Multidimensional convolutional codes in the behavioral approach

Ettore Fornasini and Maria Elena Valcher
Dipartimento di Elettronica ed Informatica, Universitá di Padova
via Gradenigo 6/a, 35131 Padova, Italy
e-mail:fornasini@paola.dei.unipd.it

Abstract

The main features of (finite) multidimensional convolutional codes are introduced as properties of the codewords supports, and connected with the polynomial matrices (encoders) adopted for their description.
Observability and local detectability are shown to be equivalent to the kernel representation of a code via some parity check matrix H^T.
The input/output representation of convolutional codes is finally discussed, and some connections between matrix primeness and the constraints every codeword imposes on the support of the corresponding information sequence are analysed.

Keywords: multidimensional systems, behavior theory, convolutional codes, polynomial matrices, parity checks, observability

1 Introduction

Since the early seventies, the pioneering work of Forney [3, 4] made it quite clear that the theory of discrete-time multidimensional linear systems over a finite field provides a very convenient setting for the analysis of convolutional codes. On the other hand, in the algebraic contex many questions concerning convolutional codes proved to have answers that seem quite illuminating and useful for systems and control applications.
However, even if both fields exhibit some common research directions and resort to similar mathematical tools, the coding point of view is somewhat different from that of linear systems. Actually, in system theory the interest centers around input-output relations, while in coding theory what is most important is the set of output sequences of the encoder, i.e. the internal structure of the code.
Quite recently, the behavioral approach, developed by J.C.Willems [10] for the analysis of dynamical systems, has been applied to the investigation of 1D and nD convolutional codes [1, 5, 7, **?**]. This new framework seems to be quite effective in the nD case, since it allows to investigate the internal properties of the code without explicitly referring to the machinery which underlies the codewords generation and, in particular, without making any assumption on the ordering of two dimensional data. So, in principle, no artificial notion of causality in $\mathbb{Z}^n$, and, consequently, no a priori restriction on the supports of the signals are needed. Indeed, the finite-support constraint we shall introduce in a while on n-dimensional codewords does not follow from causality considerations, but corresponds to the fact that most of nD information sequences encountered in the applications do not infinitely extend in $\mathbb{Z}^n$.
In this comunication we aim to exploit the behavioral approach for analysing the algebraic properties of n-dimensional convolutional codes with finite support codewords. Particular attention has been devoted to the supports of information signals and codewords, as well as to certain elementary operations (restriction, extension and concatenation) which have a concrete meaning from the signal processing standpoint. Actually, several "internal" properties of a code can be introduced in terms of these operations, and expressed as possibilities of "cutting and pasting together" pieces of different trajectories into a new one.

As each of these features mirrors into a particular polynomial matrix representation, an explicit link between the parity checks description of an nD convolutional code and the concept of observability is derived. The codewords of an observable code can be expressed as the solutions of a system of multidimensional difference equations, and hence can be recognized by means of local testing procedures.

A point of view somewhat complementary to detection calls for an input/output analysis of the way the codewords are generated, and their supports are related to the corresponding information sequences. This problem appears particularly relevant when the codewords are injectively generated, and hence a given trajectory is produced by a unique input. Although no general statement can be made on the way these supports are related, specific assumptions on the structure of the generating matrices allow to uniformly confine the support of each input signal into a suitable extension of the support of

the associated codeword.

2 Basic properties of finite convolutional codes

Let $\mathbb{F}$ be a finite field and denote by $\mathbf{z}$ the n-tuple $(z_1, z_2, ..., z_n)$, so that $\mathbb{F}[\mathbf{z}]$ and $\mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]$ are shorthand notations for the polynomial and the Laurent polynomial (L-polynomial) rings in the indeterminates $z_1, ..., z_n$, respectively.
For any sequence $\mathbf{w} = \{\mathbf{w}(\mathbf{h})\}_{\mathbf{h}\in\mathbb{Z}^n}$, taking values in $\mathbb{F}^p$, the *support* of $\mathbf{w}$ is the set of points where $\mathbf{w}$ is nonzero, i.e., $\mathrm{supp}(\mathbf{w}) := \{\mathbf{h} = (h_1, h_2, ..., h_n) \in \mathbb{Z}^n : \mathbf{w}(\mathbf{h}) \neq 0\}$. Also, $\mathbf{w}$ can be represented via a formal power series

$$\sum_{h_i\in\mathbb{Z}} \mathbf{w}(h_1, h_2, ..., h_n)\ z_1^{h_1} z_2^{h_2} \cdots z_n^{h_n} = \sum_{\mathbf{h}\in\mathbb{Z}^n} \mathbf{w}(\mathbf{h})\ \mathbf{z}^{\mathbf{h}},$$

where $\mathbf{h}$ stands for the n-tuple $(h_1, h_2, ..., h_n)$ and $\mathbf{z}^{\mathbf{h}}$ for the term $z_1^{h_1} z_2^{h_2} ... z_n^{h_n}$. On the other hand, power series can be viewed as representing vectors with entries in $\mathcal{F}_\infty := \mathbb{F}^{\mathbb{Z}^n}$, thus setting a bijective map between nD sequences taking values in $\mathbb{F}^p$ and formal power series with coefficients in $\mathbb{F}^p$. This allows us to identify nD sequences with the associated power series, in particular, finite support nD signals with L-polynomial vectors, and to denote both of them with the same symbol $\mathbf{w}$. Sometimes, mostly when a power series $\mathbf{w}$ is obtained as a Cauchy product, it will be useful to denote the coefficient of $\mathbf{z}^{\mathbf{h}}$ in $\mathbf{w}$ as $(\mathbf{w}, \mathbf{z}^{\mathbf{h}})$.
The support of a matrix $G \in \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]^{p\times m}$ is the union of the supports of its elements.

