Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Linear Algebra and its Applications 392 (2004) 119–158
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

Matrix fraction descriptions in convolutional coding

Ettore Fornasini [a,*], Raquel Pinto [b,1]

[a] *Dipartimento di Ingegneria dell'Informazione, Università di Padova, 35131 Padova, Italy*
[b] *Departamento de Matemática, Universidade de Aveiro, Aveiro, Portugal*

Received 29 September 2003; accepted 7 June 2004

Submitted by J. Rosenthal

Abstract

In this paper, polynomial matrix fraction descriptions (MFDs) are used as a tool for investigating the structure of a (linear) convolutional code and the family of its encoders and syndrome formers. As static feedback and precompensation allow to obtain all minimal encoders (in particular, polynomial encoders and decoupled encoders) of a given code, a simple parametrization of their MFDs is provided. All minimal syndrome formers, by a duality argument, are obtained by resorting to output injection and postcompensation. Decoupled encoders are finally discussed as well as the possibility of representing a convolutional code as a direct sum of smaller ones.

Keywords: Matrix fraction descriptions; Convolutional codes; Minimal encoders; Syndrome formers; Feedback group

1. Introduction

Polynomial matrices provide a powerful format for analyzing and compensating the dynamical behavior of a linear system, and constitute an indispensable tool for

* Corresponding author.
E-mail addresses: fornasini@dei.unipd.it (E. Fornasini), raquel@mat.ua.pt (R. Pinto).
[1] Supported by Fundacão para a Ciência e a Tecnologia.

0024-3795/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2004.06.007

the analysis of convolutional codes. Underlying the widespread application of polynomial matrices is often a motivation to achieve a complete parametrization of the solutions to a given problem, and to obtain an information on their properties basing only on some elementary matrix manipulations. In particular, as the codewords of a convolutional code are the solutions of a system of difference equations, a convolutional code can be seen as the image of a polynomial matrix, and polynomial matrix transformations allow to enlighten its structure.

Matrix fraction descriptions (MFDs) of rational matrices provide a very convenient way of representing rational matrices—and hence multi-input/multi-output linear transformations—as the "ratio" of two polynomial matrices. By far the most useful concept is that of an irreducible MFD, which allows to achieve a close analogy with an irreducible (scalar) ratio of two polynomials. In fact, the dynamical complexity of a multivariable system cannot be easily evaluated from a rational matrix, even when its entries are irreducible, while it is when numerator and denominator matrices of an irreducible MFD are considered. However, the set of "trivial" common factors that do not affect irreducibility of a MFD is rather large, as it includes all unimodular matrices. The elimination of an unimodular common factor possibly induces conspicuous structural modifications on the numerator and denominator matrices, and in particular on the degrees of their entries, a phenomenon that obviously cannot take place in the scalar case. In this respect, it is often useful to arrive at irreducible MFDs with denominator and/or numerator matrices having minimal row (or column) degrees.

The relevance of polynomial matrices and MFDs in convolutional coding is twofold. First, they allow to analyze the set of all codewords (the "code") per se, without explicit reference to an encoding map. This point of view is somehow typical of coding theorists [1–7], but is currently undertaken also in Willems's behavioral approach to dynamical systems [8,9]. On the other hand, a convolutional code can be viewed as the output space of a linear sequential circuit (the "encoder") over some finite field $\mathbb{F}$, and the whole family of its encoders can be investigated by resorting to the MFDs of their input/output maps, and to their state space realizations.

In this paper we shall mostly concentrate on the second aspect. Accordingly, an important concern will consist in showing that various concepts in convolutional coding theory have a neat description when encoders and syndrome formers of a code $\mathscr{C}$ are represented by MFDs, and some classical results [2,5] can be given new and perhaps simpler proofs. In this respect the first sections of the paper partly exhibit a tutorial character, and are devoted to reviewing several concepts from the algebra of polynomial and rational matrices. The results of the second part are based on an efficient characterization of those MFDs that represent minimal encoders of $\mathscr{C}$, and include some connections among minimal, canonical and basic encoders. A simple parametrization of minimal polynomial encoders is provided, as well as a feedback realization procedure for all minimal rational encoders. Moreover, duality methods allow to extend the results, without further effort, to the structure of syndrome formers of $\mathscr{C}$. Finally, we tackle the problem of analyzing and realizing decoupled encoders. To that purpose, we give first an algorithm, providing a maximally decoupled encoder,

i.e. an encoder associated with the finest decomposition of the code. Next, we obtain a canonical decoupled encoder, and parametrize, via MFDs, all minimal decoupled encoders realizing the finest decomposition of the code.

2. Matrix fraction descriptions of rational matrices

In the following we shall adopt the usual notations $\mathbb{F}[d]$ and $\mathbb{F}(d)$ for denoting the ring of polynomial and the field of rational functions with coefficients in an arbitrary field $\mathbb{F}$. Sometimes, we shall also consider the ring $\mathbb{F}[d, d^{-1}]$ of *Laurent* polynomials, i.e., polynomials in which both positive and negative powers are allowed. If

$$p(d, d^{-1}) = \sum_{m \leqslant i \leqslant M} p_i d^i, \quad p_m p_M \neq 0,$$

is an element of this ring, $m = \mathrm{ord}(p)$ and $M = \deg(p)$ are called the *order* and the *degree* of $p(d, d^{-1})$, respectively, and $\mathbb{F}[d, d^{-1}]$ is an Euclidean domain w.r.t. the difference $M - m$. The order (resp. the degree) of a vector of Laurent polynomials is the minimum order (resp. the maximum degree) of its entries.

Some definitions and results on polynomial and rational matrices are summarized below, for future reference: more details can be found in many textbooks (see, for instance [10–14]). As subsequent developments do not require higher generality, the polynomial matrices we consider are full (row or column) rank, unless otherwise specified; moreover, all statements on "left" factors, "left" matrix fraction descriptions, etc can be couched in "right" terms upon taking transposes. Furthermore, according to a well-established use in coding theory, all vectors are row vectors and, consequently, matrices representing linear transformations are applied on the right side.

Unimodular matrices, i.e., square polynomial matrices with polynomial inverse, when applied on the left and on the right of a matrix $Q(d) \in \mathbb{F}[d]^{m \times p}$, induce row and column operations on it. Two matrices $P(d)$ and $Q(d)$ of $\mathbb{F}[d]^{m \times p}$ are $\mathbb{F}[d]$-*equivalent* if there exist unimodular matrices $U(d)$ and $V(d)$ such that

$$Q(d) = U(d)P(d)V(d).$$

Let $P(d)$ be a full row rank polynomial matrix in $\mathbb{F}[d]^{m \times p}$

(i) $P(d)$ is $\mathbb{F}[d]$-equivalent to its *Smith form*

$$S(d) = \left[\begin{array}{cccc|c} \gamma_1(d) & & & 0 & \\ & \gamma_2(d) & & & 0 \\ & & \ddots & & \\ 0 & & & \gamma_m(d) & \end{array}\right],$$

where $\gamma_1(d), \gamma_2(d), \ldots, \gamma_m(d)$ are monic polynomials satisfying $\gamma_{i+1}(d) | \gamma_i(d)$, $i = 1, \ldots, m-1$. They are uniquely determined by $P(d)$ and are called the *invariant polynomials* of $P(d)$.

(ii) There exists an unimodular matrix $U(d) \in \mathbb{F}[d]^{p\times p}$ such that

$$H(d) = P(d)U(d) = \left[\begin{array}{cccc|c} h_{11}(d) & h_{12}(d) & & h_{1m}(d) & \\ & h_{22}(d) & & h_{2m}(d) & \\ & & \ddots & & 0 \\ 0 & & & h_{mm}(d) & \end{array}\right]$$

where $h_{ii}(d), i = 1, \ldots, m$, are monic polynomials satisfying $\deg h_{ii} > \deg h_{ji}$, $j < i$. $H(d)$ is called (column) *Hermite form* of $P(d)$.

$P(d) \in \mathbb{F}[d]^{m\times p}$ is *left prime* if in all factorizations

$$P(d) = \Delta(d)\bar{P}(d), \quad \Delta(d) \in \mathbb{F}[d]^{m\times m}, \tag{1}$$

the left factor $\Delta(d)$ is unimodular. Left primeness is equivalent to any one of the following statements:

(i) (SMITH FORM) the Smith form of $P(d)$ is $[I_m \mid 0]$;
(ii) (HERMITE FORM) the (column) Hermite form of $P(d)$ is $[I_m \mid 0]$;
(iii) (ROW COMPLETION) there exists $C(d) \in \mathbb{F}[d]^{(p-m)\times p}$ such that $\left[\begin{smallmatrix} P(d) \\ C(d) \end{smallmatrix}\right]$ is unimodular;
(iv) (RIGHT INVERSE) there exists a polynomial matrix $M(d) \in \mathbb{F}[d]^{p\times m}$ such that $P(d)M(d) = I_m$;
(v) (MAXIMUM ORDER MINORS) the greatest common divisor (GCD) of the mth order minors of $P(d)$ is 1;
(vi) (POLYNOMIAL OUTPUT/ POLYNOMIAL INPUT) $\forall \hat{\mathbf{r}}(d) \in \mathbb{F}(d)^{1\times m}$, $\hat{\mathbf{r}}(d)P(d) \in \mathbb{F}[d]^{1\times p}$ implies $\hat{\mathbf{r}}(d) \in \mathbb{F}[d]^{1\times m}$;
(vii) (RANK CONDITION) $P(\alpha)$ has rank m, for all $\alpha \in \bar{\mathbb{F}}$, $\bar{\mathbb{F}}$ the algebraic closure of $\mathbb{F}$.

$\Delta(d) \in \mathbb{F}[d]^{m\times m}$ is a *left maximal divisor* (lMD) of $P(d)$ if (1) holds and $P(d) = \tilde{\Delta}(d)\tilde{P}(d)$, $\tilde{\Delta}(d) \in \mathbb{F}[d]^{m\times m}$, implies $\Delta(d) = \tilde{\Delta}(d)F(d)$. The submatrix $H(d) \in \mathbb{F}[d]^{m\times m}$ in the Hermite form $P(d) = [H(d) \mid 0]U(d)$ provides an lMD of $P(d)$; any other lMD of $P(d)$ is given by $H(d)V(d)$, where $V(d)$ sweeps over all $m \times m$ unimodular matrices.

Two polynomial matrices $M_1(d) \in \mathbb{F}[d]^{m\times p_1}$ and $M_2(d) \in \mathbb{F}[d]^{m\times p_2}$ are *left coprime* if all left common factors of $M_1(d)$ and $M_2(d)$ are unimodular, which amounts to say that $[M_1(d) \mid M_2(d)]$ is left prime. Obviously, the *left greatest common divisors* (lGCD's) of $M_1(d)$ and $M_2(d)$ are the lMD's of $[M_1(d) \mid M_2(d)]$.

Suppose that $P(d)$ has full row rank, with row degrees $k_1, k_2, \ldots, k_m$, so that we can write

$$P(d) = \left[\begin{array}{cccc} d^{k_1} & & & \\ & d^{k_2} & & \\ & & \ddots & \\ & & & d^{k_m} \end{array}\right] P_{\mathrm{hr}} + P_{\mathrm{rem}}(d). \tag{2}$$

The *leading* (*or high order*) *row coefficient matrix* $P_{\rm hr} \in \mathbb{F}^{m\times p}$ is a matrix whose ith row comprises the coefficients of d^{k_i} in the ith row of $P(d)$, and the "reminder" $P_{\rm rem}(d)$ is a polynomial matrix with row degrees strictly less than those of $P(d)$.

$P(d)$ is *row reduced* if its *external degree* $\sum_{i=1}^{m} k_i$ coincides with the *internal degree*, i.e. with the maximum degree of its mth order minors. This happens if and only if $P_{\rm hr}$ has rank m, or, equivalently, if and only if $P(d)$ exhibits the *predictable degree property* [2]

$$\deg(\hat{\mathbf{v}}P) = \max_{i:\hat{v}_i(d)\neq 0} \{k_i + \deg \hat{v}_i\}, \tag{3}$$

for all nonzero polynomial vectors $\hat{\mathbf{v}}(d) \in \mathbb{F}[d]^m$ (and, obviously, for all nonzero Laurent polynomial vectors $\hat{\mathbf{v}}(d, d^{-1}) \in \mathbb{F}[d, d^{-1}]^m$). Elementary row operations allow to transform a full row rank polynomial matrix $P(d)$ into a row reduced one. If $P_1(d)$ and $P_2(d)$ are row reduced, and $P_1(d) = U(d)P_2(d)$, $U(d)$ unimodular, then—modulo a permutation—the row degrees of $P_1(d)$ and $P_2(d)$ are the same. As a consequence, when transforming $P(d)$ into a row reduced matrix, the final row degrees are uniquely determined, up to a permutation.

If $P(d) \in \mathbb{F}[d]^{m\times m}$ is row reduced, with row degrees $k_1 \geqslant \cdots \geqslant k_m$ and invariant polynomials $\psi_1(d), \ldots, \psi_m(d)$, $\psi_{i+1}|\psi_i$, then we have

$$\begin{aligned} &\deg(\psi_1 \cdots \psi_t) \geqslant k_1 + \cdots + k_t, \quad t = 1, \ldots, m-1, \\ &\deg(\psi_1 \cdots \psi_m) = k_1 + \cdots + k_m. \end{aligned} \tag{4}$$

Vice-versa, a Smith form $\mathrm{diag}\{\psi_1(d), \ldots, \psi_m(d)\}_{m\times m}$ whose row degrees satisfy (4) is equivalent to a row reduced matrix with row degrees $k_1, \ldots, k_m$. This is part of the content of a remarkable theorem due to Rosenbrock [11].

