# Equations for hidden Markov models

Alexander Schönhuth

**Abstract**

We will outline novel approaches to derive model invariants for hidden Markov and related models. These approaches are based on a theoretical framework that arises from viewing random processes as elements of the vector space of string functions. Theorems available from that framework then give rise to novel ideas to obtain model invariants for hidden Markov and related models.

## 1 Introduction

In the following, we will outline how to obtain invariants for hidden Markov and related models, based on an approach which, in its most prevalent application, served to solve the identifiability problem for hidden Markov processes (HMPs) in 1992 [13]. Some of its foundations had been layed in the late 50's and early 60's in order to get a grasp of problems related to that of identifying HMPs [5, 11, 6, 7, 8, 12]. The approach can be viewed as being centered around the definition of *finite-dimensional* discrete-time, discrete-valued stochastic processes (referred to as *discrete random processes* in the following)[1]. It Examples of finite-dimensional discrete random processes other than HMPs are quantum random walks (QRWs). QRWs have been brought up mostly to emulate Markov chain related algorithms (e.g. Markov Chain Monte Carlo techniques) on quantum computers [1].

In the following, we will introduce finite-dimensional string functions and formally describe how to view discrete random processes as string functions.

[1] In the literature, finite-dimensional discrete random processes are alternatively referred to as *finitary* [12] or *linearly dependent* [13] processes. In the following, we will stay with the term finite-dimensional (discrete random processes) in accordance with the latest contributions on the topic [14, 10, 18, 19, 20]

We will further provide helpful characterizations and related theorems. In sec. 3 we will determine polynomials that generate the ideal of the invariants of the finite-dimensional model, in the usual sense of algebraic statistics. In sec. 4, we will prove a theorem from which, as a corollary, one obtains a proof of conjecture 11.9 in [3]. This corollary will be listed in sec. 5 where we will draw the connections to the hidden Markov model in more detail. In sec. 6 we will show how to obtain invariants for the Markov models, based on the results of the preceding sections. In sec. 7 we will briefly demonstrate that trace algebras, as well, can be viewed as certain finite-dimensional string functions. Invariants of the finite-dimensional model are relatively easy to obtain

# 2 Preliminaries: String Functions

Detailed proofs and explanations of the following results can be found from [18]. Let $\Sigma^* = \cup_{n\geq 0}\Sigma^n$ denote the set of all strings of finite length over the finite alphabet $\Sigma$ where the word $\Box \in \Sigma^0$ of length $|\Box| = 0$ is the *empty string*. Single letters are usually denoted by $a, b$ whereas strings of arbitrary length are denoted by $v, w$ (for example, $v = a_1...a_n \in \Sigma^n, w = b_1...b_m \in \Sigma^m$ where $a_i, b_j \in \Sigma$). We have the concatenation operation:

$$w \in \Sigma^m, v \in \Sigma^n \quad \Longrightarrow \quad wv \in \Sigma^{m+n}. \tag{1}$$

We denote the *length* of $v \in \Sigma^n$ by $|v| = n$. We now direct our attention to real-valued string functions

$$p: \Sigma^* \longrightarrow \mathbb{R} \tag{2}$$

and further to $\mathbb{R}^{\Sigma^*}$, that is, to the real vector space of string functions over $\Sigma$. The notation $p$ is due to that discrete random processes will be viewed as string functions, which will be described in the following.

## 2.1 Discrete Random Processes as String Functions

Given a discrete random process $(X_t)$ with values in the alphabet $\Sigma$, the prescription

$$p_X(v = a_1...a_n) = \mathbf{P}(\{X_1 = a_1, ..., X_n = a_n\})$$

gives rise to a string function $p_X$ associated with the random process. $p_X(a_1...a_n)$ then just is the probability that the associated random process emits the string $a_1...a_n$ at periods $t = 1, ..., n$. String functions associated with discrete random processes can be characterized as follows.

**Theorem 2.1.** *A string function $p : \Sigma^* \to \mathbb{R}$ is associated with a discrete random process iff the following conditions hold.*

*(a)* $p(v) \geq 0$ *for all* $v \in \Sigma^*$.

*(b)* $\sum_{a\in\Sigma} p(va) = p(v)$ *for all* $v \in \Sigma^*$.

*(c)* $p(\Box) = 1$.

Note that $(b)$ in combination with $(c)$ implies

$$\forall n \geq 0 : \quad \sum_{v\in\Sigma^n} p(v) = 1. \tag{3}$$

**Definition 2.2.** A string function $p : \Sigma^* \to \mathbb{R}$ is called

- *stochastic string function (SSF)* if it is associated with a discrete random process, that is, iff $(a), (b)$ and $(c)$ of theorem 2.1 apply,
- *unconstrained stochastic string function (USSF)* if only $(a)$ and $(b)$ apply (in accordance with the terminology of [16]) and
- *generalized unconstrained stochastic string function (GUSSF)* if only $(b)$ applies.

In the following, the terms (generalized unconstrained) random process and (GU)SSF will be used interchangeably. Furthermore, note that $p(a_1...a_n)$ just is a different notation for $p_{a_1...a_l}$ which was used in [16].

## 2.2 Dimension of String Functions

The following definitions are fundamental for this work.

**Definition 2.3.** Let $p : \Sigma^* \to \mathbb{R}$ be a string function over $\Sigma$. Then

$$\mathcal{P}_p := [p(wv)_{v,w\in\Sigma^*}] \in \mathbb{R}^{\Sigma^*\times\Sigma^*} \tag{4}$$

is called the *Hankel matrix* of $p$ (also called *prediction matrix* in case of a SSF $p$). We define

$$\dim p := \text{rk } \mathcal{P}_p \tag{5}$$

to be the *dimension* of $p$. In case of $\dim p < \infty$ the string function $p$ is said to be *finite-dimensional*.

**Example 2.4.** Let $p : \Sigma^* \to \mathbb{R}$ be a string function over the binary alphabet $\Sigma = \{0, 1\}$.

$$\mathcal{P}_p = \begin{pmatrix} p(\square) & p(0) & p(1) & p(00) & p(01) & p(10) & p(11) & \dots \\ p(0) & p(00) & p(10) & p(000) & p(010) & p(100) & p(110) & \dots \\ p(1) & p(01) & p(11) & p(001) & p(011) & p(101) & p(111) & \dots \\ p(00) & p(000) & p(100) & p(0000) & p(0100) & p(1000) & p(1100) & \dots \\ p(01) & p(001) & p(101) & p(0001) & p(0101) & p(1001) & p(1101) & \dots \\ p(10) & p(010) & p(110) & p(0010) & p(0110) & p(1010) & p(1110) & \dots \\ p(11) & p(011) & p(111) & p(0011) & p(0111) & p(1011) & p(1111) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

then is the Hankel matrix where strings of finite length have been ordered lexicographically. Note that within a row values refer to strings that have the same suffix whereas within a column values refer to strings that have the same prefix. See also [4] for an example of a Hankel matrix.

The following characterization of finite-dimensional string functions is the major source of motivation for this work.

**Theorem 2.5** ([13, 14, 18])**.** *Let $p : \Sigma^* \to \mathbb{R}$ be a string function. Then the following conditions are equivalent.*

*(i) $p$ has dimension at most $d$.*

*(ii) There exist vectors $x, y \in \mathbb{R}^d$ as well as matrices $T_a \in \mathbb{R}^{d\times d}$ for all $a \in \Sigma$ such that*

$$\forall v \in \Sigma^* : \quad p(v = a_1...a_n) = \langle y|T_{a_n}...T_{a_1}|x\rangle. \tag{6}$$

Fully elaborated proofs of theorem 2.5 can be found in [14, 18]. Note that (6) can be transformed to

$$p(v) = \text{tr } T_{a_n}...T_{a_1}C \tag{7}$$

where $C = xy^T \in \mathbb{R}^{d\times d}$.

