Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Learning Paths from Signature Tensors

by

Max Pfeffer, Anna Seigal, and Bernd Sturmfels

Preprint no.: 78 2018

Learning Paths from Signature Tensors

Max Pfeffer, Anna Seigal, and Bernd Sturmfels

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 signature tensors of order three. We establish identifiability results and recovery algorithms for piecewise linear paths, polynomial paths, and generic dictionaries. A detailed analysis of the relevant condition numbers is presented. We also compute the shortest path with a given signature tensor.

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.

This article is motivated by the *signature method in machine learning* [7]. Here the geometric object is a path $[0,1] \to \mathbb{R}^d$. 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 [6], and they play an important role in stochastic analysis [10, 17]. We refer to [18, 19] for the recovery problem, and to [8, 11, 15] for algorithms and applications. Our point of departure is the approach 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 [14], 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 [18], we restrict to paths which lie in a particular family. We consider paths whose coordinates can be written as a linear combination 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.

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. Section 3 investigates the stabilizer of the congruence action. This leads to Section 4, which features 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 5 we study paths coming from generic dictionaries, and show that the associated paths can be uniquely determined by their third order signature tensors. In Section 6 we show that identifiability holds for piecewise linear paths. This partially proves [1, Conjecture 6.10]. In Section 7 we discuss path recovery by solving polynomial equations. For exact data we use Gröbner bases. Section 8 addresses the numerical analysis of path recovery from noisy signature data. Both upper bounds and lower bounds are given for the condition numbers of the core tensors of interest. In Section 9 we turn to numerical optimization and we present experimental results on unique recovery of low-complexity paths. Section 10 addresses the problem of finding the shortest path with given third signature.

2 Dictionaries and their Core Tensors

We fix a *dictionary* $\psi = (\psi_1, \psi_2, \ldots, \psi_m)$ of piecewise differentiable functions $\psi_i : [0,1] \to \mathbb{R}$. The dictionary corresponds to a path in $\mathbb{R}^m$, also denoted ψ, whose ith coordinate is ψ_i. The path ψ 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 kth term is a tensor in $(\mathbb{R}^m)^{\otimes k}$ with entries that are iterated integrals of ψ:

$$(\sigma^{(k)}(\psi))_{i_1 i_2 \cdots i_k} \;=\; \int_0^1 \int_0^{t_k} \cdots \int_0^{t_2} \mathrm{d}\psi_{i_1}(t_1)\, \mathrm{d}\psi_{i_2}(t_2) \cdots \mathrm{d}\psi_{i_k}(t_k). \tag{1}$$

Evaluating (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 a skew-symmetric matrix. Its entry q_{ij} is the *Lévy area* of the projection of ψ 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, 7, 8, 10, 17, 18, 19].

This article is based on the following two premises, to be discussed and justified below:

(a) We study the images of a fixed reference path ψ under linear maps.

(b) We focus our attention on the third order signature $\sigma^{(3)}(\psi)$.

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

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

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 (1), bearing in mind that integration is a linear operation.

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

$$\sigma(X\psi) \;=\; X(\sigma(\psi)). \tag{2}$$

The action of the linear map X on the signature $\sigma(\psi)$ is as follows. The kth order signature of ψ 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 multiplication is a tensor in $(\mathbb{R}^d)^{\otimes k}$. Using notation from the theory of tensor decomposition [14], the identity (2) can be written as

$$\sigma^{(k)}(X\psi) \;=\; [\![\, \sigma^{(k)}(\psi)\,;\, X, X, \ldots, X \,]\!] \qquad \text{for } \; k = 1, 2, 3, \ldots \tag{3}$$

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:

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

Lemma 2.1 means that, once the signature of the dictionary ψ is known, integrals no longer need to be computed. Signature tensors of a path arising from ψ 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 kth 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 discuss dictionaries in more detail for $k = 3$. We fix a dictionary ψ consisting of m functions. We refer to its third signature $C_\psi = \sigma^{(3)}(\psi)$ as the *core tensor* of ψ. This tensor has format $m \times m \times m$, and its entries are

$$c_{ijk} \;=\; \int_0^1 \int_0^{t_3} \int_0^{t_2} \mathrm{d}\psi_i(t_1)\, \mathrm{d}\psi_j(t_2)\, \mathrm{d}\psi_k(t_3) \qquad \text{for all } \; 1 \leq i, j, k \leq m. \tag{4}$$

The first and second signature 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]:

$$c_i c_j \;=\; c_{ij} + c_{ji} \quad \text{and} \quad c_i c_{jk} \;=\; c_{ijk} + c_{jik} + c_{jki}, \tag{5}$$

where c_i is the ith element of the first signature, while c_{ij} is the (i, j)th element of the second signature. This observation allows us to focus on third order signature tensors.

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 (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 (α, β, γ) is

$$[\![\, C_\psi \,;\, X, X, X \,]\!]_{\alpha\beta\gamma} \;=\; \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} c_{ijk} x_{\alpha i} x_{\beta j} x_{\gamma k}. \tag{6}$$

The expression (6) for the signature tensor in terms of the core tensor is closely related to the *Tucker decomposition* [14, 23] that arises frequently in tensor compression. Our application differs from the usual setting of the Tucker decomposition 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 families of paths they encode. We begin with the two dictionaries studied in [1]. The first dictionary is $\psi(t) = (t, t^2, \ldots, t^m)$. 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 ψ is denoted by C_{mono} to reflect the monomial functions in the dictionary. By [1, Example 2.2], its entries are

$$c_{ijk} \;=\; \frac{j}{i+j} \cdot \frac{k}{i+j+k}. \tag{7}$$

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

$$\psi_i(t) \;=\; \begin{cases} 0 & \text{if } \; t \leq \frac{i-1}{m}, \\ mt - (i-1) & \text{if } \; \frac{i-1}{m} < t < \frac{i}{m}, \\ 1 & \text{if } \; t \geq \frac{i}{m}. \end{cases} \tag{8}$$

Following [1, Example 2.1], the associated core tensor C_{axis} is "upper-triangular", with entries

$$c_{ijk} \;=\; \begin{cases} 1 & \text{if } \; i < j < k, \\ \frac{1}{2} & \text{if } \; i < j = k \text{ or } i = j < k, \\ \frac{1}{6} & \text{if } \; i = j = k, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

The signature tensors C_{mono} and C_{axis} are two real points in the *universal variety* $\mathcal{U}_{m,3} \subset (\mathbb{C}^m)^{\otimes 3}$, which consists of all third order signature tensors of paths in $\mathbb{C}^m$, or equivalently all possible 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 universal 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 this paper, $\mathcal{U}_{m,3}$ denotes the affine variety, whereas [1] uses projective varieties. The dimension of $\mathcal{U}_{m,3}$ is the number of Lyndon words on m letters of length 1, 2 or 3. These three numbers are m, $\binom{m}{2}$ and $\frac{1}{3}(m^3 - m)$. $\square$

The polynomials in the tensor entries that define $\mathcal{U}_{m,3}$ are obtained by eliminating the unknowns c_i and c_{ij} from the equations (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 $Y_1, Y_2, \ldots, Y_M$ in $\mathbb{R}^m$, where M exceeds $\frac{1}{3}m^2 + \frac{1}{2}m + \frac{1}{6} = \dim(\mathcal{U}_{m,3})$, and take the piecewise linear path with steps $Y_1, Y_2, \ldots, Y_M$. By [1, Example 5.4], the resulting *generic core tensor* equals

$$C_{\mathrm{gen}} \;=\; \frac{1}{6} \cdot \sum_{i=1}^{M} Y_i^{\otimes 3} \;+\; \frac{1}{2} \cdot \sum_{1 \leq i < j \leq M} \left(Y_i^{\otimes 2} \otimes Y_j + Y_i \otimes Y_j^{\otimes 2} \right) \;+ \sum_{1 \leq i < j < k \leq M} Y_i \otimes Y_j \otimes Y_k. \tag{10}$$

The coefficients in (10) match the tensor entries in (9). By Chen's Formula [1, eqn. (38)], C_{gen} is the degree 3 component in the tensor series $\sigma(\psi) = \exp(Y_1) \otimes \exp(Y_2) \otimes \cdots \otimes \exp(Y_M)$.

