = C_{\text{axis}}$  under the congruence action  $C \mapsto \llbracket C; X, X, X \rrbracket$  by matrices  $X \in \text{GL}(m, \mathbb{K})$  is trivial.*

*Proof.* Let  $X$  be in  $\text{Stab}_{\mathbb{K}}(C)$ . Evaluating  $C = \llbracket C; X, X, X \rrbracket$  at coordinate  $(i, j, k)$  implies  $c_{ijk}$  equals

$$\sum_{1 \leq \alpha \leq m} \frac{1}{6} x_{i\alpha} x_{j\alpha} x_{k\alpha} + \sum_{1 \leq \alpha < \beta \leq m} \frac{1}{2} x_{i\alpha} x_{j\alpha} x_{k\beta} + \sum_{1 \leq \alpha < \beta \leq m} \frac{1}{2} x_{i\alpha} x_{j\beta} x_{k\beta} + \sum_{1 \leq \alpha < \beta < \gamma \leq m} x_{i\alpha} x_{j\beta} x_{k\gamma}.$$

Here the constants in (2.9) were substituted for the entries  $c_{\alpha\beta\gamma}$  of  $C$ . We can express this equation as the dot product  $c_{ijk} = f_{jk} \cdot X_i^\top = \sum_{\alpha=1}^m f_{jk}(\alpha) X_i(\alpha)$ , where  $X_i$  is the  $i$ th row of  $X$  and  $f_{jk}$  denotes the row vector with  $\alpha$ -coordinate

$$f_{jk}(\alpha) = \frac{1}{6} x_{j\alpha} x_{k\alpha} + \sum_{\beta > \alpha} \frac{1}{2} x_{j\alpha} x_{k\beta} + \sum_{\beta > \alpha} \frac{1}{2} x_{j\beta} x_{k\beta} + \sum_{\gamma > \beta > \alpha} x_{j\beta} x_{k\gamma}.$$

When  $j > k$ , the entry  $c_{ijk}$  vanishes for all  $1 \leq i \leq m$ . Hence the vector  $f_{jk}$  for  $j > k$  has dot product zero with all rows of  $X$ . Since the rows of  $X$  span  $\mathbb{K}^m$ , we see that  $f_{jk}$  is the zero vector, and the last entry  $f_{jk}(m) = \frac{1}{6} x_{jm} x_{km}$  vanishes for all  $j \neq k$ .

We can express the entries  $c_{ijk}$  as a different dot product. Namely, factoring out the terms involving the  $j$ th row, we obtain  $c_{ijk} = g_{ik} \cdot X_j^\top$ , where  $g_{ik}$  is the row vector with  $\beta$ -coordinate

$$g_{ik}(\beta) = \frac{1}{6} x_{i\beta} x_{k\beta} + \sum_{\gamma > \beta} \frac{1}{2} x_{i\beta} x_{k\gamma} + \sum_{\alpha < \beta} \frac{1}{2} x_{i\alpha} x_{k\beta} + \sum_{\gamma > \beta > \alpha} x_{i\alpha} x_{k\gamma}.$$

For all  $i > k$ , the entry  $c_{ijk}$  vanishes. This means that the dot product of  $g_{ik}$  with all rows of  $X$  is zero, hence  $g_{ik}$  is the zero vector. Its  $m$ th entry  $g_{ik}(m)$  equals

$$\frac{1}{6} x_{im} x_{km} + \sum_{\alpha=1}^{m-1} \frac{1}{2} x_{i\alpha} x_{k\alpha} = \frac{x_{km}}{2} \left( \sum_{\alpha=1}^m x_{i\alpha} - \frac{2}{3} x_{im} \right).$$

Since  $X$  stabilizes the first order signature, up to scaling by third root of unity  $\eta$ , the rows of  $X$  sum to  $\eta$ , hence  $g_{ik}(m) = \frac{\eta}{2} x_{km} - \frac{1}{3} x_{km} x_{im}$  for all  $i > k$ . By the previous paragraph, the second term vanishes and, setting  $i = m$ , we deduce that  $x_{km} = 0$  for all  $1 \leq k < m$ . This implies that the last column of  $X$  is parallel to the  $m$ th standard basis vector  $e_m$  and hence that  $e_m$  is an eigenvector of  $X$ .

By Lemma 6.1,  $e_m$  is also an eigenvector of  $X^\top$ . Thus, all entries in the last row of  $X$  vanish except the last. This means that  $X$  has the block diagonal structure

$$X = \begin{bmatrix} * & \cdots & * & 0 \\ \vdots & & \vdots & \vdots \\ * & \cdots & * & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}.$$

The  $*$  entries represent unknowns in an  $(m-1) \times (m-1)$  block which we call  $X'$ .

We now observe that  $X'$  stabilizes  $C'$ , the axis core tensor in  $\mathbb{K}^{(m-1) \times (m-1) \times (m-1)}$  arising from  $C$  by restricting to indices  $1 \leq i, j, k \leq m-1$ . From  $C = \llbracket C; X, X, X \rrbracket$  we have  $c_{ijk} = \sum_{\alpha, \beta, \gamma=1}^m c_{\alpha\beta\gamma} x_{i\alpha} x_{j\beta} x_{k\gamma}$ . Since  $x_{uv} = 0$  whenever  $u < m = v$ , this simplifies to

$$c_{ijk} = \sum_{\alpha, \beta, \gamma=1}^{m-1} c_{\alpha\beta\gamma} x_{i\alpha} x_{j\beta} x_{k\gamma} \quad \text{for } 1 \leq i, j, k \leq m-1.$$

Hence  $X'$  is in the stabilizer of  $C'$ . The proof of Theorem 6.2 is concluded by induction on  $m$ , given that the assertion can be proved for small  $m$  by a direct computation.  $\square$

We deduce from Theorem 6.2 and Corollary 4.4 that piecewise linear paths in  $\mathbb{R}^d$ , consisting of at most  $d$  steps, can be uniquely recovered from the third order signature. This uses the fact that the upper triangular tensor  $C_{\text{axis}}$  is symmetrically concise.

