Centrum voor Wiskunde en Informatica
**REPORT***RAPPORT*

# The behavior of convolutional codes

J. Rosenthal, J.M. Schumacher and E.V. York

Department of Operations Reasearch, Statistics, and System Theory

**BS-R9533 1995**

Report BS-R9533
ISSN 0924-0659

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

# The Behavior of Convolutional Codes

Joachim Rosenthal*
*Department of Mathematics, University of Notre Dame*
*Notre Dame, IN 46556-5683, USA*
Joachim.Rosenthal@nd.edu

J.M. Schumacher
*CWI*
*P.O. Box 94079, 1090 GB Amsterdam, The Netherlands, and*
*Tilburg University, CentER and Department of Economics,*
*P.O. Box 90153, 5000 LE Tilburg, The Netherlands*
Hans.Schumacher@cwi.nl

Eric V. York
*Department of Mathematics, University of Notre Dame*
*Notre Dame, IN 46556-5683, USA*
Eric.V.York.2@nd.edu

**Abstract**

It is well known that a convolutional code can be viewed as a linear system over a finite field. In this paper we develop this viewpoint for convolutional codes using several recent innovations from the systems theory literature. In particular we define codes as behaviors of a set of compact support time trajectories over a vector space. We also consider several different representations of codes, in particular generalized first order representations. As an application of these ideas, we present a BCH construction technique for convolutional codes that yields optimal high rate codes.

*AMS Subject Classification (1991)*: 94B10, 93B20.
*Keywords & Phrases*: convolutional codes, linear systems over finite fields, system representations, duality, code constructions.
*Note*: This report has been submitted for publication elsewhere.

## 1 Introduction

Viewpoint in mathematics can be everything. A problem, that in one setting may seem difficult if not impossible to solve, suddenly becomes feasible when considered from an alternate perspective. In this paper we take a detailed look at convolutional codes from the perspective of linear systems theory with an emphasis on the different representations of these codes. Using these representations, we present a construction of convolutional codes with distance lower bounded by the overall constraint length of the encoder.

Throughout the relatively short history of the theory of convolutional codes, there have been several authors that have made the link between convolutional codes and linear systems theory. Among the first authors to do this were Massey and Sain. They published a series of papers [21, 22, 34], containing a systems theoretic analysis of convolutional codes and encoders. After this, Omura in [25] considered

*This author was supported in part by NSF grant DMS-94-00965. The research for this paper was carried out in part while he was a visitor at CWI.

Viterbi decoding and its relationship to dynamic programming and later applications of control theory to optimal receiver design for convolutional codes [26]. In several landmark papers, [3, 4] Forney started to lay the foundation for the algebraic structure of convolutional codes. In the words of the author [5]:

> These papers were a success in providing a linear-systems structure theory for convolutional codes, in settling certain questions concerning invertibility, and in defining canonical classes of convolutional encoders. They were generally failures, however, with regard to leading to any new classes of codes or decoding algorithms, or stimulating very much further work.

Since these papers were written, there have been significant advances in the theory of linear systems. One notable advance has been the behavioral approach to linear systems of Willems, championed in a series of papers [37, 38, 39, 40, 41, 42]. The behavioral approach to linear systems looks at system equations as just one possible way of representing a 'behavior', that is, a set of time trajectories of a vector of selected variables. This point of view allows the development of an extensive theory of alternative system representations which we demonstrate to be useful within the coding context. Actually some basic results from the behavioral theory were developed independently for codes by L. Staiger [36]. Another useful tool developed recently in the systems theory literature is the class of first order representations by Kuijper and Schumacher [16] and Kuijper [15].

In this paper we will explain those recent advances in systems theory in the context of convolutional codes defined over a Galois field $\mathbb{F}$. This point of view then allows us to introduce a BCH construction for convolutional codes, one of the main results of this paper.

The paper is structured as follows: In Section 2 we begin by considering the behavioral view point for convolutional codes. In particular, we present convolutional codes as duals to autoregressive systems [38], i.e. as duals to linear shift invariant behaviors which are closed in the topology of pointwise convergence. We also connect in this section to the work of L. Staiger [36]. We also show how the notion of input observability, a very natural notion in coding theory, relates to the classical state observability notion in systems theory.

In Section 3 we derive generalized first order representations for convolutional codes and discuss minimality of these representations. We then show how driving variable and input-state-output representations can be derived by partitioning the state and output of first order representations.

In Section 4 we use the representations from Section 3 to construct a class of convolutional codes whose free distance is lower bounded by the overall constraint length + 1 of the encoder. Some of the results presented in Sections 2 and 3 of this paper appeared in abbreviated form in [33, 43].

## 2 Codes as dynamical systems

Below we will explain some basic notions concerning dynamical systems and their relations to coding theory. We start with a general definition of a dynamical system as presented in [42]. For a more detailed treatment of systems theory we refer to any of the following: [6, 12, 14, 13, 35, 38, 42].

Definition 2.1 A dynamical system $\sum$ is defined as a triple

$$\sum = (T, W, \mathcal{B})$$

where $T \subseteq \mathbb{R}$ is the time axis, $W$ is an abstract set called the signal alphabet, $\mathcal{B} \subseteq W^T$ is called the behavior, and the elements of $\mathcal{B}$ are called trajectories of the system.

This definition is quite abstract and leaves us with some leeway as to how we can fit convolutional codes into this frame work. By considering only the trajectories of a dynamical systems without forcing any a priori structure as to what the inputs and outputs of the system are, considerable freedom is gained in how you choose to represent your system. It is exactly this freedom that is the key to the construction of convolutional codes as presented in Section 4.

In systems theory one is often concerned with how to model a given behavior $\mathcal{B}$, in other words, write down a set of equations in a given operator that every element of the behavior must satisfy. The equations may be differential equations when $T = \mathbb{R}$ with operator $\frac{d}{dt}$, or possibly difference equations for $T = \mathbb{Z}_+$, with the left shift operator $\sigma$, defined by

$$\sigma(w_0, w_1, w_2, \ldots) = (w_1, w_2, w_3, \ldots). \tag{2.1}$$

See [42] for many more detailed examples.

In the sequel we will develop a theory which nicely fits the class of convolutional codes into this framework. For this let $\mathbb{F} := \mathbb{F}_q$ be the Galois field with $q$ elements. As a signal space we will take the vectorspace $W = \mathbb{F}^n$ and as time axis we will use $T = \mathbb{Z}_+$. In other words the universum $W^T$ consists of all one sided infinite sequences of vectors in $\mathbb{F}^n$.

On the sequence space $W^T$ define the right shift operator $\tau$ by

$$\tau(w_0, w_1, w_2, \ldots) = (0, w_0, w_1, w_2 \ldots). \tag{2.2}$$

We say a subset $\mathcal{C} \subseteq W^T$ is right shift invariant if $\tau\mathcal{C} \subseteq \mathcal{C}$ and we say $\mathcal{C} \subseteq W^T$ has compact support if every element $b \in \mathcal{C}$ has at most finite many nonzero components. With this we define:

DEFINITION 2.2 A subset $\mathcal{C} \subseteq W^T$ is called a (time invariant) convolutional code if $\mathcal{C}$ is linear, right shift invariant and has compact support.

We would like to comment on some of the requirements we impose. First note that the condition of right shift invariance is a condition of *time invariance*. Note that right shift invariance implies that any message that is sent at time zero, can be sent at any time $t > 0$.

The fact that we impose compact support seems to be new. Note that many authors (see e.g. the monograph [27] and the references therein) do not require that a convolutional code has compact support and they instead substitute this condition with other conditions like e.g. "remergability" (for this see Subsection 2.2 about the work of Staiger [36]). On the other hand one has to say that there seems to be no universal agreement as to how to define a convolutional code and we refer to the recent paper [9] to emphasize this point.

It is our intention to show in this paper that Definition 2.2 leads to a very clear and straightforward theory which ties well together with systems theory and algebraic geometry. It is our opinion that Definition 2.2 gives a clearer picture of codes from the systems theory point of view by simplifying many of the arguments that follow, without losing any of the relevant mathematical structure concerning codes. Moreover it seems that from an applications point of view nothing new is gained by allowing infinite code words and we will say more about this in Subsection 2.2.

First we would like to note that we naturally can identify our universum $W^T$ with the vector space $\mathbb{F}^n[[\mathcal{D}]]$ and we make use of this identification throughout the paper. Note also that the right shift operator $\tau$ corresponds to multiplication by $\mathcal{D}$ in $\mathbb{F}^n[[\mathcal{D}]]$. With this in mind we have the Lemma:

LEMMA 2.3 *$\mathcal{C} \subseteq W^T$ is a convolutional code if and only if $\mathcal{C}$ is a $\mathbb{F}[\mathcal{D}]$ submodule of $\mathbb{F}^n[\mathcal{D}]$.*

PROOF To say $\mathcal{C}$ is a submodule just means that $\mathcal{C}$ is linear and $\mathcal{D}\mathcal{C} \subseteq \mathcal{C}$. The condition $\mathcal{C} \subseteq \mathbb{F}^n[\mathcal{D}]$ is equivalent to the condition that $\mathcal{C}$ has compact support. □

COROLLARY 2.4 *If $\mathcal{C} \subseteq \mathbb{F}^n[\mathcal{D}]$ is a convolutional code then there exists a unique positive integer $k$ and an injective module homomorphism*

$$\varphi : \ \mathbb{F}^k[\mathcal{D}] \longrightarrow \mathbb{F}^n[\mathcal{D}]; \quad \ell(\mathcal{D}) \longmapsto w(\mathcal{D})$$

*having the property that* $\mathrm{Im}\,(\varphi) = \mathcal{C}$. *Equivalently there is a full rank $n \times k$ polynomial matrix $G(\mathcal{D})$ such that*

$$\mathcal{C} = \{w(\mathcal{D}) \mid \exists \ell(\mathcal{D}) \in \mathbb{F}^k[\mathcal{D}], w(\mathcal{D}) = G(\mathcal{D})\ell(\mathcal{D})\}. \tag{2.3}$$

PROOF $\mathbb{F}[\mathcal{D}]$ is a principal ideal domain and $\mathbb{F}^n[\mathcal{D}]$ is a free module over $\mathbb{F}[\mathcal{D}]$. Since $\mathcal{C}$ is a submodule, $\mathcal{C}$ is free as well [8, Chapter IV, Theorem 6.1] and has a well defined rank $k$. Let $g_1(\mathcal{D}), \ldots, g_k(\mathcal{D}) \in \mathbb{F}^n[\mathcal{D}]$ be a basis for the free module $\mathcal{C}$ and let $G(\mathcal{D})$ be the $n \times k$ matrix whose $i$-th column is the vector $g_i(\mathcal{D})$. Then necessarily $G(\mathcal{D})$ has full rank and the representation of $w(\mathcal{D})$ in (2.3) is unique. □

Basing ourselves on this corollary we define:

DEFINITION 2.5 Given a convolutional code $\mathcal{C} \subseteq \mathbb{F}^n[\mathcal{D}]$. A module homomorphism $\varphi : \mathbb{F}^k[\mathcal{D}] \longrightarrow \mathbb{F}^n[\mathcal{D}]$ (respectively a $n \times k$ polynomial matrix $G(\mathcal{D})$ whose columns form a basis of the free submodule $\mathcal{C}$) will be called an encoder of the convolutional code $\mathcal{C}$. If $k = \text{rank}\mathcal{C}$ we say $\mathcal{C}$ has rate $\frac{k}{n}$.