An nD (finite) *convolutional code* $\mathcal{C}$ *of length* p is a set of finite support signals (trajectories, codewords) taking values in $\mathbb{F}^p$ and endowed with the following properties:

(L) [Linearity] *If* $\mathbf{w}_1$ *and* $\mathbf{w}_2$ *belong to* $\mathcal{C}$*, then* $\alpha\mathbf{w}_1 + \beta\mathbf{w}_2 \in \mathcal{C}$*, for all* α*,* β *in* $\mathbb{F}$*;*

(SI) [Shift-Invariance] $\mathbf{w} \in \mathcal{C}$ *implies* $\mathbf{v} = \mathbf{z}^{\mathbf{h}}\mathbf{w} \in \mathcal{C}$ *for every* $\mathbf{h} \in \mathbb{Z}^n$*, i.e.,* $\mathcal{B}$ *is invariant w.r.t. the shifts along the coordinate axes in* $\mathbb{Z}^n$*.*

As every nD code $\mathcal{C}$ can be viewed as an $\mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]$-submodule of $\mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]^p$, which is a Noetherian module [6], $\mathcal{C}$ is finitely generated, i.e., there exists a finite set of column vectors $\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_m$ in $\mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]^p$ such that

$$\begin{aligned} \mathcal{C} &\equiv \left\{\textstyle\sum_{i=1}^m \mathbf{g}_i u_i : u_i \in \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]\right\} \\ &= \{\mathbf{w} = G\mathbf{u} : \mathbf{u} \in \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]^m\} =: \mathrm{Im}G. \end{aligned} \tag{1}$$

The L-polynomial matrix $G := \mathrm{row}\{\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_m\}$ is called *encoder* (or generator matrix) of $\mathcal{C}$.

$G_1 \in \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]^{p\times m_1}$ and $G_2 \in \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]^{p\times m_2}$ are encoders of the same code if and only if there exist $P_1 \in \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]^{m_1\times m_2}$ and $P_2 \in \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]^{m_2\times m_1}$ such that $G_1P_1 = G_2$ and $G_2P_2 = G_1$. Consequently, G_1 and G_2 have the same rank r over the field of rational functions $\mathbb{F}(\mathbf{z})$. Being an invariant w.r.t. all encoders of $\mathcal{C}$, r is called the *rank of* $\mathcal{C}$. It somehow represents a complexity index of the cod, as r independent codewords can be found in $\mathcal{C}$, while $r + 1$ trajectories $(\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_{r+1})$ always satisfy an autoregressive equation $\mathbf{w}_1p_1 + \mathbf{w}_2p_2 + ... + \mathbf{w}_{r+1}p_{r+1} = \mathbf{0}$, with at least one nonzero $p_i \in \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]$.

A code $\mathcal{C}$ of rank r is *free* if it admits a full column rank encoder, that is an encoder G with r columns. This amounts to saying that each trajectory $\mathbf{w}$ in $\mathcal{C}$ is uniquely expressed as a linear combination $\mathbf{w} = \mathbf{g}_1u_1 + \mathbf{g}_2u_2 + \cdots + \mathbf{g}_ru_r$, $u_i \in \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]$, of the columns of G.

The main properties of a finite code $\mathcal{C}$ are connected with certain elementary operations we can perform on its trajectories. These operations essentially reduce to "pasting" pieces of different codewords into legal codewords, or to "cutting" a set of samples out of a given trajectory, so as to obtain a new codeword.

One of the pillars of Willems behavior theory is the notion of (external) controllability. For 1D controllable behaviors the past has no lasting implications about the future [10], which means that the restriction of a 1D trajectory to $(-\infty, t]$ does not provide any information about the values the trajectory takes on $[t+\delta, +\infty)$, when $\delta > 0$ is properly chosen. In a multidimensional context the notions of "past" and "future" are quite elusive and, in many cases, unsuitable for classifying and processing the available data. What seems more reasonable, instead, is to investigate to what extent the values a trajectory $\mathbf{w}$ assumes on a subset $\mathcal{S}_1 \subset \mathbb{Z}^n$ influence the values on a subset $\mathcal{S}_2$, disjoint from $\mathcal{S}_1$, and to check if there exists a lower bound on the distance

$$d(\mathcal{S}_1, \mathcal{S}_2) := \min\left\{\sum_{i=1}^n |h_i - k_i| : \mathbf{h} \in \mathcal{S}_1, \mathbf{k} \in \mathcal{S}_2\right\}, \tag{2}$$

which guarantees that $\mathbf{w}|\mathcal{S}_2$, the *restriction to* $\mathcal{S}_2$ *of the sequence* $\mathbf{w}$, is independent of $\mathbf{w}|\mathcal{S}_1$. This point of view led to the following definition [8].

($\mathbf{C}_1$) [Controllability] *A convolutional code* $\mathcal{C}$ *is* controllable *if there exists an integer* $\delta > 0$ *such that, for any pair of nonempty subsets* $\mathcal{S}_1, \mathcal{S}_2$ *of* $\mathbb{Z}^n$*, with* $d(\mathcal{S}_1, \mathcal{S}_2) \geq \delta$*, and any pair of codewords* $\mathbf{w}_1$ *and* $\mathbf{w}_2 \in \mathcal{C}$*, there exists* $\mathbf{v} \in \mathcal{C}$ *such that*

$$\mathbf{v}|\mathcal{S}_1 = \mathbf{w}_1|\mathcal{S}_1 \qquad \text{and} \qquad \mathbf{v}|\mathcal{S}_2 = \mathbf{w}_2|\mathcal{S}_2. \tag{3}$$

While definition (C_1) requires pasting together different signals into a new one, the following statement refers to the possibility of finding a legal extension for every portion $\mathbf{w}|\mathcal{S}$ of a codeword $\mathbf{w}$, by adjusting the sample values in a small area surrounding $\mathcal{S}$. More precisely, by introducing for $\varepsilon \geq 0$ the *ε-extension* of the set $\mathcal{S}$

$$\mathcal{S}^\varepsilon := \{\mathbf{h} \in \mathbb{Z}^n : d(\mathbf{h}, \mathcal{S}) \leq \varepsilon\},$$

one can give the following definition.