**COROLLARY 6.3.** *Let  $m \leq d$ , let  $C = C_{\text{axis}}$ , and  $X \in \mathbb{R}^{d \times m}$  of rank  $m$ . The only real solution to the tensor equation  $\llbracket C; X, X, X \rrbracket = \llbracket C; Y, Y, Y \rrbracket$  is the matrix  $Y = X$ .*

**7. Numerical identifiability.** A path in  $\mathbb{R}^d$ , coming from a dictionary of size  $m$ , can be recovered from its signature tensor  $S \in \mathbb{R}^{d \times d \times d}$  by solving a system of  $d^3$  equations in  $md$  unknowns (see section 2). This can be done in principle using Gröbner basis methods. However, such methods are infeasible when  $m$  and  $d$  are large, or when the data  $S$  is inexact or noisy. In such cases we instead minimize the distance between  $S$  and the set of tensors  $\llbracket C; X, X, X \rrbracket$  as  $X$  ranges over  $\mathbb{R}^{d \times m}$ . We seek the global minimum of the cost function

$$(7.1) \quad g : \mathbb{R}^{d \times m} \rightarrow \mathbb{R}, \quad X \mapsto \|\llbracket C; X, X, X \rrbracket - S\|^2,$$

where  $\|\cdot\|$  denotes the Euclidean norm in tensor space, the *Frobenius norm*. We first comment on the algebraic complexity of this cost function. Then, we quantify the numerical identifiability of recovering paths from third order signatures.

The number of critical points of (7.1) is the *ED degree* [11] of the orbit of an  $m \times m \times m$  tensor  $C$  under multiplication by  $d \times m$  matrices. For  $m = d$ , it is the ED degree of an orbit of the congruence action of  $\text{GL}(m, \mathbb{C})$  on the space  $\mathbb{C}^{m \times m \times m}$ .

*Example 7.1* ( $d = m = 2$ ). Fix the core tensor  $C_{\text{axis}}$  of format  $2 \times 2 \times 2$ . Its orbit under the congruence action is the degree 6 variety  $\mathcal{L}_{2,3,2}$  in [1, Table 3], defined by 9 quadrics. A computation reveals that the ED degree of  $\mathcal{L}_{2,3,2}$  is 15. For generic data  $S \in \mathbb{R}^{2 \times 2 \times 2}$ , our minimization problem has  $45 = 15 \times 3$  critical points  $X$  in  $\mathbb{C}^{2 \times 2}$ . Each critical point on  $\mathcal{L}_{2,3,2}$  corresponds to a triple of matrices  $\eta X$ , with  $\eta^3 = 1$ .

The ED degree specifies the algebraic degree of the coordinates of the optimal solution to (7.1), given rational data  $S$ . This degree can drop for special tensors  $S$ .

*Example 7.2* (the skyline path). This path has 13 steps in  $\mathbb{R}^2$ , the columns of

$$Y = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 0 & 2 & 0 & -2 & 0 & 1 & 0 & -1 & 0 \end{bmatrix}.$$

Its third order signature tensor is  $\llbracket C_{\text{axis}}; Y, Y, Y \rrbracket$ , where  $C_{\text{axis}}$  has format  $13 \times 13 \times 13$ :

$$S_{\text{skyline}} = \llbracket C_{\text{axis}}; Y, Y, Y \rrbracket = \frac{1}{6} \left[ \begin{array}{cc|cc} 343 & 0 & -84 & 18 \\ 84 & 18 & -36 & 0 \end{array} \right].$$

Using Gröbner bases, we compute the best approximation of  $S_{\text{skyline}}$  by a piecewise linear path with  $m = 2$  steps. The solution is the path given by the two columns of

$$X^* = \begin{bmatrix} a & a \\ b & -b \end{bmatrix},$$

where  $a = 3.4952680660622583405\dots$  and  $b = 1.218447154323916453\dots$ . These are algebraic numbers of degree 12. The Euclidean distance between the tensors is

$$(7.2) \quad \|\llbracket C_{\text{axis}}; X^*, X^*, X^* \rrbracket - S_{\text{skyline}}\| \approx 3.36.$$