An alternative method for sampling uses the Gröbner basis in [1, Theorem 4.10]. We write σ_{lyndon} for the vector of all signatures σ_i, σ_{ij} and σ_{ijk} whose indices are Lyndon words. This includes all m first order signatures σ_i, all $\binom{m}{2}$ second order signatures σ_{ij} with $i < j$, and all $\frac{1}{3}(m^3 - m)$ third order signatures σ_{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 σ_{lyndon}. The non-Lyndon signatures σ_{ijk} are then computed by evaluating $\phi_{ijk}(\sigma_{\mathrm{lyndon}})$, where ϕ_{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_{axis}, C_{mono} or C_{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 ψ, i.e. it is the image of ψ 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

$$[\![\, C\,;\, X, X, X \,]\!] \;=\; S.$$

This is a system of d^3 cubic equations in md unknowns x_{ij}, written explicitly as

$$\sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} c_{ijk} x_{\alpha i} x_{\beta j} x_{\gamma k} \;=\; s_{\alpha\beta\gamma} \qquad \text{for } 1 \leq \alpha, \beta, \gamma \leq d. \tag{11}$$

The system (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 also include 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].

A special case that will occupy us a lot is the action of invertible $m \times m$ matrices on $m \times m \times m$ tensors. We conclude this section by mentioning two group actions, closely related to ours, which have been studied extensively. The first action concerns homogeneous polynomials $f(x_1, \ldots, x_m)$. Matrices Z in $\mathrm{GL}(m, \mathbb{C})$ act by linear change of variables, i.e. $f(x) \mapsto f(Z \cdot x)$. This is precisely our congruence action $C \mapsto [\![C; Z, Z, Z]\!]$ in the special case where C is a *symmetric* tensor that corresponds to the cubic polynomial

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

The stabilizer of f is trivial modulo scaling if the projective hypersurface $V(f)$ is smooth [20].

Another well-studied group action (cf. [5, 16]) concerns tensors of any size $m_1 \times \cdots \times m_k$. The group $\mathrm{GL}(m_1, \mathbb{C}) \times \cdots \times \mathrm{GL}(m_k, \mathbb{C})$ acts via $C \mapsto [\![C; Z_1, \ldots, Z_k]\!]$ where $Z_i \in \mathrm{GL}(m_i, \mathbb{C})$. For $k = 3$ this is the action $C \mapsto [\![C; Z_1, Z_2, Z_3]\!]$. Our action is the restriction to the diagonal $Z := Z_1 = Z_2 = Z_3$. There is considerable literature on the above two group actions, but much less on the *congruence action* of $\mathrm{GL}(m, \mathbb{C})$ on $(\mathbb{C}^m)^{\otimes k}$ which is needed here.

3 Stabilizers under Congruence

From now on, $\mathbb{K}$ denotes 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 tensors $T \in (\mathbb{K}^m)^{\otimes k}$ via

$$T \;\mapsto\; [\![T; X, X, \ldots, X]\!].$$

Writing $X = (x_{ij})$, $T = (t_{\alpha_1 \ldots \alpha_k})$, the entry of the transformed tensor in position $\beta_1, \ldots, \beta_k$ is

$$[\![T; X, X, \ldots, X]\!]_{\beta_1 \ldots \beta_k} \quad = \sum_{\alpha_1, \ldots, \alpha_k} t_{\alpha_1 \ldots \alpha_k} x_{\beta_1 \alpha_1} x_{\beta_2 \alpha_2} \cdots x_{\beta_k \alpha_k}.$$

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

In the next section, we 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 *non-trivial*, i.e. $\mathrm{Stab}_{\mathbb{K}}(T)$ strictly contains $\{\eta I : \eta^k = 1\}$.

We introduce an important notion for stabilizers under congruence, which we will call symmetrically concise. Essentially, it means that for $T \in (\mathbb{K}^m)^{\otimes k}$, there is no subspace $W \subsetneq \mathbb{K}^m$ such that $T \in W^{\otimes k}$. The tensor T has m^k entries and k principal flattenings, which are matrices of size $m \times m^{k-1}$. The ith flattening matrix $T^{(i)}$ has its rows labeled by the ith index of T and its columns labeled by a multi-index from all remaining indices.

Recall that a tensor $T \in (\mathbb{K}^m)^{\otimes k}$ is *concise* if it has flattening ranks $(m, m, \ldots, m)$ [16, Definition 3.1.3.1]. We concatenate the k flattening matrices to form a single matrix of size $m \times km^{k-1}$. A tensor is called *symmetrically concise* if this matrix has full rank m. This notion is weaker than concise for non-symmetric 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 $T \in W^{\otimes 3}$. For symmetric tensors, concise and symmetrically concise are equivalent, because the $m \times km^{k-1}$ matrix consists of k identical blocks of size $m \times m^{k-1}$.

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 non-zero 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 $\mathbb{K}^m$.

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 non-trivial.*

Proof. Since T is not symmetrically concise, there exists a vector $v \in \mathbb{K}^m$ of norm one such that $v^\mathsf{T} T^{(i)} = 0$ for all $i = 1, \ldots, k$. This condition implies $[\![T; I+vv^\mathsf{T}, I+vv^\mathsf{T}, \ldots, I+vv^\mathsf{T}]\!] = T$, and therefore the invertible matrix $I + vv^\mathsf{T}$ is in the stabilizer of the tensor 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 2×4 flattening matrix of T is rank-deficient. This means that T has flattening ranks $(1, 1, 1)$, so T is not concise. However, the 2×12 matrix we obtain by concatenating the three flattening matrices has full rank, hence the tensor T is symmetrically concise. The stabilizer of T is trivial.

We next derive a Jacobian criterion for the stabilizer of a tensor under the congruence action to be finite. This criterion is a sufficient condition. For notational simplicity we state the Jacobian criterion only for order three tensors.

The Jacobian $\nabla f(X) \in \mathbb{K}^{m^3 \times m^2}$ of the function $f(X) = [\![T; X, X, X]\!]$ has entries:

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

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

$$\nabla f(I)_{(i,j,k),(u,v)} \quad = \quad t_{vjk}\delta_{ui} + t_{ivk}\delta_{uj} + t_{ijv}\delta_{uk}. \tag{12}$$

Consider the $m^2 \times m^2$ submatrix J_1 of the Jacobian obtained by setting $k = 1$ in (12). 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 given 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 $f : Z \mapsto [\![T; Z, Z, Z]\!]$ has positive-dimensional fibers. To check that the stabilizer is finite, it suffices to check that ∇f has full rank at a fixed matrix. If the matrix J_1 is invertible then the Jacobian ∇f has full rank at the identity $Z = I$. This implies that the stabilizer of T is a finite subgroup. $\square$

The same conclusion holds if any of the maximal minors of the Jacobian in $\mathbb{K}^{m^k \times m^2}$ is non-zero. We can apply Proposition 3.3 to see that the tensor C_{mono} with entries (7) has finite stabilizer under the congruence action for $m \leq 30$. For example, for $m = 10$ we compute $\det(J_1)^{-1} = 2^{288}3^{160}5^{81}7^{75}11^{96}13^{86}17^{52}19^{35}$. We will see that this implies 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 towards the proof of [1, Conjecture 6.10]. We will also see that, since the determinant of J_1 is very small, there are numerical challenges for learning paths, to be discussed in Section 8.