(C_2) [Zero-controllability] *A convolutional code $\mathcal{C}$ is* zero-controllable *if there exists an integer $\varepsilon > 0$ such that, for any nonempty set $\mathcal{S}$ of $\mathbb{Z}^n$ and any $\mathbf{w} \in \mathcal{C}$, there exists $\mathbf{v} \in \mathcal{C}$ satisfying $\mathbf{v}|\mathcal{S} = \mathbf{w}|\mathcal{S}$ and $\mathrm{supp}(\mathbf{v}) \subseteq \mathcal{S}^{\varepsilon}$.*

Properties (C_1) and (C_2) make sense, and are equivalent, both for finite and infinite support behaviors, and the proof of (C_1) $\Leftrightarrow$ (C_2) given below holds for both of them. However, while for an infinite behavior controllability constitutes an additional constraint w.r.t. linearity and shift invariance [8, 9], conditions (C_1) and (C_2) are always met by a finite convolutional code $\mathcal{C}$, as a consequence of its module structure [2].

Given two disjoint sets $\mathcal{S}_1$ and $\mathcal{S}_2$ which are far enough apart, controllability expresses the possibility of steering any code sequence known in $\mathcal{S}_1$ into another element of $\mathcal{C}$ assigned on $\mathcal{S}_2$, meanwhile producing a legal codeword. Like controllability, also observability will be introduced without reference to the concept of state, according to some recent works of Forney et al. [5, 7]. Observability formalizes the possibility of pasting into a codeword any pair of trajectories that take the same values on a sufficiently large subset of $\mathbb{Z}^n$. This is equivalent to saying that, however a codeword $\mathbf{w} \in \mathcal{C}$ and a subset $\mathcal{S} \subset \mathbb{Z}^n$ are chosen, the possible extensions of $\mathbf{w}|\mathcal{S}$ only depend on the values of $\mathbf{w}$ on a boundary region of $\mathcal{S}$.

Under this viewpoint, observability endows a convolutional code with a "separation property" that allows to take into account only a small amount of data in order to extend a portion of codeword.

(O_1) [Observability] *A convolutional code $\mathcal{C}$ is* observable *if there exists an integer $\delta > 0$ such that, for any pair of nonempty subsets $\mathcal{S}_1, \mathcal{S}_2$ of $\mathbb{Z}^n$, with $d(\mathcal{S}_1, \mathcal{S}_2) \geq \delta$, and any pair of codewords $\mathbf{w}_1$, $\mathbf{w}_2 \in \mathcal{C}$, satisfying $\mathbf{w}_1|\mathcal{C}(\mathcal{S}_1 \cup \mathcal{S}_2) = \mathbf{w}_2|\mathcal{C}(\mathcal{S}_1 \cup \mathcal{S}_2)$, the sequence*

$$\mathbf{v}(\mathbf{h}) = \begin{cases} \mathbf{w}_1(\mathbf{h}) & \mathbf{h} \in \mathcal{S}_1 \\ \mathbf{w}_1(\mathbf{h}) = \mathbf{w}_2(\mathbf{h}) & \mathbf{h} \in \mathcal{C}(\mathcal{S}_1 \cup \mathcal{S}_2) \\ \mathbf{w}_2(\mathbf{h}) & \mathbf{h} \in \mathcal{S}_2 \end{cases} \tag{4}$$

is a codeword.

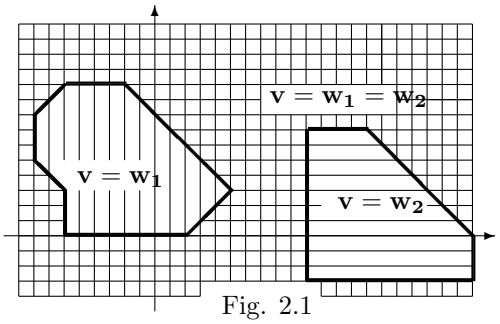

Fig. 2.1

Observability can be equivalently restated as follows: if the support of a codeword $\mathbf{w}$ can be partitioned into a pair of disjoint subsets, which are far enough apart, the restrictions of $\mathbf{w}$ to each subset represent codewords.

(O_2) [Zero-observability] *A convolutional code $\mathcal{C}$ is* zero-observable *if there exists an integer $\varepsilon > 0$ such that for any $\mathbf{w} \in \mathcal{C}$ satisfying $\mathbf{w}|(\mathcal{S}^{\varepsilon} \setminus \mathcal{S}) = \mathbf{0}$, $\mathcal{S}$ a nonempty set in $\mathbb{Z}^n$, the sequence*

$$\mathbf{v}(\mathbf{h}) = \begin{cases} \mathbf{w}(\mathbf{h}) & \mathbf{h} \in \mathcal{S} \\ 0 & \textit{elsewhere} \end{cases} \tag{5}$$

belongs to $\mathcal{C}$.