**Example 2.6.** The most prominent example for finite-dimensional SSFs are hidden Markov chains. Let $p$ be an SSF associated with a hidden Markov chain on $d$ hidden states and output alphabet $\Sigma$. Let $A = (\mathbf{P}(i \to j))_{1 \le i,j \le d}$ be the transition probability matrix, $E_{ia}, 1 \le i \le d, a \in \Sigma$ be the emission probabilities and $\pi$ be the initial probability distribution. We define

$$O_a := \text{diag } (E_{ia}, i = 1, ..., d) \in \mathbb{R}^{d \times d} \tag{8}$$

and further

$$T_a := A^T O_a \in \mathbb{R}^{d \times d}.$$

The $T_a$ together with $y := (1, ..., 1) \in \mathbb{R}^d$, $x := \pi \in \mathbb{R}^d$ then provide a representation corresponding to (6).

We will be particularly interested in finite-dimensional GUSSFs (we recall definition 2.2). The following theorem provides a characterization.

**Theorem 2.7.** *Let $p : \Sigma^* \to \mathbb{R}$ be a string function such that $\dim p \le d$. Then the following two statments are equivalent:*

*(i) $p$ is a GUSSF, that is, $\sum_{a \in \Sigma} p(va) = p(v)$ for all $v \in \Sigma^*$.*

*(ii) There exist vectors $x, y \in \mathbb{R}^d$ as well as matrices $T_a \in \mathbb{R}^{d \times d}$ for all $a \in \Sigma$ such that*

$$\forall v \in \Sigma^* : \quad p(v = a_1 ... a_n) = \langle y | T_{a_n} ... T_{a_1} | x \rangle. \tag{9}$$

*as well as*

$$y^T \sum_{a \in \Sigma} T_a = y^T \tag{10}$$

*translating to that $y$ is an eigenvector of the eigenvalue 1 of the transpose of $\sum_{a \in \Sigma} T_a$.*

In the following, we will write

$$T_v := T_{a_n} ... T_{a_1}, T_w = T_{b_m} ... T_{b_1} \tag{11}$$

in case of $v = a_1 ... a_n \in \Sigma^n, w = b_1 ... b_m \in \Sigma^m$.

*Proof.* The obvious direction is "$\Leftarrow$":

$$\begin{aligned}\sum_{a\in\Sigma} p(va) &\overset{(9)}{=} \sum_{a\in\Sigma}\langle y|T_aT_v|x\rangle = \langle y|\sum_{a\in\Sigma} T_a|T_vx\rangle \\ &\overset{(10)}{=} \langle y|T_v|x\rangle = p(v).\end{aligned} \tag{12}$$

For "$\Rightarrow$", let $d^* := \dim p \leq d$ be the actual dimension of $p$. According to theorem 2.5, we find matrices $\tilde{T}_a \in \mathbb{R}^{d^*\times d^*}, a\in\Sigma$ and vectors $\tilde{x},\tilde{y}\in\mathbb{R}^{d^*}$ such that

$$\forall z \in \mathbb{R}^{d^*}: \quad \langle\tilde{y}|\sum_{a\in\Sigma}\tilde{T}_a|z\rangle = \langle\tilde{y}|z\rangle. \tag{13}$$

In case of $d^* = d$ we will have proven the claim by putting $T_a := \tilde{T}_a, x = \tilde{x}, y = \tilde{y}$. In case of $d^* < d$ we will obtain suitable matrices $T_a \in \mathbb{R}^{d\times d}$ and vectors $x, y \in \mathbb{R}^d$ by putting

$$(T_a)_{ij}, x_i, y_i := \begin{cases} (\tilde{T}_a)_{ij}, \tilde{x}_i, \tilde{y}_i & 1 \leq i,j \leq d^* \\ 0 & \text{else}\end{cases}. \tag{14}$$

From theorem 2.5 we obtain matrices $\tilde{T}_a \in \mathbb{R}^{d^*\times d^*}, a\in\Sigma$ and vectors $\tilde{x},\tilde{y}\in\mathbb{R}^{d^*}$ such that

$$p(v = a_1...a_n) = \langle y|T_v|x\rangle. \tag{15}$$

Condition $(i)$ then implies that

$$\langle y|(\sum_{a\in\Sigma}T_a)T_v|x\rangle = \sum_{a\in\Sigma}\langle y|T_aT_v|x\rangle \overset{(15)}{=} \sum_{a\in\Sigma}p(va) \overset{(i)}{=} p(v) = \langle y|T_v|x\rangle. \tag{16}$$

It remains to show that

$$\text{span}\{T_vx \,|\, v\in\Sigma^*\} = \mathbb{R}^{d^*}. \tag{17}$$

However, assuming the contrary would lead to the contradiction

$$\begin{aligned} d^* = \dim p = \text{rk }[p(wv)]_{v,w\in\Sigma^*} &= \text{rk }[\langle y|T_vT_w|x\rangle]_{v,w\in\Sigma^*} \\ &\leq \dim\text{span}\{T_wx\,|\,w\in\Sigma^*\} < d^*. \quad \diamond\end{aligned} \tag{18}$$

Matrices $T_a$ can be computationally determined according to a procedure which we will describe in the following. Therefore, for a string function $p$, we introduce the notation

$$\begin{array}{rccc} p_v: & \Sigma^* & \to & \mathbb{R} \\ & w & \mapsto & p(wv)\end{array} \text{ resp. } \begin{array}{rccc} p^w: & \Sigma^* & \to & \mathbb{R} \\ & v & \mapsto & p(wv)\end{array}. \tag{19}$$

That is, the $p_v$ resp. $p^v$ are the row resp. column vectors of the Hankel matrix $\mathcal{P}_p$. These are string functions in their own right. Note that $p_\Box = p^\Box = p$. Moreover, note that in case of a stochastic process $p$ s.t. $p(w = b_1 ... b_m) \neq 0$ it holds that

$$\frac{1}{p(w)} p^w(v = a_1 ... a_l) \\ = \mathbf{P}(\{X_{l'+1} = a_1, ..., X_{l'+l} = a_l\} \,|\, \{X_1 = b_1, ..., X_{l'} = b_{l'}\}). \quad (20)$$

Therefore, $\frac{1}{p(w)} p^w$ is just the discrete random process being governed by the probabilities of $p$ conditioned on that $w$ has already been emitted.

The following is a generic algorithmic strategy to infer matrices $T_a \in \mathbb{R}^{d \times d}$ and $x, y \in \mathbb{R}^d$ corresponding to (6) from a finite-dimensional Hankel matrix. At this point, the algorithm needs the entire string function $p$ as an input. We will explain how to obtain a practical version of this generic strategy later in this section.

**Algorithm 2.8.**

**Input**: A string function $p$ such that $\dim p = d < \infty$.
**Output**: Matrices $T_a \in \mathbb{R}^{d \times d}, a \in \Sigma$ and vectors $x, y \in \mathbb{R}^d$ such that

$$p(v = a_1 ... a_n) = \text{tr } T_{a_n} ... T_{a_1} x y^T. \quad (21)$$

1. Determine words $v_1, ..., v_d$ resp. $w_1, ..., w_d$ such that the $f_{v_i}$ resp. the $g_{w_j}$ span the row resp. column space of $\mathcal{P}_p$. Hence the matrix

$$V := [p(w_j v_i)]_{1 \leq i,j \leq d} \quad (22)$$

has full rank $d = \dim p$.

2. Denote by $V_i$ resp. $V^j$ the $i$-th row resp. the $j$-th column of $V$ and define

$$x = (x_1, ..., x_d)^T := (p(v_1), ..., p(v_d))^T \quad (23)$$

and $y = (y_1, ..., y_d) \in \mathbb{R}^d$ such that

$$(p(v_1), ..., p(v_d)) = \sum_{i=1}^{d} y_i V_i \quad (24)$$

which can be done as $p_\Box = p$ (the uppermost row of the Hankel matrix) is linearly dependent of the $p_{v_i}$ (the basis of the row space of the Hankel matrix).