### *2.1 Convolutional codes and autoregressive systems*

In this subsection we will show that the class of convolutional codes as introduced in Definition 2.2 is in one-one correspondence with the class of left shift invariant linear and complete behaviors in $\mathbb{F}^n[[\mathcal{D}]]$. In order to prove this result we need a fundamental theorem due to Willems [38, Theorem 5]. For this first recall the notion of *completeness* in our context (for the general definition see [38]). If $T = \mathbb{Z}_+$ and $W^T = \mathbb{F}^n[[\mathcal{D}]]$ then a subseteq $\mathcal{B} \subseteq \mathbb{F}^n[[\mathcal{D}]]$ is complete if $w : T \longrightarrow W$ is in $\mathcal{B}$ whenever $w : \{0, \ldots, N\} \longrightarrow W$ belongs to $\mathcal{B}\,|_{\{0,\ldots,N\}}$ for each positive integer $N$. With this we have:

THEOREM 2.6 [38, Theorem 5] *$\mathcal{B} \subseteq W^T$ is a linear, complete and left shift invariant behavior if and only if $\mathcal{B}$ has a kernel representation, i.e. there is a $\ell \times n$ polynomial matrix $R(s)$ such that*

$$\mathcal{B} = \{w \in W^T \mid R(\sigma)w = 0\}$$

In order to connect Definition 2.2 with this class of important behaviors we will introduce a bilinear form [24]:

$$\begin{aligned} (,): \quad \mathbb{F}^n[[\mathcal{D}]] \times \mathbb{F}^n[\mathcal{D}] &\longrightarrow \mathbb{F} \\ (w, v) &\mapsto \sum_{i=0}^{\infty} \langle w_i, v_i \rangle , \end{aligned} \tag{2.4}$$

where $<,>$ represents the standard dot product on $\mathbb{F}^n$. Note that the infinite sum is indeed well defined since at most finitely many terms are nonzero. Using this bilinear form we define for every convolutional code $\mathcal{C} \subseteq \mathbb{F}^n[\mathcal{D}]$ a corresponding behavior $\mathcal{C}^\perp \subseteq \mathbb{F}^n[[\mathcal{D}]]$ through:

$$\mathcal{C}^\perp = \{w \in \mathbb{F}^n[[\mathcal{D}]] \mid (w, v) = 0, \forall v \in \mathcal{C}\}. \tag{2.5}$$

Similarly if $\mathcal{B} \subseteq \mathbb{F}^n[[\mathcal{D}]]$ is a linear, left shift invariant and complete behavior we define

$$\mathcal{B}^\perp = \{v \in \mathbb{F}^n[\mathcal{D}] \mid (w, v) = 0, \forall w \in \mathcal{B}\}. \tag{2.6}$$

Note that the bilinear form $(\ ,\ )$ is nonsingular in the sense that $(w, v) = 0$ for all $v \in \mathbb{F}^n[\mathcal{D}]$ implies that $w = 0$, and $(w, v) = 0$ for all $w \in \mathbb{F}^n[[\mathcal{D}]]$ implies that $v = 0$.

We use $\sigma$ to denote the left shift operator on $\mathbb{F}^n[[\mathcal{D}]]$. It is easy to check that, if $w \in \mathbb{F}^n[[\mathcal{D}]]$ and $v \in \mathbb{F}^n[\mathcal{D}]$, one has $(w, \mathcal{D}v) = (\sigma w, v)$. Also, if $A$ is a matrix over $\mathbb{F}$ of size $p \times n$ and $w \in \mathbb{F}^p[[\mathcal{D}]]$, $v \in \mathbb{F}^n[\mathcal{D}]$, then $(w, Av) = (A^t w, v)$. By bilinearity, it follows that for every polynomial matrix $A(s) \in \mathbb{F}^{p\times n}[s]$ we have

$$(w, A(\mathcal{D})v) \ = \ (A^t(\sigma)w, v).$$

Let $R(s) \in \mathbb{F}^{p\times n}[s]$ be a polynomial matrix. The set of all $w \in \mathbb{F}^n[[\mathcal{D}]]$ such that $R(\sigma)w = 0$ will be denoted by $\mathcal{B}(R)$. The matrix $R(s)$ is said to provide a *kernel representation* for the linear, left shift invariant and complete behavior $\mathcal{B}(R)$. Similarly, we write $\mathcal{C}(G)$ for the convolutional code generated by the generator matrix $G(\mathcal{D})$; the matrix $G(\mathcal{D})$ provides an *image representation* for the code $\mathcal{C}(G)$.

THEOREM 2.7 *If $\mathcal{C} \subseteq \mathbb{F}^n[\mathcal{D}]$ is a convolutional code with generator matrix $G(\mathcal{D})$, then $\mathcal{C}^\perp$ is a linear, left shift invariant and complete behavior with kernel representation $R(\sigma) = G^t(\sigma)$. Conversely, if $\mathcal{B} \subseteq \mathbb{F}^n[[\mathcal{D}]]$ is a linear, left shift invariant and complete behavior with kernel representation $R(\sigma)$, then $\mathcal{B}^\perp$ is a convolutional code with generator matrix $G(\mathcal{D}) = R^t(\mathcal{D})$.*

PROOF Let $\mathcal{C}$ be a convolutional code with generator matrix $G(\mathcal{D})$. An element $w \in \mathbb{F}^n[[\mathcal{D}]]$ belongs to $\mathcal{C}^\perp$ if and only if $(w, G(\mathcal{D})v) = 0$ for all $v \in \mathbb{F}^n[\mathcal{D}]$. This is equivalent to $(G^t(\sigma)w, v) = 0$ for all $v \in \mathbb{F}^n[\mathcal{D}]$, which in turn is equivalent to $G^t(\sigma)w = 0$.

For the second part of the proof, let $\mathcal{B}$ be a linear, left shift invariant and complete behavior with kernel representation $R(s)$. Take $v \in \mathcal{C}(R^t)$ so that $v = R^t(\mathcal{D})v'$ for some $v' \in \mathbb{F}^n[\mathcal{D}]$. For any $w \in \mathcal{B}$, we then have $(w, v) = (w, R^t(\mathcal{D})v') = (R(\sigma)w, v') = 0$. So it follows that $\mathcal{C}(R^t) \subseteq \mathcal{B}^\perp$ and the rest of the proof will be devoted to the reverse inclusion.

First assume that the matrix $R(s)$ can be completed to a unimodular matrix, so there is a matrix $\hat{R}(s)$ such that

$$U(s) := \begin{bmatrix} R(s) \\ \hat{R}(s) \end{bmatrix}$$

is unimodular, i.e. has a polynomial inverse. Write

$$U^{-1}(s) =: [T(s) \mid \hat{T}(s)]$$

where the partitioning is conformable to that of $U(s)$. We claim that $R(\sigma)w = 0$ for $w \in \mathbb{F}^n[[\mathcal{D}]]$ if and only $w = \hat{T}(\sigma)w'$ for some $w' \in \mathbb{F}^p[[\mathcal{D}]]$. Indeed, if $w = \hat{T}(\sigma)w'$ then we can also write $w = [T(\sigma) \mid \hat{T}(\sigma)]\left[\begin{smallmatrix} 0 \\ w' \end{smallmatrix}\right]$ which implies $U(\sigma)w = \left[\begin{smallmatrix} 0 \\ w' \end{smallmatrix}\right]$ and so $R(\sigma)w = 0$. Conversely if $R(\sigma)w = 0$ then $w = \hat{T}(\sigma)w'$ for $w' = \hat{R}(\sigma)w$. Now take $v \in \mathcal{B}^\perp$. It follows that $(\hat{T}(\sigma)w', v) = 0$ for all $w' \in \mathbb{F}^p[[\mathcal{D}]]$, so that $\hat{T}^t(\mathcal{D})v = 0$. Define $v'$ by $v' = T^t(\mathcal{D})v$, then

$$v \;=\; [R^t(\mathcal{D}) \mid \hat{R}^t(\mathcal{D})] \begin{bmatrix} T^t(\mathcal{D}) \\ \hat{T}^t(\mathcal{D}) \end{bmatrix} v \;=\; [R^t(\mathcal{D}) \mid \hat{R}^t(\mathcal{D})] \begin{bmatrix} v' \\ 0 \end{bmatrix} \;=\; R^t(\mathcal{D})v'$$

so that $v \in \mathcal{C}(R^t)$.

Now consider a general kernel representation $R(s)$. We may assume without loss of generality that $R(s)$ has full row rank. We may then write $R(s) = T(s)Q(s)$ where $T(s)$ is a square and nonsingular polynomial matrix, and $Q(s)$ can be completed to a unimodular matrix. (This follows by an application of the Smith form, which is valid over a general Euclidean domain and so in particular for matrices over $\mathbb{F}[s]$.) From the above it already follows that

$$\mathcal{C}(R^t) \;\subset\; \mathcal{B}^\perp \;\subset\; \mathcal{C}(Q^t).$$

To prove that actually $\mathcal{B}^\perp = \mathcal{C}(R^t)$, it suffices to show that the quotient spaces $\mathcal{C}(Q^t)/\mathcal{B}^\perp$ and $\mathcal{C}(Q^t)/\mathcal{C}(R^t)$ are both finite-dimensional vector spaces and that the dimensions of these spaces agree. This is what we shall do now.

As is well-known, the behavior $\mathcal{B}(T)$ determined by the nonsingular matrix $T(s)$ is a finite-dimensional vector space over $\mathbb{F}$ with dimension $r := \deg \det T(s)$. Also the mapping $Q(\sigma)$ from $\mathbb{F}^n[[\mathcal{D}]]$ to $\mathbb{F}^p[[\mathcal{D}]]$ is surjective so we can find elements $w_1, \ldots, w_r \in \mathbb{F}^p[[\mathcal{D}]]$ such that the elements $\tilde{w}_i$ defined by $\tilde{w}_i = Q(\sigma)w_i$ form a basis for $\mathcal{B}(T)$. Then $\mathcal{B}(R)$ is spanned by $\mathcal{B}(Q)$ together with the elements $w_i$, and so $v \in \mathcal{B}^\perp$ for $v \in \mathbb{F}^n[\mathcal{D}]$ if and only if $v \in \mathcal{C}(Q^t)$ and $(w_i, v) = 0$ for all $i = 1, \ldots, r$. To show that these extra restrictions are independent, assume that $(\sum_{i=1}^r \alpha_i w_i, v) = 0$ for some $\alpha_i \in \mathbb{F}$ and for all $v \in \mathcal{C}(Q^t)$. It then follows that for all $v' \in \mathbb{F}^n[\mathcal{D}]$ we have

$$(\sum_{i=1}^r \alpha_i \tilde{w}_i, v') = (\sum_{i=1}^r \alpha_i Q(\sigma)w_i, v') = (\sum_{i=1}^r \alpha_i w_i, Q^t(\mathcal{D})v') = 0$$

so that $\sum_{i=1}^{r} \alpha_i \tilde{w}_i = 0$ and hence all $\alpha_i$ are zero because the $\tilde{w}_i$ are independent. It follows that the quotient space $\mathcal{C}(Q^t)/\mathcal{B}^\perp$ is a finite-dimensional vector space with dimension $r = \deg \det T(s)$.