4 Criteria for Identifiability

We next relate the stabilizer of a core tensor C under the congruence action to the identifiability of path recovery from the signature tensor $[\![C; X, X, \ldots, X]\!]$. Here 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 the 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 expressed 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 the block of non-zero entries. 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 $m \times m \times \cdots \times m$, 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 $T = \sum_{l=1}^r v_l^{(1)} \otimes v_l^{(2)} \otimes \cdots \otimes v_l^{(k)}$ be a minimal decomposition. Assume that a rank one term is non-zero outside of the block, i.e. the coordinate $v_l^{(j)}(\alpha)$ is non-zero for some l, some j, and some index α not contained in the block. The terms $v_l^{(1)} \otimes \ldots \otimes v_l^{(j)}(\alpha) \otimes \cdots \otimes v_l^{(k)}$ sum to zero. However, the order $k-1$ tensors in a minimal decomposition, resulting from removing the jth vector from each rank one term, are linearly independent. This implies $v_l^{(j)}(\alpha) = 0$ for all l, a contradiction. $\square$

We use Lemma 4.1 to relate the stabilizer of the core tensor C under the congruence action to the set of paths with the same signature tensor $[\![C; X, X, \ldots, X]\!]$.

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

$$\left\{ Y \in \mathbb{K}^{d \times m} : [\![C; X, X, \ldots, X]\!] = [\![C; Y, Y, \ldots, Y]\!] \right\} \quad = \quad \left\{ XZ : Z \in \mathrm{Stab}_{\mathbb{K}}(C) \right\}. \quad (13)$$

Proof. Suppose $[\![C; X, \ldots, X]\!] = [\![C; Y, \ldots, Y]\!]$. Let $\tilde{C}$ be the $d \times d \times \cdots \times d$ tensor with entries

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

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

By Lemma 4.1, all minimal decompositions of $\tilde{C}$ come from those of C by adjoining zeros. Let $\tilde{C} = \sum_{l=1}^{r} \tilde{T}_l$ be a minimal decomposition, where $\tilde{T}_l = \tilde{v}_l^{(1)} \otimes \tilde{v}_l^{(2)} \otimes \cdots \otimes \tilde{v}_l^{(k)}$ 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 applying the congruence action by $\tilde{Z}$. This decomposition has rank one terms $[\![\tilde{T}_l; \tilde{Z}, \tilde{Z}, \ldots, \tilde{Z}]\!] = (\tilde{Z}\tilde{v}_l{}^{(1)}) \otimes (\tilde{Z}\tilde{v}_l{}^{(2)}) \otimes \cdots \otimes (\tilde{Z}\tilde{v}_l{}^{(k)})$. 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. □

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 kth order signature $[\![C; X, X, \ldots, X]\!]$. If the stabilizer of C under congruence by $\mathrm{GL}(m, \mathbb{K})$ is trivial up to scale (resp. finite) then rank m matrices X are identifiable (resp. algebraically identifiable) from the signature tensor $[\![C; X, X, \ldots, X]\!]$.*

Proof. Let Y be a matrix in $\mathbb{K}^{d \times m}$ that satisfies $[\![C; X, X, \ldots, X]\!] = [\![C; Y, Y, \ldots, Y]\!]$. 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. If the stabilizer consists of matrices that are trivial up to scale, this implies that Y and X agree up to scale. □

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 4.4. Consider 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

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

The left hand side of (13) is the set of all matrices of the form Y. This set strictly contains the right hand side of (13), 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. The following result gives conditions under which algebraic identifiability of a minimal decomposition implies algebraic identifiability under congruence.

Theorem 4.5. *Let ψ 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, has* $\operatorname{rank}(C) = r$, *and the number δ of minimal decompositions of C is finite. Given a generic matrix $X \in \mathbb{R}^{d\times m}$ with $m \le d$, 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

$$\sigma^{(3)}(X) \;=\; [\![C; X, X, X]\!] \quad = \quad [\![C; Y, Y, Y]\!] \;=\; \sigma^{(3)}(Y). \tag{14}$$

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 as vectors 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 the $m \times m$ matrices Z which stabilize $C' = [\![C; W, W, W]\!]$. This is the third order signature tensor of the path $W\psi$, and it also has δ 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 Λ is diagonal and invertible. Since ψ 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 Λ. $\square$

The following example attains the bound in Theorem 4.5 non-trivially.

Example 4.6. Let $m = d = 2$ and consider the dictionary $\psi = (\psi_1, \psi_2)$ with basis functions $\psi_1(t) = t - 10t^2 + 10t^3$ and $\psi_2(t) = 11t - 20t^2 + 10t^3$. Using (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 computations in `Macaulay2` [13], 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. From the stabilizer, we see that C_ψ is unchanged under swapping the two coordinates ψ_1 and ψ_2.

5 Identifiability for Generic Dictionaries

Our primary motivation for working with third order signatures is that, under reasonable hypotheses on the dictionary ψ and the ambient dimension d, the path $X\psi$ can be recovered uniquely from $\sigma^{(3)}(X)$. We begin by placing this into 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 consider a fixed dictionary ψ and the family of paths represented by the dictionary, i.e. 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 non-uniqueness of path recovery is quantified by the stabilizer of the $m \times m$ core matrix $C = C_\psi$ under the matrix congruence action:

$$\mathrm{Stab}_{\mathbb{R}}(C) \;=\; \big\{\, X \in \mathrm{GL}(m,\mathbb{R}) \;:\; XCX^{\mathsf{T}} = C \,\big\}.$$

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

Proof. The symplectic group $\mathrm{Sp}(m,\mathbb{R})$ 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 semi-simple Lie groups. If m is odd then $\mathrm{Sp}(m,\mathbb{R})$ can be realized by extending the matrices in $\mathrm{Sp}(m-1,\mathbb{R})$ 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 $C = uu^{\mathsf{T}} + Q$ 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 $XQX^{\mathsf{T}} = Q$. 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 $\mathcal{U}_{m,2}$ of signature matrices. We now consider an $m\times m\times m$ core tensor C that is a generic point in the universal variety $\mathcal{U}_{m,3}$. This is the third signature of a dictionary ψ which is generic in the sense of Example 2.3. In the following identifiability result, we allow the field $\mathbb{K}$ to 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 universal 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 ηI, 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 $\mathbb{C}^m$, let Q be a generic skew-symmetric $m \times m$ matrix, and let L be a generic element in the space $\mathrm{Lie}^{[3]}(\mathbb{C}^m)$ of homogeneous Lie polynomial of degree 3. Then

$$C \;=\; \frac{1}{6}P^{\otimes 3} \,+\, \frac{1}{2}(P \otimes Q + Q \otimes P) \,+\, L. \tag{15}$$

Indeed, $P+Q+L$ is a general Lie polynomial of degree ≤ 3, and the right hand side in (15) 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 and extracting the degree components. In particular, the stabilizer of C is contained in the stabilizer of L.