3. For each $a \in \Sigma$, determine matrices
$$W_a := [p(w_j a v_i)]_{1 \leq i,j \leq d}. \in \mathbb{R}^{d \times d}. \tag{25}$$

4. One can then show that $x, y$ and $T_a := (W_a V^{-1}), a \in \Sigma$ are as needed for theorem 2.5.

Clearly, the driving question behind algorithm 2.8 is its practicability. A first clue to this is the following theorem. Therefore, we set
$$\Sigma^{\leq n} := \cup_{t=0}^{n} \Sigma^t \tag{26}$$
to be the set of all strings of length at most $n$ and
$$\mathcal{P}_{p,n,m} := [p(wv)]_{|v| \leq n, |w| \leq m} \in \mathbb{R}^{\Sigma^{\leq n} \times \Sigma^{\leq m}}. \tag{27}$$
to be the finite minor of the Hankel matrix referring to row resp. column vectors indexed by strings of length at most $n$ resp. $m$.

**Theorem 2.9.** *Let $p : \Sigma \to \mathbb{R}$ be a string function such that $\dim p \leq d$. Then it holds that*
$$\dim p = \text{rk } \mathcal{P}_{p,d-1,d-1}. \tag{28}$$

This means that, given an upper bound $d$ on the dimension of $p$, the dimension of $p$ can be determined by inspecting the finite-dimensional matrix $\mathcal{P}_{p,d-1,d-1}$. See [18] for a proof. Note, however, that the size of $\mathcal{P}_{p,d-1,d-1}$ is exponential in $d$ such that naive approaches to determining $V$ (22) would result in exponential runtime. The final clue to the practicability of algorithm 2.8 is an efficient algorithm to determine $V$ which has recently been presented [19]. The algorithm applies in case one is provided with an arbitrary generating system of the row or column space of $\mathcal{P}_p$. Corresponding generating systems emerge naturally for finite-dimensional processes of interests, in particular for hidden Markov processes and also for quantum random walks.

A consequence of theorem 2.9 is

**Theorem 2.10** ([18])**.** *Let $p$ be a string function such that* $\dim p \leq d$*. Then $p$ is uniquely determined by the values*

$$p(v), \quad |v| \leq 2d-1. \tag{29}$$

*Proof Sketch.* The idea is, given two string functions $p_1, p_2$ where $\dim p_1, \dim p_2 \leq d$ which coincide on strings of length up to $2d-1$, to determine matrices $T_a$ and vectors $x, y$ as in theorem 2.5 according to algorithm 2.8 for both $p_1$ and $p_2$. Thanks to theorem 2.9, in algorithm 2.8, $V$ can be determined by inspecting values of $p$ at strings of length at most $2d-2$ in $\mathcal{P}_{p,d-1,d-1}$ and, subsequently, by inspecting strings of length at most $2(d-1)+1 = 2d-1$ in order to obtain the $W_a$. As $p_1$ and $p_2$ coincide on strings of length $2d-1$, this will result in the same $V$ and $W_a$. Hence $p_1 = p_2$.$\diamond$

The following corollary is an obvious consequence of theorem 2.10 due to property $(b)$ from theorem 2.1. However, it had been well-known already before. See e.g. [15, 4] and the references therein.

**Corollary 2.11** ([15, 4])**.** *A GUSSF $p$ such that* $\dim p \leq d$ *is uniquely determined by the values*

$$p(v), \quad |v| = 2d-1. \tag{30}$$

*In other words, a discrete random process whose dimension can be upper bounded by $d$ is uniquely determined by its probability distribution over the strings of length $2d-1$.*

**Remark 2.12.** Note that for a string function $p$ with $\dim p \leq d < \infty$, rows and columns of the Hankel matrix indexed by strings of length at least $d$ must necessarily be linearly dependent of their counterparts referring to strings of length at most $d-1$. These observations are crucial for the core result of the following section.

# 3 Finite-Dimensional Models

Finite-dimensional models over $\Sigma$ are defined to be the polynomial maps

$$\begin{array}{cccc} \mathbf{g}_{n,d}: & \mathcal{S}^d \subset \mathbb{C}^{|\Sigma|d^2+2d} & \longrightarrow & \mathbb{C}^{|\Sigma|^n} \\ & ((T_a)_{a\in\Sigma}), x, y) & \mapsto & (\langle y|T_{a_n}...T_{a_1}|x\rangle)_{v=a_1...a_n\in\Sigma^n}. \end{array} \tag{31}$$

where

$$((T_a)_{a\in\Sigma}), x, y) \in \mathcal{S}^d \quad :\Leftrightarrow \quad y^T \sum_{a\in\Sigma} T_a = y^T. \tag{32}$$

According to theorem 2.7, $\mathcal{S}^d$ comprises precisely the parameterizations of the generalized unconstrained random processes of dimension at most $d$. Obviously,

$$\mathcal{S}^d \cong \mathbb{C}^{(|\Sigma|-1)d^2+d(d-1)+2d}. \tag{33}$$

Therefore, the Zariski closure of image $(\mathbf{g}_{n,d})$ is an irreducible variety.

In the following, we will make use of the polynomial map (31) to derive a set-theoretic theorem with a strong view towards the invariants of the Zariski closure of the image of $\mathbf{g}_{n,d}$. In case of $n \geq 2d-1$, invariants for the image of $\mathbf{g}_{n,d}$ can be derived by inspection of the Hankel matrix. As in (27), let $\mathcal{P}_{p,n,m}$ be the partial Hankel matrix that is filled with all values $p(wv)$ such that $|v| \leq n, |w| \leq m$.

**Theorem 3.1.** *Let $n \geq 2d-1$ and $(p(v))_{v\in\Sigma^n}$ be an (unconstrained) probability distribution. Then it holds that*

$$(p(v))_{v\in\Sigma^n} \in \text{image } (\mathbf{g}_{n,d})$$

*if and only if the following two conditions apply where, in case of $|u| < n$,*

$$p(u) = \sum_{u\in\Sigma^{n-k}} p(uv). \tag{34}$$

*(a)*

$$\det\ [p(w_j v_i)]_{1\leq i,j\leq d+1} = 0 \tag{35}$$

*for all choices of words $v_1, ...v_{d+1}, w_1, ...w_{d+1}$ of length at most $d-1$, which can be equivalently put as*

$$\text{rk } \mathcal{P}_{p,d-1,d-1} \leq d \tag{36}$$

*(b)*

$$\text{rk } \mathcal{P}_{p,\lceil\frac{n}{2}\rceil,\lfloor\frac{n}{2}\rfloor} = \text{rk } \mathcal{P}_{p,\lfloor\frac{n}{2}\rfloor,\lceil\frac{n}{2}\rceil} = \text{rk } \mathcal{P}_{p,d-1,d-1} \tag{37}$$

(37) states that rows resp. columns in $\mathcal{P}_{p,\lceil\frac{n}{2}\rceil,\lfloor\frac{n}{2}\rfloor}$ and $\mathcal{P}_{p,\lfloor\frac{n}{2}\rfloor,\lceil\frac{n}{2}\rceil}$ referring to row strings $v$ resp. column strings $w$ where $|v|, |w| \geq d$ are linearly dependent of their counterparts in $\mathcal{P}_{p,\lceil\frac{n}{2}\rceil,\lfloor\frac{n}{2}\rfloor}$ and $\mathcal{P}_{p,\lfloor\frac{n}{2}\rfloor,\lceil\frac{n}{2}\rceil}$ that refer to row resp. column strings of length at most $d-1$.

*Proof.* "$\Rightarrow$": Let $(p(v))_{v\in\Sigma^n}$ be in the image of $\mathbf{g}_{n,d}$. Theorem 2.5 states that the Hankel matrix $\mathcal{P}_p$ of $p$ has rank at most $d$. This implies $(a)$ as it just expresses that some Hankel matrix minors of size $d+1$ do not have full rank.