To complete the proof, we note that it is a standard fact from polynomial module theory that the quotient module $\mathcal{C}(Q^t)/\mathcal{C}(R^t) = \mathcal{C}(R^t T^t)/\mathcal{C}(R^t)$ is finite-dimensional as a vector space over $\mathbb{F}$ with dimension given by $\deg \det T(s)$. □

In conclusion we can say that behaviors definable by a kernel representation and the class of convolutional behaviors introduced in Definition 2.2 can be thought of as dual to each other. One has finite support and right shift invariance on the one hand and infinite support and left shift invariance on the other. The fact in itself that there is a one-one relation between complete linear left shift-invariant behaviors on the one hand and finitely generated polynomial modules is well-known; what is newly established here is that this relation can be implemented through the 'perp' operation. Analogous results were obtained by Nieuwenhuis and Willems [24] in the context of behaviors defined over all of $\mathbb{Z}$.

### *2.2 The work of Ludwig Staiger*

The behavioral approach to convolutional codes, is not a new one. In fact, some basic results from the behavioral theory were developed independently for codes by L. Staiger [36] and more recent work involving convolutional codes over groups can also be found in [18]. In [36], Staiger considers the class of behaviors† that are given by finitely generated $\mathbb{F}[[\mathcal{D}]]$ submodules in $\mathbb{F}^n[[\mathcal{D}]]$ as compared to the $\mathbb{F}[\mathcal{D}]$ submodules considered in Section 2. Staiger uses a definition equivalent to the following definition of a convolutional code:

DEFINITION 2.8 Consider a $n \times k$ full rank matrix $G(\mathcal{D})$ defined over the polynomial ring $\mathbb{F}[\mathcal{D}]$. $G(\mathcal{D})$ generates a rate $\frac{k}{n}$ infinite input convolutional code $\tilde{\mathcal{C}}$ through

$$\tilde{\mathcal{C}} := \{w(\mathcal{D}) \mid \exists \ell(\mathcal{D}) \in \mathbb{F}^k[[\mathcal{D}]], w(\mathcal{D}) = G(\mathcal{D})\ell(\mathcal{D})\} \tag{2.7}$$

and we say $G(\mathcal{D})$ is an encoder of the code $\tilde{\mathcal{C}}$.

Going from $\mathbb{F}[\mathcal{D}]$ modules to $\mathbb{F}[[\mathcal{D}]]$ requires two additional conditions on the behavior. The first is that the behavior be closed in the topology of pointwise convergence and the second requirement is that the behavior be "remergable".

In order to help clarify this topology in the present context, we define a metric on $\mathbb{F}^n[[\mathcal{D}]]$ as it was introduced by Staiger in [36]:

DEFINITION 2.9 Let $w, v \in \mathbb{F}^n[[\mathcal{D}]]$ then define $\rho(w, v)$ as:

$$\rho(w,v) := \begin{cases} 0 & \text{if } w = v \\ \frac{1}{s} & \text{where } s := \min\{t | w_t \neq v_t\}. \end{cases} \tag{2.8}$$

Clearly $\rho$ satisfies the basic axioms for a metric and induces therefore a topology. The metric induced by $\rho$ is in fact equivalent to giving $W$ the discrete topology and $W^T$ the product topology hence, equivalent to the topology of pointwise convergence as mentioned previously (see e.g. [23, Chapter 7.4]).

The second requirement is that the behavior be "remergable", which we now define. Let $A \subseteq W^T$.

DEFINITION 2.10 Let $A \subseteq W^T$. A trajectory $a \in A$ is said to be remergable if for every $t \in \mathbb{Z}_+$, there exists a $\gamma \in \mathbb{Z}_+$ and a sequence of vectors $\{v_{t+1}, v_{t+2}, \cdots, v_{t+\gamma}\} \subseteq \mathbb{F}^n$ such that

$$(a_0, a_1, a_2, \ldots, a_t, v_{t+1}, v_{t+2}, \cdots, v_{t+\gamma}, 0, 0, \ldots) \in A.$$

†The class of behaviors Staiger considers is more general: in his work he sets $W = \mathbb{F}$ and $T = \mathbb{Z}_+$, the length of the code words are then derived from the characteristics of the behavior.

The set $A \subseteq W^T$ is said to be remergable if every element in $A$ is remergable.

We now state the main result of interest for us in [36] which gives a classification of the subsets of $W^T$ that represent infinite input convolutional codes.

THEOREM 2.11 *$\tilde{\mathcal{C}} \subseteq W^T$ represents an infinite input convolutional code if and only if $\tilde{\mathcal{C}}$ is linear, closed, time invariant and remergable.*

Let the class of behaviors that satisfy Theorem 2.11 be denoted by $\mathcal{S}^n$. The extra property of remergability gives us the following relationship between the behaviors of $\mathcal{S}^n$ and the set of convolutional codes $\mathcal{C} \subseteq \mathbb{F}^n[\mathcal{D}]$:

THEOREM 2.12 *For any $\tilde{\mathcal{C}} \in \mathcal{S}^n$, there exists a convolutional code $\mathcal{C} \subseteq W^T$, such that $\tilde{\mathcal{C}}$ is the closure of $\mathcal{C}$.*

PROOF Let $G(\mathcal{D})$ be a generator matrix for the convolutional code determined by $\tilde{\mathcal{C}}$. Let $\mathcal{C}$ be the compact support behavior defined by $G(\mathcal{D})$. Remergability implies that for any $\tilde{w} \in \tilde{\mathcal{C}}$ and any $0 < t \in T$, there exists a $w \in \mathcal{C}$ such that $\rho(\tilde{w}, w) \leq \frac{1}{t}$. Hence, every $\tilde{w} \in \tilde{\mathcal{C}}$ is the limit point of a sequence of elements in $\mathcal{C}$, which implies that $\tilde{\mathcal{C}} = \text{closure}(\mathcal{C})$. □

Staigers notion of remergability is in fact equivalent (under the conditions of linearity and right shift invariance) to the systems theoretic notion of controllability which we now recall:

DEFINITION 2.13 Given a dynamical system $\sum = (T, W, \mathcal{B})$, we call $\sum$ *controllable* if for any $w, v \in \mathcal{B}$, and any $s, t \in T$, there exists a $\{u_1, u_2, u_3, \ldots, u_j\}$ such that

$$(w_0, w_1, \ldots, w_s, u_1, u_2, u_3, \ldots, u_j, v_t, v_{t+1}, v_{t+2}, \ldots) \in \mathcal{B}$$

The class of convolutional codes can easily be seen to be controllable if one considers they all have an image representation. In fact one can limit the number of $u$ vectors needed in Definition 2.13 by the memory of the code. As stated previously, we have:

LEMMA 2.14 *Let $\tilde{\mathcal{C}} \subseteq \mathbb{F}[[\mathcal{D}]]$ be linear and time-invariant, then $\tilde{\mathcal{C}}$ is remergable if and only if $\tilde{\mathcal{C}}$ is controllable.*

PROOF The implication $\Leftarrow$ is clear. To prove $\Rightarrow$ let $w, v \in \tilde{\mathcal{C}}$. For any $t \in \mathbb{Z}_+$, $w_{[0,t]}$ remerges to the all zero trajectory in finite time. Let $\tilde{w}$ denote the remerged trajectory. Then by time invariance and linearity the trajectory $\tilde{w} + \mathcal{D}^m v \in \tilde{\mathcal{C}}$ for any $m \in \mathbb{Z}_+$. □

With this in mind, one can see Theorem 2.11 and Theorem 2.12 as dual versions of the following theorem of Willems[‡]:

THEOREM 2.15 [42, Proposition 4.3] *Let $\sum = (\mathbb{Z}_+, \mathbb{F}, \mathcal{B})$ be a dynamical system with $\mathcal{B}$ linear, left shift invariant and closed in the topology of pointwise convergence, then the following conditions are equivalent:*

1. *$\sum$ is controllable.*
2. *$\mathcal{B} = closure(\tilde{\mathcal{B}})$, where $\tilde{\mathcal{B}} = \{w \in \mathcal{B} \mid w$ has compact support $\}$.*
3. *$\mathcal{B} = \operatorname{Im} M(\sigma)$, where $M(s) \in \mathbb{F}^{k \times n}[s]$.*

[‡]The results were stated for $W = \mathbb{R}^n, T = \mathbb{Z}$, here we state the equivalent results in the present framework.

*2.3 Complexity, observability and Kronecker indices*

In Definition 2.5 we defined the rate of a convolutional code $\mathcal{C} \subseteq \mathbb{F}^n[\mathcal{D}]$. In this section we define some further basic properties of convolutional codes and we relate them both to Systems theory and Coding theory. As a preliminary we recall a standard result from module theory:

LEMMA 2.16 *$G(\mathcal{D})$ and $G'(\mathcal{D})$ define the same behavior if and only if there exists a $k \times k$ unimodular matrix $U(\mathcal{D})$ such that*

$$G(\mathcal{D})U(\mathcal{D}) = G'(\mathcal{D}).$$

One important consequence of Lemma 2.16 and Definition 2.2 is that the class of equivalent encoders is given a finer structure than if we were to use Definition 2.8.

EXAMPLE 2.17 Consider the codes with generator matrix defined over $\mathbb{F}_2[\mathcal{D}]$ by

$$G(\mathcal{D}) = \begin{pmatrix} \mathcal{D}^2+1 \\ \mathcal{D}+1 \end{pmatrix}, G'(\mathcal{D}) = \begin{pmatrix} \mathcal{D}+1 \\ 1 \end{pmatrix}.$$

Then $G(\mathcal{D})$ is not externally equivalent to $G'(\mathcal{D})$, however, if one allows infinite input then by considering the input $\frac{1}{\mathcal{D}+1}$ , they are easily seen to be externally equivalent.

Based on Lemma 2.16 we will be able to assume that $G(D)$ has ordered Kronecker (column) indices

$$\nu_1 \geq \cdots \geq \nu_k$$

where the indices $\nu_i$ are formally defined through:

$$\nu_i = \max\{\deg(g_{ij}) \mid 1 \leq j \leq n\} \quad i = 1, \ldots, k.$$

Note that the indices $\nu_1 \geq \cdots \geq \nu_k$ depend only on the particular convolutional code $\mathcal{C} \subseteq W^T$ and are independent of the particular encoder matrix $G(\mathcal{D})$. Also note that the sum of the Kronecker indicies is equal to the highest degree of the full size minors of any encoder $G(\mathcal{D})$. Based on this we define:

DEFINITION 2.18 The complexity of a convolutional code $\mathcal{C}$ is defined as the highest degree of the full size minors of an encoder $G(\mathcal{D})$.