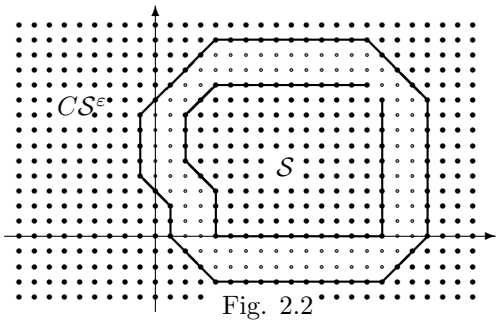

Fig. 2.2

Proposition 2.2 *Observability and zero observability are equivalent.*

PROOF (O_1) $\Rightarrow$ (O_2) Assume that $\mathcal{C}$ fulfills condition (O_1). Given $\mathcal{S} \subset \mathbb{Z}^n$ and $\mathbf{w} \in \mathcal{C}$ such that $\mathbf{w}|(\mathcal{S}^{\delta} \setminus \mathcal{S}) = \mathbf{0}$, take in ($O_1$) $\mathbf{w}_1 = \mathbf{w}$, $\mathbf{w}_2 = \mathbf{0}$, $\mathcal{S}_1 = \mathcal{S}$ and $\mathcal{S}_2 = \mathcal{C}\mathcal{S}^{\delta}$. The trajectory $\mathbf{v} \in \mathcal{C}$ satisfying (4), satisfies also (5) with $\varepsilon = \delta$.

(O_2) $\Rightarrow$ (O_1) Assume that $\mathcal{C}$ fulfills condition (O_2). Given $\mathcal{S}_1, \mathcal{S}_2 \subset \mathbb{Z}^n$, with $d(\mathcal{S}_1, \mathcal{S}_2) > \varepsilon$, and $\mathbf{w}_1$, $\mathbf{w}_2 \in \mathcal{C}$ satisfying $\mathbf{w}_1|\mathcal{C}(\mathcal{S}_1 \cup \mathcal{S}_2) = \mathbf{w}_2|\mathcal{C}(\mathcal{S}_1 \cup \mathcal{S}_2)$, the sequence $\mathbf{w}_1 - \mathbf{w}_2 \in \mathcal{C}$ satisfies $(\mathbf{w}_1 - \mathbf{w}_2)|\mathcal{C}(\mathcal{S}_1 \cup \mathcal{S}_2) = \mathbf{0}$. As a consequence, the sequence $\mathbf{w}$ given by

$$\mathbf{w}(\mathbf{h}) = \begin{cases} \mathbf{w}_1(\mathbf{h}) - \mathbf{w}_2(\mathbf{h}) & \mathbf{h} \in \mathcal{S}_1 \\ 0 & \text{elsewhere} \end{cases}$$

is in $\mathcal{C}$, and $\mathbf{v} := \mathbf{w} + \mathbf{w}_2 \in \mathcal{C}$ fulfills (4). So, (O_1) holds for $\delta = \varepsilon + 1$. ■

3 Codeword recognition

Observability is related with the issue of recognizing whether a given sequence $\mathbf{v} \in \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]^p$ is an element of $\mathcal{C}$. This problem can be managed by resorting to a linear filter, a *syndrome former*, that produces an identically zero output signal when the input is a codeword of $\mathcal{C}$. From a mathematical point of view, this requires to find a set of sequences (parity checks) endowed with the property that their convolution with every element of $\mathcal{C}$ is zero.

So, for a given code $\mathcal{C} \subseteq \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]^p$, a (finite) *parity check* is a column vector $\mathbf{s} \in \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]^p$ that satisfies $\mathbf{s}^T\mathbf{w} = \mathbf{0}$, for all $\mathbf{w} \in \mathcal{C}$. The set $\mathcal{C}^{\perp}$ of all finite parity checks of $\mathcal{C}$ is the *orthogonal code*, and as a submodule of

$\mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$, it is generated by the columns of some matrix $H \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^{p\times q}$, that is

$$\mathcal{C}^\perp = \{\mathbf{s} \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p : \mathbf{s} = H\mathbf{x}, \mathbf{x} \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^q\} = \mathrm{Im}H. \quad (6)$$

Condition $\mathbf{s}^T\mathbf{w} = \mathbf{0}$, $\forall\ \mathbf{s} \in \mathcal{C}^\perp$, however, need not imply $\mathbf{w} \in \mathcal{C}$. In general

$$\mathcal{C}^{\perp\perp} := \{\mathbf{w} \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p : \mathbf{s}^T\mathbf{w} = \mathbf{0}, \forall\ \mathbf{s} \in \mathcal{C}^\perp\} \quad (7)$$

properly includes $\mathcal{C}$, and is the set of all L-polynomial vectors obtained by combining the columns of G over the field of rational functions $\mathbb{F}(\mathbf{z})$. It is clear that $\mathcal{C}$ can be identified via a finite set of parity checks if and only if $\mathcal{C} = \mathcal{C}^{\perp\perp}$ or, equivalently,

$$\mathcal{C} = \ker H^T := \{\mathbf{w} \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p : H^T\mathbf{w} = \mathbf{0}\}. \quad (8)$$

In this setting, observability finds a somewhat more substantial interpretation. Actually, if $\mathcal{C} = \ker H^T$, the restriction of a code sequence to a set $\mathcal{S}$ still provides a legal codeword every time the distance between $\mathcal{S}$ and the remaining support of the sequence exceeds the range of action of the parity check matrix H.

Proposition 3.1 below shows that kernel representations are possible, as it can be expected, only for observable code, and makes it clear that observability induces further constraints on the structure of $\mathcal{B}$, in addition to linearity and shift invariance.

Proposition 3.1 *A convolutional code $\mathcal{C} \subseteq \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$ is observable if and only if there exist an integer $h > 0$ and an L-polynomial matrix $H^T \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^{h\times p}$ such that $\mathcal{B} = ker H^T$.*

The proof of the proposition depends on the following technical lemma.