In the rest of this section, we address the numerical identifiability of recovering a path from its signature tensor. We fix a core tensor  $C$  with trivial stabilizer under congruence action. Consider the set  $\mathcal{N}(C, d)$  of signature tensors  $S$  for which  $\{Y \in \mathbb{R}^{d \times m} : S = \llbracket C; Y, Y, Y \rrbracket\}$  has cardinality at least 2. These are instances for which path recovery is nonidentifiable. However, even if a path is identifiable from its signature tensor in the exact sense of sections 4–6, different paths may lead to numerically indistinguishable signatures. We quantify this via the distance to the set  $\mathcal{N}(C, d)$  of *ill-posed instances*. We define the *numerical nonidentifiability* of a pair  $(C, X)$  to be

$$\kappa(X, C) = \frac{\|X\|^3 \cdot \|C\|}{\inf_{S \in \mathcal{N}(C, d)} \|\llbracket C; X, X, X \rrbracket - S\|},$$

where  $\|\cdot\|$  denotes the Frobenius norm. When the numerical nonidentifiability is large, this reflects the proximity of  $\llbracket C; X, X, X \rrbracket$  to a nonidentifiable tensor. Conversely, a small value of the numerical nonidentifiability means that all close-by tensors  $\llbracket C; X, X, X \rrbracket$  are also identifiable. We give upper and lower bounds on  $\kappa(X, C)$ , in terms of the flattenings of  $C$  and the condition number of the rectangular matrix  $X$ ,  $\kappa(X) = \|X\| \cdot \|X^+\|$ , where  $X^+$  denotes the pseudoinverse.

We first remark on connections between the numerical nonidentifiability and the condition number. Following [5, section O.2], and setting  $S = \llbracket C; Y, Y, Y \rrbracket$ , the condition number of our recovery problem is

$$(7.3) \quad \text{cond}(X, C) = \lim_{\delta \rightarrow 0} \sup_{\|\llbracket C; X, X, X \rrbracket - S\| \leq \delta} \frac{\|X - Y\|}{\|\llbracket C; X, X, X \rrbracket - S\|} \cdot \frac{\|\llbracket C; X, X, X \rrbracket\|}{\|X\|}.$$

The condition number records how much the recovered matrix can change with small changes to the signature tensor. When the condition number is finite, the matrix can be recovered uniquely using symbolic computations. However, when the condition number is large, small changes in the signature induce large changes in the recovered matrix, a problem for numerical computations. Following the approach introduced by [27] in the context of linear programming, the condition number is often determined

by the inverse distance to the set of instances with infinite condition number [5, 9]. On the set of ill-posed instances  $\mathcal{N}(C, d)$  the condition number is infinite, because the problem is nonidentifiable. Hence the numerical nonidentifiability gives a lower bound on the inverse distance to the instances with infinite condition number. Condition numbers for algebraic identifiability can be defined using the local set-up described in [3]. Proving a *condition number theorem* to relate (7.3) to our inverse distance would be an interesting topic for further study.

**THEOREM 7.3.** *The numerical nonidentifiability of the pair  $(X, C)$ , for a matrix  $X \in \mathbb{R}^{d \times m}$  and a tensor  $C \in \mathbb{R}^{m \times m \times m}$  with trivial stabilizer, satisfies the upper bound*

$$(7.4) \quad \kappa(X, C) \leq \kappa(X)^3 \left( \frac{\|C\|}{\max(\zeta_m^{(1)}, \zeta_m^{(2)}, \zeta_m^{(3)})} \right),$$

where  $\zeta_m^{(i)}$  denotes the smallest singular value of the  $i$ th flattening of the tensor  $C$ , and  $\kappa(X)$  is the condition number of the matrix  $X$ , defined as above.

*Proof.* We aim to bound  $\kappa(X, C)^{-1}$ , the distance of  $\llbracket C; X, X, X \rrbracket$  to the locus of nonidentifiable tensors, from below. Since  $C$  has trivial stabilizer, Corollary 4.4 implies that all nonidentifiable tensors must be of the form  $\llbracket C; Y, Y, Y \rrbracket$ , where  $Y \in \mathbb{R}^{d \times m}$  has rank strictly less than  $m$ . The flattenings of such tensors have rank strictly less than  $m$ , so it suffices to lower-bound the distance of the flattenings to the much larger set  $\{A \in \mathbb{R}^{d \times d^2} : \text{rank}(A) < m\}$ . We have

$$\begin{aligned} \|X\|^3 \cdot \|X^+\|^3 \cdot \|C\| \cdot \kappa(X, C)^{-1} &\geq \min_{\text{rank}(A) < m} \|XC^{(i)}(X \otimes X)^T - A\| \cdot \|X^+\|^3 \\ &= \min_{\text{rank}(A') < m} \|X(C^{(i)} - A')(X \otimes X)^T\| \cdot \|X^+\|^3 \\ &\geq \min_{\text{rank}(A') < m} \|C^{(i)} - A'\| \\ &= \zeta_m^{(i)}, \end{aligned}$$

where  $X \otimes X \in \mathbb{R}^{d^2 \times m^2}$  is the Kronecker product of the matrix  $X$  with itself,  $A' \in \mathbb{R}^{m \times m^2}$ , and  $\|\cdot\|$  is the Frobenius norm. The chain of inequalities holds for  $i = 1, 2, 3$ , and the claim follows.  $\square$

We quantify the suitability of a core tensor  $C$ , with trivial stabilizer under congruence, for path recovery. We define the numerical nonidentifiability of  $C$  to be the smallest number  $\kappa(C)$  satisfying

$$\kappa(C) \geq \frac{\kappa(X, C)}{\kappa(X)^3}$$

for all  $X \in \mathbb{R}^{d \times m}$  of rank  $m$  and all  $d$ . From Theorem 7.3, we obtain the following.