Theorem 2.9 then states that bases of the row resp. the column space of $\mathcal{P}_p$ can be obtained by inspecting row resp. column vectors referring to strings of length at most $d-1$ which implies $(b)$.

"$\Leftarrow$": Let $(p(u))_{u\in\Sigma^n}$ s.t. $(a), (b)$ apply. In order to prove that $(p(u))_{u\in\Sigma^n} \in$ image $\mathbf{g}_{n,d}$, we have to provide a parameterization $((T_a)_{a\in\Sigma}, x, y) \in \mathcal{S}^d$ such that

$$p(u = a_1...a_n) = y^T T_{a_n}...T_{a_1} x \tag{38}$$

for all strings $u \in \Sigma^n$. Therefore, we will provide a parameterization $((T_a)_{a\in\Sigma}, x, y) \in \mathbb{C}^{|\Sigma|d^2+2d}$ such that

$$p(u = a_1...a_k) = y^T T_{a_k}...T_{a_1} x \tag{39}$$

for all strings $u$ such that $|u| \leq n$ where $p(u)$ is defined according to (34) in case of $|u| < n$. By this definition of $p(u), |u| < n$, it is straightforward to show that $((T_a)_{a\in\Sigma}, x, y) \in \mathcal{S}^d$ which completes the proof. Furthermore, note that it suffices to provide a parameterization $((T_a)_{a\in\Sigma}, x, y) \in \mathcal{S}^{d^*}$ for arbitrary $d^* \leq d$ since, in case of $d^* < d$, we extend the $T_a$ as well as $x, y$ by zero entries to obtain a $d$-dimensional parametrization from $\mathcal{S}^d$. Combining these facts, we have to show that, for suitable $d^* \leq d$, there are matrices $T_a \in \mathbb{R}^{d^*\times d^*}$ and vectors $x, y \in \mathbb{R}^{d^*}$ such that (39) holds.

We obtain the desired parameterization $((T_a)_{a\in\Sigma}, x, y)$ according to the ideas of algorithm 2.8. First, determine strings $v_1, ..., v_{d^*}$ and $w_1, ..., w_{d^*}$ of length at most $d-1$ such that

$$V := [p(w_j v_i)]_{1\leq i,j\leq d^*} \tag{40}$$

has full rank $d^* := \text{rk } \mathcal{P}_{p,d-1,d-1} \leq d$. We define

$$x = (x_1, ..., x_{d^*})^T := (p(v_1), ..., p(v_{d^*}))^T \tag{41}$$

and $y = (y_1, ..., y_{d^*}) \in \mathbb{R}^{d^*}$ such that

$$(p(v_1), ..., p(v_{d^*})) = \sum_{i=1}^{d^*} y_i V_i \tag{42}$$

where $V_i = (p(v_i w_1), ..., p(v_i w_{d^*})^T$ is the $i$-th row of $V$ which can be done since the uppermost row of $\mathcal{P}_{p,n,d-1}$ is linearly dependent of the rows referring to the strings $v_i$. Furthermore, for each $a \in \Sigma$, we determine matrices

$$W_a := [p(w_j a v_i)]_{1 \leq i,j \leq d^*}. \in \mathbb{R}^{d^* \times d^*} \tag{43}$$

Note that probabilities in $W_a$ may refer to strings of length up to $2d - 1$ which establishes the necessity of the assumption $n \geq 2d - 1$. We then claim that defining

$$T_a := W_a V^{-1} \tag{44}$$

gives rise to the desired parametrization in terms of (39). We will obtain an easy proof of this claim by three elementary lemmata.

**Lemma 3.2.** *For all $v, w \in \Sigma^*$ such that $|wv| \leq \lceil \frac{n}{2} \rceil$ ($T_v = T_{a_k} ... T_{a_1}, v = a_1 ... a_k \in \Sigma^k$):*

$$T_v \begin{pmatrix} p(wv_1) \\ \vdots \\ p(wv_{d^*}) \end{pmatrix} = \begin{pmatrix} p(wvv_1) \\ \vdots \\ p(wvv_{d^*}) \end{pmatrix} \tag{45}$$

**Proof of lemma 3.2**: Note first that $|v_i| \leq d - 1 \leq \frac{2d-1}{2} \leq \frac{n}{2}$ which implies $|wvv_i| \leq n$. As $(p(wv_1), ..., p(wv_{d^*}))^T$ is contained in the column space of $V$ it suffices to show the statement for $w = w_j$. We do this by induction on $|v|$:
$|v| = 1$:

$$T_a \begin{pmatrix} p(w_j v_1) \\ \vdots \\ p(w_j v_{d^*}) \end{pmatrix} = W_a V^{-1} \begin{pmatrix} p(w_j v_1) \\ \vdots \\ p(w_j v_{d^*}) \end{pmatrix} = W_a e_j = \begin{pmatrix} p(w_j a v_1) \\ \vdots \\ p(w_j a v_{d^*}) \end{pmatrix}. \tag{46}$$

$|v| \to |v|+1$: Let $\tilde{v} = av$ where $a \in \Sigma$.

$$\begin{aligned} T_{\tilde{v}} \begin{pmatrix} p(w_j v_1) \\ \vdots \\ p(w_j v_{d^*}) \end{pmatrix} &= T_v T_a \begin{pmatrix} p(w_j v_1) \\ \vdots \\ p(w_j v_{d^*}) \end{pmatrix} \overset{|v|=1}{=} T_v \begin{pmatrix} p(w_j a v_1) \\ \vdots \\ p(w_j a v_{d^*}) \end{pmatrix} \\ &\overset{(*)}{=} \begin{pmatrix} p(w_j v a v_1) \\ \vdots \\ p(w_j v a v_{d^*}) \end{pmatrix} = \begin{pmatrix} p(w_j \tilde{v} v_1) \\ \vdots \\ p(w_j \tilde{v} v_{d^*}) \end{pmatrix} \end{aligned} \tag{47}$$

where $(*)$ follows from the induction hypothesis. $\diamond$

**Lemma 3.3.** *For all $v, w \in \Sigma^*$ such that $|w|, |v| \leq \lceil \frac{n}{2} \rceil, |wv| \leq n$ ($T_v = T_{a_k} \ldots T_{a_1}, v = a_1 \ldots a_k \in \Sigma^k$):*

$$y^T T_v \begin{pmatrix} p(wv_1) \\ \vdots \\ p(wv_{d^*}) \end{pmatrix} = p(wv). \tag{48}$$

**Proof of lemma 3.3**: Note that the columns in $\mathcal{P}_{p, \lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$ resp. $\mathcal{P}_{p, \lceil \frac{n}{2} \rceil, \lfloor \frac{n}{2} \rfloor}$ referring to $w$ is contained in the span of the columns referring to the $w_j$'s, according to the choice of the $w_j$. Therefore, it suffices to show the statement for $w = w_j$. We do this by induction on $|v|$:
$|v| = 0$ ($v = \Box, T_\Box = Id$):

$$y^T T_\Box \begin{pmatrix} p(w_j v_1) \\ \vdots \\ p(w_j v_{d^*}) \end{pmatrix} = y^T \begin{pmatrix} p(w_j v_1) \\ \vdots \\ p(w_j v_{d^*}) \end{pmatrix} = p(w_j) \tag{49}$$

follows from the choice of $y$.
$|v| \to |v|+1$: Let $\tilde{v} = av, a \in \Sigma$.