Lemma 3.2 [2] *Let $m(\mathbf{z})$ be in $\mathbb{F}[\mathbf{z}]$. For any integer $\rho > 0$ there is $p(\mathbf{z}) \in \mathbb{F}[\mathbf{z}]$ s.t. $m(\mathbf{z})p(\mathbf{z}) \in \mathbb{F}[\mathbf{z}^\rho] := \mathbb{F}[z_1^\rho, ..., z_n^\rho]$.* ■

Proof of Proposition 3.1 Assume that $\mathcal{C} = \mathrm{Im}G$, $G \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^{p\times m}$, is an observable code, and let $\mathcal{C}^\perp = \mathrm{Im}\ H$, $H \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^{p\times q}$, denote the orthogonal code introduced in (6). We will show that $\mathcal{C} \equiv \ker H^T$. Since $H^TG = \mathbf{0}$, it is clear that $\ker H^T \supseteq \mathcal{C}$. To prove the converse, express $\mathbf{w} \in \ker H^T$ as $\mathbf{w} = G\mathbf{n}/d(\mathbf{z})$, $d \in \mathbb{F}[\mathbf{z}], \mathbf{n} \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^{m\times 1}$. By Lemma 3.2, for every integer $\rho > 0$ there is a suitable polynomial $p(\mathbf{z})$ such that $p(\mathbf{z})d(\mathbf{z}) \in \mathbb{F}[z_1^\rho, \ldots, , z_n^\rho]$. If property (O_2) holds for $\varepsilon > 0$, and $r > 0$ is an integer such that $\mathrm{supp}(\mathbf{w}) \subseteq B(\mathbf{0}, r)$, we choose $\rho > 2r + \varepsilon$. So, the codeword $p(\mathbf{z})d(\mathbf{z})\mathbf{w} = G\mathbf{n}p(\mathbf{z})$ can be written as $\sum_{i_1,i_2,\ldots,i_n} c_{i_1,i_2,\ldots,i_n}\, z_1^{\rho i_1} z_2^{\rho i_2} \cdots z_n^{\rho i_n}\, \mathbf{w}$, and thus is the sum of disjoint shifted copies of $\mathbf{w}$, and the distance between two arbitrary copies exceeds ε. So, by (O_2), each copy of $\mathbf{w}$, and hence $\mathbf{w}$ itself, is in $\mathcal{C}$.

Conversely, let $\mathcal{C} = \ker H^T$, and set $\varepsilon = 2s$, with $s > 0$ an integer s.t. $B(\mathbf{0}, s) \supseteq \mathrm{supp}(H^T)$. If $\mathcal{S}$ is a subset of $\mathbb{Z}^n$ and $\mathbf{w} \in \mathcal{C}$ satisfies $\mathbf{w}|(\mathcal{S}^\varepsilon \setminus \mathcal{S}) = \mathbf{0}$, the sequence

$$\mathbf{v}(\mathbf{h}) = \begin{cases} \mathbf{w}(\mathbf{h}) & \mathbf{h} \in \mathcal{S} \\ 0 & \text{elsewhere} \end{cases}$$

is in $\ker H^T$ and hence in $\mathcal{C}$. Consequently, $\mathcal{C}$ is zero-observable. ■

The kernel description given in Proposition 3.1 leads to new insights into the internal structure of an observable code. Observability, indeed, expresses the "local nature" of the code or, equivalently, the existence of a bound on the size of all windows (in $\mathbb{Z}^n$) we have to look at when deciding whether a signal belongs to $\mathcal{C}$. Denoting by $\mathcal{C}|\mathcal{S} := \{\mathbf{w}|\mathcal{S} : \mathbf{w} \in \mathcal{C}\}$ the set of all restrictions to $\mathcal{S}$ of code trajectories, the above localization property finds a formal statement as follows:

(O_3) [Local-detectability] *A convolutional code $\mathcal{C}$ is* locally-detectable *if there is an integer $\nu > 0$ such that every signal $\mathbf{w}$ satisfying $\mathbf{w}|\mathcal{S} \in \mathcal{C}|\mathcal{S}$ for every $\mathcal{S} \subset \mathbb{Z}^n$ with $\mathrm{diam}(\mathcal{S}) \leq \nu$, is in $\mathcal{C}$.*

Proposition 3.3 *Local detectability and observability are equivalent.*

Proof Assume that $\mathcal{C}$ satisfies (O_3) for a certain $\nu > 0$. Given $\mathcal{S} \subset \mathbb{Z}^n$ and $\mathbf{w} \in \mathcal{C}$ such that $\mathbf{w}|(\mathcal{S}^\nu \setminus \mathcal{S}) = \mathbf{0}$, define $\mathbf{v}$ as follows

$$\mathbf{v}(\mathbf{h}) = \begin{cases} \mathbf{w}(\mathbf{h}) & \mathbf{h} \in \mathcal{S}^\nu \\ 0 & \text{elsewhere.} \end{cases} \quad (9)$$

Consider any window $\mathcal{W}$, with $\mathrm{diam}(\mathcal{W}) \leq \nu$. If $\mathcal{W}$ is included in $\mathcal{S}^\nu$, then $\mathbf{v}|\mathcal{W} = \mathbf{w}|\mathcal{W} \in \mathcal{C}|\mathcal{W}$, otherwise we have $\mathcal{W} \cap \mathcal{S} = \emptyset$, and therefore $\mathbf{v}|\mathcal{W} = \mathbf{0}|\mathcal{W} \in \mathcal{C}|\mathcal{W}$. So, by (O_3), $\mathbf{v}$ is a codeword, and (O_2) holds for $\varepsilon = \nu$.