Any rational matrix $G(d) \in \mathbb{F}(d)^{m\times p}$ admits a left

$$G(d) = D_{\rm L}(d)^{-1} N_{\rm L}(d) \tag{5}$$

and a right matrix fraction description (MFD)

$$G(d) = N_{\rm R}(d) D_{\rm R}(d)^{-1}, \tag{6}$$

where $N_{\rm L}$, $N_{\rm R}$, $D_{\rm L}$, $D_{\rm R}$ are polynomial matrices of suitable dimensions.

The MFDs (5) and (6) are *irreducible* if $D_{\rm L}(d)$ and $N_{\rm L}(d)$ are left coprime and $D_{\rm R}(d)$ and $N_{\rm R}(d)$ are right coprime, respectively. Any rational matrix $G(d) \in \mathbb{F}(d)^{m\times p}$ has an irreducible lMFD $\bar{D}_{\rm L}(d)^{-1}\bar{N}_{\rm L}(d)$ [7] and any other lMFD of $G(d)$, $\hat{D}_{\rm L}(d)^{-1}\hat{N}_{\rm L}(d)$ satisfies

$$[\hat{D}_{\rm L}(d) \mid \hat{N}_{\rm L}(d)] = \Delta(d)[\bar{D}_{\rm L}(d) \mid \bar{N}_{\rm L}(d)]$$

for a suitable matrix $\Delta(d)$. In case $\Delta(d)$ is unimodular, $\hat{D}_{\rm L}(d)^{-1}\hat{N}_{\rm L}(d)$ is irreducible too. Some irreducible lMFDs of $G(d)$ have the additional property that

$$[D_{\rm L}(d) \mid N_{\rm L}(d)] \tag{7}$$

is row reduced; in this case, the row degrees of (7) are uniquely determined, up to a permutation. If (5) and (6) are irreducible lMFDs and rMFDs of $G(d) \in \mathbb{F}(d)^{m\times p}$,

respectively, then there exist suitable polynomial matrices $X(d)$, $Y(d)$, $W(d)$ and $Z(d)$ such that the generalized Bézout identity [7]

$$\begin{bmatrix} X(d) & Y(d) \\ -N_{\mathrm{L}}(d) & D_{\mathrm{L}}(d) \end{bmatrix} \begin{bmatrix} D_{\mathrm{R}}(d) & W(d) \\ N_{\mathrm{R}}(d) & Z(d) \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_m \end{bmatrix} \tag{8}$$

holds.

After commuting in (8) the left-hand matrices, it follows that

$$\begin{bmatrix} I_p & 0 \\ N_{\mathrm{L}}(d) & I_m \end{bmatrix} \begin{bmatrix} I_p & W(d) \\ 0 & I_m \end{bmatrix} \begin{bmatrix} D_{\mathrm{R}}(d) & 0 \\ 0 & I_m \end{bmatrix} \begin{bmatrix} X(d) & Y(d) \\ -N_{\mathrm{L}}(d) & D_{\mathrm{L}}(d) \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & D_{\mathrm{L}}(d) \end{bmatrix},$$

which implies that $D_{\mathrm{R}}(d)$ and $D_{\mathrm{L}}(d)$ have the same nonunit invariant polynomials and, up to a nonzero constant factor, the same determinant. Moreover, any two complementary maximal order minors in $[N_{\mathrm{L}}(d) \mid D_{\mathrm{L}}(d)]$ and in $\begin{bmatrix} D_{\mathrm{R}}(d) \\ N_{\mathrm{R}}(d) \end{bmatrix}$ are associate, so that the two matrices have the same internal degree [3].

A rational function $p(d)/q(d) \in \mathbb{F}(d)$ is *causal* if there exists a formal power series $\sum_{t=0}^{\infty} a_t d^t$ such that

$$q(d) \sum_{t=0}^{\infty} a_t d^t = p(d).$$

In case $a_0 = 0$, the rational function is *strictly causal*. A rational matrix $G(d)$ is causal (strictly causal) if all its elements are causal (strictly causal).

Given any Laurent formal power series $\hat{A}(d) = \sum_t A_t d^t \in \mathbb{F}^{m\times p}((d))$ and an integer $T \in \mathbb{Z}$, the truncation operator $\mathscr{P}_T$ is defined via:

$$\mathscr{P}_T : \mathbb{F}^{m\times p}((d)) \to \mathbb{F}^{m\times p}((d)) : \sum_t A_t d^t \mapsto \sum_{t<T} A_t d^t. \tag{9}$$

The following are equivalent:

(i) $G(d)$ is causal;
(ii) in any irreducible lMFD $G(d) = D(d)^{-1}N(d)$ the matrix $D(0)$ is nonsingular;
(iii) for all $\hat{\mathbf{u}}(d) \in \mathbb{F}^{1\times m}((d))$,

$$\mathscr{P}_1(\hat{\mathbf{u}}G) = \mathscr{P}_1\big((\mathscr{P}_0\hat{\mathbf{u}})G\big).$$

3. Basic properties of convolutional codes

Convolutional codes are families of sequences (*codewords*) endowed with particular structural constraints that can be specified in algebraic terms through equivalent

sets of conditions. What set to choose is somehow a matter of taste: in this paper we refer to some natural operations on codewords that underlie the properties of controllability, observability [8,11,15,16], shift-invariance and superposition. Afterwards, it will be proved that these properties confer a convolutional code with a particular structure of vector space.

Let $\mathbb{F}$ be a finite field, and denote by

$$\mathbf{w} : \mathbb{Z} \to \mathbb{F}^p : t \mapsto \mathbf{w}_t \tag{10}$$

any *discrete time trajectory* with values in $\mathbb{F}^p$. Clearly, $\mathbf{w}$ can be represented either as a bilateral sequence indexed in $\mathbb{Z}$ or as a *bilateral formal power series* with vector coefficients, $\hat{\mathbf{w}}(d) := \sum_{t\in\mathbb{Z}} \mathbf{w}_t d^t$. In the sequel we shall use the sequence and the corresponding series interchangeably, depending on the problem we are dealing with. The *support* and the *span* of a trajectory $\mathbf{w}$ (and of the corresponding series $\hat{\mathbf{w}}(d)$) are the sets

$$\begin{aligned}\operatorname{supp}(\mathbf{w}) &= \{\mathbf{t} \in \mathbb{Z} : \mathbf{w_t} \neq \mathbf{0}\}\\ \operatorname{span}(\mathbf{w}) &= [\inf \operatorname{supp}(\mathbf{w}), \sup \operatorname{supp}(\mathbf{w})],\end{aligned}$$

respectively. The restriction $\mathbf{w}|_I$ of a sequence $\mathbf{w}$ to a certain time interval $I \subset \mathbb{Z}$ is the function

$$\mathbf{w}|_I : I \to \mathbb{F}^p : t \mapsto \mathbf{w}_t. \tag{11}$$

The "universe" of all trajectories $(\mathbb{F}^p)^{\mathbb{Z}}$ is endowed with an $\mathbb{F}$-linear structure, which allows for superposition of two trajectories and scalar multiplication of a trajectory by elements of $\mathbb{F}$.

The *one-step forward* (resp *backward*) *shift* of a codeword $\mathbf{w}$, $\sigma\mathbf{w}$ ($\sigma^{-1}\mathbf{w}$):

$$\begin{aligned}\sigma\mathbf{w} &: \mathbb{Z} \to \mathbb{F}^p : t \mapsto \mathbf{w}_{t-1}\\ \sigma^{-1}\mathbf{w} &: \mathbb{Z} \to \mathbb{F}^p : t \mapsto \mathbf{w}_{t+1}\end{aligned}$$

is obtained through the multiplication by d (resp d^{-1}) of the corresponding series $\hat{\mathbf{w}}(d)$:

$$\begin{aligned}\hat{\mathbf{w}}(d) \mapsto d\ \hat{\mathbf{w}}(d) &= \sum \mathbf{w}_{t-1}\, d^t,\\ \hat{\mathbf{w}}(d) \mapsto d^{-1}\ \hat{\mathbf{w}}(d) &= \sum \mathbf{w}_{t+1}\, d^t.\end{aligned}$$

The *concatenation* $\mathbf{w}^{(1)} \bigwedge_\theta \mathbf{w}^{(2)}$ of two trajectories $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$ at time θ is defined as follows:

$$(\mathbf{w}^{(1)} \bigwedge_\theta \mathbf{w}^{(2)})_t := \begin{cases} \mathbf{w}_t^{(1)} & \text{if } t < \theta\\ \mathbf{w}_t^{(2)} & \text{if } t \geqslant \theta.\end{cases}$$

Definition 3.1. Let $\mathscr{B}$ be a subset of $(\mathbb{F}^p)^{\mathbb{Z}}$.

(i) $\mathscr{B}$ is *N-controllable* (for some $N \in \mathbb{N}$) if, given any two trajectories $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$ in $\mathscr{B}$ and an arbitrary time instant θ, there exists a suitable $\mathbf{r} \in \mathscr{B}$ such that

$$\mathbf{w}^{(1)} \bigwedge_{\theta} \mathbf{r} \bigwedge_{\theta+N} \mathbf{w}^{(2)} \in \mathscr{B}.$$

If there is an $N \in \mathbb{N}$ such that $\mathscr{B}$ is N-controllable then $\mathscr{B}$ is said to be *strongly controllable*.

(ii) $\mathscr{B}$ is *L-observable* (for some $L \in \mathbb{N}$) if given any two trajectories $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$ of $\mathscr{B}$, such that $\mathbf{w}^{(1)}|_{[j,j+L)} = \mathbf{w}^{(2)}|_{[j,j+L)}$ for some $j \in \mathbb{Z}$, the concatenation $\mathbf{w}^{(1)} \bigwedge_j \mathbf{w}^{(2)}$ is in $\mathscr{B}$.
$\mathscr{B}$ is *strongly observable* if there is an $L \in \mathbb{N}$ such that $\mathscr{B}$ is L-observable.

A trajectory $\mathbf{w}$ is *left compact* if there exists $h \in \mathbb{Z}$ such that $\mathbf{w}_t = 0$, $\forall t < h$. A left compact trajectory corresponds to a *Laurent power series* (with vector coefficients) $\hat{\mathbf{w}}(d) = \sum_{t \geqslant h} \mathbf{w}_t d^t$, and we are allowed to multiply any left compact support trajectory $\hat{\mathbf{w}}(d)$ by an arbitrary scalar Laurent power series $s(d) = \sum_\tau s_\tau d^\tau$:

$$\hat{\mathbf{w}}(d) = \sum \mathbf{w}_t d^t \mapsto s(d)\hat{\mathbf{w}}(d) = \sum_t \left(\sum_i s_{t-i} \mathbf{w}_i \right) d^t.$$

Given a nonzero Laurent power series $\hat{\mathbf{w}}(d) = \sum_{t=h}^{+\infty} \mathbf{w}_t d^t$, $\mathbf{w}_h \neq 0$, we call h the *order* of $\tilde{\mathbf{w}}(d)$.

The set $\mathbb{F}((d))$ of scalar Laurent power series with coefficients in $\mathbb{F}$ is a field and the universe of all left compact trajectories $\mathbb{F}((d))^p$ has the structure of a p-dimensional vector space over $\mathbb{F}((d))$. When dealing with a family of trajectories $\mathscr{B}$ which is an $\mathbb{F}((d))$-subspace of $\mathbb{F}((d))^p$, strong controllability and strong observability are equivalent properties, as shown in the following proposition.

Proposition 3.2. *Let $\mathscr{B}$ be an $\mathbb{F}((d))$-subspace of $\mathbb{F}((d))^p$. The following are equivalent*:

(i) *$\mathscr{B}$ is strongly observable.*
(ii) *$\mathscr{B}$ is strongly controllable.*
(iii) *$\mathscr{B}$ admits a polynomial basis.*

Proof. (i) $\Rightarrow$ (ii) Suppose that $\mathscr{B}$ is N-observable, for some $N \in \mathbb{N}$. Denote by $\mathscr{B}^{(i)}$ the $\mathbb{F}$-subspace of $\mathscr{B}$ constituted by all trajectories in $\mathscr{B}$ with support in $[i, +\infty)$. Clearly

$$\cdots \supseteq \mathscr{B}^{(-2)} \supseteq \mathscr{B}^{(-1)} \supseteq \mathscr{B}^{(0)}$$

and consequently the same inclusions hold for the restriction subspaces

$$\cdots \supseteq \mathscr{B}^{(-2)}|_{[0,N)} \supseteq \mathscr{B}^{(-1)}|_{[0,N)} \supseteq \mathscr{B}^{(0)}|_{[0,N)}.$$

We prove first that $\mathscr{B}^{(-r)}|_{[0,N)} = \mathscr{B}^{(-r-1)}|_{[0,N)}$, for some $r \in \mathbb{N}$ implies $\mathscr{B}^{(-r)}|_{[0,N)} = \mathscr{B}^{(-k)}|_{[0,N)} \forall k \geqslant r$. In fact, suppose that $\mathscr{B}^{(-r)}|_{[0,N)} = \mathscr{B}^{(-r-1)}|_{[0,N)}$

and let $s \in \mathscr{B}^{(-r-2)}|_{[0,N)}$, i.e., $s = \mathbf{w}|_{[0,N)}$ for some $\mathbf{w} \in \mathscr{B}^{(-r-2)}$. As $(d\mathbf{w})|_{[0,N)} \in \mathscr{B}^{(-r-1)}|_{[0,N)} = \mathscr{B}^{(-r)}|_{[0,N)}$ we have that $(d\mathbf{w})|_{[0,N)} = \tilde{\mathbf{w}}|_{[0,N)}$ for some $\tilde{\mathbf{w}} \in \mathscr{B}^{(-r)}$. The N-observability of $\mathscr{B}$ implies that $\tilde{\mathbf{w}} \bigwedge_0 d\mathbf{w} \in \mathscr{B}^{(-r)}$, and consequently $s = (d^{-1}(\tilde{\mathbf{w}} \bigwedge_0 d\mathbf{w}))|_{[0,N)} \in \mathscr{B}^{(-r-1)}|_{[0,N)}$. Therefore

$$\mathscr{B}^{(-r)}|_{[0,N)} = \mathscr{B}^{(-r-1)}|_{[0,N)} \Rightarrow \mathscr{B}^{(-r-1)}|_{[0,N)} = \mathscr{B}^{(-r-2)}|_{[0,N)}$$

and $\mathscr{B}^{(-r)}|_{[0,N)} = \mathscr{B}^{(-k)}|_{[0,N)} \forall\, k \geqslant r$.