**COROLLARY 7.4.** *The numerical nonidentifiability of  $C \in \mathbb{R}^{m \times m \times m}$ , with trivial stabilizer under congruence, satisfies*

$$\kappa(C) \leq \frac{\|C\|}{\max(\zeta_m^{(1)}, \zeta_m^{(2)}, \zeta_m^{(3)})}.$$

*Proof.* Divide (7.4) by  $\kappa(X)^3$ . The supremum of the left-hand side, as  $X$  ranges over  $\mathbb{R}^{d \times m}$  for all  $d$ , is equal to  $\kappa(C)$ . Hence  $\kappa(C)$  is bounded by the right-hand side.  $\square$



We now bound the numerical nonidentifiability of the piecewise linear dictionary.

**COROLLARY 7.5.** *The numerical nonidentifiability of  $C_{\text{axis}}$  is at most  $6\|C_{\text{axis}}\|$ .*

*Proof.* We show that the singular values of the second flattening  $C^{(2)} \in \mathbb{R}^{m \times m^2}$  are at least  $\frac{1}{6}$ . The  $j \times (i, k)$  entry is  $c_{ijk}$ . Since the entries of  $C$  are zero unless  $i \leq j \leq k$ , the flattening has an  $m \times m$  block, indexed by  $j \times (i, i)$ , which equals  $\frac{1}{6}$  times the identity matrix  $I$ . Let  $B$  denote the  $m \times (m^2 - m)$  matrix obtained by removing these  $m$  columns. Then  $C^{(2)}(C^{(2)})^\top = \frac{1}{36}I + BB^\top$ . The singular values of  $C^{(2)}$  are the square roots of the eigenvalues of  $C^{(2)}(C^{(2)})^\top$ . Consider an eigenvector  $v$  of  $BB^\top$  with eigenvalue  $\lambda$ . Then  $\lambda \geq 0$  because  $BB^\top$  is positive semidefinite and  $v$  is an eigenvector of  $C^{(2)}(C^{(2)})^\top$  with eigenvalue  $\frac{1}{36} + \lambda$ . Hence the singular values of  $C^{(2)}$  are bounded from below by  $\frac{1}{6}$ .  $\square$

Proposition 3.1 shows that the recovery problem is ill-posed if the tensor is not symmetrically concise. What follows is a numerical analogue to Proposition 3.1.

**PROPOSITION 7.6.** *Let  $C \in \mathbb{R}^{m \times m \times m}$  and  $C^{(\text{all})}$  be the  $m \times 3m^2$  matrix obtained by concatenating the three flattening matrices  $C^{(i)}$ . If  $\varsigma_m$  is the smallest singular value of  $C^{(\text{all})}$ , then*

$$\kappa(C) \geq \frac{\|C\|}{7m^{3/2}\varsigma_m}.$$

*Proof.* We compute the distance to a tensor  $S$  in the orbit of  $C$  that is not symmetrically concise. This gives an upper bound for the minimal distance to the set of ill-posed instances. Consider  $S = \llbracket C; I - vv^\top, I - vv^\top, I - vv^\top \rrbracket$ , where  $v$  is the left singular vector corresponding to the singular value  $\varsigma_m$  of  $C^{(\text{all})}$ . Then  $v$  is in the kernel of all three flattenings of  $S$ , which means that  $S$  is not symmetrically concise.

We have  $v^\top C^{(\text{all})} = \varsigma_m w^\top$ , where  $w$  is the right singular vector of length  $3m^2$ , corresponding to singular value  $\varsigma_m$ . We define  $w_i$  such that  $w$  is the stacking of  $w_1, w_2, w_3$  with each  $w_i$  of length  $m^2$ . Then  $v^\top C^{(i)} = \varsigma_m w_i^\top$ , hence  $\|v^\top C^{(i)}\| = \varsigma_m \|w_i\| \leq \varsigma_m \|w\| \leq \varsigma_m$ . We use this to upper bound the distance from  $C$  to  $S$ , as follows:

$$\begin{aligned} \|C - S\| &= \|\llbracket C; vv^\top, I, I \rrbracket + \llbracket C; I, vv^\top, I \rrbracket + \llbracket C; I, I, vv^\top \rrbracket - \llbracket C; vv^\top, vv^\top, I \rrbracket \\ &\quad - \llbracket C; I, vv^\top, vv^\top \rrbracket - \llbracket C; vv^\top, I, vv^\top \rrbracket + \llbracket C; vv^\top, vv^\top, vv^\top \rrbracket\| \\ &\leq \left( \sum_{i=1}^3 \|v^\top C^{(i)}\| + \|v^\top C^{(i)}\| \|vv^\top\| \right) + \|v^\top C^{(1)}\| \|vv^\top\|^2 \leq 7\varsigma_m. \end{aligned}$$

We have  $\kappa(C) \geq \frac{\kappa(X, C)}{\kappa(X)^3}$  for all matrices  $X$  of rank  $m$ . In particular,  $\kappa(C) \geq \frac{\kappa(I, C)}{\kappa(I)^3}$ . By definition of the numerical nonidentifiability

$$\kappa(I, C) = \frac{\|I\|^3 \|C\|}{\inf_{\tilde{S} \in \mathcal{N}(C, m)} \|C - \tilde{S}\|} \geq \frac{m^{3/2} \|C\|}{\|C - S\|} \geq \frac{m^{3/2} \|C\|}{7\varsigma_m},$$

since  $\|I\| = \sqrt{m}$ . The condition number of  $I$  is  $m$ , so the claim follows.  $\square$