A basis for $\mathrm{Lie}^{[3]}(\mathbb{C}^m)$ is given by taking Lie brackets of all Lyndon words of length three on the alphabet $\{1,2,\ldots,m\}$. 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

$$\mathrm{Lie}^{[3]}(\mathbb{C}^m) \;\simeq\; S_{(2,1)}(\mathbb{C}^m). \tag{16}$$

The dimension $\frac{1}{3}(m^3-m)$ of this irreducible representation 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 (16) is surjective, since the homogeneous Lie polynomial L in (15) can be chosen arbitrarily.

We now apply Popov's classification [2, 22] of irreducible G-modules with non-trivial generic stabilizer. A recent extension to arbitrary fields is due to Garibaldi and Guralnick [12]. 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 C in G is trivial. $\square$

Core tensors of generic dictionaries C_{gen} are symmetrically concise, since the non-symmetrically concise tensors form a positive codimension set. Hence Corollary 4.3 can be applied to generic dictionaries, and we conclude from Theorem 5.2 that the paths 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 a matrix $X \in \mathbb{R}^{d\times m}$ of rank m, the only real solution to $[\![C;X,X,X]\!] = [\![C;Y,Y,Y]\!]$ is $Y = X$.*

We close this section with a remark about equations defining $\mathcal{U}_{m,3}$. We derive these from the parametrization (15). 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

$$\begin{array}{rcl} p_k q_{ij} & = & \frac{1}{2}(c_{kij}+c_{ikj}+c_{ijk}) \;-\; \frac{1}{2}(c_{kji}+c_{jki}+c_{jik}), \\ p_i p_j p_k & = & \quad c_{kij}+c_{ikj}+c_{ijk} \;\;+\;\; \quad c_{kji}+c_{jki}+c_{jik}. \end{array} \tag{17}$$

Here (i,j,k) runs over triples in $\{1,2,\ldots,m\}$. The linear forms on the right hand side (17) span the shuffle linear forms, i.e. linear equations that cut out (16) 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 $\mathrm{vec}(A)$ for its vectorization, the row vector whose coordinates are the entries of A. We vectorize 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:

$$H \quad = \quad P\cdot\big[\,\mathrm{vec}(PP^{\mathsf{T}}) \;\; \mathrm{vec}(Q)\,\big].$$

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 (17).

We now write $H[C]$ for the matrix that is obtained from H by replacing each monomial by the corresponding linear form on the right hand side of (17). Thus $H[C]$ is an $m\times 2m^2$ matrix whose entries are linear forms in C. We have shown that the 2×2 minors of $H[C]$ cut out the variety $\mathcal{U}_{m,3}$. A vast generalization of this fact appears in recent work by Galuppi [11]. His results also imply that the 2×2 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 signature tensor. As before, we take $\mathbb{K}$ to be $\mathbb{R}$ or $\mathbb{C}$. Let C_{axis} be the piecewise linear core tensor in (9) and S any tensor in the orbit

$$\left\{ [\![C_{\text{axis}}; X, X, X]\!] \in \mathbb{K}^{d\times d\times d} \,:\, X \in \mathbb{K}^{d\times m} \right\}.$$

We show that there is a unique matrix X, up to scaling by third root of unity, such that $S = [\![C_{\text{axis}}; X, X, X]\!]$. This proves Conjectures 6.10 and 6.12 in [1] for $m \leq d$. In particular, when 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 $C = C_{\text{axis}} \in \mathbb{K}^{m\times m\times m}$ under congruence. If $e_m = (0, \dots, 0, 1)^\mathsf{T}$ is an eigenvector of X then e_m is also an eigenvector of X^T.*

Proof. Any matrix X in the stabilizer of C also stabilizes the first and second order signatures, up to scaling by third root of unity. The core tensor C corresponds to a path from $(0, \dots, 0)$ to $(1, \dots, 1)$, so the first order signature is $b = (1, \dots, 1)^\mathsf{T}$. 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 equals

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

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

$$XC_2X^\mathsf{T} \;=\; \eta' C_2 \quad\Longrightarrow\quad \eta' C_2^{-1} \;=\; X^\mathsf{T} C_2^{-1} X \quad\Longrightarrow\quad \eta' C_2^{-1} b \;=\; \eta X^\mathsf{T} C_2^{-1} b. \tag{18}$$

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

Consider the matrix obtained from C by multiplying by v along the second index, $D_{ik} = \sum_j c_{ijk} v_j$. 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 $i = k$, 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, $D = \frac{\eta'}{\eta} X D X^\mathsf{T}$.

Suppose that e_m is an eigenvector of X. By the same argument as in (18) we find that $D^{-1} e_m = 6 e_m$ is an eigenvector of X^T. We conclude that e_m is an eigenvector of X^T. $\square$

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

Proof. Let X be in $\mathrm{Stab}_{\mathbb{K}}(C)$. Evaluating $C = [\![C; X, X, X]\!]$ at coordinate (i, j, k) implies

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

Here the constants in (9) were substituted for the entries $c_{\alpha\beta\gamma}$ of C. We can express this equation as the dot product $c_{ijk} = f_{jk} \cdot X_i^\mathsf{T} = \sum_{\alpha=1}^m f_{jk}(\alpha)X_i(\alpha)$, where X_i is the ith row of X and f_{jk} denotes the row vector with α-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 $\mathbb{K}^m$, this implies that f_{jk} is the zero vector, and in particular 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 jth row, we obtain $c_{ijk} = g_{ik} \cdot X_j^\mathsf{T}$, where g_{ik} is the row vector with β-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 mth entry $g_{ik}(m)$ equals

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

Since X stabilizes the first order signature, up to scaling by third root of unity η, the rows of X sum to η, 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 mth 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^T, and therefore the last row of X also has all entries vanishing except the last. This means that X has the block diagonal structure

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

where 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 $\mathbb{K}^{(m-1)\times(m-1)\times(m-1)}$ which arises from C by restricting to indices $1 \leq i, j, k \leq m-1$. Indeed, from $C = [\![C; X, X, X]\!]$ 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

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

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

We deduce from Theorem 6.2 and Corollary 4.3 that real piecewise linear paths in $\mathbb{R}^d$, consisting of at most d steps, can be uniquely recovered from their third order signature. Here we are using the fact that the upper triangular tensor C_{axis} is symmetrically concise.

Corollary 6.3. *Let $m \le d$, let $C = C_{\text{axis}}$, and take $X \in \mathbb{R}^{d\times m}$ of rank m. Then the only real solution to the tensor equation $[\![C; X, X, X]\!] = [\![C; Y, Y, Y]\!]$ is the matrix $Y = X$.*

7 Solving Polynomial Equations

The remainder of this paper is devoted to practical methods for learning a real path from its third order signature. The framework is given by premises (a) and (b) from Section 2. Our model is a fixed dictionary ψ, represented by its core tensor $C = C_\psi$ of format $m \times m \times m$.

The observed data is a tensor S of format $d \times d \times d$. We begin by assuming that S is in the orbit of C, that is, there exists a $d \times m$ matrix X such that $[\![C; X, X, X]\!] = S$. Later we consider approximations of S by a tensor in the orbit. Our aim is to recover the matrix X. This encodes the path $X\psi$ with third order signature $S = \sigma^{(3)}(X\psi)$.

Our earlier sections focused on theoretical results. We proved identifiability for generic dictionaries and piecewise linear paths for $m \le d$, in Corollaries 5.3 and 6.3. In light of these results, and the computations in [1, Section 6], we know that identifiability holds for all values of m and d seen in Table 1. Our experiments confirm these identifiability results.