Second, note that there exists a trajectory $\mathbf{w} \in \mathscr{B}^{(0)}$ that does not belong to $\mathscr{B}^{(1)}$, and

$$\mathbf{w}|_{[0,N)}, (d\mathbf{w})|_{[0,N)}, \ldots, (d^{N-1}\mathbf{w})|_{[0,N)}$$

are linearly independent over $\mathbb{F}$, which implies that $\dim_{\mathbb{F}} \mathscr{B}|_{[0,N)} \geqslant N$. We have therefore shown that $r \leqslant N(p-1)$ and

$$\mathscr{B}|_{[0,N)} = \bigcup_{i=0}^{N(p-1)} \mathscr{B}^{(-i)}|_{[0,N)}. \tag{12}$$

Finally, consider any two trajectories $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$ in $\mathscr{B}$. Given any $k \in \mathbb{Z}$, time-invariance and linearity of $\mathscr{B}$ imply

$$\mathbf{w}^{(1)}|_{[k,k+N)} - \mathbf{w}^{(2)}|_{[k,k+N)} \in \mathscr{B}|_{[0,N)}$$

and, by (12), there exists $\mathbf{w}^{(3)} \in \mathscr{B}$, with support in $[k - N(p-1), +\infty)$ such that

$$\mathbf{w}^{(3)}|_{[k,k+N)} = \mathbf{w}^{(1)}|_{[k,k+N)} - \mathbf{w}^{(2)}|_{[k,k+N)}.$$

Since $\mathbf{w}^{(2)} + \mathbf{w}^{(3)}$ and $\mathbf{w}^{(1)}$ coincide on the interval $[k, k+N)$ and $\mathscr{B}$ is N-observable, the signal given by

$$\mathbf{w}_t = \begin{cases} (\mathbf{w}^{(2)} + \mathbf{w}^{(3)})_t & \text{if } t < k \\ \mathbf{w}^{(1)}_t & \text{if } t \geqslant k \end{cases}$$

is a trajectory of $\mathscr{B}$. Moreover

$$(\mathbf{w}^{(2)} + \mathbf{w}^{(3)})|_{(-\infty,k-N(p-1))} = \mathbf{w}^{(2)}|_{(-\infty,k-N(p-1))}$$

gives

$$\mathbf{w} = \mathbf{w}^{(2)} \bigwedge_{k-N(p-1)} (\mathbf{w}^{(2)} + \mathbf{w}^{(3)}) \bigwedge_{k} \mathbf{w}^{(1)}$$

which proves that $\mathscr{B}$ is $N(p-1)$-controllable.

(ii) $\Rightarrow$ (iii) Suppose $\mathscr{C}$ is N-controllable, and let $G(d) \in \mathbb{F}((d))^{m\times p}$ be a *generator matrix* of $\mathscr{C}$, i.e., a matrix whose rows constitute a basis for $\mathscr{C}$. As premultiplication of $G(d)$ by a nonsingular $M(d) \in \mathbb{F}((d))^{m\times m}$ still gives a generator matrix, we can assume that each row of $G(d)$ includes only nonnegative powers of d and has nonzero constant term.

If $G(0)$ is not full rank, let $\hat{\mathbf{g}}_k(d)$, $k > 1$, be the first row of $G(d)$ with the property that $\hat{\mathbf{g}}_k(0)$ linearly depends on the previous rows of $G(0)$ and consider the space $\mathscr{S}$ of $\mathbb{F}((d))$-linear combinations of the first $k-1$ rows of $G(d)$

$$\hat{\mathbf{c}}(d) = \sum_j \mathbf{c}_j d^j = \hat{\mathbf{a}}(d) \begin{bmatrix} \hat{\mathbf{g}}_1(d) \\ \dots \\ \hat{\mathbf{g}}_{k-1}(d) \end{bmatrix}, \quad \hat{\mathbf{a}}(d) \in \mathbb{F}((d))^{k-1}. \tag{13}$$

Because of the $\mathbb{F}$-linear independence of the first $k-1$ rows of $G(0)$, the order of the series $\hat{\mathbf{c}}(d)$ in (13) coincides with the order of $\hat{\mathbf{a}}(d)$. This implies that two series $\hat{\mathbf{c}}^{(1)}(d)$ and $\hat{\mathbf{c}}^{(2)}(d)$ in $\mathscr{S}$ coincide up to the degree ℓ if and only if the same holds for the corresponding $\hat{\mathbf{a}}^{(1)}(d)$ and $\hat{\mathbf{a}}^{(2)}(d)$.

Clearly $\mathscr{S}$ includes some power series in $\mathbb{F}[[d]]^p$ that fits at least the constant term of $\hat{\mathbf{g}}_k(d)$, and possibly its higher terms up to some finite degree ν. However, the value of ν is uniformly bounded, as $\hat{\mathbf{c}}(d)$ varies in $\mathscr{S}$. Otherwise, we could find an infinite sequence of polynomial vectors $\hat{\mathbf{a}}^{(1)}(d)$, $\hat{\mathbf{a}}^{(2)}(d), \dots$, with $\deg(\hat{\mathbf{a}}^{(i)}) = i$, such that the corresponding $\hat{\mathbf{c}}^{(i)}(d) \in \mathscr{S}$ fit $\hat{\mathbf{g}}_k(d)$ up to the degree i. As $\hat{\mathbf{a}}^{(i)}(d)$ and $\hat{\mathbf{a}}^{(i+1)}(d)$ agree up to the degree i, $i = 1, 2, \dots$, we could define the series $\hat{\mathbf{a}}(d) := \lim_{i\to\infty} \hat{\mathbf{a}}^{(i)}(d) \in \mathbb{F}[[d]]^{k-1}$, which allows to express $\hat{\mathbf{g}}_k(d)$ as a combination of the first $k-1$ rows of $G(d)$, a contradiction.

If $\bar{\nu}$ denotes the maximum value of ν, corresponding to some $k-1$-tuple $\hat{\bar{\mathbf{a}}}(d) = [\hat{\bar{a}}_1(d) \dots \hat{\bar{a}}_{k-1}(d)]$, in the generator matrix

$$G'(d) := \begin{bmatrix} & 1 & & & & & & \\ & & \ddots & & & & & \\ & & & & 1 & & & \\ -[d^{-\bar{\nu}}\hat{\bar{a}}_1(d)] & \dots & -[d^{-\bar{\nu}}\hat{\bar{a}}_{k-1}(d)] & d^{-\bar{\nu}} & & & & \\ & & & & & 1 & & \\ & & & & & & \ddots & \\ & & & & & & & 1 \end{bmatrix} G(d)$$

the first k rows of $G'(0)$ are independent over the field $\mathbb{F}$. Upon iterating the above procedure, if further rows of $G'(0)$ linearly depend on the previous ones, we can ultimately assume that the generator matrix

$$G(d) = \begin{bmatrix} \hat{\mathbf{g}}_1(d) \\ \dots \\ \hat{\mathbf{g}}_m(d) \end{bmatrix}$$

does not include negative powers, and $G(0)$ has rank m.

As $\mathscr{C}$ is N-controllable, there exist sequences $\mathbf{r}_1, \dots, \mathbf{r}_m$ such that

$$\begin{aligned} \mathbf{p}_1 &= \mathbf{g}_1 \bigwedge_1 \mathbf{r}_1 \bigwedge_{N+1} \mathbf{0} \\ &\dots \quad \dots \quad \dots \quad \dots \quad \dots \\ \mathbf{p}_m &= \mathbf{g}_m \bigwedge_1 \mathbf{r}_m \bigwedge_{N+1} \mathbf{0} \end{aligned}$$

are finite support elements of $\mathscr{C}$, and the degrees of the corresponding polynomial vectors $\hat{\mathbf{p}}_1(d), \ldots, \hat{\mathbf{p}}_m(d)$ in $\mathbb{F}[d]^p$ do not exceed N. As

$$P(d) := \begin{bmatrix} \hat{\mathbf{p}}_1(d) \\ \cdots \\ \hat{\mathbf{p}}_m(d) \end{bmatrix}$$

satisfies $P(0) = G(0)$, the polynomial matrix $P(d)$ is full row rank and, hence, a generator matrix of $\mathscr{C}$.

(iii) $\Rightarrow$ (i) The hypothesis implies that there exists an $m \times p$ polynomial generator matrix, $G(d)$, of $\mathscr{B}$, such that

$$\mathscr{B} = \left\{\hat{\mathbf{w}}(d) \in \mathbb{F}((d))^p : \hat{\mathbf{w}}(d) = \hat{\mathbf{u}}(d)G(d),\ \exists \hat{\mathbf{u}}(d) \in \mathbb{F}((d))^m\right\}.$$

Consider two unimodular matrices $U(d)$ and $V(d)$ such that

$$\Gamma(d) = U(d)G(d)V(d),$$

where $\Gamma(d) = [\tilde{\Gamma}(d) \mid 0]$ is the Smith form of $G(d)$. Clearly, the polynomial matrix $\tilde{G}(d) := U(d)G(d)$ is a generator matrix of $\mathscr{B}$, too.

From

$$\hat{\mathbf{w}}(d) = \hat{\mathbf{u}}(d)\tilde{G}(d),$$

it follows that

$$\begin{aligned} \hat{\mathbf{w}}(d)V(d) &= \hat{\mathbf{u}}(d)\tilde{G}(d)V(d) \\ &= \hat{\mathbf{u}}(d)\Gamma(d) \\ &= \hat{\mathbf{u}}(d)[\tilde{\Gamma}(d) \mid 0]. \end{aligned}$$

Upon partitioning $V(d)$ into $[V^{(1)}(d) \mid V^{(2)}(d)]$, where $V^{(1)}(d) \in \mathbb{F}[d]^{p\times m}$ and $V^{(2)}(d) \in \mathbb{F}[d]^{p\times(p-m)}$, we have that

$$\hat{\mathbf{w}}(d) \in \mathscr{B} \Leftrightarrow \hat{\mathbf{w}}(d)V^{(2)}(d) = 0.$$

The polynomial matrix $V^{(2)}(d)$ can be expressed as

$$V^{(2)}(d) = V_0 + V_1 d + \cdots + V_N d^N,$$

$V_i \in \mathbb{F}^{p\times(p-m)}$ and $N \in \mathbb{N}$, and therefore we have

$$\hat{\mathbf{w}}(d) \in \mathscr{B} \Leftrightarrow \sum_{i=0}^{N} \mathbf{w}_{t-i} V_i = 0 \quad \forall t.$$

If $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$ are any two trajectories of $\mathscr{B}$ such that

$$\mathbf{w}^{(1)}|_{[k,k+N]} = \mathbf{w}^{(2)}|_{[k,k+N]}$$

for some $k \in \mathbb{Z}$, the trajectory $\mathbf{w}^{(1)} \bigwedge_k \mathbf{w}^{(2)} \in \mathbb{F}((d))^p$ satisfies

$$(\mathbf{w}^{(1)} \bigwedge_k \mathbf{w}^{(2)})_t := \begin{cases} \mathbf{w}_t^{(1)} & \text{if } t < k, \\ \mathbf{w}_t^{(1)} = \mathbf{w}_t^{(2)} & \text{if } k \leqslant t \leqslant k+N, \\ \mathbf{w}_t^{(2)} & \text{if } t > k+N, \end{cases}$$

and consequently,

$$\sum_{i=0}^{N}(\mathbf{w}^{(1)} \bigwedge_{k} \mathbf{w}^{(2)})_{t-i} V_i = 0 \quad \forall t.$$

This implies $\mathbf{w}^{(1)} \bigwedge_k \mathbf{w}^{(2)} \in \mathscr{B}$, i.e. $\mathscr{B}$ is $(N+1)$-observable. □

Remark. The equivalence between strong observability and strong controllability stated in Proposition 3.2 does not hold anymore in Willems' behavior theory [8,9], where "bilateral" signals (i.e., signals whose support can be any subset of $\mathbb{Z}$) are considered. If we restrict to Willems "complete" behaviors, i.e., to families of bilateral trajectories that can be described as kernels of polynomial matrices, controllable behaviors are kernels of right prime matrices (or, equivalently, images of polynomial matrices) while all complete behaviors are observable. So, for complete bilateral behaviors, controllability always implies observability, but the converse does not hold.