Conversely, assume that $\mathcal{C}$ is observable. By Proposition 3.1, there exists an L-polynomial matrix $H \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^{p\times q}$ such that $\mathcal{B} = \ker H^T$. Let $\nu > 0$ be an integer such that $\mathrm{supp}(H^T) \subseteq B(\mathbf{0}, \nu)$, and suppose that $\mathbf{v}$ is any signal satisfying $\mathbf{v}|\mathcal{S} \in \mathcal{C}|\mathcal{S}$ for every $\mathcal{S} \subset \mathbb{Z}^n$ with $\mathrm{diam}(\mathcal{S}) \leq 2\nu$. If $\bar{\mathcal{S}} := -\mathrm{supp}(H^T)$, the computation of the coefficient of $\mathbf{z}^\mathbf{k}$ in $H^T\mathbf{v}$ involves only samples of $\mathbf{v}$ indexed in

$$\mathbf{k} + \bar{\mathcal{S}} := \{\mathbf{h} \in \mathbb{Z}^n : \mathbf{h} - \mathbf{k} \in \bar{\mathcal{S}}\} = -\mathrm{supp}(\mathbf{z}^\mathbf{k} H^T). \quad (10)$$

On the other hand, since $\mathrm{diam}(\mathbf{k} + \bar{\mathcal{S}}) \leq 2\nu$, there exists $\mathbf{w_k} \in \mathcal{C}$ which satisfies $\mathbf{v}|(\mathbf{k} + \bar{\mathcal{S}}) = \mathbf{w_k}|(\mathbf{k} + \bar{\mathcal{S}})$, and this result holds for every $\mathbf{k} \in \mathbb{Z}^n$. So, the coefficient of $\mathbf{z}^\mathbf{k}$ in $H^T\mathbf{v}$ is the same as in $H^T\mathbf{w_k} \equiv \mathbf{0}$, and hence $\mathbf{v} \in \ker H^T = \mathcal{C}$. ■

The equivalent descriptions of observability given in $(\mathrm{O}_1) \div (\mathrm{O}_3)$ rely on the codewords' supports, whereas Proposition 3.1 involves parity checks and kernel representations. A different approach to this notion consists of regarding convolutional codes of length p as elements in the lattice of submodules of $\mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$, and investigating whether observable elements enjoy some special ordering properties.

Keeping in with the same spirit, we may investigate how an observable code is affected by certain "extension operations" that merge lattice elements into larger

ones. There are essentially two natural ways to perform these extensions: one consists of embedding $\mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$, and therefore each of its submodules, in the rational vector space $\mathbb{F}(\mathbf{z})^p$, the other of considering $\mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$ as a submodule of $\mathcal{F}_\infty^p$, the set of nD codewords of length p, whose supports possibly extend to the whole space $\mathbb{Z}^n$.

Once a convolutional code $\mathcal{C}$ of length p is given, in the first case we have to consider the smallest vector subspace of $\mathbb{F}(\mathbf{z})^p$ including $\mathcal{C}$

$$\mathcal{C}_{\rm rat} := \Big\{\sum_{i=1}^{r} \mathbf{w}_i a_i : \mathbf{w}_i \in \mathcal{C},\ a_i \in \mathbb{F}(\mathbf{z}),\ r \in \mathbb{N}\Big\}, \quad (11)$$

and restrict our attention to the submodule $\mathcal{C}_{\rm rat} \cap \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$ of finite support codewords. In general this properly includes $\mathcal{C}$, and hence is a larger element of the lattice. In the other case, we merge $\mathcal{C}$ in

$$\mathcal{C}_\infty := \Big\{\sum_{i=1}^{r} \mathbf{w}_i a_i : \mathbf{w}_i \in \mathcal{C},\ a_i \in \mathcal{F}_\infty,\ r \in \mathbb{N}\Big\}, \quad (12)$$

the smallest $\mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]$-submodule of $\mathcal{F}_\infty^p$ which includes $\mathcal{C}$. Again we have to confine ourselves to the set of its finite elements $\mathcal{C}_\infty \cap \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$, which clearly includes all codewords of $\mathcal{C}$.

Proposition 3.4 *Let $\mathcal{C} \subseteq \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$ be a convolutional code of rank r. The following statements are equivalent:*

(1) $\mathcal{C}$ is observable;

(2) $\mathcal{C} \equiv \mathcal{C}_\infty \cap \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$;

(3) $\mathcal{C} \equiv \mathcal{C}_{\rm rat} \cap \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$;

(4) $\mathcal{C}$ is maximal in the set of all submodules of $\mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$ of rank r;

(5) $s\mathbf{w} \in \mathcal{C} \ \Rightarrow\ \mathbf{w} \in \mathcal{C}$, for every $\mathbf{w} \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$ and every nonzero $s \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]$;

(6) $\mathcal{C} = \mathcal{C}^{\perp\perp}$.

PROOF (1) $\Rightarrow$ (2) As $\mathcal{C}$ is observable, there exists $H \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^{p\times q}$ such that $\mathcal{C} = \ker H^T = \{\mathbf{w} \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p : H^T\mathbf{w} = \mathbf{0}\}$. If $\mathbf{w} \in \mathcal{C}_\infty \cap \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$, then $\mathbf{w} = \sum_i \mathbf{w}_i a_i$, $a_i \in \mathcal{F}_\infty, \mathbf{w}_i \in \mathcal{C}$, and therefore $H^T\mathbf{w} = H^T\Big(\sum_i \mathbf{w}_i a_i\Big) = \sum_i (H^T\mathbf{w}_i)a_i = \mathbf{0}$. Thus $\mathbf{w} \in \ker H^T = \mathcal{C}$, which implies $\mathcal{C} \supseteq \mathcal{C}_\infty \cap \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$. The reverse inclusion is obvious.

(2) $\Rightarrow$ (3) Follows immediately from $\mathcal{C} \subseteq \mathcal{C}_{\rm rat} \cap \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p \subseteq \mathcal{C}_\infty \cap \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$.

(3) $\Rightarrow$ (4) If $\mathcal{C}' \supseteq \mathcal{C}$ and rank$\mathcal{B}'$ = rank$\mathcal{C}$, it is clear that $\mathcal{B}$ and $\mathcal{C}'$ generate the same $\mathbb{F}(\mathbf{z})$-subspace of $\mathbb{F}(\mathbf{z})^p$ and, consequently, $\mathcal{C}_{\rm rat} \cap \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p = \mathcal{C}'_{\rm rat} \cap \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$. So, the inclusions chain $\mathcal{C}_{\rm rat} \cap \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p \supseteq \mathcal{C}' \supseteq \mathcal{C}$ and assumption (3) together imply $\mathcal{C}' = \mathcal{C}$, which means that $\mathcal{C}$ is maximal.