We first discuss the use of *Gröbner bases* for solving polynomial equations. The tensor identity $[\![C; X, X, X]\!] = S$ is a system of d^3 cubic equations in md unknowns x_{ij} which we solve. Table 1 displays the numbers md and d^3, to illustrate the input size of each instance. To appreciate the complexity of our problem, note that the relevant Bézout number equals 3^{md}. We begin by assuming that the tensor S comes from the dictionary ψ and its entries s_{ijk} are given exactly. Therefore, the equations have a solution X in $\mathbb{R}^{d\times m}$.

The $m\times m\times m$ core tensor C is fixed. 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

$$[\![C; X, X, X]\!] \;=\; [\![C; X_0, X_0, X_0]\!].$$

For each pair (m, d), and each choice of X_0, we computed two Gröbner bases, the first for the core tensor C_{axis} in (9) and the second for the core tensor C_{mono} in (7). 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 2. In all cases for

$m\backslash d$	2	3	4	5	6	7	8
2	4_8	6_{27}	8_{64}	10_{125}	12_{216}	14_{343}	16_{512}
3	6_8	9_{27}	12_{64}	15_{125}	18_{216}	21_{343}	24_{512}
4	8_8	12_{27}	16_{64}	20_{125}	24_{216}	28_{343}	32_{512}
5	10_8	15_{27}	20_{64}	25_{125}	30_{216}	35_{343}	40_{512}
6	12_8	18_{27}	24_{64}	30_{125}	36_{216}	42_{343}	48_{512}
7	14_8	21_{27}	28_{64}	35_{125}	42_{216}	49_{343}	56_{512}

Table 1: Number of unknowns in a system of cubic equations. The index counts the cubics.

$m\backslash d$	2	3	4	5	6	7	8
2	$0,0$	$0,0$	$0,0$	$0,0$	$0,1$	$0,1$	$1,2$
3	NI	$0,0$	$0,0$	$0,1$	$0,1$	$1,3$	$2,5$
4	NI	$0,0$	$0,0$	$1,1$	$2,4$	$5,9$	$12,21$
5	NI	NI	$4,22$	$16,43$	$72,188$	601, F	F, F
6	NI	NI	F, F	F, F	F, F	F, F	F, F
7	NI	NI	F, F	F, F	F, F	F, F	F, F

Table 2: Timings for our Gröbner basis computations in `maple 16`. The first entry is for C_{axis}, the second for C_{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.

which the computation succeeded, there were three standard monomials, corresponding to the matrices ηX_0, where $\eta^3 = 1$. This confirms rational identifiability for the core tensor.

Solving the cubic equations becomes infeasible when the dimensions m and d get larger, or when the data S is inexact or noisy. In these situations we minimize the distance between S and the variety of tensors $[\![C; X, X, X]\!]$ where X ranges over the matrix space $\mathbb{R}^{d\times m}$. We seek the global minimum of the cost function

$$g : \mathbb{R}^{d\times m} \to \mathbb{R}\,, \quad X \mapsto \big\| \, [\![C; X, X, X]\!] \, - \, S \, \big\|^2. \tag{19}$$

Here the norm is the Euclidean norm in tensor space, the *Frobenius norm*. The algebraic complexity of this problem is given by the *ED degree* [9] of the orbit of an $m\times m\times m$ tensor C under multiplication by $d \times m$ matrices. In the special case $m = d$, this is the ED degree of an orbit of the congruence action of $\mathrm{GL}(m, \mathbb{C})$ on the tensor space $V = \mathbb{C}^{m\times m\times m}$. It is an interesting algebra problem to study both the degree and the ED degree of such orbits.

Example 7.1 ($d = m = 2$)**.** Fix the piecewise linear core tensor C_{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. This means that, 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 ηX, with $\eta^3 = 1$.

For general rational data $S \in \mathbb{Q}^{m\times m\times m}$, the ED degree specifies the algebraic degree of the coordinates of the optimal solution to (19). This degree can drop for special tensors S.

Example 7.2 (The skyline path)**.** This path has 13 steps in $\mathbb{R}^2$, given by 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 $[\![C_{\text{axis}}; Y, Y, Y]\!]$, where C_{axis} has format $13 \times 13 \times 13$:

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

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

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

where $a = 3.4952680660622583405476023 9....$ and $b = 1.21844715432391645318468062.....$ The Euclidean distance between the tensors is

$$\|\, [\![C_{\text{axis}}; X^*, X^*, X^*]\!] \,-\, S_{\text{skyline}} \,\| \;\approx\; 3.36. \tag{20}$$

We note that the numbers a and b are algebraic over $\mathbb{Q}$ and hence so is the optimal value in (20). The degree 45 in Example 7.1 is an upper bound on their degree over $\mathbb{Q}$. It turns out that the true degree of both a and b is 12. Their minimal polynomials are found to be

$$\begin{aligned} 200a^{12} - 105375a^9 + 18911292a^6 - 633591952a^3 + 112021056 &\;=\; 0, \\ 350b^{12} - 173035b^9 + 83611584b^6 + 1904373936b^3 - 3717439488 &\;=\; 0. \end{aligned}$$

These polynomials have solvable Galois group, so a and b can be written in radicals over $\mathbb{Q}$.

8 Numerical Identifiability and Condition Numbers

This section addresses the numerical analysis of our problem. We study condition numbers for recovering paths from signature tensors. Our condition numbers will be defined in accordance with the geometric theory in [4]. Our results give both lower bounds and upper bounds on these condition numbers. In Section 9 we shall turn to numerical experiments.

Even if a path is identifiable from its signature tensor in the exact sense of Sections 4–7, different paths may lead to numerically indistinguishable signatures. The condition number of the recovery problem describes how far a path is from being non-identifiable. Following [4], the condition number is determined by the distance to the set of *ill-posed instances*. In our setting, the set $\mathcal{N}(C, d)$ of ill-posed instances consists of the signature tensors $S = [\![C; X, X, X]\!]$ for which $\{Y \in \mathbb{R}^{d\times m} : S = [\![C; Y, Y, Y]\!]\}$ has cardinality greater than 1.

Our optimization problem is to minimize the function g in (19). Beside instabilities due to C being ill-conditioned, numerical optimization has other well-documented drawbacks. Since the objective function g is highly non-convex, 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 in global optimization is out of the scope of this article. See [21].

The condition number of path recovery from $[\![C; X, X, X]\!]$ depends on the matrix X and the tensor C. Let $\|\cdot\|$ denote the Frobenius norm. In order to obtain scale-invariance, we define the *condition number* as

$$\kappa(X, C) \quad = \quad \frac{\|X\|^3 \cdot \|C\|}{\inf_{S \in \mathcal{N}(C,d)} \|[\![C; X, X, X]\!] - S\|}.$$

Writing the condition number as an inverse distance to an ill-posed set is usually stated as a *condition number theorem*, but it can also be used as a definition [4, Section 6.1].

According to Corollary 4.3, any matrix $X \in \mathbb{R}^{d \times m}$ with full rank m is uniquely identifiable from the third order signature $[\![C; X, X, X]\!]$ if C has trivial stabilizer under congruence. For small condition numbers, we expect recovery of X to within good accuracy and the path is called *numerically identifiable*. 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. This poses a problem for numerical computations. When the condition number is infinite, multiple $d \times m$ matrices have the same signature $[\![C; X, X, X]\!]$ and X cannot be recovered uniquely.

We next give an upper bound on the condition number $\kappa(X, C)$, in terms of the singular values of the flattenings of C, and the usual spectral condition number $\kappa(X) = \|X\| \cdot \|X^+\|$ of the rectangular matrix X. Here X^+ denotes the pseudo-inverse of X.