$$\begin{aligned} y^T T_{\tilde{v}} \begin{pmatrix} p(w_j v_1) \\ \vdots \\ p(w_j v_{d^*}) \end{pmatrix} &= y^T T_v T_a \begin{pmatrix} p(w_j v_1) \\ \vdots \\ p(w_j v_{d^*}) \end{pmatrix} \\ &\overset{L.\,3.2}{=} y^T T_v \begin{pmatrix} p(w_j a v_1) \\ \vdots \\ p(w_j a v_{d^*}) \end{pmatrix} \overset{(*)}{=} p(wav) = p(w\tilde{v}) \end{aligned} \tag{50}$$

where $(*)$ follows from the induction hypothesis. $\diamond$

**Proof of theorem 3.1 (cont.)**: Let $u \in \Sigma^*$ such that $|u| \leq n$. Split $u = wv$ into two strings $w, v$ such that $|w|, |v| \leq \lceil \frac{n}{2} \rceil$. We compute

$$\begin{aligned} y^T T_u x = y^T T_v T_w x = y^T T_v T_w \begin{pmatrix} p(v_1) \\ \vdots \\ p(v_{d^*}) \end{pmatrix} &= y^T T_v T_w y^T T_v T_w \begin{pmatrix} p(\Box v_1) \\ \vdots \\ p(\Box v_{d^*}) \end{pmatrix} \\ \overset{L.\ 3.2, |w\Box| \leq \lceil \frac{n}{2} \rceil}{=} y^T T_v \begin{pmatrix} p(wv_1) \\ \vdots \\ p(wv_{d^*}) \end{pmatrix} &\overset{L.\ 3.3}{=} p(wv) = p(u) \end{aligned} \tag{51}$$

where we have replaced $v$ resp. $w$ of lemma 3.2 by $w$ resp. $\Box$ here in order to obtain the fourth equation. $\diamond$

Due to theorem 3.1, invariants that are induced by conditions $(a)$ and $(b)$ fully describe the finite-dimensional model $\mathbf{g}_{n,d}$ for $n \geq 2d - 1$, hence generate the ideal of model invariants.

**Example 3.4.** Consider

$$\mathcal{P}_{p,4,2} = \begin{pmatrix} p(\Box) & p(0) & p(1) & p(00) & p(01) & p(10) & p(11) \\ p(0) & p(00) & p(10) & p(000) & p(010) & p(100) & p(110) \\ p(1) & p(01) & p(11) & p(001) & p(011) & p(101) & p(111) \\ p(00) & p(000) & p(100) & p(0000) & p(0100) & p(1000) & p(1100) \\ p(01) & p(001) & p(101) & p(0001) & p(0101) & p(1001) & p(1101) \\ p(10) & p(010) & p(110) & p(0010) & p(0110) & p(1010) & p(1110) \\ p(11) & p(011) & p(111) & p(0011) & p(0111) & p(1011) & p(1111) \end{pmatrix}$$

where $\Sigma = \{0, 1\}$. Condition $(a)$ then translates to the only equation

$$\det \begin{pmatrix} p(\Box) & p(0) & p(1) \\ p(0) & p(00) & p(10) \\ p(1) & p(01) & p(11) \end{pmatrix} = 0.$$

The column conditions in $(b)$ can be stated as follows:

$$\begin{pmatrix} p(00) \\ p(000) \\ p(001) \\ p(0000) \\ p(0001) \\ p(0010) \\ p(0011) \end{pmatrix}, \begin{pmatrix} p(01) \\ p(010) \\ p(011) \\ p(0100) \\ p(0101) \\ p(0110) \\ p(0111) \end{pmatrix}, \begin{pmatrix} p(10) \\ p(100) \\ p(101) \\ p(1000) \\ p(1001) \\ p(1010) \\ p(1011) \end{pmatrix}, \begin{pmatrix} p(11) \\ p(110) \\ p(111) \\ p(1100) \\ p(1101) \\ p(1110) \\ p(1111) \end{pmatrix}$$

$$\in \text{span}\{ \begin{pmatrix} p(\Box) \\ p(0) \\ p(1) \\ p(00) \\ p(01) \\ p(10) \\ p(11) \end{pmatrix}, \begin{pmatrix} p(0) \\ p(00) \\ p(01) \\ p(000) \\ p(001) \\ p(010) \\ p(011) \end{pmatrix} \begin{pmatrix} p(1) \\ p(10) \\ p(11) \\ p(100) \\ p(101) \\ p(110) \\ p(111) \end{pmatrix} \}$$

The row conditions are completely analogous to the column conditions.

Clearly, invariants induced by $(b)$ refer to polynomial rings

$$K[X_{ij}, Y_i, 1 \leq i \leq M, 1 \leq j \leq N] \tag{52}$$

and the smallest varieties therein that contain all points $x_{ij}, y_i$ such that $(y_1, ..., y_M)$ is linearly dependent of $(x_{11}, ..., x_{M1}), ..., (x_{1N}, ..., x_{MN})$. The Zariski closure of the image of $\mathbf{g}_{n,d}$ being an irreducible variety leads us to the following conjecture.

**Conjecture 3.5.** *Let $n \geq 2d - 1$.*

$$(p(v))_{v \in \Sigma^n} \in \overline{\text{image } (\mathbf{g}_{n,d})}$$

*if and only if*

$$\det\ [p(w_j v_i)]_{1 \leq i,j \leq d+1} = 0$$

*for all choices of words $v_1, ...v_{d+1}, w_1, ...w_{d+1}$ such that $|w_j v_i| \leq n$.*

**Remark 3.6.** The finite-dimensional models have to be handled with certain *care.* Even if a (unconstrained) probability distribution is in the image of $\mathbf{g}_{n,d}$ the finite-dimensional string function giving rise to it might not be an (unconstrained) stochastic process, meaning that the string function is not necessarily non-negative, since values referring to longer strings as computed according to (6) might be negative. It is one of the *big open problems* of the theory of finite-dimensional processes how to algorithmically determine whether a set of matrices as in (6) gives rise to a non-negative string function.

# 4 String Length Complexity

In this section, we will prove a set-theoretical theorem an ideal-theoretical counterpart of which would yield, as a corollary, a proof of conjecture 11.9, [3]. The theorem may be of interest in its own right, as the assumptions to be met by the models under considerations are fairly mild.

Roughly speaking, an ideal-theoretical extension of the theorem would be about how to lift sets of generators for models describing distributions over strings of length $n$ to generators for distributions over strings of length $n+1$, given that $n$ is greater than the *string length complexity* of the underlying models.

In the following,

$$\mathcal{M} \subset \mathbb{R}^{\Sigma^*} \tag{53}$$

is a class of USSFs.

**Definition 4.1.** Let $\mathcal{M} \subset \mathbb{R}^{\Sigma^*}$ be a class of USSFs. We define the *string length complexity* of $\mathcal{M}$ to be

$$\begin{aligned}&\text{SLC }(\mathcal{M}) := \\ &\qquad \inf\{N \in \mathbb{N} \mid p_1, p_2 \in \mathcal{M} :\ (p_1)_{|\Sigma^n} = (p_2)_{|\Sigma^n} \ \Rightarrow\ p_1 = p_2\}.\end{aligned} \tag{54}$$

That is, members of $\mathcal{M}$ are uniquely determined by their distributions over strings of length SLC $(\mathcal{M})$.

Given a class of USSFs, let

$$\mathcal{M}_n := \{(p(v))_{v\in\Sigma^n} \,|\, p \in \mathcal{M}\} \tag{55}$$

be the set of distributions over strings of length $n$ that are induced by the members of $\mathcal{M}$. In case of SLC $(\mathcal{M}) = n$ the map

$$\begin{array}{rccc}\pi_{\Sigma^n}: & \mathcal{M} & \longrightarrow & \mathcal{M}_n \\ & p & \mapsto & p_{|\Sigma^n} = (p(v))_{v\in\Sigma^n}\end{array} \tag{56}$$

is one-to-one.