Corollary 3.3. *If $\mathscr{C} \subseteq \mathbb{F}((d))^p$ is an $\mathbb{F}((d))$-subspace, N-controllable but not $(N-1)$-controllable, then $\mathscr{C}$ admits a polynomial basis of degree N, but it does not admit any one of degree $N-1$.*

Proof. From the proof of Proposition 3.2, it follows that the N-controllability of $\mathscr{C}$ implies that $\mathscr{C}$ admits a polynomial basis of degree N. To see that it doesn't admit a polynomial basis of degree $N-1$, suppose that $P(d) \in \mathbb{F}[d]^{m\times p}$ is a polynomial generator matrix for $\mathscr{C}$, with row degrees not greater than $N-1$, and consider two arbitrary elements of $\mathscr{C}$, say $\mathbf{w}^{(1)}, \mathbf{w}^{(2)}$. Then $\hat{\mathbf{w}}^{(1)}(d) = \hat{\mathbf{u}}^{(1)}(d)P(d)$ and $\hat{\mathbf{w}}^{(2)}(d) = \hat{\mathbf{u}}^{(2)}(d)P(d)$, for suitable $\hat{\mathbf{u}}^{(1)}(d)$ and $\hat{\mathbf{u}}^{(2)}(d)$ in $\mathbb{F}((d))^m$. Defining $\mathbf{u} := \mathbf{u}^{(1)} \bigwedge_{\theta} \mathbf{u}^{(2)}$, it follows that $\hat{\mathbf{w}}(d) := \hat{\mathbf{u}}(d)P(d)$ is in $\mathscr{C}$ and, for all $\theta \in \mathbb{Z}$, $\mathbf{w}$ satisfies $\mathbf{w} = \mathbf{w}^{(1)} \bigwedge_{\theta} \mathbf{r} \bigwedge_{N-1+\theta} \mathbf{w}^{(2)}$ for a suitable $\mathbf{r}$, i.e., $\mathscr{C}$ is $(N-1)$ -controllable. □

Definition 3.4. A $[p, m]$-convolutional code $\mathscr{C}$ is a strongly controllable (or, equivalently, a strongly observable) m-dimensional $\mathbb{F}((d))$-subspace of $\mathbb{F}((d))^p$.

Remark. Different definitions of a convolutional code have been considered in the literature. In most cases they are equivalent each other; sometimes, however, new approaches provide interesting generalizations. An useful survey can be found in [16]. Convolutional codes were first introduced as images of polynomial or rational matrices [17]. In the late 1960s, Massey and Sain [18] described a convolutional code as the output space of a linear, time-invariant system, thus establishing the first connection between systems theory and convolutional coding. This approach was largely reinforced by Forney [1–4], and it was used thereafter in most of the coding literature. The behavioral approach to linear systems, introduced in the late 1980's,

seems to represent a very natural setting for investigating a convolutional code, as a code is simply a set of trajectories (codewords). Loeliger and Mittelholzer [15] were probably the first who adopted this point of view, when they defined a convolutional code over a group as a time-invariant, strongly controllable and strongly observable group code. In Proposition 3.2, we show that when considering convolutional codes constituted by left compact sequences over a finite field, strong observability and controllability properties are equivalent to the existence of a polynomial basis, i.e., to the existence of a polynomial generator matrix, which was the primitive definition of convolutional code.

Some basic properties a convolutional code is endowed with are an immediate consequence of the above definition. First of all, being closed under scalar multiplication by elements of $\mathbb{F}((d))$, $\mathscr{C}$ is closed under forward and backward shifts, and is an $\mathbb{F}[d]$ and an $\mathbb{F}[d^{-1}]$-module as well. Moreover, as shown in Proposition 3.2 above, $\mathscr{C}$ admits a polynomial basis, and consequently all codewords can be viewed as outputs of some moving average (i.e. "feedback free") linear model.

Definition 3.5. Any $m \times p$ rational (in particular, polynomial) matrix $G(d)$ whose rows provide an $\mathbb{F}((d))$-basis for a $[p, m]$-convolutional code $\mathscr{C}$ is called an *encoder* of $\mathscr{C}$, and $\mathscr{C}$ is the *image of* $G(d)$, in the sense that

$$\mathscr{C} = \left\{\hat{\mathbf{w}}(d) : \hat{\mathbf{w}}(d) = \hat{\mathbf{u}}(d)G(d), \hat{\mathbf{u}}(d) \in \mathbb{F}((d))^m\right\}.$$

4. Encoder structure

Let $\mathscr{C}$ denote a $[p, m]$-convolutional code and $G(d)$ any encoder of $\mathscr{C}$. Then

$$\tilde{G}(d) = T(d)G(d) \tag{14}$$

parametrizes all the (rational) encoders of $\mathscr{C}$, as $T(d)$ ranges over the linear group $GL(m, \mathbb{F}(d))$ of nonsingular rational $m \times m$ matrices. Two $m \times p$ rational matrices are *equivalent encoders* if the codes they generate are the same. As a consequence of (14), two encoders are equivalent if and only if they differ each other by a rational nonsingular left factor, which amounts to say that the sets of their rows provide two rational basis for the same rational subspace $V_{\mathscr{C}}$ of $\mathbb{F}(d)^p$.

We first restrict our attention to polynomial encoders of a given code $\mathscr{C}$. Basing on the results of Section 2 it is easy to prove that $\mathscr{C}$ admits

- *basic encoders*, i.e. encoders with $G(d)$ left prime. They are related to each other via (14), where $T(d)$ describes the group of $m \times m$ polynomial unimodular matrices;
- *row reduced encoders*;
- *canonical encoders* [6,19], i.e. encoders with $G(d)$ left prime and row reduced. In Forney's terminology, the rows of a canonical encoder constitute a *minimal basis* for the rational space $V_{\mathscr{C}}$. Up to a permutation, the row degrees ϕ_i do not depend

on the particular canonical encoder. They are called [6] the *Forney indices of* $\mathscr{C}$, and $\sum_i \phi_i$ is, by definition, the *degree of the code*.

The above polynomial encoders realize some peculiar connections between the spans of the information sequences and the corresponding codewords.

A polynomial encoder $G(d)$ is basic if and only if for any information signal $\hat{\mathbf{u}}(d)$, the support of $\hat{\mathbf{u}}(d)$ and $\hat{\mathbf{u}}(d)G(d)$ have the same minimum point, and, moreover, there exists a positive integer δ, such that, for all $\hat{\mathbf{u}}(d) \in \mathbb{F}((d))^m$

$$\sup \ \mathrm{span}(\hat{\mathbf{u}}) \leqslant \sup \mathrm{span}(\hat{\mathbf{u}}G) + \delta. \tag{15}$$

In fact, if $G(d)$ is basic, it has a right polynomial inverse $Q(d) = [q_{ij}(d)]$, and $\hat{\mathbf{u}}(d) = [\hat{\mathbf{u}}(d)G(d)]Q(d)$ implies (15), with $\delta = \max_{i,j:q_{ij}(d)\neq 0}\{\deg q_{ij}\}$. Moreover, since $G(0)$ has full row rank, the minimum points of the support of $\hat{\mathbf{u}}(d)$ and $\hat{\mathbf{u}}(d)G(d)$ coincide.

Vice-versa, if $G(d)$ fails to be basic, we consider its Smith form

$$G(d) = V(d)\left[\begin{array}{ccc|c} \psi_1(d) & & & \\ & \ddots & & 0 \\ & & \psi_m(d) & \end{array}\right] W(d),$$

where $V(d)$ and $W(d)$ are unimodular matrices and $\deg \psi_1 > 0$. If $\psi_1(d) = d^k,\ k > 0$, the minimum point of the support of $[\,1 \quad \cdots \quad 0\,]\,V(d)^{-1}$ is 0, but the corresponding codeword starts at $t = k$. If $\psi_1(0) \neq 0$, the information signal $\hat{\mathbf{u}}(d) = [\,\frac{1}{\psi_1(d)} \quad 0 \quad \cdots \quad 0\,]\,V^{-1}(d)$ has infinite support while the corresponding codeword has not.

On the other hand, when $G(d)$ is row reduced, with row degrees $k_1, k_2, \ldots, k_m$, a precise estimate of the maximum point of the support of $\hat{\mathbf{u}}(d)G(d)$ can be obtained via the *predictable degree property*. As we have seen in Section 2, when multiplying a polynomial vector into a row reduced polynomial matrix, (3) allows to "predict" the degree of the product independently of the particular values of the coefficients of the polynomial vector. So we have

$$\deg(\hat{\mathbf{u}}G) = \max_{i:u_i(d)\neq 0}\{k_i + \deg u_i\}, \tag{16}$$

and a finite support information signal $\hat{\mathbf{u}}(d, d^{-1}) = [\,\hat{u}_1(d, d^{-1}) \quad \cdots \quad \hat{u}_m(d, d^{-1})\,] \in \mathbb{F}[d, d^{-1}]^m$ produces a codeword $\hat{\mathbf{u}}(d, d^{-1})G(d)$ with support in $(-\infty, 0]$ if and only if $\deg \hat{u}_i \leqslant -k_i$, $i = 1, \ldots, m$.

When dealing with rational encoders, it is quite useful to consider their (left) matrix fraction descriptions

$$G(d) = D(d)^{-1}N(d), \tag{17}$$

where $D(d) \in \mathbb{F}[d]^{m\times m}$ and $N(d) \in \mathbb{F}[d]^{m\times p}$. It is worth noticing that

- the numerator matrix $N(d)$ is a polynomial encoder of $\mathscr{C}$: just put $T(d) = D(d)$ in (14);

- if $D(d)^{-1}N(d)$ is an irreducible lMFD, $G(d)$ is a *causal encoder* if and only if $D(0)$ is nonsingular;
- given a basic encoder $G_b(d) \in \mathbb{F}[d]^{m\times p}$, all equivalent encoders of $\mathscr{C}$ have MFDs

$$\bar{G}(d) = [\bar{D}(d)]^{-1}[\Delta(d)G_b(d)] \tag{18}$$

where $\Delta(d)$ and $\bar{D}(d)$ are nonsingular $m \times m$ polynomial matrices. In particular (18) is irreducible if and only if $\bar{D}(d)^{-1}\Delta(d)$ is irreducible too. So, all causal encoders of $\mathscr{C}$ are represented by (18), when $\bar{D}(d)^{-1}\Delta(d)$ is irreducible and $\bar{D}(0)$ is invertible.

Remark. Matrix fraction descriptions of the encoders are strongly connected to controllability system matrices considered by Forney in [4]. Every input/output pair $[\hat{\mathbf{w}}(d) \mid \hat{\mathbf{u}}(d)] \in \mathbb{F}((d))^{p+m}$ satisfies

$$\begin{aligned}[\hat{\mathbf{w}}(d) \mid \hat{\mathbf{u}}(d)] &= \hat{\mathbf{u}}(d)[G(d) \mid I_m] \\ &= \hat{\mathbf{u}}(d)D(d)^{-1}[N(d) \mid D(d)] \\ &= \hat{\mathbf{v}}(d)[N(d) \mid D(d)]\end{aligned}$$

and vice-versa, given $\hat{\mathbf{v}}(d) \in \mathbb{F}((d))^m$, $\hat{\mathbf{v}}(d)[N(d) \mid D(d)]$ is an input/output pair. In case $[N(d) \mid D(d)]$ is left prime, $[\hat{\mathbf{w}}(d) \mid \hat{\mathbf{u}}(d)]$ is polynomial if and only if $\hat{\mathbf{v}}(d)$ is polynomial, and the rows $[\hat{\mathbf{n}}_i(d) \mid \hat{\mathbf{d}}_i(d)]$, $i = 1, \ldots, m$, of $[N(d) \mid D(d)]$ provide a basis for the $\mathbb{F}[d]$-module of all polynomial input/output pairs.

An encoder $G(d)$ of a $[p, m]$-convolutional code $\mathscr{C}$ is *noncatastrophic* if it maps every infinite support information series $\hat{\mathbf{u}}(d) \in \mathbb{F}((d))^m$ into an infinite support codeword $\hat{\mathbf{w}}(d) = \hat{\mathbf{u}}(d)G(d)$.

Proposition 4.1 [2]. *Given a causal encoder $G(d)$ of $\mathscr{C}$, the following are equivalent*:

(i) *$G(d)$ is noncatastrophic*;
(ii) *in any irreducible left MFD $G(d) = D(d)^{-1}N(d)$ the numerator matrix $N(d)$ factorizes into $N(d) = \Delta(d)\bar{N}(d)$, where $\bar{N}(d)$ is a basic encoder and* $\det \Delta(d) = \alpha d^k$, $0 \neq \alpha \in \mathbb{F}$ *and* $k \in \mathbb{N}$.
(iii) *$G(d)$ admits a right inverse $A(d)B(d)^{-1} \in \mathbb{F}(d)^{p\times m}$, with* $\det B(d) = \beta d^h$, $0 \neq \beta \in \mathbb{F}$ *and* $h \in \mathbb{N}$, *or, equivalently, there exists a polynomial matrix $M(d) \in \mathbb{F}[d]^{p\times m}$ such that $G(d)M(d) = d^s I_m$, $s \in \mathbb{N}$.*

Proof. The proof of the equivalence (i) $\Leftrightarrow$ (ii) is similar to, but somehow easier than that of Proposition 9.2, and will be omitted for sake of brevity.

For the implication (ii) $\Rightarrow$ (iii), consider a polynomial right inverse $\bar{L}(d)$ of $\bar{N}(d)$, so that $G(d)\bar{L}(d)\Delta(d)^{-1}D(d) = I_m$. If $\bar{a}(d)B(d)^{-1}$ denotes any right MFD of $\Delta(d)^{-1}D(d)$, just assume $A(d) := \bar{L}(d)\bar{A}(d)$.