Theorem 8.1. *The condition number of the pair (X, C), consisting of 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*

$$\kappa(X, C) \quad \leq \quad \kappa(X)^3 \left(\frac{\|C\|}{\max(\varsigma_m^{(1)}, \varsigma_m^{(2)}, \varsigma_m^{(3)})} \right), \tag{21}$$

where $\varsigma_m^{(i)}$ denotes the smallest singular value of the ith flattening of the tensor C.

Proof. We aim to bound $\kappa(X, C)^{-1}$, the distance of $[\![C; X, X, X]\!]$ to the locus of non-identifiable tensors, from below. Since C has trivial stabilizer, Corollary 4.3 implies that all non-identifiable tensors must be of the form $[\![C; Y, Y, Y]\!]$ where $Y \in \mathbb{R}^{d \times m}$ is rank-deficient. The flattenings of these tensors are rank deficient, so it suffices to lower-bound the distance of the flattenings $C^{(i)}$ to the much larger set $\{A \in \mathbb{R}^{m \times m^2} : \operatorname{rank}(A) < m\}$. We have

$$\begin{aligned}
\|X\|^3 \cdot \|X^+\|^3 \cdot \|C\| \cdot \kappa(X, C)^{-1} \;&\geq \min_{\operatorname{rank}(A) < m} \|X C^{(i)} (X \otimes X)^\mathsf{T} - A\| \cdot \|X^+\|^3 \\
&= \min_{\operatorname{rank}(A) < m} \|X (C^{(i)} - A)(X \otimes X)^\mathsf{T}\| \cdot \|X^+\|^3 \\
&\geq \min_{\operatorname{rank}(A) < m} \|C^{(i)} - A\| \\
&= \quad \varsigma_m^{(i)},
\end{aligned}$$

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

Let C be a core tensor with trivial stabilizer under congruence. We define the condition number of the tensor C to be the smallest number $\kappa(C)$ satisfying

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

for all $X \in \mathbb{R}^{d \times m}$ of rank m and all d. From Theorem 8.1, we obtain the following result.

Corollary 8.2. *The condition number of an order three core tensor C, with trivial stabilizer under congruence, satisfies*

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

Proof. Divide the inequality (21) by $\kappa(X)^3$. The supremum of the left hand side, as X ranges over all matrices, is equal to $\kappa(C)$. Hence $\kappa(C)$ is bounded by the right hand side. □

We next give an upper bound on the condition number of the piecewise linear dictionary.

Corollary 8.3. *The condition number of the core tensor C_{axis} is bounded above by $6\|C_{\text{axis}}\|$.*

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 $C^{(2)}(C^{(2)})^\mathsf{T} = \frac{1}{36}I + BB^\mathsf{T}$. The singular values of $C^{(2)}$ are the square roots of the eigenvalues of $C^{(2)}(C^{(2)})^\mathsf{T}$. Consider an eigenvector v of BB^T with eigenvalue λ. Then $\lambda \geq 0$ because BB^T is positive semi-definite and v is an eigenvector of $C^{(2)}(C^{(2)})^\mathsf{T}$ with eigenvalue $\frac{1}{36} + \lambda$. Hence the singular values of $C^{(2)}$ are bounded from below by $\frac{1}{6}$. □

Proposition 3.1 shows that the recovery problem is ill-posed if the tensor is not symmetrically concise. Similarly, we expect the condition number to be large if the core tensor is close to not being symmetrically concise. What follows is a numerical analogue to Proposition 3.1.

Proposition 8.4. *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 ς_m is the smallest singular value of $C^{(\text{all})}$ then*

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

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 $S = [\![C; I - vv^\mathsf{T}, I - vv^\mathsf{T}, I - vv^\mathsf{T}]\!]$, where v is the left singular vector corresponding to the singular value ς_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 $v^\mathsf{T} C^{(\text{all})} = \varsigma_m w^\mathsf{T}$ where w is the right singular vector of length $3m^2$, corresponding to singular value ς_m. We define w_i such that w is the stacking of vectors w_1, w_2, w_3 with each w_i of length m^2. Then $v^\mathsf{T} C^{(i)} = \varsigma_m w_i^\mathsf{T}$ hence $\|v^\mathsf{T} C^{(i)}\| = \varsigma_m \|w_i\| \leq \varsigma_m \|w\| \leq \varsigma_m$. We use this to upper bound the distance from C to S, as follows:

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

The condition number satisfies $\kappa(C) \geq \frac{\kappa(X,C)}{\kappa(X)^3}$ for all matrices X of rank m. In particular, setting $X = I$, we have $\kappa(C) \geq \frac{\kappa(I,C)}{\kappa(I)^3}$. By definition of the condition number, we have

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

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

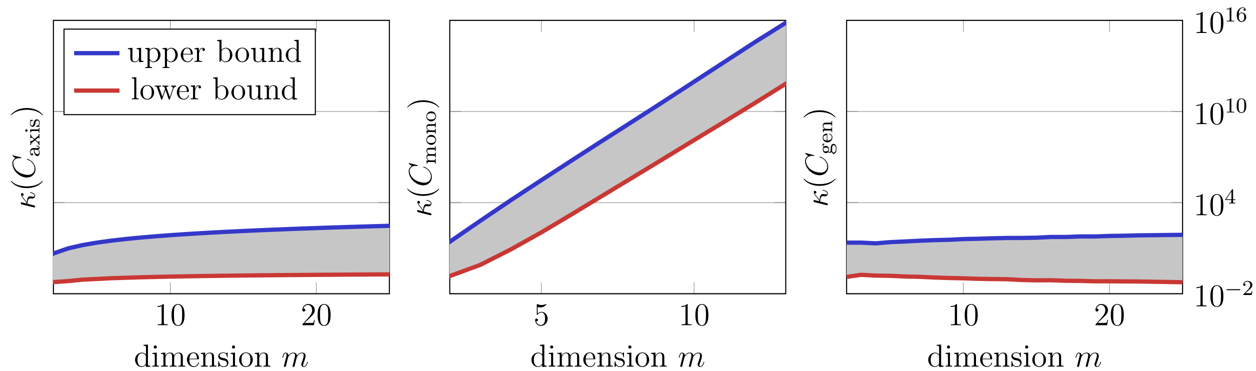

Figure 1: Lower bounds and upper bounds on the condition numbers for the piecewise linear core tensor (left), the monomial core tensor (middle) and generic core tensors (right).

The condition number $\kappa(C)$ quantifies the numerical identifiability of path recovery from paths represented with respect to the dictionary with core tensor C. We have derived informative upper and lower bounds on $\kappa(C)$ 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 8.2. The upper bound is that in Proposition 8.4. We see that the condition number of the monomial dictionary grows exponentially with m. On the other hand, the piecewise linear dictionary is much more stable. Corollary 8.3 shows that the condition number of the piecewise linear dictionary remains below $6\|C_{\text{axis}}\| = 6\sqrt{\frac{m}{36} + \frac{1}{2}\binom{m}{2} + \binom{m}{3}}$. This is seen on the left in Figure 1. Generic dictionaries behave similarly: their condition numbers seem to remain below 100, independently of m. The right diagram in Figure 1 shows the average for 100 generic signature tensors C_{gen}. These were created using the first method in Example 2.3.

We conclude that piecewise linear paths do not have the problem of ill-conditioning, and the same holds for paths from generic dictionaries, in a certain range of m. However, polynomial paths are ill-conditioned even for relatively small values of their degree m. These theoretical results are confirmed by the numerical experiments in the next section.