**Theorem 4.2.** *Let $\mathcal{M}$ be a class of unconstrained random processes such that*

*(i)*

$$\text{SLC } (\mathcal{M}) \leq n - 1 < \infty. \tag{57}$$

*(ii)*

$$p \in \mathcal{M} \quad \Rightarrow \quad \forall a \in \Sigma : \ p^a \in \mathcal{M}. \tag{58}$$

*Then it holds that*

$$(p(u), u \in \Sigma^{n+1}) \in \mathcal{M}_{n+1} \quad \Leftrightarrow \quad \begin{cases} (p(av), v \in \Sigma^n) \in \mathcal{M}_n & \forall a \in \Sigma \\ (p(v), v \in \Sigma^n) \in \mathcal{M}_n \end{cases} \tag{59}$$

*where $p(v) = \sum_{a \in \Sigma} p(va)$.*

**Remark 4.3.** Theorem 4.2 is meant to be a first step to obtain an analogous theorem resulting from replacing $\mathcal{M}_n, \mathcal{M}_{n+1}$ by their Zariski closures $\overline{\mathcal{M}_n}, \overline{\mathcal{M}_{n+1}}$. Generators for the ideal of invariants of $\overline{\mathcal{M}_{n+1}}$, given generators for the ideal of invariants of $\overline{\mathcal{M}_n}$, could be obtained by the following idea. If $h \in \mathbb{C}[X_v, v \in \Sigma^n]$ is one of the generators for $\overline{\mathcal{M}_n}$ where the $X_v$ are indeterminates for the probabilities $p(v), v \in \Sigma^n$, one obtains $|\Sigma| + 1$ generators for $\overline{\mathcal{M}_{n+1}}$ by replacing the indeterminates $X_v, v \in \Sigma^n$ by indeterminates $X_{av}, v \in \Sigma^n$ for all $a \in \Sigma$ which results in new generators

$$h_a \in \mathbb{C}[X_{av}, v \in \Sigma^n] \subset \mathbb{C}[X_u, u \in \Sigma^{n+1}] \tag{60}$$

as well as replacing each $X_v$ by the polynomials $\sum_a X_{va} \in \mathbb{C}[X_u, u \in \Sigma^{n+1}]$ resulting in another generator

$$h_+ \in \mathbb{C}[\sum_a X_{va}, v \in \Sigma^n] \subset \mathbb{C}[X_u, u \in \Sigma^{n+1}]. \tag{61}$$

The theorem would state that the generators obtained by this procedure generate the ideal of invariants of $\mathcal{M}_{n+1}$.

Note that in particular the maximum degree of the generators of $\overline{\mathcal{M}_{n+1}}$ would be at most that of $\overline{\mathcal{M}_n}$.

*Proof.* "$\Rightarrow$": From (58) we obtain that $(p^a(v), v \in \Sigma^n) \in \mathcal{M}_n$ for each $a \in \Sigma$. The second part is just the trivial observation that $(p(u), u \in \Sigma^{n+1}) \in$

$\mathcal{M}_{n+1}$ implies $(p(v), v \in \Sigma^n) \in \mathcal{M}_n$.

"$\Leftarrow$": From the second case in (59) we obtain that $(p(v), v \in \Sigma^n) \in \mathcal{M}_n$. As elements of $\mathcal{M}$ are uniquely determined by their values for strings of length at least $m$ and $n \geq m+1$ we obtain a USSF $\tilde{p} \in \mathcal{M}$ such that

$$p(v) = \tilde{p}(v) \quad \text{for all } v \in \Sigma^n. \tag{62}$$

It remains to show that also

$$p(w) = \tilde{p}(w) \quad \text{for all } w \in \Sigma^{n+1} \tag{63}$$

which amounts to showing that

$$p(av) = \tilde{p}(av) = \tilde{p}^a(v) \quad \text{for all } (a, v) \in \Sigma \times \Sigma^n. \tag{64}$$

We further observe that

$$(p^a(v), v \in \Sigma^n) \in \mathcal{M}_n \tag{65}$$

for all $a \in \Sigma$, because of $n \geq m+1 > m$, implies the existence of a unique $q^a \in \mathcal{M}$ s.t.

$$q^a(v) = p^a(v) \quad \text{for all } v, |v| \leq n. \tag{66}$$

As $\tilde{p} \in \mathcal{M}$, we have that $\tilde{p}^a \in \mathcal{M}$ for all $a \in \Sigma$, due to (58). Moreover, for $u \in \Sigma^{n-1}$,

$$\tilde{p}^a(u) = \tilde{p}(au) \overset{(62)}{=} p(au) \overset{(66)}{=} q^a(u). \tag{67}$$

As $n-1 \geq m$ and $\tilde{p}^a\mathcal{M}$ and $q^a\mathcal{M}$ coincide on strings of length $n-1 \geq m$, we obtain

$$\tilde{p}^a = q^a \tag{68}$$

because of $(i)$. We finally compute

$$p(av) = p^a(v) \overset{(66)}{=} q^a(v) \overset{(68)}{=} \tilde{p}^a(v) = \tilde{p}(av) \tag{69}$$

which establishes (64). $\diamond$

## 4.1 Finite-Dimensional Models

Theorem 4.2 applies for the finite-dimensional models. Eq. 57 is established by theorem 2.10 in subsection 2.2 whose statement is that finite-dimensional processes $p$ of dimension at most $d$ are uniquely determined by the values $p(v), |v| = 2d - 1$.

In terms of the language introduced here, we can restate theorem 2.10 as follows.

**Theorem 4.4.** *Let*

$$\mathcal{M}_d := \{p \in \mathbb{R}^{\Sigma^*};\, |\ p \textit{ is USSF and } \dim p \leq d\}$$

*be the class of unconstrained processes of dimension at most d. Then it holds that*

$$\text{SLC } (\mathcal{M}_d) = 2d - 1.$$

Furthermore observe that

$$(p^a)^w(v) = p^a(wv) = p(awv) = p^{aw}(v) \tag{70}$$

for all $a \in \Sigma, v, w \in \Sigma^*$ which translates to

$$(p^a)^w = p^{aw}. \tag{71}$$

Hence the column space of $\mathcal{P}_{p^a}$ is contained in that of $\mathcal{P}_p$ which yields

$$\dim p^a \leq \dim p \tag{72}$$

as $\dim p$ is just the dimension of the column space of $\mathcal{P}_p$.

This observation in combination with theorem 4.4 make the assumptions of theorem 4.2 hold for $\mathcal{M}_d$, which yields the following corollary.

**Corollary 4.5.** *Let $n \geq 2d$. Then it holds that*

$$(p(u), u \in \Sigma^{n+1}) \in \text{image } \mathbf{g}_{n+1,d} \quad \Leftrightarrow \quad \begin{cases} (p(av), v \in \Sigma^n) \in \text{image } \mathbf{g}_{n,d} & \forall a \in \Sigma \\ (p(v), v \in \Sigma^n) \in \text{image } \mathbf{g}_{n,d} \end{cases} \tag{73}$$

Again, an analogous ideal-theoretical result referring to the Zariski closures of image $\mathbf{g}_{n,d}$, image $\mathbf{g}_{n+1,d}$ would yield that the maximum degree of the generators would not increase for $n \geq 2d$.

# 5 Hidden Markov Models

In the following, let

$$\begin{array}{rccc} \mathbf{f}_{n,l}: & \mathbb{C}^{l(l-1)+l(|\Sigma|-1)+l} & \longrightarrow & \mathbb{C}^{|\Sigma|^n} \\ & ((T_a = A^T O_a)_{a\in\Sigma}), x) & \mapsto & (\text{tr } T_{a_n}...T_{a_1} x(1,...,1)^T)_{v=a_1...a_n\in\Sigma^n}. \end{array} \tag{74}$$

where $A$ and the $O_a$ as in example 2.6, be the polynomial map associated with the unconstrained (constrained if and only if $\sum_{i=1}^l x_i = 1$) hidden Markov model referring to hidden Markov models acting on $l$ hidden states and distributions over strings of length $n$, as described in [16].