In the systems literature the complexity is also known as the McMillan degree and we will use this convention as well.

We shall now discuss a notion of *observability* for convolutional codes. The classical notion of observability as it was developed by Kalman relates to the reconstructibility of the state from observations of inputs and outputs. In particular for a linear system it means that if inputs and outputs are zero for a sufficiently long period, then the state must be zero. The definition that we shall give for observability of convolutional codes is similar to this interpretation. Note that the definition below gives observability as a property of the code rather than as a property of its representation; this is in marked contrast with the standard behavioral theory [42] where observability is a property of the representation rather than of the behavior.

DEFINITION 2.19 Let a convolutional code $\mathcal{C}$ be given. Consider the polynomials $w(\mathcal{D})$ that have the following property: there exists an integer $J$ such that for all $j \geq J$ there is a polynomial $v_j(\mathcal{D})$ such that $w(\mathcal{D}) + \mathcal{D}^j v_j(\mathcal{D}) \in \mathcal{C}$. The code $\mathcal{C}$ is said to be *observable* if all polynomials with this property belong to $\mathcal{C}$.

EXAMPLE 2.20 The code given by the generator matrix $G(\mathcal{D})$ of Example 2.17 is not observable since $\left[\begin{smallmatrix} \mathcal{D}+1 \\ 1 \end{smallmatrix}\right] + \mathcal{D}^j \left[\begin{smallmatrix} \mathcal{D}+1 \\ 1 \end{smallmatrix}\right]$ belongs to the code for all $j \geq 0$ but $\left[\begin{smallmatrix} \mathcal{D}+1 \\ 1 \end{smallmatrix}\right]$ is not in the code. The generator matrix $G'(\mathcal{D})$ in the same example gives a code which is observable, as is easily seen from the criteria that will be given below.

The relevance of the definition for decoding is obvious, since it means that after a certain number of zeros have been received following a certain non-zero sequence, one can be sure that the received sequence is a message. The number of zeros that one has to wait is allowed by the definition to depend on the message, but we shall see below that actually a uniform bound can be given that depends only on the code. A somewhat more concise reformulation of the definition can be given as follows.

LEMMA 2.21 *A code $\mathcal{C}$ is observable if and only if*

$$\bar{\mathcal{C}} \cap \mathbb{F}^n[\mathcal{D}] \;=\; \mathcal{C}$$

*where the overbar denotes closure in the topology of pointwise convergence.*

PROOF The 'if' part is clear since the sequence of polynomials $w(\mathcal{D})+\mathcal{D}^j v_j(\mathcal{D}) \in \mathcal{C}$ converges to $w(\mathcal{D})$ in the topology of pointwise convergence as $j$ tends to infinity. Suppose now that $\mathcal{C}$ is observable. The inclusion $\mathcal{C} \subseteq \bar{\mathcal{C}} \cap \mathbb{F}^n[\mathcal{D}]$ is obvious. To prove the reverse inclusion, let $w(\mathcal{D}) \in \mathbb{F}^n[\mathcal{D}]$ be such that $w^j(\mathcal{D}) \to w(\mathcal{D})$ for some sequence of polynomials $w^j(\mathcal{D})$ in $\mathcal{C}$. Then there is for any given $\ell$ a $J$ such that $\sum_{i=0}^{\ell} w_i\mathcal{D}^i = \sum_{i=0}^{\ell} w_i^j\mathcal{D}^i$ for $j \geq J$. By the definition of observability, this implies that $w \in \mathcal{C}$. □

Before we come to an algebraic characterization of observability we insert the following lemma which states that "nontrivial 1/1 codes are never observable".

LEMMA 2.22 *Let $p(\mathcal{D})$ be a scalar polynomial of the form $p_\ell D^\ell + p_{\ell+1}D^{\ell+1} + \cdots$, where $p_\ell \neq 0$. Then there exist for all $j \geq 0$ polynomials $r_j(\mathcal{D})$ and $q_j(\mathcal{D})$ such that*

$$p(\mathcal{D})r_j(\mathcal{D}) \;=\; \mathcal{D}^\ell + \mathcal{D}^{\ell+j}q_j(\mathcal{D}). \tag{2.9}$$

*However there exists no polynomial $r(\mathcal{D})$ such that $p(\mathcal{D})r(\mathcal{D}) = \mathcal{D}^\ell$ unless $p(\mathcal{D}) = p_\ell D^\ell$.*

PROOF Define field elements $r_0, \ldots, r_{j-1}$ by solving the system of equations $p_\ell r_0 = 1$, $p_{\ell+1}r_0 + p_\ell r_1 = 0$, …, $p_{\ell+j-1}r_0 + \cdots + p_\ell r_{j-1} = 0$; note that these equations can indeed be solved by the assumption that $p_\ell \neq 0$. Then the polynomial $r_j(\mathcal{D}) := \sum_{i=0}^{j-1} r_i\mathcal{D}^i$ is such that (2.9) holds for some polynomial $q_j(\mathcal{D})$. The second part of the statement is obvious. □

THEOREM 2.23 *A code $\mathcal{C}$ with encoder $G(\mathcal{D})$ is observable if and only if the greatest common divisor of the full size minors of $G(\mathcal{D})$ is of the form $\mathcal{D}^\ell$ for some $\ell \geq 0$.*

PROOF Suppose first that the condition on the minors of $G(\mathcal{D})$ holds. After postmultiplication by a unimodular matrix if necessary, we may assume that $G(\mathcal{D}) = R(\mathcal{D})T(\mathcal{D})$ where $R(\mathcal{D})$ has a left polynomial inverse and $T(\mathcal{D})$ is a diagonal matrix with diagonal entries of the form $\mathcal{D}^j$. We can find a matrix $R'(\mathcal{D})$ such that $[R'(\mathcal{D}) \mid R(\mathcal{D})]$ is unimodular. Define $H(\mathcal{D})$ and $H'(\mathcal{D})$ by

$$\begin{bmatrix} H(\mathcal{D}) \\ H'(\mathcal{D}) \end{bmatrix} \;=\; [R'(\mathcal{D}) \mid R(\mathcal{D})]^{-1}$$

so that in particular $H'(\mathcal{D})R(\mathcal{D}) = I$ and $H(\mathcal{D})R(\mathcal{D}) = 0$. A polynomial $w(\mathcal{D})$ belongs to $\mathcal{C}(G)$ if and only if $H(\mathcal{D})w(\mathcal{D}) = 0$ and $H'(\mathcal{D})w(\mathcal{D}) \in \mathcal{C}(T)$. Suppose now that $w(\mathcal{D})$ is such that for all sufficiently large $j$ there exists a $v_j(\mathcal{D}) \in \mathbb{F}^n[\mathcal{D}]$ such that $w + \mathcal{D}^j v_j(\mathcal{D}) \in \mathcal{C}$; then it follows that $H(\mathcal{D})w(\mathcal{D})+\mathcal{D}^j H(\mathcal{D})v_j(\mathcal{D}) = 0$ and $H'(\mathcal{D})w(\mathcal{D})+\mathcal{D}^j H'(\mathcal{D})v_j(\mathcal{D}) \in \mathcal{C}(T)$. The first condition (applied with $j$ larger than the degree of $w(\mathcal{D})$) implies that $H(\mathcal{D})w(\mathcal{D}) = 0$. As for the second condition, note that membership of $\mathcal{C}(T)$ is decided on the basis of only the lowest coefficients, and so again by taking $j$ sufficiently large it follows that $H'(\mathcal{D})w(\mathcal{D}) \in \mathcal{C}(T)$. Consequently, we have $w \in \mathcal{C}(G)$.

For the converse part of the proof, suppose now that the full size minors of $G(\mathcal{D})$ have a common divisor that is not of the form $p_\ell D^\ell$ for some $\ell$. We can then write $G(\mathcal{D}) = R(\mathcal{D})T(\mathcal{D})$ where $T(\mathcal{D})$

is diagonal and at least one of the diagonal elements is not of the form $p_\ell D^\ell$. It then follows from the preceding lemma that $T(\mathcal{D})$ generates a non-observable behavior, so there exists a polynomial $w'(\mathcal{D})$ such that for all sufficiently large $j$ there is a $v'_j(\mathcal{D})$ such that $w'(\mathcal{D}) + \mathcal{D}^j v'_j(\mathcal{D}) \in \mathcal{C}(T)$, but $w'(\mathcal{D}) \notin \mathcal{C}(T)$. Since $R(\mathcal{D})$ has full column rank, this implies that $R(\mathcal{D})w'(\mathcal{D}) \notin \mathcal{C}$. On the other hand, we do have $R(\mathcal{D})w'(\mathcal{D}) + \mathcal{D}^j R(\mathcal{D})v'_j(\mathcal{D}) \in \mathcal{C}$. It follows that $\mathcal{C}$ is unobservable. □

It can be inferred from the proof that for an observable code $\mathcal{C}$, membership of a polynomial $w(\mathcal{D})$ can be decided by verifying whether $H(\mathcal{D})w(\mathcal{D}) = 0$ where $H(\mathcal{D})$ is some given polynomial matrix (a syndrome former, see below), plus a finite test which involves only the first coefficients of $w(\mathcal{D})$. Moreover it follows that there is a uniform bound on the number of zeros that have to be received before one can conclude that a message has been completed.

The theorem above shows that our concept of 'observability' is equivalent to the existence of a feedforward inverse with delay, as studied by Massey and Sain [22]. There is a conceptual difference, however, since observability is defined above purely in terms of the code itself, where the term 'code' is understood in the behavioral sense as the set of all possible messages.

### *2.4 Syndrome formers*

Recall that in Section (2) we defined an encoder for a code as an injective module homomorphism from $\varphi : \mathbb{F}^k[\mathcal{D}] \longrightarrow \mathbb{F}^n[\mathcal{D}]$ whose image is the code and we associated to this map the matrix $G(\mathcal{D})$. With this in mind, we define:

DEFINITION 2.24 A syndrome former for a code as a module homomorphism given by

$$\psi : \mathbb{F}^n[\mathcal{D}] \longrightarrow \mathbb{F}^{n-k}[\mathcal{D}]$$

with the property that $\mathrm{Im}\,(\varphi) \subseteq \ker(\psi)$.

We will say that a code $\mathcal{C}$ is *shift invariant* if $\mathcal{D}w(\mathcal{D}) \in \mathcal{C}$ implies $w(\mathcal{D}) \in \mathcal{C}$. The next lemma provides a simple criterion for representability by a syndrome former.

LEMMA 2.25 *A convolutional code is shift invariant and observable if and only if there exist an encoder $\varphi$ and a syndrome former $\psi$ such that the following sequence is exact:*

$$0 \longrightarrow \mathbb{F}^k[\mathcal{D}] \xrightarrow{\varphi} \mathbb{F}^n[\mathcal{D}] \xrightarrow{\psi} \mathbb{F}^{n-k}[\mathcal{D}].$$

Note that a syndrome former $\psi$ can also be presented through a $(n-k) \times n$ polynomial matrix $H(\mathcal{D})$, where the definition just gives that $H(\mathcal{D})G(\mathcal{D}) = 0$. In particular

$$\mathrm{Im}\, G(\mathcal{D}) \subseteq \ker H(\mathcal{D})$$

with equality holding if and only if $\mathcal{C}$ is observable.