On the other hand, suppose that (iii) holds. Taking into account that $D(d)^{-1}N(d)$ is irreducible, from $D(d)^{-1}\Delta(d)\bar{N}(d)A(d)B(d)^{-1} = I_m$ we get an irreducible left

MFD $\Delta(d)^{-1}D(d)$ of $\bar{N}(d)A(d)B(d)^{-1}$. Consequently, det Δ divides det $B = \beta d^h$. This proves implication (ii) $\Leftarrow$ (iii). □

As a consequence of the above proposition, a noncatastrophic encoder $G(d)$ has the characteristic property that the span of each information sequence does not exceed "too much" that of the corresponding codeword. In fact, part (iii) is equivalent to the existence of a Laurent polynomial inverse $L(d, d^{-1}) = \sum_{m \leqslant i \leqslant M} P_i d^i$, $P_m \neq 0$, $P_M \neq 0$ of $G(d)$ and

$$\text{span } (\hat{\mathbf{u}}) \subseteq [\text{infspan } (\hat{\mathbf{u}}G) + m,\ \text{sup span } (\hat{\mathbf{u}}) + M].$$

Systematic encoders are rational matrices that reduce to the following structure:

$$G(d) = [I_m \mid G_2(d)]$$

up to a column permutation. Clearly they constitute a special class of noncatastrophic encoders. Every convolutional code $\mathscr{C}$ admits systematic causal encoders: just take a basic encoder $G_b(d)$ of $\mathscr{C}$, select any $m \times m$ submatrix $D(d)$ of $G_b(d)$ with nonsingular $D(0)$, and consider the equivalent encoder $G(d) = D(d)^{-1}G_b(d)$. In general, however, they fail to be polynomial. A necessary and sufficient condition for the existence of a polynomial systematic encoder is that some (and therefore every) basic encoder of $\mathscr{C}$ has a nonzero constant minor of order m.

5. State space realization and minimal encoders

State space models for convolutional encoders have been considered since many years [18], and provide a neat framework for classifying encoders complexity by resorting to the dimension of their minimal realizations.

A linear, discrete time, dynamical system $\Sigma = (A, B, C, J)$ [4,10,11]

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{x}_t A + \mathbf{u}_t B \\ \mathbf{w}_t &= \mathbf{x}_t C + \mathbf{u}_t J \end{aligned} \tag{19}$$

$A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{m\times n}$, $C \in \mathbb{F}^{n\times p}$, $J \in \mathbb{F}^{m\times p}$ is an *n-dimensional realization* of a $[p, m]$ causal encoder $G(d)$ of $\mathscr{C}$ if, starting from zero initial conditions, Σ encodes every information series $\hat{\mathbf{u}}(d) = \sum_t \mathbf{u}_t d^t$ into the corresponding codeword produced by $G(d)$, namely

$$\hat{\mathbf{w}}(d) := \sum_t \mathbf{w}_t d^t = \hat{\mathbf{u}}(d)G(d)$$

This happens if and only if

$$G(d) = J + Bd(I - dA)^{-1}C$$

Every causal encoder $G(d)$ can be realized by a linear dynamical system (19). The following procedure is an adaptation of similar algorithms available in the literature [4,20,21].

1. Rewrite $G(d)$ as

$$G_{sp}(d) + J \tag{20}$$

$G_{sp}(d)$ strictly causal, and consider a left MFD

$$G_{sp}(d) = D_{\mathrm{L}}(d)^{-1} N_{\mathrm{L}}(d)$$

such that $D_{\mathrm{L}}(0)$ is non singular and

$$[D_{\mathrm{L}}(d) \mid N_{\mathrm{L}}(d)] \tag{21}$$

is row reduced, with row degrees $k_1, \ldots, k_m$. Suppose for the moment that all row degrees are strictly positive and let $n := \sum_{i=1}^{m} k_i$.

2. Denote by M_i the $k_i \times k_i$ nilpotent Jordan block

$$M_i = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix},$$

and introduce the following matrices:

$$\bar{M} := M_{k_1} \oplus M_{k_2} \oplus \cdots \oplus M_{k_m}, \quad \bar{B} := \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_{1+k_1} \\ \cdots \\ \mathbf{e}_{1+k_1+\cdots+k_{m-1}} \end{bmatrix},$$

of dimension $n \times n$ and $m \times n$, respectively. It is clear that $S(d) := d\bar{B}(I_n - \bar{M}d)^{-1}$ has the following structure:

$$S(d) = \begin{bmatrix} d & d^2 & \ldots & d^{k_1} & & & & & & & & \\ & & & & d & d^2 & \ldots & d^{k_2} & & & & \\ & & & & & & & & \ddots & & & \\ & & & & & & & & & d & d^2 & \ldots & d^{k_m} \end{bmatrix}$$

and, consequently, there exists $C \in \mathbb{F}^{n \times p}$ such that

$$N_{\mathrm{L}}(d) = S(d)C \tag{22}$$

3. Rewrite $D_{\mathrm{L}}(d)$ as $(I_m - S(d)\bar{A})D_{\mathrm{L}}(0)$, for a suitable $\bar{A} \in \mathbb{F}^{n \times m}$. Upon defining

$$A := \bar{M} + \bar{A}\bar{B}, \quad B := D_{\mathrm{L}}(0)^{-1}\bar{B}, \tag{23}$$

it is easy check that $S(d)(I_n - dA) = (I_m - S(d)\bar{A})\bar{B}d$, which implies $D_{\mathrm{L}}(d)^{-1}N_{\mathrm{L}}(d) = Bd(I_n - dA)^{-1}C$.
Thus (20), (22) and (23) provide an n-dimensional state space realization of the encoder $G(d)$.

4. In case $k_i = 0$ for some i, the procedure is the same as above; however the ith row in $\bar{B}$ and in $S(d)$ has to be zero, and the ith diagonal block M_{k_i} is empty.

In case we start from an irreducible MFD $D_{\mathrm{L}}(d)^{-1}N_{\mathrm{L}}(d)$ of $G_{\mathrm{sp}}(d)$, the above procedure provides a *minimal realization*, in the sense that any other state space

realization of the encoder has dimension greater than or equal to n. Suppose, in fact, that $\tilde{\Sigma} = (\tilde{A}, \tilde{B}, \tilde{C})$ is any realization of $G_{\mathrm{sp}}(d)$, with dimension $\tilde{n} < n$. Then $G_{\mathrm{sp}}(d)$ can be represented as

$$\tilde{B}d(I_{\tilde{n}} - \tilde{A}d)^{-1}\tilde{C} = R(d)Q(d)^{-1}\tilde{C} = \tilde{D}(d)^{-1}\tilde{N}(d)\tilde{C} = D_{\mathrm{L}}(d)^{-1}N_{\mathrm{L}}(d),$$

where $R(d)Q(d)^{-1}$ and $\tilde{D}(d)^{-1}\tilde{N}(d)$ are irreducible MFDs of $\tilde{B}d(I_{\tilde{n}} - \tilde{A}d)^{-1}$ with

$$\begin{bmatrix} Q(d) \\ R(d) \end{bmatrix} \quad \text{and} \quad [\tilde{D}(d) \mid \tilde{N}(d)] \tag{24}$$

column (resp. row) reduced. As we have shown in section 2, both matrices in (24) have the same internal degree, and therefore their external degrees coincide, too. Since the external degree of $\begin{bmatrix} d\tilde{B} \\ I - \tilde{A}d \end{bmatrix}$ does not exceed $\tilde{n}$, this is also true for the degrees of (24) and, consequently, for the (external and internal) degrees of $[\tilde{D}(d) \mid \tilde{N}(d)\tilde{C}]$ and of (21). This, however, gives a contradiction, as the external degree of (21) is $n > \tilde{n}$.

We summarize the above discussion in the following proposition.

Proposition 5.1. *Suppose that $D(d)^{-1}N(d)$ is an irreducible left MFD of a causal encoder $G(d)$ such that*

$$[D(d) \mid N(d)]$$

is row reduced, with row degrees $k_1, k_2, \ldots, k_m$. The minimal dimension $\mu(G)$ of a state realization of $G(d)$ is called the "McMillan degree of $G(d)$" [6,19], *and is given by $n = \sum_i k_i$.*

A convolutional code $\mathscr{C}$ admits infinitely many different encoders. So a natural problem is that of characterizing which encoders of $\mathscr{C}$ have minimal McMillan degree, and hence can be realized by linear sequential circuits with minimum number of delay elements. They are called *minimal encoders (of $\mathscr{C}$)*.

It is easy to check that the McMillan degree of a canonical encoder $G_c(d)$ coincides with the degree of its code $\mathscr{C}$. In fact $I^{-1}G_c(d)$ is an irreducible MFD of $G_c(d)$ and $[I \mid G_c(d)]$ is row reduced, the row degrees being the Forney indices $\phi_1, \ldots, \phi_m$ of $\mathscr{C}$.

On the other hand, any other causal encoder $G(d)$ admits an irreducible left MFD

$$G(d) = D(d)^{-1}[\Delta(d)G_c(d)]$$

with $D(0)$ invertible and $\Delta(d)$ nonsingular. Moreover, in case $[D(d) \mid \Delta(d)G_c(d)]$ is not row reduced, left multiplication by a suitable unimodular $V(d)$ produces a row reduced matrix

$$[V(d)D(d) \mid V(d)\Delta(d)G_c(d)]$$

with row degrees $k_1, k_2, \ldots, k_m$ and $(V(d)D(d))^{-1}[V(d)\Delta(d)G_c(d)]$ is still an irreducible MFD of $G(d)$. Consequently

$$\mu(G) = \sum_i k_i = \text{extdeg}[V D \mid V \Delta G_c] \geqslant \text{extdeg}(V \Delta G_c) \geqslant \text{intdeg}(V \Delta G_c)$$
$$\geqslant \text{intdeg}(G_c) = \text{extdeg}(G_c) = \sum_i \phi_i$$

and $\deg \mathscr{C} = \sum \phi_i$ provides the minimum McMillan degree of all causal encoders of $\mathscr{C}$. We have therefore proved the following proposition.

Proposition 5.2 [2]. *A causal encoder $G(d)$ of $\mathscr{C}$ is minimal if and only if its McMillan degree coincides with* $\deg \mathscr{C}$.

It is clear that canonical encoders are minimal and it is easy to check that minimal polynomial encoders are basic. In fact, if $G(d)$ is polynomial and nonbasic, there exists a nonunimodular left factor $\Delta(d)$ such that $G(d) = \Delta(d)G_c(d)$, $G_c(d)$ a canonical encoder. On the other hand, if $[I_m \mid G(d)]$ fails to be row reduced, there exists an unimodular matrix $V(d)$ such that $[V(d) \mid V(d)G(d)]$ is row reduced. Then

$$\mu(G) = \text{extdeg}[V \mid VG] = \text{intdeg}[V \mid VG] \geqslant \text{intdeg}(VG) = \text{intdeg}(G)$$
$$= \text{intdeg}(\Delta G_c) > \text{intdeg}(G_c) = \sum \phi_i.$$

The following examples show that the converse inclusions do not hold.

Example 5.1. The canonical encoder

$$G_c(d) = \begin{bmatrix} d^4+1 & d^4 & d \\ d^3 & 1 & d+1 \end{bmatrix}$$

has McMillan degree 7. The left MFD $U(d)^{-1}G_c(d)$, with

$$U(d) = \begin{bmatrix} d^2+1 & d^2 \\ d^2 & d^2-1 \end{bmatrix}$$

is an irreducible representation of the polynomial encoder

$$G_b(d) = \begin{bmatrix} -d^6+d^5+d^4-d^2+1 & -d^6+d^4+d^2 & d^2+d \\ d^6-d^5-d^3+d^2 & d^6-d^2-1 & -d^2-d-1 \end{bmatrix}. \tag{25}$$

Clearly $G_b(d)$ is basic, noncanonical, since (25) fails to be row reduced, and minimal, since $[U(d)\ G_c(d)]$ is row reduced with external row degree 7.

Example 5.2. The canonical encoder

$$G_c(d) = \begin{bmatrix} d+1 & d & d \\ -d & -d+1 & 1 \end{bmatrix}$$

has McMillan degree 2. The equivalent encoder

$$G(d) = U(d)^{-1}G_c(d) = \begin{bmatrix} d^2+1 & d^2 \\ -1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} d+1 & d & d \\ -d & -d+1 & 1 \end{bmatrix}$$

is basic, as $U(d)$ is unimodular, and nonminimal. In fact

$$[U(d)\ G_c(d)] = \begin{bmatrix} d^2+1 & d^2 & d+1 & d & d \\ -1 & -1 & -d & -d+1 & 1 \end{bmatrix}$$

is row reduced and the sum of the row degrees is 3, so that $\mu(G) = 3 > \mu(G_c)$.

6. Structure of minimal encoders

The purpose of this section is to characterize the structure of all minimal encoders of a code $\mathscr{C}$, and to provide a complete parametrization based on their MFDs. The first Proposition and the subsequent Corollary are based on a result on polynomial invertibility that traces back to a classical paper [2] by Forney.