The following theorem of Heller resulted from the attempts set off in the late 50's [11, 6, 7, 8] to give novel characterizations of hidden Markov processes. Many of those results are based on the idea that HMPs have finite dimension, which was noticed earlier in that series of papers without explicitly stating it. We give a version of Heller's theorem that is adapted to the language in use here. Heller's version is formulated in the language of homological algebra—without string functions and Hankel matrices. In his paper, discrete random processes are viewed as modules over certain rings. This language later has never been used in the theory of stochastic processes or related areas, probably as the required amount of prior knowledge unfamiliar to statisticians and probabilists is high. In the following we define

$$C_p := \text{span}\{p^w \,|\, w \in \Sigma^*\} \tag{75}$$

to be the column space of the Hankel matrix $\mathcal{P}_p$ of a string function $p$.

**Theorem 5.1** (Heller, 1965)**.** *A string function $p : \Sigma^* \to \mathbb{R}$ is associated with a (unconstrained) hidden Markov process if and only if there are (U)SSFs $p_i \in C_p, i = 1, ..., l$ s.t.*

*(a)* $p \in \text{cone } \{p_i \,|\, i = 1, ..., l\}$,

*(b)* $\forall w \in \Sigma^* : \ (p_i)^w \in \text{cone } \{p_i \,|\, i = 1, ..., l\}$.

Note first that this again points out that hidden Markov processes $p$ are finite-dimensional as $C_p \subset \text{span}\{p_i \,|\, i = 1, ..., l\}$ hence $\dim p \leq l$. Note further that $(a)$ in combination with $(b)$ implies that $p^v \in \text{cone } \{p_i\}$ for all $v \in \Sigma^*$ which renders cone $\{p_i\}$ to be full-dimensional. It is closed due to being polyhedral and pointed due to being generated by SSFs which are strictly positive string functions. Collecting properties results in cone $\{p_i\}$ being a proper, polyhedral cone.

Given a hidden Markov process $p$, the $p_i$ can be obtained as the random processes starting from the hidden states (i.e. having initial probability distribution $e_i$). The other direction requires more work. A translation of Heller's proof [12] to the language of string functions can be found in [17]. A rather straightforward consequence of Heller's theorem is the following corollary.

**Corollary 5.2.** *Let $p$ be a USSF of dimension of at most 2. Then $p$ is associated with an unconstrained hidden Markov process acting on 2 hidden states.*

*Proof Sketch*: As all $p^w \geq 0$ the cone generated by all column vectors

$$\text{cone}\,\{p^w \,|\, w \in \Sigma^*\} \tag{76}$$

is pointed hence its closure is generated by its extremal rays. In two dimensions this is equivalent to the closure of cone $\{p^w \,|\, v \in \Sigma^*\}$ being polyhedral. It's a routine exercise to check for the assumptions of Heller's theorem to hold for this cone. ◇

One might be tempted to infer that the ideal of model invariants of $\mathbf{f}_{n,2}$ can be computed by computing the invariants of the 2-dimensional model, as provided by theorem 3.1. However, a 2-dimensional process need not be associated with a hidden Markov process acting on 2 hidden states. According to the proof of theorem 5.1, one might need up to $2|\Sigma|$ many hidden states to describe an arbitrary 2-dimensional process by means of a hidden Markov parameterization.

## 5.1 Degree of Invariants

Heller's theorem gives rise to an application of theorem 4.2 to hidden Markov processes where $n \geq 2l$. Assumption $(i)$ of theorem 4.2 is met since hidden Markov processes on $l$ hidden states, as finite-dimensional random processes of dimension $\leq l$, are determined by their distributions over the strings of length $2l-1$. Assumption $(ii)$ is met due to Heller's theorem.[2] The only

[2] Proofs for this can also be formulated in terms of the hidden Markov processes' parameterizations. However, such proofs are lengthy and technical exercises.

thing one has to be aware of is that the dimension of the column space of $\mathcal{P}_{p^a}$ can be lower than that of $\mathcal{P}_p$ itself. In this case, one obtains the necessary cone generators by projecting $C_p$ onto $C_{p^a}$ (we recall that $C_{p^a} \subset C_p$). In sum, the class of unconstrained hidden Markov processes meet the assumptions of theorem 4.2, which yields

**Corollary 5.3.** *Let $n \geq 2d$. Then it holds that*

$$(p(u), u \in \Sigma^{n+1}) \in \text{image } \mathbf{f}_{n+1,l} \quad \Leftrightarrow \quad \begin{cases} (p(av), v \in \Sigma^n) \in \text{image } \mathbf{f}_{n,l} & \forall a \in \Sigma \\ (p(v), v \in \Sigma^n) \in \text{image } \mathbf{f}_{n,l} \end{cases} \tag{77}$$

Note that an ideal-theoretic equivalent of theorem 5.3 would yield a proof of conjecture 11.9 from [3] as a corollary. However, an ideal-theoretical equivalent of theorem 5.3 would be a stronger result:

**Conjecture 5.4.** *Let $\mathbf{f}_{n,l}$ be the unconstrained hidden Markov model for $l$ hidden states and strings of length $n$. Then the maximum degree of the invariants $d(n,l)$ of $\mathbf{f}_{n,l}$ does not increase for $n \geq 2l$, that is,*

$$...d(n+1,l) \leq d(n,l) \leq d(n-1,l) \leq ... \leq d(2l,l). \tag{78}$$

As $d(5,2) = 1$ (see [3], table 11.1 (?)), we would obtain that $d(n,l) = 1$ for $n \geq 5$, that is, the ideal of invariants would be generated by linear equations exclusively.

# 6 The Markov model

In the following, let (U)SSFs $p$ be induced by Markov chains. That is,

$$p(v = a_1...a_n) = \pi(a_1) \prod_{i=2}^{n} M_{a_{i-1}a_i} \tag{79}$$

where $\pi \in \mathbb{R}^\Sigma$ is a strictly positive vector (with entries not necessarily summing up to one in case of a USSF $p$) and $M \in \mathbb{R}^{\Sigma^2}$ is a matrix with the entries of a row summing up to one. Moreover, in this section, let

$$\begin{array}{cccc} \mathbf{f}_{n,l=|\Sigma|}: & \mathbb{C}^{l+l(l-1)} & \longrightarrow & \mathbb{C}^{|\Sigma|^n} \\ & (\pi, M) & \mapsto & (\pi(a_1)\prod_{i=2}^n M_{a_{i-1}a_i})_{v=a_1...a_n\in\Sigma^n}. \end{array} \tag{80}$$

be the polynomial map (associated with the Markov model in case of $\pi, M$ being in accordance with the laws from above) with alphabet $\Sigma$ on strings of length $n$. In the language of string functions and Hankel matrices, we have the following theorem.

**Theorem 6.1.** *A (U)SSF $p$ is associated with a Markov chain iff*

$$\forall a \in \Sigma : \quad \dim \operatorname{span}\{p^{va} \,|\, v \in \Sigma^*\} \leq 1. \tag{81}$$

A proof can be found in [17], for example.

This can be straightforwardly exploited to obtain invariants of $\mathbf{f}_{n,l}$.

**Theorem 6.2.** *Let $(p(v), v \in \Sigma^n)$ be a (unconstrained) probability distribution such that $n \geq 2|\Sigma| - 1$. Then $(p(v), v \in \Sigma^n)$ lies in the image of $\mathbf{f}_{n,l=|\Sigma|}$ if and only if*

$$\det \begin{bmatrix} p(vau) & p(wau) \\ p(vau') & p(wau') \end{bmatrix} = 0 \tag{82}$$

*for all choices $u, u', v, w \in \Sigma^*, a \in \Sigma$ such that $|vau|, |vau'|, |wau|, |wau'| \leq n$ and, as usual, $p(v) := \sum_{w \in \Sigma^{n-|v|}} p(vw)$ for strings $v$ such that $|v| < n$.*

*Proof.* "$\Rightarrow$" is obvious as for a Markov chain $p$, (82) is a necessary consequence of (81) in theorem 6.1.