$H(\mathcal{D})$ induces a representation for the convolutional code $\mathcal{C}$ introduced in (2.3) through:

$$\mathcal{C} = \{w(\mathcal{D}) \mid H(\mathcal{D})w(\mathcal{D}) = 0\}. \tag{2.10}$$

DEFINITION 2.26 We say a syndrome former is a minimal syndrome former if $\psi$ is onto, i.e. if

$$\mathbb{F}^n[\mathcal{D}] \xrightarrow{\psi} \mathbb{F}^{n-k}[\mathcal{D}] \longrightarrow 0.$$

is exact.

In particular we have:

LEMMA 2.27 *Let $H(\mathcal{D})$ be a syndrome former. Then $H(\mathcal{D})$ is minimal if and only if the full size minors are coprime.*

Note that the lemma justifies our use of the word minimal. In conclusion we have that:

THEOREM 2.28 *A convolutional code is shift invariant and observable if and only there are an associated encoder and syndrome former $\varphi$ and $\psi$ such that the following sequence is exact:*

$$0 \longrightarrow \mathbb{F}^k[\mathcal{D}] \xrightarrow{\varphi} \mathbb{F}^n[\mathcal{D}] \xrightarrow{\psi} \mathbb{F}^{n-k}[\mathcal{D}] \longrightarrow 0.$$

EXAMPLE 2.29 Let $\mathcal{C}$ be defined by the generator matrix $G(\mathcal{D}) = (\mathcal{D}+1, \mathcal{D}+1)'$. We can associate to this a minimal syndrome former $H(\mathcal{D}) = (1,1)$. The code generated by $G(\mathcal{D})$ is not observable; the code recognized by $H(\mathcal{D})$ is slightly bigger and is observable.

Note that the definitions presented here are very general indeed. In fact, there is no need to restrict to modules over polynomial rings in this setting. Any mathematical object where the notion of a factor space is well defined can serve for this theory. For this let $\widetilde{C}$ be such a space and assume $C$ is a subset of $\widetilde{C}$ where $\widetilde{C}/C$ is well defined. Let $M$ be a set isomorphic to $C$. We can think of $M$ as our set of messages, $C$ our encoded messages and $\widetilde{C}$ the set of all possible transmitted messages. With the proper restrictions, we can define observablity, controllability and minimality in terms of the associated encoder and syndrome former $\varphi$ and $\psi$ with

$$0 \longrightarrow M \xrightarrow{\varphi} \widetilde{C} \xrightarrow{\psi} \widetilde{C}/C \longrightarrow 0$$

being exact. Here you would require that $\varphi$ and $\psi$ be defined as before, respecting the underlying structure of the objects considered ( e.g. groups, rings, fields, modules, vector spaces etc...).

## 3 First order representations of convolutional codes

In this section we will show how a convolutional code has in a canonical way a first order representation. We derive these results on a purely module theoretic level. It is possible to make use of the duality between codes and autoregressive systems as presented in Section 2 to develop the theory presented in this section.

THEOREM 3.1 (Realization Theorem I) *Assume $\mathcal{C} \subseteq \mathbb{F}^n[\mathcal{D}]$ is a rate $\frac{k}{n}$ convolutional code of complexity $c$. Then there exists $(c+n-k)\times c$ matrices $K, L$ and a $(c+n-k)\times n$ matrix $M$ (all defined over $\mathbb{F}$) such that (2.3) is equivalently described through*

$$\mathcal{C} := \{w(\mathcal{D}) \mid \exists x(\mathcal{D}) : Kx(\mathcal{D}) + \mathcal{D}Lx(\mathcal{D}) + Mw(\mathcal{D}) = 0\}. \tag{3.11}$$

*Moreover the following properties hold:*

1. $\begin{pmatrix} K + \mathcal{D}_0 L \mid M \end{pmatrix}$ *has rank $c+n-k$ for all $\mathcal{D}_0 \in \overline{\mathbb{F}}$.*
2. *$(K + \mathcal{D}_0 L)$ has rank $c \Leftrightarrow G(\mathcal{D}_0)$ has rank $k$.*

PROOF Assume $G(\mathcal{D})$ has column indices $\mu_1 \geq \cdots \geq \mu_m$ and McMillan degree $c := \sum_{i=0}^{m} \mu_i$. Let

$$X(\mathcal{D}) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \mathcal{D} & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \mathcal{D}^{\mu_1-1} & 0 & & 0 \\ 0 & 1 & \vdots & \\ \vdots & \mathcal{D} & & \\ \vdots & & & \\ & \mathcal{D}^{\mu_2-1} & & \\ & 0 & \ddots & 1 \\ \vdots & & & \mathcal{D} \\ \vdots & & \ddots & \\ 0 & 0 & & \mathcal{D}^{\mu_m-1} \end{pmatrix} \tag{3.12}$$

In [31] the matrix $X(\mathcal{D})$ with was called a 'basis matrix'. The matrix $X(\mathcal{D})$ has dimensions $c \times m$, $X(\mathcal{D})$ is of full rank for all $\mathcal{D} \in \overline{\mathbb{F}}$ (the algebraic closure of $\mathbb{F}$) and has the property that for every polynomial vector

$$f(\mathcal{D}) = (f_1(\mathcal{D}), \ldots, f_m(\mathcal{D})) \in \mathbb{F}^m[\mathcal{D}], \deg f_i(\mathcal{D}) \leq \mu_i - 1$$

there exists a unique vector $v \in \mathbb{F}^c$ such that $vX(\mathcal{D}) = f(\mathcal{D})$. Next identify all polynomial vectors in $f(\mathcal{D}) \in \mathbb{F}^m[\mathcal{D}]$ with $\deg f_i(\mathcal{D}) \leq \mu_i$ with a scalar row vector in the space $\mathbb{F}^{m+c}$ and consider the map

$$\begin{array}{rccc} \Phi: & \mathbb{F}^{2c+n} & \longrightarrow & \mathbb{F}^{c+k} \\ & \Phi(v) & \longmapsto & v \begin{bmatrix} X(\mathcal{D}) \\ \mathcal{D}X(\mathcal{D}) \\ G(\mathcal{D}) \end{bmatrix} \end{array} \tag{3.13}$$

Since $X(\mathcal{D})$ is of full rank one verifies that there are $(c+n-k)$ linearly independent constant vectors in the left kernel of this matrix, i.e. there is a full rank matrix $(K\ L\ M)$ of size $(c+n-k) \times (2c+n)$ which can be expressed in pencil from as:

$$\begin{pmatrix} K + \mathcal{D}L & | M \end{pmatrix} \begin{pmatrix} X(\mathcal{D}) \\ G(\mathcal{D}) \end{pmatrix} = 0. \tag{3.14}$$

First we prove that the representation has property (1). Suppose that for some $\mathcal{D}_0 \in \overline{\mathbb{F}}$, and some $y \in \mathbb{F}^{c+n-k}$ we have that $y \begin{pmatrix} K + \mathcal{D}_0 L & M \end{pmatrix} = 0$. This implies that $yK = -\mathcal{D}_0 yL$ so there exists constants $\alpha, \beta$ and a $v \in \mathbb{F}^c$ such that $yK = \alpha v$ and $yL = \beta v$. By (3.14) we obtain $(yK + \mathcal{D}yL)\, X(\mathcal{D}) = (\alpha + \beta\mathcal{D})vX(\mathcal{D}) = 0$ which implies that $(\alpha, \beta) = (0,0)$ or $v = 0$ since $X(\mathcal{D})$ has full rank. Either condition implies that $y\,(K\ L\ M) = 0$ which gives the desired result, $y = 0$. This and the fact that$(G(\mathcal{D}), X(\mathcal{D}))'$ is a minimal basis (i.e. column proper and no rank drop over the algebraic closure) imply that

$$\mathrm{Im}_{\mathbb{F}[\mathcal{D}]} \begin{pmatrix} X(\mathcal{D}) \\ G(\mathcal{D}) \end{pmatrix} = \ker_{\mathbb{F}[\mathcal{D}]} \begin{pmatrix} K + \mathcal{D}L & M \end{pmatrix}.$$

It follows that

$$w(\mathcal{D}) \in \mathcal{C} \Leftrightarrow \exists u(\mathcal{D}), \quad G(\mathcal{D})u(\mathcal{D}) = w(\mathcal{D})$$

$$\Leftrightarrow \begin{pmatrix} X(\mathcal{D})u(\mathcal{D}) \\ G(\mathcal{D})u(\mathcal{D}) \end{pmatrix} = \begin{pmatrix} x(\mathcal{D}) \\ w(\mathcal{D}) \end{pmatrix} \Leftrightarrow \begin{pmatrix} K + \mathcal{D}L & | M \end{pmatrix} \begin{pmatrix} x(\mathcal{D}) \\ w(\mathcal{D}) \end{pmatrix} = 0,$$

hence $\mathcal{C}$ is described by (3.11). Next we prove property (2). From equation (3.14) we obtain $(K + \mathcal{D}L)\, X(\mathcal{D}) = -MG(\mathcal{D})$ From this it follows that if $y \in \ker\,(K + \mathcal{D}_0 L)$then $-yM \in \ker G(\mathcal{D}_0)$. By property (1) we know that $\ker\,(K + \mathcal{D}_0 L) \cap \ker M = \vee$, hence we obtain the following inequality:

$$\dim\,(\ker G(\mathcal{D}_0)) \geq \dim\,(\ker\,(K + \mathcal{D}_0 L)) \geq n - k. \tag{3.15}$$

Therefore, if $\mathrm{rank} G(\mathcal{D}_0) = k$ then $\mathrm{rank}\,(K + \mathcal{D}_0 L) = c$. The converse follows from the fact that $(K + \mathcal{D}_0 L)\, X(\mathcal{D}_0) + MG(\mathcal{D}_0) = 0$ and the matrices $\left(\begin{array}{c|c} K + \mathcal{D}_0 L & M \end{array}\right), X(\mathcal{D}_0)$have full rank imply that $G(\mathcal{D}_0)$ does as well. □

EXAMPLE 3.2 Consider the rate $\frac{2}{3}$ code over $\mathbb{F}_2$ given by the generator matrix:

$$G[\mathcal{D}] := \begin{pmatrix} \mathcal{D}^2 & \mathcal{D}+1 \\ \mathcal{D}^2+\mathcal{D}+1 & 1 \\ 1 & \mathcal{D} \end{pmatrix}$$

The Kronecker indicies are given by $\mu := [\,2, 1\,]$, and a basis matrix is:

$$X(\mathcal{D}) := \begin{pmatrix} 1 & 0 \\ \mathcal{D} & 0 \\ 0 & 1 \end{pmatrix}.$$

The scalar matrix corresponding to $(X(\mathcal{D})', \mathcal{D}X(\mathcal{D})', G(\mathcal{D})')'$ is given by:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

The kernel is given by:

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix},$$

hence,

$$K := \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad L := \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad M := \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Image representations and kernel representations are dual in a certain sense and, as there are canonical first order representations for image representations, there are also canonical first order representations for kernel representations. The following theorem clarifies this.