(4) $\Rightarrow$ (5) Suppose $s\mathbf{w} \in \mathcal{C}$, $s \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]$. The code $\mathcal{C}'$ generated by $\mathcal{C}$ and $\mathbf{w}$ has the same rank of $\mathcal{B}$, and hence, by the maximality assumption, coincides with $\mathcal{C}$.

(5) $\Rightarrow$ (6) As $\mathcal{C}$ and $\mathcal{C}^{\perp\perp}$ have the same rank r and $\mathcal{C}^{\perp\perp} \supseteq \mathcal{C}$, both codes generate the same $\mathbb{F}(\mathbf{z})$-subspace of $\mathbb{F}(\mathbf{z})^p$. In particular, $\mathbf{w} \in \mathcal{C}^{\perp\perp}$ implies $\mathbf{w} \in (\mathcal{C}^{\perp\perp})_{\rm rat} = \mathcal{C}_{\rm rat}$. So, there exist $p_i, s_i \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]$ and $\mathbf{w}_i \in \mathcal{B}$, such that $\mathbf{w} = \sum_{i=1}^{r} \mathbf{w}_i\ p_i/s_i$, which implies $s\mathbf{w} \in \mathcal{C}$, $s = \ell$.c.m.$\{s_i\}$. By assumption (5), also $\mathbf{w}$ is in $\mathcal{C}$.

(6) $\Rightarrow$ (1) Since $\mathcal{C}^\perp$ is a submodule of $\mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p$, there exists a suitable L-polynomial matrix H such that $\mathcal{C}^\perp = {\rm Im}H$. So

$$\mathcal{C}^{\perp\perp} = \{\mathbf{w} \in \mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^p : \mathbf{v}^T\mathbf{w} = \mathbf{0}, \forall \mathbf{v} \in {\rm Im}H\} = \ker H^T.$$

By assumption (6), $\mathcal{C}$ coincides with $\ker H^T$, and hence is observable. ■

4 Codeword generation

The analysis we carried out in the previous sections focused on the properties of the codewords, without concern for the way they are generated. Once a code $\mathcal{C}$ is represented via a finite set of generators $\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_m$, however, it is natural to look at $G := [\mathbf{g}_1\ \mathbf{g}_2\ ...\ \mathbf{g}_m]$ as a transfer matrix, and hence to consider $\mathcal{C}$ as the image of an input-output map acting on $\mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^m$. When the input/output point of view is adopted, it is often imperative to associate codewords of $\mathcal{C}$ and information sequences bijectively. The meaning of this requirement is that information messages need to be unambiguously retrieved from the received codewords at the decoding stage, which amounts to say that the encoder G induces a 1-1 map.

Throughout this section we steadily assume that $\mathcal{C}$ has a full column rank encoder G, and hence is a free module. Under this assumption, G admits (possibly infinitely many) rational left inverses G^{-1}. Each of them, when applied to a codeword $\mathbf{w} = G\mathbf{u}$, uniquely retrieves the (finite) information sequence $\mathbf{u}$. This implies that every estimate $\hat{\mathbf{w}} \in \mathcal{B}$ of a codeword $\mathbf{w}$ produces a finite error $\mathbf{e_u} := \mathbf{u} - G^{-1}\hat{\mathbf{w}} = G^{-1}(\mathbf{w} - \hat{\mathbf{w}})$ in reconstructing the information sequence $\mathbf{u}$. Consequently, no catastrophic error can arise [1, 3]. However, if we apply the "decoder" G^{-1} directly to the noisy sequence $\mathbf{v} = \mathbf{w} + \mathbf{r}$, as $\mathbf{r}$ generally is not an element of $\mathcal{C}$, the decoding algorithm possibly gives an infinite support sequence, which differs from the correct input in infinitely many points and clearly is not even an admissible information sequence. This drawback can be avoided if and only if G^{-1} is an L-polynomial matrix.

Proposition 4.1 below provides equivalent conditions for the existence of an L-polynomial inverse, and in particular shows that such an inverse exists if and only if G is left zero-prime (ℓZP).

Definition *Let G be in $\mathbb{F}[\mathbf{z},\mathbf{z}^{-1}]^{p\times m}$ and $\hat{G} = \mathbf{z}^{\mathbf{h}}G = z_1^{h_1}\cdots z_n^{h_n}G$ in $\mathbb{F}[\mathbf{z}]^{p\times m}$ for some $\mathbf{h} \in \mathbb{N}^n$. If $\mathbb{K}$ denotes the algebraic closure of $\mathbb{F}$, the* L-variety $\mathcal{V}^L(G)$ of the

maximal order minors of G *is the algebraic set*

$$\mathcal{V}^L(G) := \mathcal{V}(\hat{G}) \setminus \Big\{(k_1, ..., k_n) : k_i \in \mathbb{K},\ \prod_i k_i = 0\Big\}, \tag{13}$$

where $\mathcal{V}(\hat{G})$ *denotes the variety (in* $\mathbb{K}$*) of the maximal order minors of* $\hat{G}$.

The above definition is well-posed, as (13) does not depend on the choice of $\hat{G}$.

Proposition 4.1 [2] *Let* G *be a* $p \times m$ *L-polynomial matrix. The following statements are equivalent:*
i) G *is right zero-prime (rZP);*
ii) there exists $P \in \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]^{m \times p}$ *s.t.* $PG = I_m$;
iii) $\mathcal{V}^L(G)$ *is empty;*
iv) $Im\ G^T = \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]^m$. ■

When the encoder G has an L-polynomial inverse, a uniform bound can be found on the support of the information sequences which correspond to the code trajectories. Actually, if P is such an inverse, $\mathbf{w} \in \mathcal{C}$ is generated by the input signal $\mathbf{u} = P\mathbf{w}$ whose support cannot exceed "too much" that of $\mathbf{w}$. This feature, we will refer to as *wrapping input property*, is quite appealing, as the mere recognition of the support of a codeword allows the derivation of a uniformly tight bound on the support of the corresponding information sequence and hence guarantees that small errors in the codeword estimate reflect into small errors in the information sequence reconstruction.