The condition number quantifies the numerical identifiability of path recovery for the dictionary with core tensor  $C$ . We have derived informative upper and lower bounds on the related notion  $\kappa(C)$ , in terms of singular values of the flattenings of  $C$ . Figure 1 shows these bounds for small  $m$ . The lower bound is that in Corollary 7.4. The upper bound is that in Proposition 7.6. We see that the numerical nonidentifiability of the monomial dictionary grows exponentially with  $m$ . We also empirically

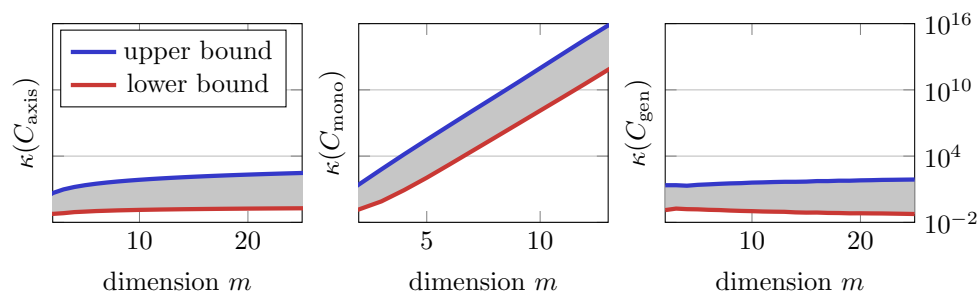


FIG. 1. Lower bounds and upper bounds on the numerical nonidentifiability for the piecewise linear core tensor (left), the monomial core tensor (middle), and generic core tensors (right).

observe such a trend for other bases of polynomial functions, such as the Chebyshev functions. On the other hand, the piecewise linear dictionary is much more stable. Corollary 7.5 shows that the numerical nonidentifiability of the piecewise linear dictionary remains below  $6\|C_{\text{axis}}\| = 6\sqrt{\frac{m}{36} + \frac{1}{2}\binom{m}{2} + \binom{m}{3}}$ , as seen on the left in Figure 1. The numerical nonidentifiability of generic dictionaries seems to remain below 100, independently of  $m$ . The right diagram in Figure 1 shows the average for 100 generic signature tensors  $C_{\text{gen}}$ , created using the first method in Example 2.3.

We can conclude that for piecewise linear paths, well-conditioned matrices  $X$  have signature tensors  $\llbracket C_{\text{axis}}; X, X, X \rrbracket$  that are reasonably far from the nonidentifiable locus. The same holds for paths from generic dictionaries, in a certain range of  $m$ . However, polynomial paths send well-conditioned matrices  $X$  to tensors  $\llbracket C_{\text{mono}}; X, X, X \rrbracket$  which are very close to being nonidentifiable, even for relatively small values of  $m$ . This suggests the possibility of numerical challenges for path recovery from such tensors, as we confirm in the numerical experiments in the next section.

**8. Path recovery via optimization.** Given a fixed dictionary, our aim is to compute a path represented by the dictionary whose signature most closely matches an input signature. In addition to the issues of numerical identifiability discussed in section 7, numerical optimization has several well-documented drawbacks. Since the objective function  $g$  is nonconvex, an abundance of local minima can be expected. The problem of local minima is inherent in almost all optimization methods, but there are some heuristic ways to overcome the problem. A thorough overview and application of state-of-the-art theory is out of the scope of this article. See [25].

We performed computational experiments for a range of values of  $m$  and  $d$ . We considered piecewise linear, polynomial, and generic paths, which were created using the first method in Example 2.3. For each pair  $(m, d)$ , we generated 100 random matrices  $X \in \mathbb{R}^{d \times m}$  with entries  $x_{ij} \sim N(0, 1)$  to represent the path  $X\psi$ . We computed  $S = \sigma^{(3)}(X)$  up to machine precision and then minimized the function in (7.1).

We implemented the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm with an Armijo backtracking line search in MATLAB 2018a. This was followed by a trust region Newton method for improved convergence, taken from the Manopt toolbox [4] which allows for direct implementation for matrix inputs. We stopped if  $\|\nabla g(X)\| < 10^{-10}$  or after 100 steps of the BFGS procedure and 1000 steps of the trust region algorithm. We allowed 10 reinitializations  $x_{ij} \sim N(0, 1)$  to try to eliminate local minima and other numerical issues that arise from the relatively high degree of the objective function. Let  $X^*$  denote the result of this computation. We declare the

TABLE 2

Percentage of successful recoveries for random piecewise linear paths (top) and random paths represented by generic dictionaries (bottom).

$m \backslash d$	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	100	100	100	100	100	100	100	100	100	100	100	100	100	100
3		100	100	100	100	100	99	100	100	100	100	100	100	100
4			100	97	100	100	100	100	100	100	100	100	100	100
5				97	99	99	100	100	100	100	100	100	100	100
6					97	95	98	96	97	100	100	100	100	100
7						91	92	95	96	97	99	99	100	100
8							90	92	95	98	99	99	98	100
9								93	90	94	98	95	95	96
10									85	96	94	97	93	93

$m \backslash d$	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	100	100	100	100	100	100	100	100	100	100	100	100	100	99
3		99	100	100	100	100	100	98	100	100	100	100	99	100
4			99	99	100	100	100	98	98	98	99	98	98	100
5				100	100	100	100	100	99	99	100	100	100	100
6					98	99	100	100	100	99	100	100	100	99
7						100	98	97	99	99	99	100	100	100
8							99	99	99	100	99	98	98	99
9								97	92	98	97	97	99	98
10									100	98	97	97	100	99

TABLE 3

The recovery rate for polynomial paths is low once the condition number becomes too big. Subscripts count the failures due to ill-conditioning.