Proposition 6.1. *Let $G(d) \in \mathbb{F}(d)^{m\times p}$ be a causal encoder of $\mathscr{C}$. The following are equivalent*:

(i) *$G(d)$ is minimal.*
(ii) *$G(d)$ admits a left MFD*

$$G(d) = D(d)^{-1}G_c(d) \tag{26}$$

where $G_c(d)$ is a canonical encoder and $\deg \text{row}_i\, D \leqslant \deg \text{row}_i\, G_c$, $i = 1, \ldots, m$.
(iii) *$G(d)$ has a right polynomial inverse $X(d) \in \mathbb{F}[d]^{p\times m}$ and a right polynomial inverse $Y(d^{-1}) \in \mathbb{F}[d^{-1}]^{p\times m}$.*

Proof. (i) $\Rightarrow$ (ii) Consider an irreducible left MFD $D(d)^{-1}N(d)$ of $G(d)$ with $[D(d) \mid N(d)]$ row reduced, so that

$$deg\ \mathscr{C} = \mu(G) = \text{extdeg}[D \mid N] \geqslant \text{extdeg}(N) \geqslant \mu(G). \tag{27}$$

All terms in (27) coincide, so $N(d)$ is a canonical encoder of $\mathscr{C}$ and the row degrees in $N(d)$ are the same as in $[D(d) \mid N(d)]$. Consequently the row degrees of $D(d)$ cannot exceed the corresponding ones in $N(d)$.

(ii) $\Rightarrow$ (iii) If $R(d)$ denotes a right polynomial inverse of $G_c(d)$, we have that

$$X(d) := R(d)D(d)$$

is an inverse of $G(d)$ with entries in $\mathbb{F}[d]$.

On the other hand, if $\phi_1, \ldots, \phi_m$ are the row degrees of $G_c(d)$,

$$\begin{aligned} G(d) &= [\text{diag}\{d^{-\phi_1}, \ldots, d^{-\phi_m}\}D(d)]^{-1}[\text{diag}\{d^{-\phi_1}, \ldots, d^{-\phi_m}\}G_c(d)] \\ &=: \tilde{D}(d^{-1})^{-1}\tilde{N}(d^{-1}) \end{aligned}$$

is a left MFD of $G(d)$ in $\mathbb{F}[d^{-1}]$. Since $G_c(d)$ is left prime and row reduced, $\tilde{N}(d^{-1})$ is full rank for every $d^{-1} \in \bar{\mathbb{F}}$, and $\tilde{N}(0) = (G_c)_{\mathrm{hr}}$ is full rank too. This implies that $\tilde{N}(d^{-1})$ is left prime and has a right inverse $\tilde{R}(d^{-1})$ in $\mathbb{F}[d^{-1}]$. So,

$$Y(d^{-1}) := \tilde{R}(d^{-1})\tilde{D}(d^{-1})$$

provides an $\mathbb{F}[d^{-1}]$ polynomial right inverse of $G(d)$.

(iii) $\Rightarrow$ (i) Suppose that $D(d)^{-1}N(d)$ is an irreducible left MFD of $G(d)$, and $[D(d) \mid N(d)]$ is row reduced with row degrees $k_1, \ldots, k_m$. Upon defining

$$[\tilde{D}(d^{-1}) \mid \tilde{N}(d^{-1})] := \mathrm{diag}\{d^{-k_1}, \ldots, d^{-k_m}\}[D(d) \mid N(d)],$$

consider also $\tilde{D}(d^{-1})^{-1}\tilde{N}(d^{-1})$, an irreducible left MFD of $G(d)$ over the ring $\mathbb{F}[d^{-1}]$, with $\tilde{D}(d^{-1})$ row reduced.

Let $M(d)$ be a polynomial right inverse of $[D(d) \mid N(d)]$ and note that the equation $D(d)^{-1}N(d)X(d) = I_m$ implies $I_m = N(d)[X(d) \mid I_p]M(d)$, showing that $N(d)$ is left prime. By a similar argument one sees that $\tilde{N}(d^{-1})$ is left prime. This guarantees that $\tilde{N}(0)$, which is equal to the leading row coefficient matrix N_{hr} of $N(d)$, has rank m. So $N(d)$ is row reduced and provides a canonical encoder of $\mathscr{C}$.

Finally, by resorting to

$$[D(d) \mid N(d)] := \mathrm{diag}\{d^{k_1}, \ldots, d^{k_m}\}[\tilde{D}(d^{-1}) \mid \tilde{N}(d^{-1})], \tag{28}$$

we have that the row degrees of $D(d)$ do not exceed the corresponding degrees of $N(d)$. So the McMillan degree of $G(d)$ is the sum of the row degrees of $N(d)$, and $G(d)$ is a minimal encoder. □

Corollary 6.2. *A systematic causal encoder is minimal and a minimal encoder is noncatastrophic.*

Proposition 6.4 shows that all minimal encoders of $\mathscr{C}$, and in particular all canonical and systematic encoders, can be represented as MFDs whose numerator is a *fixed* canonical encoder $G_c(d)$. The proof depends on the following technical lemma.

Lemma 6.3. *Suppose that both $[N(d) \mid D(d)]$ and its block $N(d)$ are row reduced, with same row degrees $k_1, \ldots, k_m$. Suppose, moreover, that $V(d)$ is unimodular, and let*

$$[\tilde{N}(d) \mid \tilde{D}(d)] = V(d)[N(d) \mid D(d)].$$

If $\tilde{N}(d)$ is row reduced, the same holds true for $[\tilde{N}(d) \mid \tilde{D}(d)]$, and both matrices have row degrees $k_1, \ldots, k_m$, up to a permutation.

Proof. As $N(d)$ and $\tilde{N}(d)$ are row reduced and differ each other by a left unimodular factor $V(d)$, the row degrees k_i of $N(d)$ and $\tilde{k}_i$ of $\tilde{N}(d)$ coincide, up to a permutation. So, possibly after multiplying $V(d)$ on the left by a permutation matrix,

we shall assume $k_i = \tilde{k}_i$, $i = 1, \ldots, m$. The predictable degree property of $N(d)$ and $\tilde{D}(d) = V(d)D(d)$ imply

$$\begin{aligned}\deg\ \mathrm{row}_i \tilde{D} &\leqslant \max_{j:v_{ij}(d)\neq 0}\{\deg\ \mathrm{row}_j D + \deg v_{ij}\}\\ &\leqslant \max_{j:v_{ij}(d)\neq 0}\{k_j + \deg v_{ij}\} = \tilde{k}_i.\end{aligned}$$

Thus $\tilde{k}_i$, $i = 1, \ldots, m$, are the row degrees of $[\tilde{N}(d) \mid \tilde{D}(d)]$, which is row reduced. □

Proposition 6.4. *Let $G_c(d)$ be a canonical encoder of $\mathscr{C}$.*

(i) *All minimal encoders of $\mathscr{C}$ can be represented as*

$$G(d) = D(d)^{-1}G_c(d),$$

upon varying the denominator in the set of $m \times m$ polynomial matrices $D(d)$ with $D(0)$ nonsingular and $\deg\ \mathrm{row}_i\ D \leqslant \deg\ \mathrm{row}_i\ G_c$, $i = 1, \ldots, m$.

(ii) *All polynomial minimal encoders of $\mathscr{C}$ are obtained by restricting denominators $D(d)$ to unimodular matrices.*

(iii) *All systematic causal encoders of $\mathscr{C}$ are given by*

$$G(d) = D(d)^{-1}G_c(d)$$

where $D(d)$ is any $m \times m$ matrix of $G_c(d)$ with $D(0)$ nonsingular.

(iv) *Suppose that the row degrees of $G_c(d)$ are non decreasing, and Forney indices assume $q \leqslant m$ distinct values $\phi'_1 < \phi'_2 < \cdots < \phi'_q$, with multiplicity d_h, $h = 1, \ldots, q$.*
Any other canonical encoder of $\mathscr{C}$, with non decreasing row degrees, is given by

$$\tilde{G}_c(d) = D(d)^{-1}G_c(d) \tag{29}$$

as $D(d)$ varies in the group of block polynomial matrices of the form

$$\begin{bmatrix} D_{11} & & & \\ D_{21}(d) & D_{22} & & \\ \vdots & \vdots & \ddots & \\ D_{q1}(d) & D_{q2}(d) & \cdots & D_{qq} \end{bmatrix}, \tag{30}$$

where $D_{hh} \in \mathbb{F}^{d_h \times d_h}$ is non singular, $h = 1, \ldots, q$, and the degree of each entry in $D_{hk}(d)$, $h > k$, does not exceed $\phi'_h - \phi'_k$.

Proof. (i) By Proposition 6.1, any minimal encoder $G(d)$ can be expressed as $G(d) = \tilde{D}(d)^{-1}\tilde{G}_c(d)$, where $\tilde{G}_c(d)$ is a canonical encoder and $\tilde{D}(d)$ is a polynomial matrix whose row degrees do not exceed the corresponding ones in $\tilde{G}_c(d)$.

Let $V(d)$ be an unimodular matrix such that $V(d)\tilde{G}_c(d) = G_c(d)$, and let $D(d) := V(d)\tilde{D}(d)$. Clearly $G(d)$ can be represented as $D(d)^{-1}G_c(d)$; moreover, by Lemma 6.3, $[D(d) \mid G_c(d)]$ is row reduced and deg $\text{row}_i\, D \leqslant$ deg $\text{row}_i\, G_c$, $i = 1, \ldots, m$.

(ii) Since $G_c(d)$ is left prime, $D(d)^{-1}G_c(d)$ is polynomial if and only if $D(d)^{-1}$ is polynomial, which amounts to say that $D(d)$ is unimodular.

(iii) Every systematic encoder $G(d)$ of $\mathscr{C}$ satisfies $G(d)P = [I_m \mid \tilde{G}_2(d)]$, where P is a suitable column permutation matrix. If $G(d)$ is causal, it has to be minimal, and consequently it can be expressed by a MFD

$$[I_m \mid \tilde{G}_2(d)]P^{-1} = D(d)^{-1}G_c(d),$$

with $D(0)$ nonsingular. So

$$D(d)[I_m \mid \tilde{G}_2(d)] = G_c(d)P$$

shows that, up to a column permutation, $D(d)$ is an invertible $m \times m$ submatrix of $G_c(d)$.

Conversely, assume that, up to a column permutation, $D(d)$ is an $m \times m$ invertible submatrix of $G_c(d)$ with $D(0)$ nonsingular. Then there exists a permutation matrix P such that $G_c(d)P = [D(d) \mid M(d)]$ and consequently

$$D(d)^{-1}G_c(d) = [I_m \mid D(d)^{-1}M(d)]P^{-1}$$

is systematic.

(iv) Suppose that the row degrees $\phi_1, \ldots, \phi_m$ of two canonical encoders $\tilde{G}_c(d)$ and $G_c(d)$ are non decreasing and consider an unimodular matrix $T(d)$ such that $\tilde{G}_c(d) = T(d)G_c(d)$. As both $\tilde{G}_c(d)$ and $G_c(d)$ are row reduced, the predictable degree property implies that

$$\phi_i = \deg(\text{row}_i(TG_c)) = \max_{j:T_{ij}(d)\neq 0} \{\phi_j + \deg(T_{ij})\} \tag{31}$$

and therefore

$$\begin{aligned} \deg(T_{ij}) \leqslant \phi_i - \phi_j \text{ or } T_{ij}(d) = 0 \quad &\text{if } \phi_i > \phi_j, \\ \deg(T_{ij}) = 0 \text{ or } T_{ij}(d) = 0 \quad &\text{if } \phi_i = \phi_j, \\ T_{ij}(d) = 0 \quad &\text{if } \phi_i < \phi_j. \end{aligned}$$

Clearly $T(d)$ is block triangular and its diagonal blocks must be constant and non singular. Moreover it is easy to show that both $T(d)$ and $D(d) := T(d)^{-1}$ satisfy the degree constraints specified in (iv), and therefore $\tilde{G}_c(d)$ can be represented as in (29).

Conversely any $D(d)$ as given in (30) is unimodular, and clearly

$$\deg \text{row}_i(G_c) = \deg \text{row}_i(D^{-1}G_c),\ i = 1, \ldots, m,$$

which implies that $D(d)^{-1}G_c(d)$ is canonical. □

Remark. A particular choice of matrix $D(d)$ in (30) is described in [4], that allows to obtain a canonical encoder in *echelon form*.

7. Abstract states

Given a causal (polynomial or rational) encoder $G(d)$, consider the map

$$s : \mathbb{F}[d^{-1}]^m \to d\mathbb{F}[[d]]^p : \hat{\mathbf{u}}(d^{-1}) \mapsto (\mathrm{id} - \mathscr{P}_1)(\hat{\mathbf{u}}G),$$

that associates to an information signal $\hat{\mathbf{u}}(d^{-1})$ with support in $(-\infty, 0]$ the restriction to $[1, +\infty)$ of the corresponding codeword. The elements of the image of s, i.e. the free evolutions of the encoder output on $[1, +\infty)$, are called the *abstract states* of the encoder [2,5]; so, an information signal $\hat{\mathbf{u}}(d) \in \mathbb{F}((d))^m$ induces at time $t = 1$ the abstract state given by the codeword restriction $(\mathrm{id} - \mathscr{P}_1)((\mathscr{P}_1\hat{\mathbf{u}})G)$. Since $\mathrm{Im}\, s$ is canonically isomorphic to $\mathbb{F}[d^{-1}]^m / \ker s$, the abstract state induced by $\hat{\mathbf{u}}(d)$ can be viewed also as the coset $\mathscr{P}_1\hat{\mathbf{u}} + \ker s$ in $\mathbb{F}[d^{-1}]^m$. In general System Theory, two inputs $\hat{\mathbf{u}}_1(d)$ and $\hat{\mathbf{u}}_2(d)$ in $\mathbb{F}[d^{-1}]^m$ are "Nerode equivalent" [22] if and only if the output sequences they induce on $[1, +\infty)$ are the same, and remain the same whenever both $\hat{\mathbf{u}}_1(d)$ and $\hat{\mathbf{u}}_2(d)$ are followed by an arbitrary input $\hat{\mathbf{v}}(d) \in d\mathbb{F}[[d]]^m$. Thus, an abstract state of an encoder can be viewed as a "Nerode equivalent class" on the information sequences ending at time 0.