"$\Leftarrow$" Clearly, (82) implies the assumptions (35) and (37) of theorem 3.1 to hold, which yields that $(p(v), v \in \Sigma^n)$ lies in the image of the finite-dimensional model. We thus find, by means of algorithm 2.8, $(T_a)_{a\in\Sigma}, x, y$ such that the probabilities $p(v)$ for all $v$ up to length $n \geq 2|\Sigma|$ can be computed according to (6). Note that $T_a$ maps $p^v$ onto $p^{va}$ where $p^v, p^{va}$ are identified with a coordinate representation induced by the basis of the column spaces that one has found according to algorithm 2.8 (see remark 3.6). In this sense, (82) translates to

$$\dim \operatorname{image} T_a \leq 1 \tag{83}$$

for all $a \in \Sigma$. Clearly, this implies (81) of theorem 6.1 from which the assertion follows. $\diamond$

**Remark 6.3.** While the assumption $n \geq 2|\Sigma| - 1$ helps to give a rather concise proof of theorem 6.2, we feel that it is not a necessary requirement. However, inference of Markov chain parameters giving rise to probability distributions $(p(v), v \in \Sigma^n)$ for which the determinantal invariants (82) apply is a much more technical undertaking. Moreover, it seems that some (potentially more involved) pecularities have to be resolved.

# 7 Trace algebras

In this section, we will draw some connections between trace algebras and the theory of finite-dimensional string functions. For a rigorous introduction to trace algebras see [9]. We recall that in Bernd's preprint [2] the quartic hidden Markov model invariant listed in [3] could be identified as a relation between trace polynomials.

Here, we shall try to shed some light on the general relationships between trace algebras and finite-dimensional models. In terms of the language of trace algebras, we will derive some defining relations for the trace algebras.

Therefore, we introduce the following definition.

**Definition 7.1.** A string function $p : \Sigma^* \to \mathbb{R}$ is called *traceable of order $r$* if there are matrices $X_a \in \mathbb{R}^{r\times r}, a \in \Sigma$ such that

$$p(v = a_1...a_n) = \text{tr } X_{a_n}...X_{a_1}. \tag{84}$$

Traceable string functions are finite-dimensional, as can be seen by application of a simple lemma.

**Lemma 7.2.** *Let $p_i, i = 1, ..., k$ be string functions of dimensions $d_i$. Let $p := \sum_{i=1}^k p_i$. Then it holds that*

$$\dim p \leq \sum_{i=1}^{k} d_i. \tag{85}$$

This gives rise to

**Theorem 7.3.** *Let $p \in \mathbb{R}^{\Sigma^*}$ be traceable of order $r$. Then*

$$\dim p \leq r^2. \tag{86}$$

*Proof.* Let $p_i \in \mathbb{R}^{\Sigma^*}, i = 1, ..., r$ be defined by

$$p_i(v = a_1...a_n) := \text{tr } X_{a_n}...X_{a_1} e_i e_i^T. \tag{87}$$

From theorem 2.5 we obtain $\dim p_i \leq r$. As $Id = \sum_i e_i e_i^T$, which yields

$$p = \sum p_i \tag{88}$$

the assertion follows from application of lemma 7.2. ◇

If the identity matrix $Id = \sum_i e_i e_i^T$ was presentable in the form $Id = xy^T$ itself, traceable string functions of order $r$ would be of dimension at most $r$, as given by theorem 2.5. As this is not the case, there are traceable string functions of order $r$ whose dimension is larger than $r$. Moreover, not every string function of dimension $r^2$ seems to be traceable. However, an example of that kind is yet to be delivered.

The consequences of theorem 7.3 for the theory of trace algebras are that invariants which can be computed for the $r^2$-dimensional models $\mathbf{f}_{n,r^2}$ also apply as defining relations for the trace algebras generated by all trace polynomials

$$\text{tr } (X_{i_n}...X_{i_1}), 1 \leq i_j \leq d, n \geq 0. \tag{89}$$

The exact relationships between trace algebras, hidden Markov as well as the finite-dimensional models are yet to be determined.

# 8 Open Questions

1. Theorem 5.1 characterizes hidden Markov chains within the theory of finite-dimensional random processes. Determine invariants that correspond to this characterization.

2. Determine the relationships between trace algebras and the models under consideration here in more detail.

3. Deliver a proof for a more general version of theorem 6.2, as discussed above.

4. Determine the peculiarities of differences between the two-dimensional models and the hidden Markov models for 2 hidden states.

5. Tropicalization of Teichmüller spaces (see [2])?

# References

[1] D. Aharonov, A. Ambainis, J. Kempe, U. Vazirani, ”Quantum walks on graphs”, in *Proc. of 33rd ACM STOC, New York*, 2001, pp. 50-59.

[2] B. Sturmfels, “Trace Algebras, Hidden Markov Models and Tropicalization of Teichmüller Spaces”, preprint, 2008.

[3] N. Bray and J. Morton, “Equations defining hidden Markov models”, in *Algebraic Statistics for Computational Biology* (L. Pachter and B. Sturmfels, eds), Cambridge University Press, pp. 235-247, 2005.

[4] L. Finesso, A. Grassi and P. Spreij, “Approximation of stationary processes by hidden Markov models”, preprint, arXiv:math.OC/0606591.

[5] D. Blackwell and L. Koopmans, “On the identifiability problem for functions of finite markov chains”, *Annals of Mathematical Statistics* vol. 28, pp. 1011–1015, 1957.

[6] S.W. Dharmadhikari, “Functions of finite markov chains”, *Annals of Mathematical Statistics*, vol. 34, pp. 1022-1032, 1963.

[7] S.W. Dharmadhikari. “Sufficient conditions for a stationary process to be a function of a finite markov chain”, *Annals of Mathematical Statistics*, vol. 34, pp. 1033-1041, 1963.

[8] S.W. Dharmadhikari, A characterization of a class of functions of finite markov chains. *Annals of Mathematical Statistics*, vol. 36, pp. 524-528, 1965.

[9] V. Drensky, Computing with matrix invariants. *Math. Balk.*, New Ser. 21, No. 1-2, pp. 101-132, 2007.

[10] U. Faigle and A. Schoenhuth, “Asymptotic mean stationarity of sources with finite evolution dimension”, *IEEE Trans. Inf. Theory*, vol. 53(7), pp. 2342-2348, 2007

[11] E.J. Gilbert, ”On the identifiability problem for functions of finite Markov chains”, *Ann. Math. Stat.*, vol. 30, pp. 688-697, 1959.

[12] A. Heller “On stochastic processes derived from Markov chains”, *Annals of Mathematical Statistics*, vol. 36(4), pp. 1286-1291, 1965

[13] H. Ito, S.-I. Amari and K. Kobayashi “Identifiability of hidden Markov information sources and their minimum degrees of freedom”, *IEEE Trans. Inf. Theory*, vol. 38(2), pp. 324-333, 1992.

[14] H. Jaeger. “Observable operator models for discrete stochastic time series”, *Neural Computation*, vol. 12(6), pp. 1371-1398, 2000.

[15] Y. Ephraim and N. Merhav, ”Hidden Markov processes”, *IEEE Trans. Inf. Theory*, vol. 48(6), pp. 1518-1569, 2002.

[16] L. Pachter and B. Sturmfels, *Algebraic Statistics for Computational Biology*, Cambridge University Press, 2005.

[17] A. Schoenhuth, “Discrete-valued stochastic vector spaces”, *PhD thesis (German), University Cologne*, 2006.

[18] A. Schönhuth and H. Jaeger, “Characterization of ergodic hidden Markov sources”, `http://www.zaik.uni-koeln.de/∼paper/index.html?show=zaik2008-573`, submitted to *IEEE Trans. Inf. Theory*, 2007.

[19] A. Schönhuth, “A simple and efficient solution of the identifiability problem for hidden Markov models and quantum random walks”, `http://arxiv.org/abs/0808.2833`, *Proc. ISITA 2008*, to appear, 2008.

[20] A. Schönhuth, “On analytic properties of entropy rate”, *IEEE Trans. Inf. Theory*, to appear.