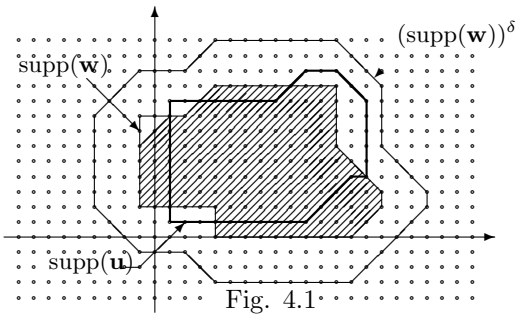

Fig. 4.1

(WI) [Wrapping input property] *A convolutional code* $\mathcal{C}$ *has the* wrapping input property *if there exist a full column rank encoder* G *and a positive integer* δ *such that* $\mathbf{w} = G\mathbf{u}$ *implies*

$$\mathrm{supp}(\mathbf{u}) \subseteq \big(\mathrm{supp}(\mathbf{w})\big)^{\delta}. \tag{14}$$

Interestingly enough, the zero primeness of G is not only sufficient but also necessary for property (WI). So, free codes satisfying property (WI) can be identified with codes that are generated by ℓZP matrices.

Proposition 4.2 *A convolutional code* $\mathcal{C}$ *has the (WI) property if and only if it admits a right zero-prime encoder.*

Proof The "if" part has already been proved. To show the converse, we need the following characterization of right zero prime matrices.

Lemma 4.3 [2] *Let* $G \in \mathbb{F}[\mathbf{z}, \mathbf{z}^{-1}]^{p \times q}$ *be a Laurent polynomial matrix and denote by* $\mathbb{F}[[\mathbf{z}, \mathbf{z}^{-1}]]^q$ *the space of bilateral scalar formal power series in the indeterminates* $z_1, \ldots, z_n$. *Then* G *is right zero prime if and only if*

$$G\mathbf{s} = \mathbf{0} \tag{15}$$

for some sequence $\mathbf{s} \in \mathbb{F}[[\mathbf{z}, \mathbf{z}^{-1}]]^q$ *implies* $\mathbf{s} = \mathbf{0}$. ■

Suppose, now, that $\mathcal{C}$ has the (WI) property w.r.t. some positive integer δ and some full column rank encoder G. We aim to prove that G is rZP. If not, there would be a sequence $\mathbf{s} \in \mathbb{F}[[\mathbf{z}, \mathbf{z}^{-1}]]^q$ satisfying (15). Let η be the radius of a ball, $B(\mathbf{0}, \eta)$, centered in the origin and including $\mathrm{supp}(G)$. If $\mathbf{k}$ is an element of $\mathrm{supp}(\mathbf{s})$, the finite support sequence

$$\mathbf{u}(\mathbf{h}) := \begin{cases} \mathbf{s}(\mathbf{h}) & \mathbf{h} \in B(\mathbf{k}, 2\delta + \eta) \\ \mathbf{0} & \text{elsewhere} \end{cases}$$

generates a codeword $\mathbf{w} := G\mathbf{u}$ that does not fulfill (14). ■

The (WI) property introduces very severe constraints on the supports of the input sequences which produce the code trajectories. So, it is not unexpected that it reflects into the strongest primeness property an encoder can be endowed with, namely zero-primeness. Obviously, weaker requirements on the supports of the generating sequences correspond to weaker primeness properties of G. In particular, minor primeness guarantees that the signal producing a codeword $\mathbf{w}$ exhibits a support which slightly exceeds a parallelepipedal box including $\mathrm{supp}(\mathbf{w})$, whereas variety primeness ensures that each projection of $\mathbf{u}$ and $\mathbf{w}$ onto a coordinate hyperplane gives a pair of signals with the (WI) property. The interested reader is referred to [2].

References

[1] E. Fornasini and M.E. Valcher. Algebraic aspects of 2D convolutional codes. *IEEE Trans.Inf.Th.*, IT 33:1210–1225, 1994.

[2] E. Fornasini and M.E. Valcher. Multidimensional systems with finite support behaviors: signal structure, generation and detection. *to appear in SIAM J. of Control & Opt.*, 1996.

[3] G.D. Forney. Convolutional codes I: algebraic structure. *IEEE Trans. Inf.Th.*, 6:720–738, 1970.

[4] G.D. Forney. Minimal bases of rational vector spaces, with applications to multivariable linear systems. *SIAM J. of Control.*, 13:493–521, 1975.

[5] G.D. Forney and M.D. Trott. The dynamics of group codes: state spaces, trellis diagrams and canonical encoders. *IEEE Trans.Inf. Th.*, IT-39:1491–1513, 1993.

[6] S. Lang. *Algebra*. Addison-Wesley Publ.Co., 1967.

[7] H.A. Loeliger, G.D. Forney, T. Mittelholzer, and M.D. Trott. Minimality and observability of group systems. *Lin. Alg. and Appl.*, 205-206:937–963, 1994.

[8] P. Rocha. *Structure and Representation of 2-D Systems*. PhD thesis, University of Groningen, The Netherlands, 1990.

[9] P. Rocha and J.C. Willems. Canonical computational forms for AR 2-D systems. *Multidimensional Systems and Signal Processing*, 2:251–278, 1990.

[10] J.C. Willems. Models for dynamics. *Dynamics reported*, 2:171–269, 1988.