$m \backslash d$	2	3	4	5	6	7	8	9	10	11	12
2	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>
3		98 <sub>2</sub>	99 <sub>1</sub>	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>	100 <sub>0</sub>
4			85 <sub>13</sub>	87 <sub>13</sub>	94 <sub>6</sub>	91 <sub>9</sub>	95 <sub>5</sub>	98 <sub>2</sub>	99 <sub>1</sub>	99 <sub>1</sub>	98 <sub>2</sub>
5				5 <sub>31</sub>	12 <sub>24</sub>	20 <sub>30</sub>	29 <sub>37</sub>	35 <sub>40</sub>	47 <sub>41</sub>	53 <sub>38</sub>	57 <sub>37</sub>
6					0 <sub>1</sub>	0 <sub>1</sub>	0 <sub>2</sub>	0 <sub>3</sub>	0 <sub>6</sub>	0 <sub>5</sub>	0 <sub>9</sub>
7						0 <sub>21</sub>	0 <sub>24</sub>	0 <sub>35</sub>	0 <sub>29</sub>	0 <sub>28</sub>	0 <sub>35</sub>

recovery successful if  $\|X^* - X\|/\|X^*\| < 10^{-5}$ . Tables 2 and 3 show the percentages of successful recoveries. The success rate for piecewise linear paths is 100% for small  $m$  but it becomes slightly worse for larger  $m$ . For paths represented by a generic dictionary, the results are also close to 100%.

Ill-conditioning has occurred when we recover a matrix  $X^*$  with a large distance to the original  $X$ , but whose signatures are the same. We call an instance a *failure due to ill-conditioning* if the relative error between the matrices  $\|X^* - X\|/\|X^*\|$  exceeds  $10^{-5}$  but the relative distance between the signatures is less than  $10^{-8}$ . This indicates a condition number exceeding 1000. Such failures never occurred for piecewise linear paths and generic paths, in over 10000 experiments. The situation is dramatically worse for polynomial paths: the subscripts in Table 3 count the failures due to ill-conditioning. For  $m \geq 6$  if a matrix with sufficiently close signature was found, then in all cases it was a failure due to ill-conditioning. The machine precision inaccuracy in the signature leads to large differences in the recovered matrix. The overall recovery rates for polynomial dictionaries are low. We remark that although many of the other failures are not counted as being due to ill-conditioning under our requirements stated above, they often yield a relatively far away matrix with close-by signature tensor.

In conclusion, our experimental findings are consistent with the theoretical results on the numerical identifiability in section 7. We find that generic paths and piecewise

linear dictionaries behave best in numerical algorithms for recovering paths. The middle diagram in Figure 1 showed that the numerical nonidentifiability of the monomial core tensor  $C_{\text{mono}}$  grows rapidly with  $m$ . Our experiments confirmed the difficulty of path recovery from the monomial dictionary.

**9. Shortest paths.** Our study has so far been concerned with paths of low complexity in a space of high dimension. Such paths are identifiable from their third signature. In this final section we shift gears. We now come to a situation where the number of functions in the dictionary,  $m$ , is much larger than the dimension of the space,  $d$ . The paths are represented by a dictionary  $\psi$ , but identifiability no longer holds for the paths  $X\psi$  because there are too many parameters to recover the matrix  $X$  from its third order signature. We impose extra constraints to select a meaningful path among those with the same signature. A natural constraint is the length of the path. This leads to the problem of finding the shortest path for a given signature.

In this section we address the task of computing shortest paths when the third signature tensor is fixed. Recall that the length of a path  $\psi : [0, 1] \rightarrow \mathbb{R}^m$  equals

$$\text{len}(\psi) = \int_0^1 \sqrt{\langle \dot{\psi}(t), \dot{\psi}(t) \rangle} dt,$$

where  $\dot{\psi}(t) = \frac{d\psi}{dt}$ . This is a rather complicated function to evaluate in general. However, things are much easier for piecewise linear paths. For the  $m$ -step path given by the dictionary in (2.8) and the matrix  $X = (x_{ij})$ , the length is given by the formula

$$\text{len}(X) := \text{len}(X\psi) = \sum_{j=1}^m \sqrt{\sum_{i=1}^d x_{ij}^2}.$$

Note that this function is piecewise differentiable. We can therefore regularize the objective function (7.1) with a length constraint. This leads to the new function

$$h(X, \lambda) = \text{len}(X) + \lambda g(X),$$

where  $\lambda$  is a parameter. A necessary condition for a minimum is that both the gradient in  $X$  and the gradient in  $\lambda$  equal zero. The latter requirement ensures that  $X$  yields the required signature. A problem with this method is that critical points are usually saddle points, which cannot be easily obtained using standard gradient-related techniques. This holds because  $h$  is not bounded from below for  $\lambda \rightarrow -\infty$ . To work around this, we use a trick from optimization; see [25]. We fix  $\lambda_0$  and minimize

$$h(X, \lambda_0)/\lambda_0 = \lambda_0^{-1} \text{len}(X) + g(X).$$

Once a minimum  $X_0$  is found, we set  $\lambda_1 = 2\lambda_0$  and minimize again with  $\lambda_1$  and  $X_0$  as a starting point for the iteration. We repeat, setting  $\lambda_n = 2\lambda_{n-1}$  until  $\lambda_n$  is sufficiently large and the impact of the length constraint is negligible. Then, for some  $X_n$ , the function  $g$  is minimal, i.e.,  $X_n$  has the correct signature up to machine precision. Local minima might occur—a guarantee that  $X_n$  gives the shortest path cannot be made. However, this method has proved to be satisfactory for our application.