In this section, we shall investigate how some properties of an encoder do reflect into the structure of its abstract state space, the final goal being a classical characterization of minimal encoders, due to Forney. In our discussion, we provide in advance a fairly complete account of different inclusions between the span of an information sequence and that of the corresponding codeword, and show how they are related to a nontrivial intersection between the code $\mathscr{C}$ and the abstract state space of the encoder.

Let $D(d)^{-1}N(d)$ be an irreducible left MFD of a causal encoder $G(d)$, with $N(d)$ row reduced and $\deg \mathrm{row}_i(N) = k_i$, $i = 1, \ldots, m$, and factorize $N(d)$ into $N(d) = \Delta(d)\bar{N}(d)$, $\bar{N}(d)$ left prime.

Lemma 7.1. *Consider the following inclusion relations*:

(I) $\inf \mathrm{span}(\hat{\mathbf{v}}) \geqslant \inf \mathrm{span}(\hat{\mathbf{v}}G)$, $\forall\, \hat{\mathbf{v}}(d) \in \mathbb{F}((d))^m$,
(S_{fin}) $\sup \mathrm{span}(\hat{\mathbf{v}}) \leqslant \sup \mathrm{span}(\hat{\mathbf{v}}G)$, $\forall\, \hat{\mathbf{v}}(d) \in \mathbb{F}[d, d^{-1}]^m$,
(S_∞) $\sup \mathrm{span}(\hat{\mathbf{v}}) = \infty \Longrightarrow \sup \mathrm{span}(\hat{\mathbf{v}}G) = \infty$, $\forall\, \hat{\mathbf{v}}(d) \in \mathbb{F}((d))^m$,
(B_{fin}) $\mathrm{span}(\hat{\mathbf{v}}) \subseteq \mathrm{span}(\hat{\mathbf{v}}G)$, $\forall\, \hat{\mathbf{v}}(d) \in \mathbb{F}[d, d^{-1}]^m$,
(B) $\mathrm{span}(\hat{\mathbf{v}}) \subseteq \mathrm{span}(\hat{\mathbf{v}}G)$, $\forall\, \hat{\mathbf{v}}(d) \in \mathbb{F}((d))^m$.

Then we have the equivalences:

$$(I)\ \&\ (S_{\mathrm{fin}}) \Longleftrightarrow (B_{\mathrm{fin}}) \tag{32}$$

$$(I)\ \&\ (S_{\mathrm{fin}})\ \&\ (S_\infty) \Longleftrightarrow (B) \tag{33}$$

Moreover

(a) (I) *holds if and only if* $\mathrm{rank}\, N(0) = m$,
(b) (S_{fin}) *holds if and only if* $\deg \mathrm{row}_i(D) \leqslant \deg \mathrm{row}_i(N)$, $i = 1, \ldots, m$,
(c) (S_∞) *holds if and only if* $\det(\Delta) = \alpha d^k$, $\alpha \in \mathbb{F}\backslash\{0\}$, $k \geqslant 0$.

Proof. (32) and (33) are obvious.

(a) rank $N(0) = m$ is equivalent to rank $G(0) = m$, which is clearly equivalent to (I).

(b) Let $\deg \operatorname{row}_i(D) \leqslant \deg \operatorname{row}_i(N) = k_i, \quad i = 1, \ldots, m$. Given $\hat{\mathbf{v}}(d) \in \mathbb{F}[d, d^{-1}]^m$, suppose $\sup \operatorname{span}(\hat{\mathbf{v}}G) = \ell \in \mathbb{N}$. Then $\hat{\mathbf{u}}(d, d^{-1}) := \hat{\mathbf{v}}(d)D(d)^{-1}$ is Laurent polynomial, as

$$\hat{\mathbf{u}}(d, d^{-1})[D(d) \mid N(d)] = [\hat{\mathbf{v}}(d) \mid \hat{\mathbf{v}}(d)G(d)]$$

is polynomial and $[D(d) \mid N(d)]$ is left prime. Furthermore, since $N(d)$ is row reduced,

$$\deg(\hat{\mathbf{u}}N) = \ell \Longrightarrow \deg \hat{u}_i \leqslant \ell - k_i, \ i = 1, \ldots, m$$

and

$$\hat{v}_i(d) = \hat{\mathbf{u}}(d, d^{-1})\operatorname{col}_i(D), \quad i = 1, \ldots, m$$

implies

$$\deg \hat{v}_i \leqslant \max_{0 \leqslant i \leqslant m} \{\deg \hat{u}_i + k_i\} \leqslant \ell.$$

We therefore have $\sup \operatorname{span}(\hat{\mathbf{v}}) \leqslant \sup \operatorname{span}(\hat{\mathbf{v}}G)$.

Vice-versa, suppose that $\deg \operatorname{row}_i(D) > k_i, \ \exists i \in \{1, \ldots, m\}$. The information sequence $\hat{\mathbf{v}}(d) := [0 \cdots d^{-k_i} \ \cdots \ 0]D(d) = d^{-k_i} \operatorname{row}_i(D)$, is polynomial with degree greater than zero, and the corresponding codeword,

$$\hat{\mathbf{v}}(d)G(d) = \hat{\mathbf{v}}(d)D(d)^{-1}N(d) = d^{-k_i} \operatorname{row}_i(N)$$

has degree zero, i.e., $\sup \operatorname{span}(\hat{\mathbf{v}}) > \sup \operatorname{span}(\hat{\mathbf{v}}G)$.

(c) has been already proved in Proposition 4.1. □

Proposition 7.2. *The code $\mathscr{C}$ does not include nonzero abstract states of the encoder $G(d)$, i.e.* $(\operatorname{Im} s) \cap \mathscr{C} = \{0\}$, *if and only if* (I), (S_{fin}) *and* (S_∞) *in Lemma* 7.1 *simultaneously hold.*

Proof. If (I) does not hold, there exists $\hat{\mathbf{v}}(d) \in \mathbb{F}((d))^m$ such that $\inf \operatorname{span}(\hat{\mathbf{v}}) \leqslant 0$ and $\inf \operatorname{span}(\hat{\mathbf{v}}G) > 0$. By the causality of $G(d)$,

$$0 = \mathscr{P}_1(\hat{\mathbf{v}}G) = \mathscr{P}_1\big((\mathscr{P}_1\hat{\mathbf{v}})G\big),$$

which implies that the nonzero codeword $(\mathscr{P}_1\hat{\mathbf{v}})G = (\mathrm{id} - \mathscr{P}_1)\big((\mathscr{P}_1\hat{\mathbf{v}})G\big)$ is an abstract state of $G(d)$.

If (S_{fin}) or (S_∞) do not hold, there exists $\hat{\mathbf{v}}(d) \in \mathbb{F}((d))^m$ such that $\sup \operatorname{span}(\hat{\mathbf{v}}) > 0$ and $\sup \operatorname{span}(\hat{\mathbf{v}}G) \leqslant 0$. Therefore

$$0 = (\mathrm{id} - \mathscr{P}_1)(\hat{\mathbf{v}}G) = (\mathrm{id} - \mathscr{P}_1)\big((\mathscr{P}_1\hat{\mathbf{v}})G\big) + (\mathrm{id} - \mathscr{P}_1)\big([(\mathrm{id} - \mathscr{P}_1)\hat{\mathbf{v}}]G\big)$$

and by causality, $(\mathrm{id} - \mathscr{P}_1)\big((\mathscr{P}_1\hat{\mathbf{v}})G\big) = -\big((\mathrm{id} - \mathscr{P}_1)\hat{\mathbf{v}}\big)G \neq 0$ belongs to $(\operatorname{Im} s) \cap \mathscr{C}$.

Vice-versa, suppose that (I), (S_{fin}) and (S_∞) hold and suppose that the abstract state of $\hat{\mathbf{u}}(d^{-1}) \in \mathbb{F}[d^{-1}]^m$ is a codeword, i.e.,

$$s(\hat{\mathbf{u}}(d^{-1})) = (\mathrm{id} - \mathscr{P}_1)(\hat{\mathbf{u}}G) = \hat{\mathbf{v}}(d)G(d) \tag{34}$$

for some $\hat{\mathbf{v}}(d) \in \mathbb{F}((d))^m$. As $\inf \operatorname{span}(\hat{\mathbf{v}}G) > 0$, (I) implies $\inf \operatorname{span}(\hat{\mathbf{v}}) > 0$ and, by causality, $\hat{\mathbf{v}}(d)G(d) = (\mathrm{id} - \mathscr{P}_1)([(\mathrm{id} - \mathscr{P}_1)\hat{\mathbf{v}}]G)$. By (34), the codeword

$$(\hat{\mathbf{u}}(d^{-1}) - \hat{\mathbf{v}}(d))G(d) = \mathscr{P}_1(\hat{\mathbf{u}}G) + (\mathrm{id} - \mathscr{P}_1)(\hat{\mathbf{u}}G) - \hat{\mathbf{v}}(d)G(d) = \mathscr{P}_1(\hat{\mathbf{u}}G)$$

has support in $(-\infty, 0]$. Thus by (S_{fin}) and (S_∞), we have $\operatorname{span}(\hat{\mathbf{u}}(d^{-1}) - \hat{\mathbf{v}}(d)) \subseteq (-\infty, 0]$ and therefore $\hat{\mathbf{v}}(d) = (\mathrm{id} - \mathscr{P}_1)(\hat{\mathbf{u}} - \hat{\mathbf{v}}) = 0$, i.e, $s(\hat{\mathbf{u}}(d^{-1})) = 0$. □

The following proposition is now an immediate consequence of Proposition 7.2.

Proposition 7.3 [2, 5, 19]. *The following are equivalent*:

(i) $(\operatorname{Im} s) \cap \mathscr{C} = \{0\}$,
(ii) $G(d)$ *is a minimal encoder*,
(iii) $\operatorname{span}(\hat{\mathbf{v}}) \subseteq \operatorname{span}(\hat{\mathbf{v}}G),\ \forall\, \hat{\mathbf{v}}(d) \in \mathbb{F}((d))^m$.

Proof. Both (i) and (iii) are equivalent to assumption (I) & (S_{fin}) & (S_∞) of Lemma 7.1.

On the other hand, represent $G(d)$ as $D(d)^{-1}N(d)$, with $N(d)$ row reduced, and write $N(d) = \Delta(d)\bar{N}(d)$, with $\bar{N}(d)$ left prime. By (a) and (c) of Lemma 7.1, conditions (I) and (S_∞) are equivalent to assume that $\Delta(d)$ is unimodular (i.e., $N(d)$ is left prime), and so, by Proposition 6.1, we conclude that (I) & (S_{fin}) & (S_∞) altogether imply and are implied by the minimality of $G(d)$. □

We restrict now our analysis to the abstract state structure of two classes of encoders, i.e., minimal encoders and polynomial reduced encoders.

Referring to the representation (26), let $G(d) = D(d)^{-1}G_c(d)$ be a minimal encoder, and $k_i \leqslant \phi_i$ be the row degrees of $D(d)$. The abstract zero state of the encoder, viewed as a coset in $\mathbb{F}[d^{-1}]^m/\ker s$,

$$\ker s = \left\{\hat{\mathbf{u}}(d^{-1}) \in \mathbb{F}[d^{-1}]^m : \hat{\mathbf{u}}(d^{-1})D(d)^{-1}G_c(d) \in \mathbb{F}[d^{-1}]^p\right\}, \tag{35}$$

can be computed as follows. If $\hat{\mathbf{u}}(d^{-1}) \in \ker s$, then $\hat{\mathbf{v}}(d, d^{-1}) := \hat{\mathbf{u}}(d^{-1})D(d)^{-1}$ must be a Laurent polynomial vector, otherwise the upper bound of the support of $\hat{\mathbf{v}}(d, d^{-1})G_c(d)$ would not be finite because of the left primeness of $G_c(d)$. Substituting $\hat{\mathbf{u}}(d^{-1}) = \hat{\mathbf{v}}(d, d^{-1})D(d)$ into (35), gives $\deg \hat{v}_i \leqslant -\phi_i$, $i = 1, \ldots, m$, and, consequently,

$$\begin{aligned}\ker s &= \left\{\hat{\mathbf{u}}(d^{-1}) = \hat{\mathbf{w}}(d^{-1})\operatorname{diag}\{d^{-\phi_1}, \ldots, d^{-\phi_m}\}D(d),\ \hat{\mathbf{w}}(d^{-1}) \in \mathbb{F}[d^{-1}]^m\right\} \\ &= \left\{\hat{\mathbf{u}}(d^{-1}) = \hat{\mathbf{w}}(d^{-1})\tilde{D}(d^{-1}),\ \hat{\mathbf{w}}(d^{-1}) \in \mathbb{F}[d^{-1}]^m\right\},\end{aligned}$$

where $\tilde{D}(d^{-1}) = \mathrm{diag}\{d^{-\phi_1}, \ldots, d^{-\phi_m}\}D(d)$. Taking the Smith form of $\tilde{D}(d^{-1})$

$$\tilde{D}(d^{-1}) = \tilde{W}(d^{-1})\mathrm{diag}\{\tilde{\psi}_1(d^{-1}), \ldots, \tilde{\psi}_m(d^{-1})\}\tilde{V}(d^{-1}),$$

$\tilde{V}(d^{-1})$ and $\tilde{W}(d^{-1})$ unimodular matrices, we have also

$$\ker s = \big\{\hat{\mathbf{u}}(d^{-1}) = \hat{\mathbf{m}}(d^{-1})\mathrm{diag}\{\tilde{\psi}_1(d^{-1}), \ldots, \tilde{\psi}_m(d^{-1})\}\tilde{V}(d^{-1}), \quad \hat{\mathbf{m}}(d^{-1}) \in \mathbb{F}[d^{-1}]^m\big\}.$$

So, the abstract states of $G(d)$ are the cosets, modulo $\ker\ s$, of the $\mathbb{F}$-linear combinations of the independent vectors $d^{-i}\mathbf{e}_j\tilde{V}(d^{-1})$, $j = 1, \ldots, m$, $0 \leqslant i < \deg\tilde{\psi}_j(d^{-1})$.