THEOREM 3.3 (Realization Theorem II) *There exist $c \times (c+k)$ matrices $P, Q$ and a $n \times (c+k)$ matrix $R$ (all defined over $\mathbb{F}$) such that (2.10) is equivalently described through*

$$\{w(\mathcal{D}) = Rz(\mathcal{D}) \mid Pz(\mathcal{D}) = \mathcal{D}Qz(\mathcal{D})\}. \tag{3.16}$$

*where $z(\mathcal{D}) = \sum_t z_t \mathcal{D}^t \in \mathbb{F}^{c+k}[\mathcal{D}]$. Furthermore, the following properties hold:*

1. $\begin{pmatrix} P + \mathcal{D}_0 Q \\ R \end{pmatrix}$ *has rank $c + k$ for all $\mathcal{D}_0 \in \overline{\mathbb{F}}$.*

*2.* $(P + \mathcal{D}_0 Q)$ *has rank* $c \Leftrightarrow H(\mathcal{D}_0)$ *has rank* $n-k$.

PROOF The proof is dual to that of Theorem 3.1 and we omit some of the details. Let $H(\mathcal{D})$ be a syndrome former for the code $\mathcal{C}$. Assume $H(\mathcal{D})$ has row indices $\mu_1 \geq \cdots \geq \mu_{n-k}$ and McMillan degree $c := \sum_{i=0}^{n-k} \mu_i$.Consider the $(n-k) \times c$ matrix of the form

$$X(\mathcal{D}) = \begin{pmatrix} 1 & \mathcal{D} & \dots & \mathcal{D}^{\mu_1-1} & 0 & \cdots & & & & & & 0 \\ 0 & & & 0 & 1 & \mathcal{D} & \dots & \mathcal{D}^{\mu_2-1} & 0 & & & \vdots \\ \vdots & & & & 0 & & & 0 & \ddots & & & 0 \\ 0 & \cdots & & & & & & & \cdots & 0 & 1 & \mathcal{D} & \dots & \mathcal{D}^{\mu_{n-k}-1} \end{pmatrix}.$$

Next identify all polynomial vectors in $f(\mathcal{D}) \in \mathbb{F}^{n-k}[\mathcal{D}]$ with $\deg f_i(\mathcal{D}) \leq \mu_i$ with a scalar column vector in the space $\mathbb{F}^{n-k+c}$ and consider the map

$$\begin{array}{rccc} \Phi: & \mathbb{F}^{2c+n} & \longrightarrow & \mathbb{F}^{n-k+c} \\ & \Phi(v) & \longmapsto & \begin{pmatrix} X(\mathcal{D}) & \mathcal{D}X(\mathcal{D}) & H(\mathcal{D}) \end{pmatrix} v \end{array} \tag{3.17}$$

Since $X(\mathcal{D})$ is of full rank over $\mathbb{F}$, one verifies that there are $c+k$ linearly independent constant vectors in the right kernel of this matrix, i.e. there is a full rank matrix $\begin{pmatrix} P & -Q & R \end{pmatrix}'$ of size $(2c+n) \times (c+k)$ . From this it follows that property (1) holds and that

$$\ker_{\mathbb{F}[\mathcal{D}]} \begin{pmatrix} X(\mathcal{D}) & H(\mathcal{D}) \end{pmatrix} = \text{image}_{\mathbb{F}[\mathcal{D}]} \begin{pmatrix} P - \mathcal{D}Q \\ R \end{pmatrix}.$$

From this we obtain the following:

$$w(\mathcal{D}) \in \mathcal{C} \Leftrightarrow H(\mathcal{D})w(\mathcal{D}) = 0 \Leftrightarrow \begin{pmatrix} X(\mathcal{D}) & H(\mathcal{D}) \end{pmatrix} \begin{pmatrix} 0 \\ w(\mathcal{D}) \end{pmatrix} = 0$$

$$\Leftrightarrow \exists z(\mathcal{D}), \quad \begin{pmatrix} P - \mathcal{D}Q \\ R \end{pmatrix} z(\mathcal{D}) = \begin{pmatrix} 0 \\ w(\mathcal{D}) \end{pmatrix},$$

hence the $\mathcal{C}$ is described by (3.16). The proof of property (2) follows similarly. □

In this representation the $(c+k)$-vector $z(\mathcal{D})$ describes the set of 'internal variables' and $w(\mathcal{D})$ describes again the set of 'external variables', i.e. the behavior of the system. The matrices $P, Q$ are linear maps from the space of internal variables to the space of state variables. Corresponding to change of coordinates one has a natural equivalence among pencil representations

$$(P, Q, R) \sim (SPT^{-1}, SQT^{-1}, RT^{-1}),$$

where $S \in Gl_c$ and $T \in Gl_{c+k}$. As in the case of image representations every equivalence class describes a parity check matrix $H(\mathcal{D})$ uniquely up to pre multiplication by a $(n-k) \times (n-k)$ unimodular matrix.

We would like to remark that the dynamics described through

$$P z_{t+1} = Q z_t; \; w_t = R z_t.$$

is usually characterized in the coding literature through graphical methods — e.g. a *trellis diagram* [27].

Below we describe two alternative representations of convolutional codes that are easily derived from the codes corresponding first order form.

*3.1 Driving variable form*

Using Property (1) of Theorem (3.16) with $\mathcal{D}_0 = 0$ guarantee that after a suitable change of basis we can express the matrices $\begin{pmatrix} P & Q & R \end{pmatrix}$as:

$$P = \begin{pmatrix} I & 0 \end{pmatrix}, Q = \begin{pmatrix} A & B \end{pmatrix}, R = \begin{pmatrix} C & D \end{pmatrix}.$$

After partitioning $z_t$ as $z_t = (x_t{}', u_t{}')'$ , the transformation gives rise to the well known *driving variable* form for convolutional codes given by the system

$$\begin{array}{rcccl} x_{t+1} & = & Ax_t & + & Bu_t \\ w_t & = & Cx_t & + & Du_t \end{array}, \tag{3.18}$$

and discussed in more detail in [21].

*3.2 Input/state/output form*

Properties (1) and (2) of Theorem (3.11) with $\mathcal{D}_0 = 0$ guarantee that after a suitable change of basis we can express the matrices $\begin{pmatrix} K & L & M \end{pmatrix}$as:

$$K = \begin{pmatrix} I \\ 0 \end{pmatrix}, L = \begin{pmatrix} -A \\ C \end{pmatrix}, M = \begin{pmatrix} 0 & B \\ -I & D \end{pmatrix}.$$

After partitioning the output $w_t$ as $w_t = (y_t, u_t)$ the transformation gives rise to the well known *input-state-output* form for convolutional codes given by the system

$$\begin{array}{rcccl} x_{t+1} & = & Ax_t & + & Bu_t \\ y_t & = & Cx_t & + & Du_t \end{array}. \tag{3.19}$$

Note that there is a difference between the systems (3.18) and (3.19). In the latter we are partitioning the behavior of our code as input and output. Hence system (3.19) corresponds to a systematic representation of the code i.e.

$$G(\mathcal{D}) = \begin{pmatrix} I \\ \widetilde{G}(\mathcal{D}) \end{pmatrix}$$

where $\widetilde{G}(\mathcal{D})$ is a matrix over $\mathbb{F}(\mathcal{D})$.

*3.3 Connections to geometry*

In this subsection we would like to shortly explain how our approach to convolutional coding theory is motivated in part by results in algebraic geometry.

A rate $k/n$ linear block code is by definition a $k$-dimensional subspace of the vector space $\mathbb{F}^n$. The set of all $k$-dimensional subspaces of the vector space $\mathbb{F}^n$ is called the Grassmann variety. In this way we can view a rate $k/n$ linear block code as a point in a Grassmann variety, which is a smooth and projective variety. This variety parametrizes therefore all rate $k/n$ linear block codes.

As it was recognized first by Martin and Hermann [19] it is well possible to identify a linear system with a morphism from the projective line to a Grassmann variety. The set of all morphisms from the projective line to a Grassmann variety, whose mapping degree (i.e. the McMillan degree of the underlying system) has a fixed value is called in the algebraic geometry literature a *Quot Scheme*.

Such a Quot Scheme is again a smooth projective variety [28] and as it turns out it naturally parameterizes all linear systems having a fixed input number, a fixed output number and a fixed McMillan degree. It is therefore our belief that this variety should also naturally parameterize all rate $k/n$ convolutional codes having a fixed complexity.

It turns out that the way we defined convolutional codes they indeed become points in a quot scheme. For this first note that the first order representation of the form (3.11) is not unique. Indeed if $T \in Gl_{p+c}$ and $S \in Gl_c$ one has natural equivalent descriptions given through the equivalence relation:

$$(K, L, M) \sim (TKS^{-1}, TLS^{-1}, TM). \tag{3.20}$$

On the other hand one verifies that if both $(K, L, M)$ and $(\tilde{K}, \tilde{L}, \tilde{M})$ are realizations of a particular rate $k/n$ behavior $\mathcal{C} \subseteq \mathbb{F}^n[\mathcal{D}]$ in the sense of Theorem 3.1 then necessarily $(K, L, M)$ and $(\tilde{K}, \tilde{L}, \tilde{M})$ are equivalent in the sense of (3.20).

The equivalence classes induces by the group action (3.20) build up a quot scheme in the following way: Consider the set of all triples $(K, L, M)$ having size $(c+n-k)\times c$, $(c+n-k)\times c$ and $(c+n-k)\times n$ respectively and which satisfy:

1. $\left( \begin{array}{c|c} K + \mathcal{D}_0 L & M \end{array} \right)$ has rank $c + n - k$ for all $\mathcal{D}_0 \in \overline{\mathbb{F}}$.
2. $\left( \begin{array}{c|c} K & M \end{array} \right)$ has rank $c + n - k$.
3. $(K + \mathcal{D}L)$ has rank $c$.

Then the main theorem in [29] states that the set of equivalence classes under the equivalence relation (3.20) forms a quot scheme and this particular quot scheme is a smooth and projective variety. Note that this variety is just the Grassmann variety if the complexity $c = 0$. In this way we can view rate $k/n$ linear block codes as convolutional codes of complexity zero.

## 4 An algebraic construction of convolutional codes

In this section we will apply the ideas developed previously to construct convolutional codes. From now on, we assume that all encoders and syndrome formers considered satisfy the minimality conditions stated in Theorem 2.28. A key problem in convolutional coding theory has been to find a method for effectively characterizing the free distance $d_f$ of a given convolutional code. Very much related to this problem is the task of designing codes of a given rate and complexity with good free distance. At present, the most effective techniques for doing this has been to make an exhaustive search of the class of codes determined by the rate and complexity,and compute the free distance of encoders in this class, until one with maximal or *near* maximal free distance is found. To date, there have been few significant construction techniques developed for convolutional codes.