For  $d \leq 3$ , the resulting shortest paths corresponding to the piecewise linear dictionary and the monomial dictionary can be easily plotted. We report on two examples, for  $d = 2$  and for  $d = 3$ .

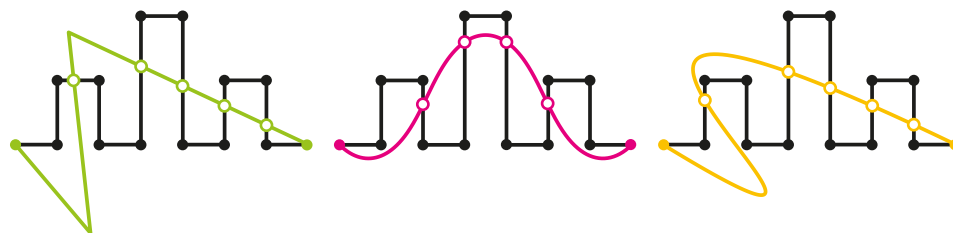


FIG. 2. The shortest path with  $m = 3$  steps (left),  $m = 100$  steps (middle), and a polynomial path of degree 3 (right). All have the same third signature as the skyline path.

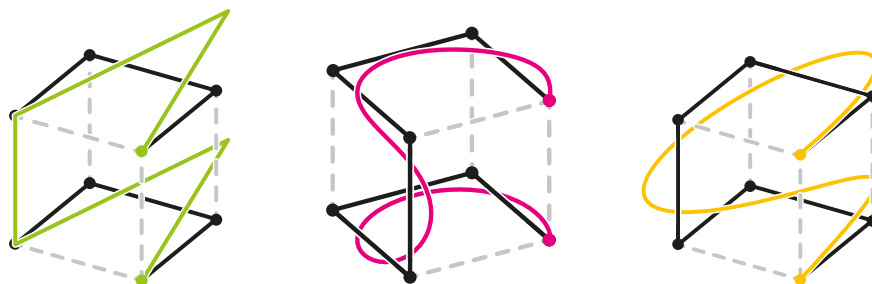


FIG. 3. The shortest path with  $m = 5$  steps (left) and  $m = 100$  steps (middle) having the same third order signature as the Klee-Minty path. On the right we see a polynomial path of degree 5 whose third order signature is close to that of the Klee-Minty path.

*Example 9.1* (skyline path). Consider the path in Example 7.2, with  $2 \times 2 \times 2$  signature tensor  $S_{\text{skyline}}$ . This path is shown in black in Figure 2. For a range of values  $m \geq 3$ , we computed the shortest piecewise linear path with  $m$  steps having signature  $S_{\text{skyline}}$ . The shortest piecewise linear path with  $m = 3$  is depicted in Figure 2 on the left. The middle image shows the shortest path with  $m = 100$ . This is an approximation to a shortest smooth path with that signature. We also learned polynomial paths but without length constraints. The right image shows a cubic path.

*Example 9.2* ( $d = 3$ ). We define the *Klee-Minty path* to be the following axis path with 7 steps in  $\mathbb{R}^3$ . It travels along the edges of a 3-cube and visits all 8 vertices:

$$X = \begin{bmatrix} 1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

The third order signature tensor of this path equals

$$S_{\text{km}} = \llbracket C_{\text{axis}}; X, X, X \rrbracket = \frac{1}{6} \left[ \begin{array}{ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -6 \\ 0 & 6 & 0 & -6 & 3 & -3 \end{array} \middle| \begin{array}{ccc} 0 & 6 & 0 \\ -6 & 3 & 3 \\ 0 & 0 & 1 \end{array} \right].$$

We expect to find piecewise linear paths with  $m = 5$  steps and third signature  $S_{\text{km}}$  because  $md = 15$  exceeds  $\dim(\mathcal{U}_{3,3}) = 14$ . Our optimization method found several such paths. The shortest among them is shown on the left of Figure 3. Next, we computed the shortest path with  $m = 100$  steps, shown in the middle of Figure 3.

Counting dimensions, polynomial paths of degree 5 are expected to fill  $\mathcal{U}_{3,3}$ . But we did not find any path with signature  $S_{\text{km}}$ . A close solution was a matrix  $X$  with

$$\| [C_{\text{mono}}; X, X, X] - S_{\text{km}} \| \approx 0.00914.$$

The associated quintic path  $X\psi$  is shown on the right in Figure 3. We believe that the issue is the distinction between the signature image and the signature variety, discussed in [1, section 2.2]. The tensor  $S_{\text{km}}$  seems to lie in the set  $\mathcal{P}_{3,3,5}^{\mathbb{R}} \setminus \mathcal{P}_{3,3,5}^{\text{im}}$ .

**Acknowledgments.** We thank Robert Guralnick and Vladimir Popov for the references used in Theorem 5.2. We are grateful to Carlos Améndola, Paul Breiding, Laura Colmenarejo, Joscha Diehl, Francesco Galuppi, Peter Bürgisser, and André Uschmajew for helpful conversations. Many thanks also to an anonymous referee whose detailed comments led to significant improvements to the paper.