9 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. This is done by minimizing the function in (19). We performed numerous computational experiments, for a wide range of values of m and d. We considered C_{axis}, C_{mono}, or C_{gen}, for piecewise linear, polynomial, and generic paths. Generic signatures were created using the first method described in Example 2.3.

In order to minimize (19), we first used a solver in `Matlab` that implements the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. This was followed by a trust region method for improved convergence. The trust region method was taken from the `Manopt` toolbox [3].

Tables 3 and 4 summarize our results. For each pair (m, d), we generate random matrices X in $\mathbb{R}^{d\times m}$, with entries $x_{ij} \sim N(0,1)$, representing a path $X\psi$, and we compute $S = \sigma^{(3)}(X)$ up to machine precision. We then minimize the function in (19) using the numerical method described above. Let X^* denote the result of this computation. We declare the recovery to be successful if $X^* - X$ has Frobenius norm less than 10^{-8}. Since m and d are small, using the relative error here would produce very similar results.

Our tables show the percentage of successful recoveries. When recovery was not successful, there are several possible explanations:

1. Convergence was too slow, i.e. the algorithm did not converge in 1000 iterations.
2. Convergence was to a local minimum.
3. Convergence was to a different solution, i.e. the problem was ill-conditioned.

In our experiments, we experienced all three of these problems. We focus on ill-conditioning, and explain Tables 3 and 4 in terms of our theoretical results in Section 8. The subscripts in Table 4 show the failure of convergence that is related to ill-conditioning.

Ill-conditioning was not a problem for piecewise linear paths: recovery of a different matrix with the same signature happened only once in over 10000 experiments. The success rate for piecewise linear paths is over 80% for small m but it becomes gradually worse for larger m. We remark that preliminary computations in `tensorlab` [24] showed excellent recovery of paths for $m = 2, 3$ and $d \geq m + 1$ when the matrix X had entries drawn from the uniform distribution on the interval $[0, 1]$, but for $m \geq 4$ or normally distributed entries, this approach was not tractable. For paths represented by a generic dictionary, the results are also rather good, as shown in the lower chart in Table 3. Just like in the piecewise linear case, we rarely see convergence to different solutions when working with generic dictionaries.

$m\backslash d$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
2	65	85	71	84	80	82	73	80	81	84	87	85	82	84	80	84	88
3		64	83	74	79	87	80	79	72	88	84	83	81	84	79	85	79
4			63	72	64	57	68	79	60	69	72	69	70	71	68	68	78
5				44	52	44	51	58	57	62	67	54	55	68	65	74	65
6					36	42	47	52	57	56	55	47	68	59	57	62	55
7						33	35	43	42	40	46	48	63	59	54	59	60
8							28	30	36	42	57	51	51	50	53	47	57
9								33	33	31	43	47	31	38	39	40	36
10									21	26	34	27	29	26	36	40	34

$m\backslash d$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
2	70	75	80	72	78	73	75	76	69	75	76	80	88	72	75	77	81
3		54	61	69	62	69	56	66	64	63	67	74	68	73	61	64	71
4			47	53	49	49	56	45	47	66	53	61	61	66	69	55	54
5				35	47	44	43	46	51	41	52	48	41	45	45	48	50
6					34	37	42	44	32	42	49	44	39	48	56	62	50
7						22	34	41	31	42	40	38	48	39	43	46	50
8							32	36	26	34	36	39	38	43	41	33	42
9								28	27	27	24	21	29	31	29	37	31
10									26	23	21	22	25	28	39	34	27

Table 3: Percentage of successful recoveries for random piecewise linear paths (top) and random paths represented by generic dictionaries (bottom).

The situation is dramatically worse for the core tensor C_{mono} representing polynomial paths. We know from the middle diagram in Figure 1 that the condition number of C_{mono} grows rapidly with m. Already for $m = 5$, the recovery problem is severely ill-conditioned. Our experiments confirmed this. Table 4 shows recovery rates that are very low, even as we increased the maximum number of iterations to 10000. If a matrix was found then in almost all cases it was different than the original matrix X but its third signature differed from the target signature S by less than 10^{-5}. In summary, the machine precision inaccuracy in the signature leads to large differences in the recovered matrix.

In conclusion, our experimental findings are consistent with the theoretical results on condition numbers in Section 8. We find that, while generic paths are acceptable, piecewise linear dictionaries behave best in numerical algorithms for recovering paths. Piecewise linear paths will also feature in the non-identifiable learning problem in the next section.

$m\backslash d$	2	3	4	5	6	7	8	9	10	11	12	13	14
2	70_6	84_1	85_0	86_0	89_0	84_0	91_0	85_0	85_0	89_0	88_0	89_0	84_0
3		67_{19}	82_8	75_8	84_5	91_2	82_2	84_2	84_1	85_2	85_0	82_2	78_1
4			35_{36}	46_{39}	43_{35}	52_{27}	62_{18}	70_{17}	67_{17}	66_{20}	70_{20}	71_{13}	68_{17}
5				0_{86}	1_{75}	2_{77}	1_{78}	2_{80}	4_{79}	7_{74}	1_{77}	12_{64}	5_{66}

Table 4: The recovery rate for polynomial paths is low once the condition number becomes too big. Subscripts indicate recovery of a matrix that is different from the original one.

10 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 ψ, 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 $\psi : [0,1] \to \mathbb{R}^m$ is given by the integral

$$\mathrm{len}(\psi) \;=\; \int_0^1 \sqrt{\langle \dot{\psi}(t), \dot{\psi}(t) \rangle}\, \mathrm{d}t.$$

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 (8) and the matrix $X = (x_{ij})$, the length is given by the simple formula

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

Note that this function is piecewise differentiable. We can therefore regularize the objective function (19) with a length constraint. This leads to the new function

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

where λ is a parameter. A necessary condition for a minimum of this function is that both the gradient in X and the gradient in λ equal zero. The latter requirement makes sure that any solution X satisfies the signature condition. 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 $\lambda \to -\infty$.

To work around this, we use a simple trick from optimization. We fix λ_0 and minimize

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

Once a minimum X_0 is found, we set $\lambda_1 = 2\lambda_0$ and minimize again with λ_1 and X_0 as a starting point for the iteration. We repeat, setting $\lambda_n = 2\lambda_{n-1}$ until λ_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 [21]. 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. The above experiments were performed many times for these two dictionaries. We report on two examples, for $d = 2$ and for $d = 3$.

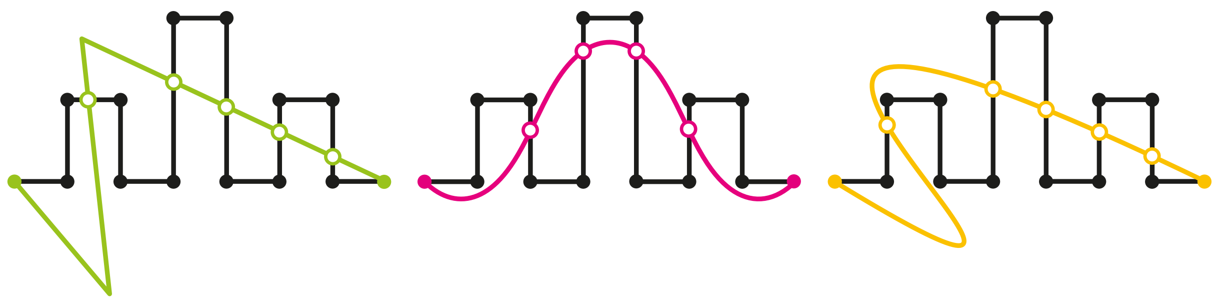

Figure 2: The shortest path with $m = 3$ steps (left), $m = 100$ steps (middle), and a polynomial path of degree 3 (right). All have the same third signature as the skyline path.