### *4.1 State observability*

The notion of state observability plays a crucial role in our code construction presented below. Recall the system given by

$$\begin{array}{rcl} x_{t+1} & = & Ax_t + Bu_t \\ y_t & = & Cx_t + Du_t \end{array}. \tag{4.21}$$

If we know the outputs and input for a given time period $[t, t+\tau]$, can we determine what the corresponding states of the machine are? Note that since we know the inputs for all time in the given interval, it is enough to determine the initial state $x_t$. By iterating the equations of system 4.21 we arrive at the following equations:

$$\begin{pmatrix} y_t \\ y_{t+1} \\ y_{t+2} \\ \vdots \\ y_{t+\tau} \end{pmatrix} = \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^\tau \end{pmatrix} x_t + \begin{pmatrix} D & 0 & \cdots & & 0 \\ CB & D & \ddots & & \vdots \\ CAB & CB & D & \ddots & \\ \vdots & \ddots & \ddots & \ddots & 0 \\ CA^{\tau-1}B & CA^{\tau-2}B & \cdots & CB & D \end{pmatrix} \begin{pmatrix} u_t \\ u_{t+1} \\ u_{t+2} \\ \vdots \\ u_{t+\tau} \end{pmatrix}.$$

Clearly, the state will be observable if the matrix $\begin{pmatrix} C \\ \vdots \\ CA^\tau \end{pmatrix}$ has full rank.

DEFINITION 4.1 Let $s$ be the dimension of the state (the $x_i$ vectors) of the system given by equation (4.21). The pair $[A, C]$ is called observable if the matrix

$$\begin{pmatrix} C \\ \vdots \\ CA^{s-1} \end{pmatrix}$$

has rank $s$.

*4.2 A new parity check matrix for convolutional codes*

From the kernel representation of a code $\mathcal{C}$, one has that $w(\mathcal{D}) \in \mathcal{C}$ iff $H(\mathcal{D})w(\mathcal{D}) = 0$. If $w(D) = w_0 + w_1 D + \ldots + w_\gamma D^\gamma$ and $H(\mathcal{D}) = H_0 + H_1\mathcal{D} + \ldots + H_m\mathcal{D}^m$, we can represent this with a the $(m+\gamma+2) \times (\gamma+1)$ sliding block matrix, (see e.g. [17]):

$$\begin{pmatrix} H_0 & 0 & 0 & \cdots & 0 \\ H_1 & H_0 & 0 & \cdots & \vdots \\ H_2 & H_1 & H_0 & \cdots & \vdots \\ \vdots & H_2 & H_1 & \ddots & \vdots \\ \vdots & \ddots & H_2 & \ddots & H_0 \\ H_m & \ddots & \ddots & \ddots & H_1 \\ 0 & H_m & \ddots & \ddots & H_2 \\ 0 & 0 & H_m & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & H_m \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_\gamma \end{pmatrix} = 0.$$

We now derive a sliding block matrix in terms of the matrices $A, B, C, D$ that correspond to the $i/s/o$ representation of the code.

THEOREM 4.2 *Given an observable code $\mathcal{C}$ and the corresponding $i/s/o$ representation given in equation (3.19) with initial state $x_0 = 0$, then $w(\mathcal{D}) = w_0 + w_1 D + \ldots + w_\gamma D^\gamma \in \mathcal{C}$ if and only if*

$$\left(\begin{array}{cccc|ccccc} 0 & \cdots & \cdots & 0 & A^\gamma B & A^{\gamma-1}B & \cdots & AB & B \\ \hline & & & & D & & & & \\ & & & & CB & D & & & \\ & & -I & & CAB & \ddots & D & & \\ & & & & \vdots & \ddots & CB & D & \\ & & & & CA^{\gamma-1}B & CA^{\gamma-2}B & \cdots & CB & D \end{array}\right) \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_\gamma \\ u_0 \\ u_1 \\ \vdots \\ u_\gamma \end{pmatrix} = 0. \tag{4.22}$$

PROOF For the bottom half of the matrix set $t = 0$ and $\tau = \gamma$ in Equation 4.1. To see the top half note that $x_0 = 0$ and $x_{t+1} = Ax_t + Bu_t$ imply that

$$x_{\gamma+1} = A^\gamma B u_0 + A^{\gamma-1} B u_1 + \cdots + AB u_{\gamma-1} + B u_\gamma.$$

Since the code is observable and $w_t = 0$ for all $t \geq \gamma + 1$ we must have $x_{\gamma+1} = 0$ and the result follows. □

*4.3 Input/state/output construction*

In the previous section we gave a factorization of the parity check matrix of a convolutional code in terms of the matrices $A, B, C, D$ corresponding to the $i/s/o$ form of the encoder. In this section we will show how this leads one to a BCH construction for convolutional codes. Let $\alpha \in \mathbb{F}$ and $s, r \in \mathbb{N}$. Consider the following two matrices.

$$A = \begin{pmatrix} \alpha^r & 0 & \cdots & 0 \\ 0 & \alpha^{2r} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \alpha^{sr} \end{pmatrix} \tag{4.23}$$

$$B = \begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{r-1} \\ 1 & \alpha^2 & \alpha^4 & \cdots & \alpha^{2(r-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \alpha^s & \alpha^{2s} & \cdots & \alpha^{s(r-1)} \end{pmatrix} \tag{4.24}$$

THEOREM 4.3 *Let $A$ and $B$ be defined as above. Choose $\alpha$, a primitive element in $\mathbb{F}_q$, where $q \geq s^2 r$. Then the matrix*

$$\begin{pmatrix} B & AB & A^2B & \cdots & A^{s^2}B \end{pmatrix}$$

*has the property that every subset of $s$ columns are linearly independent.*

PROOF By construction the above matrix is a Vandermonde matrix with first row given by

$$\begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{rs^2} \end{pmatrix}.$$

□

This gives us the following:

THEOREM 4.4 *Let $q \geq s^2 r$ and $A$ and $B$ be given by (4.23) and (4.24) respectively. Choose a matrix $C$, so that the pair $[A, C]$ is observable. Then, for any matrix $D$, the convolutional code $\mathcal{C}$ determined by the $i/s/o$ representation given by $[A, B, C, D]$ has*

$$d_f \geq s + 1.$$

REMARK 4.5 That such a matrix $C$ can be chosen and what in particular is a good choice of $C$, as well as the fact that this construction can be extended to the case where the base field is $\mathbb{F}_2$ is discussed in detail in [32].

PROOF Let $w(\mathcal{D}) = w_0 + w_1 D + \ldots + w_\gamma D^\gamma \in \mathcal{C}$ , where $w_t = \begin{pmatrix} y_t \\ u_t \end{pmatrix}$. Then

$$\begin{pmatrix} y_0 & y_1 & \ldots & y_\gamma \\ u_0 & u_1 & & u_\gamma \end{pmatrix}$$

is an input output sequence of the FSM determined by $[A, B, C, D]$ and in particular

$$(y_0, y_1, \ldots, y_\gamma, u_0, u_1, \ldots u_\gamma)'$$

is in the kernel of (4.22). We will show that the weight of $w(\mathcal{D}) \geq s+1$. There are two fundamental cases to consider: 1) $\gamma \leq s^2$, or 2) $\gamma > s^2$.

**case1:** If $\gamma \leq s^2$ then $(u_0, u_1, \cdots u_\gamma)^{'}$ is in the kernel of $\begin{pmatrix} B & AB & A^2B & \cdots & A^{s^2}B \end{pmatrix}$ hence, by construction, $\mathrm{wt}((u_0, u_1, \cdots u_\gamma)) \geq s+1$.

**case2:** If $\gamma > s^2$, then either $\mathrm{wt}((u_0, u_1, \cdots u_{s^2})) \geq s+1$ or $\mathrm{wt}((u_0, u_1, \cdots u_{s^2})) < s+1$. In the first case we are done, so assume that $\mathrm{wt}((u_0, u_1, \cdots u_{s^2})) = b < s+1$. This implies that there are $s-b$ subsequences of length $s$ or bigger containing only zero vectors. Let $u_{t+1}, u_{t+2}, \cdots u_{t+s}$ be one such subsequence. The state of the FSM at time $t+1$ must be non-zero, since if it were zero, this would imply that $(u_0, u_1, \cdots u_t)^{'}$ is in the kernel of $\begin{pmatrix} B & AB & A^2B & \cdots & A^{s^2}B \end{pmatrix}$ where $\mathrm{wt}((u_0, u_1, \cdots u_t)) < s+1$ a contradiction. By observability, this implies that there must be at least one non-zero output $y_i$ for $t+1 \leq i \leq t+s$ . Since we have at least $s-b$ such subsequences, we must have at least $s-b$ non-zero outputs, hence $\mathrm{wt}((w_0, w_1, \cdots w_{s^2})) \geq s-b+b = s$ hence $w(\mathcal{D}) \geq s+1$. □

EXAMPLE 4.6 Let $n = 3, k = 2$ and $s = 4$. Then set $37 = p \geq s^2 k = 32$. Choose $\alpha = 2$, where 2 is a generator for the group of units of $\mathbb{F}_{37}$. The corresponding $A$ and $B$ matrices are:

$$A = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 16 & 0 & 0 \\ 0 & 0 & 27 & 0 \\ 0 & 0 & 0 & 34 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 8 \\ 1 & 16 \end{pmatrix}$$

Next, choose a $C$ so that the system is observable and any $D$. One such choice is as follows:

$$C = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 0 \end{pmatrix}$$

A minimal generator of the code corresponding to this i/s/o representation is:

$$\begin{pmatrix} 28\,\mathcal{D}^2 + 16\,\mathcal{D} + 34 & 21\,\mathcal{D}^2 + 36\,\mathcal{D} + 17 \\ 14 + 31\,\mathcal{D} & 4\,\mathcal{D}^2 + 7\,\mathcal{D} + 4 \\ 4\,\mathcal{D}^2 + \mathcal{D} & 27\,\mathcal{D}^2 + 3\,\mathcal{D} \end{pmatrix}$$

The designed distance is 5, however one can easily show that the actual distance is greater than or equal to 6.

Several authors have presented techniques for constructing convolutional codes [1, 10, 11, 20], however the above construction differs from these in the following ways:

- knowledge of optimal block codes is not required
- works for any rate
- by letting $q = 2^m$, $m \in \mathbb{Z}_+$ one can extend the construction technique to the binary case [32].