#### REFERENCES

- [1] C. AMÉNDOLA, P. FRIZ, AND B. STURMFELS, *Varieties of Signature Tensors*, Forum Math. Sigma, to appear.
- [2] E. M. ANDREEV AND V. L. POPOV, *Stationary subgroups of points of general position in the representation space of a semisimple Lie group*, Funct. Anal. Appl., 5 (1971), pp. 265–271.
- [3] P. BREIDING AND N. VANNIEUWENHOVEN, *Convergence analysis of Riemannian Gauss-Newton methods and its connection with the geometric condition number*, Appl. Math. Lett., 78 (2018), pp. 42–50.
- [4] N. BOUMAL, B. MISHRA, P.-A. ABSIL, AND R. SEPULCHRE, *Manopt, a Matlab Toolbox for optimization on manifolds*, J. Mach. Learn. Res., 15 (2014), pp. 1455–1459.
- [5] P. BÜRGISSEER AND F. CUCKER, *Condition: The Geometry of Numerical Algorithms*, Grundlehren Math. Wiss. 349, Springer-Verlag, Berlin, 2013.
- [6] P. BÜRGISSEER, A. GARG, R. OLIVEIRA, M. WALTER, AND A. WIGDERSON, *Alternating Minimization, Scaling Algorithms, and the Null-Cone Problem from Invariant Theory*, arXiv:1711.08039, 2017.
- [7] K.-T. CHEN, *Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula*, Ann. of Math., 65 (1957), pp. 163–178.
- [8] I. CHEVYREV AND A. KORMILITZIN, *A Primer on the Signature Method in Machine Learning*, arXiv:1603.03788, 2016.
- [9] J. DEMMEL, *On condition numbers and the distance to the nearest ill-posed problem*, Numer. Math., 51 (1987), pp. 251–289.
- [10] J. DIEHL AND J. REIZENSTEIN, *Invariants of Multidimensional Time Series Based on their Iterated-Integral Signature*, arXiv:1801.06104, 2018.
- [11] J. DRAISMA, E. HOROBET, G. OTTAVIANI, B. STURMFELS, AND R. THOMAS, *The Euclidean distance degree of an algebraic variety*, Found. Comput. Math., 16 (2016), pp. 99–149.
- [12] P. FRIZ AND M. HAIRER, *A Course on Rough Paths. With an Introduction to Regularity Structures*, Universitext, Springer-Verlag, Berlin, 2014.
- [13] F. GALUPPI, *The Rough Veronese Variety*, arXiv:1809.02522, 2018.
- [14] S. GARIBALDI AND R. GURALNICK, *Generically Free Representations II: Irreducible Representations*, arXiv:1711.06400, 2017.
- [15] D. GRAYSON AND M. STILLMAN, *Macaulay2, A Software System for Research in Algebraic Geometry*, <https://www.math.uiuc.edu/Macaulay2/>.
- [16] T. KOLDA AND B. BADER, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500.
- [17] A. KORMILITZIN, K. SAUNDERS, P. HARRISON, J. GEDDES AND T. LYONS, *Application of the Signature Method to Pattern Recognition in the CEQUEL Clinical Trial*, arXiv:1606.02074, 2016.
- [18] A. KORMILITZIN, K. SAUNDERS, P. HARRISON, J. GEDDES, AND T. LYONS, *Detecting Early Signs of Depressive and Manic Episodes in Patients with Bipolar Disorder using the Signature-Based Model*, arXiv:1708.01206, 2017.
- [19] J. KRUSKAL, *Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra Appl., 18 (1977), pp. 95–138.

- [20] J. M. LANDSBERG, *Tensors: Geometry and Applications*, Grad. Stud. Math. 128, AMS, Providence, RI, 2012.
- [21] D. LEVIN, T. LYONS, AND H. NI, *Learning from the Past, Predicting the Statistics for the Future, Learning an Evolving System*, arXiv:1309.0260, 2018.
- [22] T. LYONS AND Z. QIAN, *System Control and Rough Paths*, Oxford University Press, Oxford, UK, 2002.
- [23] T. LYONS AND W. XU, *Hyperbolic development and inversion of signature*, J. Funct. Anal., 272 (2017), pp. 2933–2955.
- [24] T. LYONS AND W. XU, *Inverting the signature of a path*, J. Eur. Math. Soc (JEMS), 20 (2018), pp. 1655–1687.
- [25] J. NOCEDAL AND S. J. WRIGHT, *Numerical Optimization*, Springer-Verlag, Berlin, 2006.
- [26] A. M. POPOV, *Finite stationary subgroups in general position of simple linear Lie groups*, Trans. Moscow Math. Soc., 48 (1988), pp. 3–63.
- [27] J. RENEGAR, *Incorporating condition measures into the complexity theory of linear programming*, SIAM J. Optim., 5 (1995) pp. 506–524.
- [28] C. REUTENAUER, *Free Lie Algebras*, London Math. Soc. Monogr. 7, Oxford University Press, New York, 1993.
- [29] L. TUCKER, *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31 (1966), pp. 279–311.
- [30] V. STRASSEN, *Relative bilinear complexity and matrix multiplication*, J. Reine Angew. Math., (1987), pp. 406–443.
- [31] N. VERVLIT, O. DEBALS, L. SORBER, M. VAN BAREL, AND L. DE LATHAUWER, *Tensorlab 3.0*, <https://www.tensorlab.net> (2016).