Example 10.1 (Skyline path)**.** Consider the skyline path in Example 7.2, with $2 \times 2 \times 2$ signature tensor S_{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_{skyline}.

The shortest piecewise linear path with $m = 3$ steps is depicted in Figure 2 on the left. The middle image shows the shortest path with $m = 100$ steps. This is an approximation to a shortest smooth path with that signature. We also learned polynomial paths using the core tensor C_{mono}, but without length constraints. The right image shows a cubic path.

Example 10.2 ($d = 3$)**.** We here examine the *Klee-Minty path* in $\mathbb{R}^3$. This is the following axis path with 7 steps. It travels along the edges of the unit 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 the Klee-Minty path equals

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

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

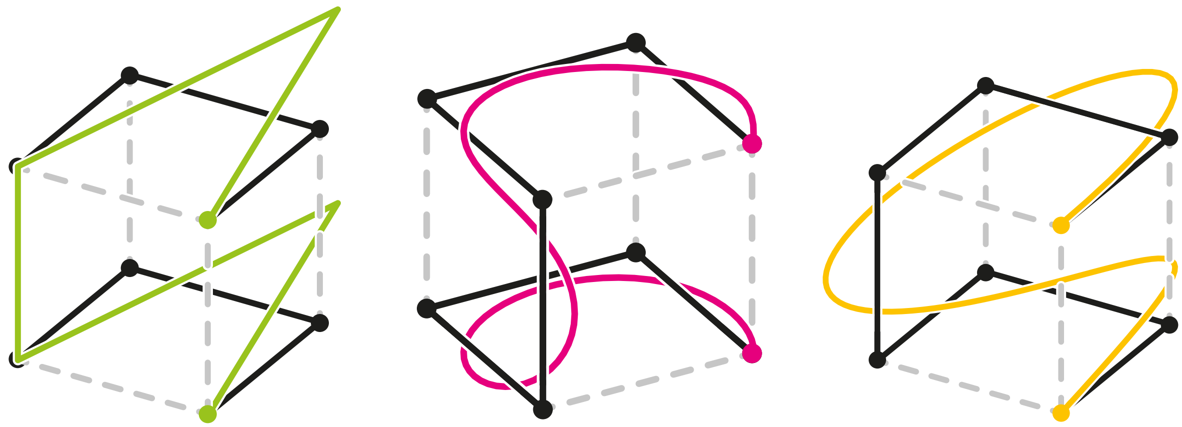

Figure 3: The shortest path with $m = 5$ steps (left) and $m = 100$ steps (middle) having the same third order signature as the Klee-Minty path. On the right we see a polynomial path of degree 5 whose third order signature is close to that of the Klee-Minty path.

For the same reason as above, polynomial paths of degree 5 are expected to fill $\mathcal{U}_{3,3}$. But we did not find any path with signature $S_{\text{kleeminty}}$. A close solution was a matrix X with

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

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 $S_{\text{kleeminty}}$ seems to lie in the set $\mathcal{P}^{\mathbb{R}}_{3,3,5} \backslash \mathcal{P}^{\text{im}}_{3,3,5}$.

Acknowledgments. We thank Robert Guralnick and Vladimir Popov for the references that are used in the proof of Theorem 5.2. We are grateful to Carlos Améndola, Laura Colmenarejo, Joscha Diehl, Francesco Galuppi and André Uschmajew for helpful conversations. Anna Seigal and Bernd Sturmfels were partially supported by the US National Science Foundation (DMS-1419018).

References

[1] C. Améndola, P. Friz and B. Sturmfels: *Varieties of signature tensors*, `arXiv:1804.08325`.

[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) 265–271.

[3] N. Boumal, B. Mishra, P.-A. Absil and R. Sepulchre: *Manopt, a Matlab Toolbox for Optimization on Manifolds*, Journal of Machine Learning Research, **15** (2014) 1455–1459, `manopt.org`.

[4] P. Bürgisser and F. Cucker: *Condition. The Geometry of Numerical Algorithms*, Fundamental Principles of Mathematical Sciences, 349, Springer-Verlag, Heidelberg, 2013.

[5] P. Bürgisser, A. Garg, R. Oliveira, M. Walter and A. Wigderson: *Alternating minimization, scaling algorithms, and the null-cone problem from invariant theory*, `arXiv:1711.08039`.

[6] K.-T. Chen: *Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula*, Annals of Mathematics **65** (1957) 163–178.

[7] I. Chevyrev and A. Kormilitzin: *A primer on the signature method in machine learning*, `arXiv:1603.03788`.

[8] J. Diehl and J. Reizenstein: *Invariants of multidimensional time series based on their iterated-integral signature*, `arXiv:1801.06104`.

[9] J. Draisma, E. Horobet, G. Ottaviani, B. Sturmfels and R. Thomas: *The Euclidean distance degree of an algebraic variety*, Foundations of Computational Mathematics **16** (2016) 99–149.

[10] P. Friz and M. Hairer: *A Course on Rough Paths. With an introduction to regularity structures*, Universitext, Springer, Cham, 2014.

[11] F. Galuppi: *The rough Veronese variety*, Manuscript, September 2018.

[12] S. Garibaldi and R. Guralnick: *Generically free representations II: irreducible representations*, `arXiv:1711.06400`.

[13] D. Grayson and M. Stillman: *Macaulay2, a software system for research in algebraic geometry*, available at `www.math.uiuc.edu/Macaulay2/`.

[14] T. Kolda and B. Bader: *Tensor decompositions and applications*, SIAM Review **51** (2009) 455–500.

[15] 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`.

[16] J.M. Landsberg: *Tensors: Geometry and Applications*, Graduate Studies in Mathematics **128**, American Mathematical Society, Providence, RI, 2012.

[17] T. Lyons and Z. Qian: *System Control and Rough Paths*, Oxford University Press, 2002.

[18] T. Lyons and W. Xu: *Hyperbolic development and inversion of signature*, J. Funct. Anal. **272** (2017) 2933–2955.

[19] T. Lyons and W. Xu: *Inverting the signature of a path*, J. Eur. Math. Soc **20** (2018) 1655–1687.

[20] H. Matsumura and P. Monsky: *On the automorphisms of hypersurfaces*, J. Math. Kyoto Univ. **3** (1963/1964) 347–361.

[21] J. Nocedal and S. J. Wright: *Numerical Optimization*, Springer Verlag, 2006.

[22] A. M. Popov: *Finite stationary subgroups in general position of simple linear Lie groups*, Trans. Mosc. Math. Soc. **48** (1988) 3–63.

[23] L. Tucker: *Some mathematical notes on three-mode factor analysis*, Psychometrika **31** (1966) 279–311.

[24] N. Vervliet, O. Debals, L. Sorber, M. Van Barel and L. De Lathauwer: Tensorlab 3.0, March 2016, available at `https://www.tensorlab.net/`.

Authors' addresses:

Max Pfeffer, MPI-MiS Leipzig `max.pfeffer@mis.mpg.de`
Anna Seigal, UC Berkeley `seigal@berkeley.edu`
Bernd Sturmfels, MPI-MiS Leipzig and UC Berkeley `bernd@mis.mpg.de`