EXAMPLE 4.7 Let $p = 1801$, $\alpha = 11$ and $s = 30$. Then

$$\begin{aligned} g_{1,1}(\mathcal{D}) = {} & 315\mathcal{D} + 749 + 75\mathcal{D}^{10} + 897\mathcal{D}^2 + 639\mathcal{D}^3 - 610\mathcal{D}^4 + 872\mathcal{D}^5 - 133\mathcal{D}^6 + 40\mathcal{D}^7 - 431\mathcal{D}^8 + \\ & 565\mathcal{D}^9 + 247\mathcal{D}^{11} + 408\mathcal{D}^{12} + 674\mathcal{D}^{13} - 11\mathcal{D}^{14} - 783\mathcal{D}^{15} \end{aligned}$$

$$\begin{aligned} g_{1,2}(\mathcal{D}) = {} & 935\mathcal{D} + 104\mathcal{D}^{10} + 838\mathcal{D}^2 + 410\mathcal{D}^3 + -340\mathcal{D}^4 - 376\mathcal{D}^5 - 141\mathcal{D}^6 + 995\mathcal{D}^7 + 322\mathcal{D}^8 - \\ & 258\mathcal{D}^9 - 529\mathcal{D}^{11} - 193\mathcal{D}^{12} - 507\mathcal{D}^{13} - 746\mathcal{D}^{14} - 552\mathcal{D}^{15} + 559 \end{aligned}$$

$$\begin{aligned} g_{2,1}(\mathcal{D}) = {} & 825\mathcal{D} + 418\mathcal{D}^{10} + 82\mathcal{D}^2 + 830\mathcal{D}^3 + 47\mathcal{D}^4 + 850\mathcal{D}^5 + 449\mathcal{D}^6 - 741\mathcal{D}^7 + 601\mathcal{D}^8 + 306\mathcal{D}^9 + \\ & 452\mathcal{D}^{11} + 524\mathcal{D}^{12} + 310\mathcal{D}^{13} + 235\mathcal{D}^{14} - 708\mathcal{D}^{15} \end{aligned}$$

$$g_{2,2}(\mathcal{D}) = 1 - 442\mathcal{D} + 672\mathcal{D}^{10} + 756\mathcal{D}^2 - 586\mathcal{D}^3 + 909\mathcal{D}^4 + 224\mathcal{D}^5 - 457\mathcal{D}^6 + 661\mathcal{D}^7 - 532\mathcal{D}^8 - 300\mathcal{D}^9 + 385\mathcal{D}^{11} - 98\mathcal{D}^{12} - 627\mathcal{D}^{13} + 281\mathcal{D}^{14}$$

$$g_{3,1}(\mathcal{D}) = 858\mathcal{D} - 424\mathcal{D}^{10} - 185\mathcal{D}^2 - 91\mathcal{D}^3 + 928\mathcal{D}^4 + 988\mathcal{D}^5 - 570\mathcal{D}^6 - 593\mathcal{D}^7 + 640\mathcal{D}^8 - 631\mathcal{D}^9 + 750\mathcal{D}^{11} + 175\mathcal{D}^{12} + 647\mathcal{D}^{13} + 895\mathcal{D}^{14} - 22\mathcal{D}^{15}$$

$$g_{3,2}(\mathcal{D}) = -22\mathcal{D} + 907\mathcal{D}^{10} + 812\mathcal{D}^2 + 550\mathcal{D}^3 - 615\mathcal{D}^4 + 324\mathcal{D}^5 - 105\mathcal{D}^6 + 435\mathcal{D}^7 + 559\mathcal{D}^8 - 679\mathcal{D}^9 - 336\mathcal{D}^{11} - 472\mathcal{D}^{12} + 544\mathcal{D}^{13} - 273\mathcal{D}^{14} + 233\mathcal{D}^{15}$$

defines a convolutional code with $d_f \geq 31$.

## 5 Conclusions

In this paper we studied convolutional codes in the behavioral framework of linear systems theory. We showed that the class of autoregressive systems can be considered as dual to the class of convolutional codes. We also developed some of the theory of convolutional codes considered as behaviors. Using these tools we were able to represent the class of convolutional codes in ways not considered in the literature previously. By making use of these different representations of codes and their corresponding systems theoretic properties we were able to derive an algebraic construction of convolutional codes where the resulting codes have free distance lower bounded by the overall constraint length + 1.

## References


[1] K. A. S. Abdel-Ghaffar (1989). Some convolutional codes whose free distances are maximal. *IEEE Trans. Inform. Theory* IT-35, No. 1, 188–191.

[2] H. Blomberg and R. Ylinen (1983). *Algebraic Theory for Multivariable Linear Systems.* Academic Press, London.

[3] G. D. Forney (1970). Convolutional codes I. Algebraic structure. *IEEE Trans. Inform. Theory* IT-16, No. 5, 720–738.

[4] G. D. Forney (1973). Structural analysis of convolutional codes via dual codes. *IEEE Trans. Inform. Theory* IT-19, No. 5, 512–518.

[5] G. D. Forney. Algebraic structure of convolutional codes, and algebraic system theory. In *Mathematical System Theory: The Influence of R.E. Kalman*, A.C. Antoulas (ed.). Springer, Berlin, 1991.

[6] P. A. Fuhrmann (1981). *Linear Systems and Operators in Hilbert Space.* McGraw-Hill, New York, NY.

[7] M. L. J. Hautus (1969). Controllability and observability conditions of linear autonomous systems. *Ned. Akad. Wetenschappen, Proc. Ser. A* 72, 443–448.

[8] T.W. Hungerford (1980). *Algebra.* Springer, New York.

[9] R. Johannesson and Z. Wan (1992). A linear algebra approach to minimal convolutional encoders. *IEEE Trans. Inform. Theory* IT-39, No. 4, 1219–1233.

[10] J. Justesen (1973). New convolutional code constructions and a class of asymptotically good time-varying codes. *IEEE Trans. Inform. Theory* IT-19, No. 2, 220–225.

[11] J. Justesen (1975). An algebraic construction of rate $1/\nu$ convolutional codes. *IEEE Trans. Inform. Theory* IT-21, No. 1, 577–580.

[12] T. Kailath (1980). *Linear Systems.* Prentice-Hall, Englewood Cliffs, N.J.

[13] R.E. Kalman, P.L. Falb and M.A. Arbib (1969). *Topics in Mathematical System Theory.* McGraw-Hill, New York.

[14] R. E. Kalman (1963). Mathematical description of linear dynamical systems. *SIAM J. Control Optim.* 1, 152–192.

[15] M. Kuijper (1994). *First-Order Representations of Linear Systems.* Birkhäuser, Boston.

[16] M. Kuijper and J. M. Schumacher (1990). Realization of autoregressive equations in pencil and descriptor form. *SIAM J. Control Optim.* 28, No. 5, 1162–1189.

[17] S. Lin and D. Costello (1983). *Error Control Coding: Fundamentals and Applications.* Prentice-Hall, Englewood Cliffs, NJ.

[18] H. A. Loeliger, G. D. Forney, T. Mittelholzer and M. D. Trott (1994). Minimality and observability of group systems. *Linear Algebra Appl.* 205/206, 937–963.

[19] C. F. Martin and R. Hermann (1978). Applications of algebraic geometry to system theory: the McMillan degree and Kronecker indices as topological and holomorphic invariants. *SIAM J. Control Optim.* 16, 743–755.

[20] J. L. Massey, D.J. Costello and J. Justesen (1973). Polynomial weights and code constructions. *IEEE Trans. Inform. Theory* IT-19, No. 1, 101–110.

[21] J. L. Massey and M. K. Sain (1967). Codes, automata, and continuous systems: explicit interconnections. *IEEE Trans. Automat. Contr.* AC-12, No. 6, 644–650.

[22] J. L. Massey and M. K. Sain (1968). Inverses of linear sequential circuits. *IEEE Trans. on Computers* C-17, No. 4, 330–337.

[23] J. Munkres (1975). *Topology: A First Course.* Prentice-Hall, New Jersey.

[24] J.W. Nieuwenhuis and J. C. Willems. Duality for linear time invariant finite dimensional systems. In *Analysis and Optimization of Systems*, J.L. Lions A. Bensoussan (ed.), Lect. Notes Contr. Inf. Sci. 111. Springer, Berlin, 1988, pp. 13–21.

[25] J. K. Omura (1969). On the Viterbi decoding algorithm. *IEEE Trans. Inform. Theory* IT-15, 177–179.

[26] J. K. Omura (1971). Optimal receiver design for convolutional codes and channels with memory via control theoretical concepts. *Information Sciences* 3, 243–266.

[27] Ph. Piret (1988). *Convolutional Codes, an Algebraic Approach.* MIT Press, Cambridge, MA.

[28] M. S. Ravi and J. Rosenthal (1994). A smooth compactification of the space of transfer functions with fixed McMillan degree. *Acta Appl. Math.* 34, 329–352.

[29] M. S. Ravi and J. Rosenthal (1995). A general realization theory for higher order linear differential equations. *Systems & Control Letters* 25, No. 5, 351–360.

[30] M. S. Ravi, J. Rosenthal and J. M. Schumacher (Aug. 1995). Homogeneous behaviors. Preprint.

[31] M. S. Ravi, J. Rosenthal and J. M. Schumacher (1995). A realization theory for homogeneous AR-systems, an algorithmic approach, in *Proc. IFAC Conference on System Structure and Control* (Nantes, France, July 1995), 183–188.

[32] J. Rosenthal and E.V. York. BCH convolutional codes. In preparation.

[33] J. ROSENTHAL AND E.V. YORK (1995). Linear systems defined over a finite field, dynamic programming and convolutional codes, in *Proc. IFAC Conference on System Structure and Control* Nantes, France, July 1995, 466–471.

[34] M.K. SAIN AND J.L. MASSEY (1969). Invertibility of linear time-invariant dynamical systems. *IEEE Trans. Automat. Contr.* AC-14, 141–149.

[35] J. M. SCHUMACHER. Linear system representations. In *Three Decades of Mathematical System Theory. A Collection of Surveys at the Occasion of the 50th Birthday of Jan C. Willems*, H. Nijmeijer and J. M. Schumacher (eds.), Lect. Notes Contr. Inform. Sci. 135. Springer, Berlin, 1989, pp. 382–408.

[36] L. STAIGER (1983). Subspaces of $GF(q)^\omega$ and convolutional codes. *Information and Control* 59, 148–183.

[37] J. C. WILLEMS (1986). Deducing the input/output and the input/state/output structure from the external behavior, in *Proc. 25th IEEE Conf. Dec. Contr.* (Athens, Dec. 1986), 1936–1937, IEEE Press, New York.

[38] J. C. WILLEMS (1986). From time series to linear system. Part I: Finite dimensional linear time invariant systems. *Automatica* 22, 561–580.

[39] J. C. WILLEMS (1986). From time series to linear system. Part II: Exact modelling. *Automatica* 22, 675–694.

[40] J. C. WILLEMS (1987). From time series to linear system. Part III: Approximate modelling. *Automatica* 23, 87–115.

[41] J. C. WILLEMS (1989). Models for dynamics. *Dynamics Reported* 2, 171–269.

[42] J. C. WILLEMS (1991). Paradigms and puzzles in the theory of dynamical systems. *IEEE Trans. Automat. Control* AC-36, No. 3, 259–294.

[43] E.V. YORK, J. ROSENTHAL AND J. M. SCHUMACHER (1995). On the relationship between algebraic systems theory and coding theory: representations of codes, in *Proc. 34th IEEE Conf. Dec. Contr.* (New Orleans, Dec. 1995), 3271–3276.