Moreover, letting $\tilde{N}(d^{-1}) := \mathrm{diag}\{d^{-\phi_1}, \ldots, d^{-\phi_m}\}G_c(d)$, the codeword induced by any information signal $\hat{\mathbf{u}}(d^{-1}) \in \mathbb{F}[d^{-1}]^m$ satisfies

$$\tilde{\psi}_1(d^{-1})\hat{\mathbf{u}}(d^{-1})G(d) = \hat{\mathbf{u}}(d^{-1})\tilde{V}(d^{-1})^{-1}\begin{bmatrix} 1 & & \\ & \ddots & \\ & & \frac{\tilde{\psi}_1(d^{-1})}{\tilde{\psi}_m(d^{-1})} \end{bmatrix}\tilde{W}(d^{-1})^{-1}\tilde{N}(d^{-1}) \in \mathbb{F}[d^{-1}]^p$$

which implies that $\tilde{\psi}_1(d^{-1})\hat{\mathbf{u}}(d^{-1}) \in \ker s$.

If $G(d)$ is a row reduced polynomial encoder with row degrees $k_1, \ldots, k_m$, the zero state $\ker\ s$ consists of all input signals $\hat{\mathbf{u}}(d^{-1})$ satisfying $\deg\hat{u}_i \leqslant -k_i$, $i = 1, \ldots, m$, and, vice-versa, this condition implies that $G(d)$ is row reduced. So, the restriction to $[1, +\infty)$ of the codeword induced by $\hat{\mathbf{u}}(d^{-1}) \in \mathbb{F}[d^{-1}]^m$ provides a complete information on the restriction of $\hat{u}_i(d)$ to $(-k_i, 0]$, $i = 1, 2, \ldots, m$, and no information on the remaining coefficients of $\hat{u}_i(d)$.

8. State feedback and parametrization of minimal encoders

In this section it will be shown that all minimal encoders of $\mathscr{C}$ can be obtained from a minimal one, by applying static feedback and static precompensation to a minimal state space realization of a canonical encoder $G_c(d)$. In the coding literature, rational minimal encoders are often synthesized via linear sequential circuits involving feedback elements [5]. Consequently, modifying feedback elements (and introducing a combinatorial precompensation circuit) allows to sweep the whole class of circuits that synthesize the minimal encoders of $\mathscr{C}$.

Suppose that $\Sigma = (A, B, C, J)$ is the minimal realization of $G_c(d) = I_m^{-1}G_c(d)$, given by (20), (22) and (23) in Section 5. As we have seen, the dimension n of the realization coincides with the degree $\sum_i \phi_i$ of $\mathscr{C}$. If the state $\mathbf{x}$ is feed-back into the system via a matrix $K \in \mathbb{F}^{n\times m}$, the input sequence becomes the sum of the information sequence $\{\mathbf{u}_t\}$ and the feedback sequence $\{\mathbf{x}_t K\}$, and the state model Σ modifies into $\Sigma^{(K)} = (A + KB, B, C + KJ, J)$, as we have (Fig. 1)

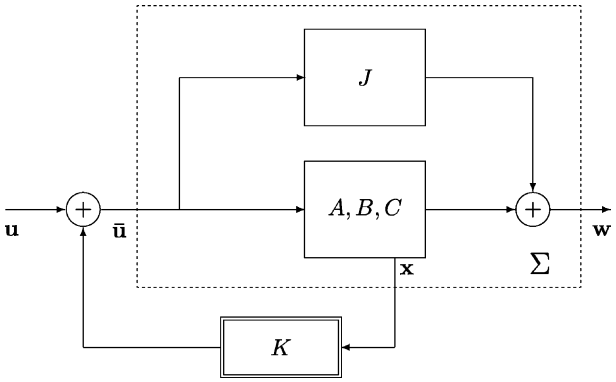

Fig. 1.

$$\mathbf{x}_{t+1} = \mathbf{x}_t A + [\mathbf{u}_t + \mathbf{x}_t K]B = \mathbf{x}_t[A + KB] + \mathbf{u}_t B$$
$$\mathbf{w}_t = \mathbf{x}_t C + [\mathbf{u}_t + \mathbf{x}_t K]J = \mathbf{x}_t[C + KJ] + \mathbf{u}_t J$$

The series $\hat{\mathbf{x}}(d) := \sum_t \mathbf{x}_t d^t$, corresponding to the forced state evolution of $\Sigma^{(K)}$, and the information series $\hat{\mathbf{u}}(d) := \sum_t \mathbf{u}_t d^t$ are connected by

$$\hat{\mathbf{x}}(d) = (\hat{\mathbf{u}}(d) + \hat{\mathbf{x}}(d)K)Bd(I_n - Ad)^{-1},$$

which implies

$$\hat{\mathbf{x}}(d) = \hat{\mathbf{u}}(d)[I_m - Bd(I_n - dA)^{-1}K]^{-1}Bd(I_n - dA)^{-1}.$$

As the output $\hat{\mathbf{w}}(d) := \sum_t \mathbf{w}_t d^t$ is given by $\hat{\mathbf{x}}(d)(C + KJ) + \hat{\mathbf{u}}(d)J$, the transfer matrix of $\Sigma^{(K)}$ is represented by the left MFD

$$\begin{aligned} G^{(K)}(d) &= [I_m - Bd(I_n - dA)^{-1}K]^{-1}[J + Bd(I_n - dA)^{-1}C] \\ &= [I_m - S(d)K]^{-1}G_c(d). \end{aligned}$$

As K varies in $\mathbb{F}^{n\times m}$, the matrix $(I_m - S(d)K)$ describes all polynomial matrices in $\mathbb{F}^{m\times m}$ having I_m as constant term and ith row degree not greater than ϕ_i, $i = 1, 2, \ldots, m$.

If the input of $\Sigma^{(K)}$ is filtered through an invertible static precompensator $M \in \mathbb{F}^{m\times m}$, the equations of the resulting state model become (Fig. 2)

$$\mathbf{x}_{t+1} = \mathbf{x}_t[A + KB] + \mathbf{u}_t MB$$
$$\mathbf{w}_t = \mathbf{x}_t[C + KJ] + \mathbf{u}_t MJ$$

and the transfer matrix of the resulting system $\Sigma^{(K,M)} = (A + KB, MB, C + KJ, MJ)$ has the following left MFD

$$\begin{aligned} G^{(K,M)}(d) &= [M^{-1} - Bd(I_n - dA)^{-1}KM^{-1}]^{-1}[J + Bd(I_n - dA)^{-1}C] \\ &= [M^{-1} - S(d)KM^{-1}]^{-1}G_c(d). \end{aligned}$$

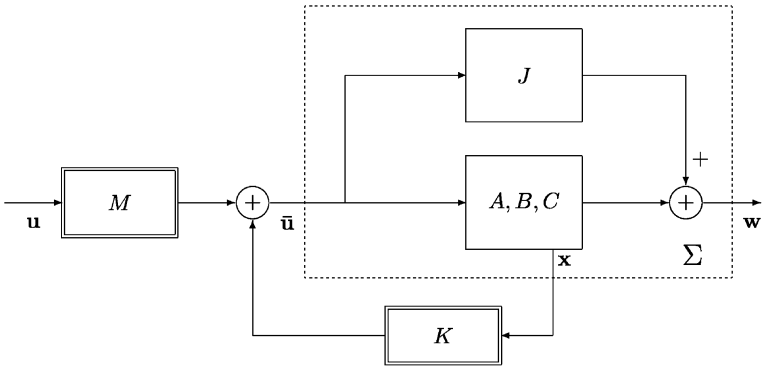

Fig. 2.

As each minimal encoder of $\mathscr{C}$ can be represented as $G(d) = D(d)^{-1}G_c(d)$, with $D(0)$ invertible and deg $\text{row}_i D \leqslant$ deg $\text{row}_i G_c$, $i = 1, \ldots, m$, it is possible to determine a unique precompensator $M = D(0)^{-1}$ and a unique state feedback matrix K such that $D(d) = M^{-1} - S(d)KM^{-1}$. We summarize the above discussion in the following proposition,

Proposition 8.1. *Let $G_c(d)$ be a canonical encoder of a $[p, m]$ convolutional code $\mathscr{C}$ of degree n. The set $\mathscr{M}$ of all minimal encoders of $\mathscr{C}$ is constituted by the transfer matrices of all systems $\Sigma^{(K,M)} = (A + KB, MB, C + KJ, MJ)$, obtained by application of static feedback and (nonsingular) precompensation to a minimal realization $\Sigma = (A, B, C, J)$ of $G_c(d)$. Therefore, the set of the pairs $(K, M) \in \mathbb{F}^{n\times m} \times \text{Gl}(m, \mathbb{F})$ biuniquely parametrizes $\mathscr{M}$.*

If the encoders are represented as MFDs in the indeterminate d^{-1}, minimal encoders of $\mathscr{C}$ are MFDs with the following structure:

$$\begin{aligned} G(d) &= \tilde{D}(d^{-1})^{-1}\tilde{N}(d^{-1}) \\ &:= \left[\text{diag}\{d^{-\phi_1}, \ldots, d^{-\phi_m}\}D(d)\right]^{-1}\left[\text{diag}\{d^{-\phi_1}, \ldots, d^{-\phi_m}\}G_c(d)\right], \end{aligned}$$

where $\tilde{D}(d^{-1})$ describes the set of all $m \times m$ row reduced polynomial matrices with row degrees $\phi_1, \ldots, \phi_m$, and $\tilde{N}(d^{-1})$ is a fixed left prime row reduced polynomial matrix in d^{-1}. Rosenbrock's theorem [11], quoted in Section 2, shows that the Smith forms of the denominator matrices $\tilde{D}(d^{-1})$ of minimal encoders comprise all strings of m monic polynomials $\psi_1(d^{-1}), \ldots, \psi_m(d^{-1})$ satisfying

$$\begin{aligned} &\psi_{i+1}|\psi_i \\ &\deg(\psi_1 \cdots \psi_t) \geqslant \phi_1 + \ldots + \phi_t \\ &\deg(\psi_1 \cdots \psi_m) = \phi_1 + \ldots + \phi_m = \deg \mathscr{C}. \end{aligned}$$

Note that the Smith form of $\tilde{D}(d^{-1})$ provides also the invariant polynomials—and in particular the minimal polynomial—of the matrix A in any minimal state space realization of $\tilde{D}(d^{-1})^{-1}\tilde{N}(d^{-1})$.

9. Syndrome formers

Every $[p, m]$-convolutional code $\mathscr{C}$ can be associated with the orthogonal (or *dual*) code [3] of dimension $p - m$,

$$\mathscr{C}_\perp := \left\{\hat{\mathbf{v}}_\perp(d) \in \mathbb{F}((d))^p : \hat{\mathbf{v}}_\perp(d)\hat{\mathbf{w}}(d)^{\mathrm{T}} = 0,\ \forall \hat{\mathbf{w}}(d) \in \mathscr{C}\right\}.$$

$\mathscr{C}_\perp$ admits a polynomial basis $\hat{\mathbf{g}}_{1\perp}(d), \ldots, \hat{\mathbf{g}}_{(p-m)\perp}(d) \in \mathbb{F}[d]^p$ and, by Proposition 3.2, it is $\bar{N}$-controllable for some $\bar{N} \in \mathbb{N}$.

It is easy to see that $\mathscr{C}_\perp$ uniquely determines $\mathscr{C}$. Actually, if $G_\perp(d) \in \mathbb{F}(d)^{(p-m)\times p}$ is any encoder of $\mathscr{C}_\perp$, then

$$\hat{\mathbf{w}}(d)G_\perp(d)^{\mathrm{T}} = 0 \Leftrightarrow \hat{\mathbf{w}}(d) \in \mathscr{C}.$$

The rational matrix $S(d) := G_\perp(d)^{\mathrm{T}} \in \mathbb{F}(d)^{p\times(p-m)}$ is called a syndrome former of $\mathscr{C}$, and

$$S(d)T(d)$$

provides all syndrome formers of $\mathscr{C}$ as $T(d)$ varies on the group of nonsingular $(p - m) \times (p - m)$ rational matrices. Once a syndrome former $S(d)$ has been selected, the syndrome of any $\hat{\mathbf{w}}(d) \in \mathbb{F}((d))^p$ is given by $\hat{\mathbf{s}}(d) := \hat{\mathbf{w}}(d)S(d)$, and $\hat{\mathbf{w}}(d)$ is in $\mathscr{C}$ if and only if its syndrome is zero.

As syndrome formers of $\mathscr{C}$ are exactly the transpose of the encoders of $\mathscr{C}_\perp$, we may expect that a discussion on syndrome formers structure could mirror that on the encoders of $\mathscr{C}$. A preliminary, fundamental connection between syndrome formers and basic encoders of $\mathscr{C}$ is provided by the following lemma. It depends on the argument that follows the generalized Bézout identity (8).

Lemma 9.1 [3]. *Suppose that $G_b(d) \in \mathbb{F}[d]^{m\times p}$ is a basic encoder of $\mathscr{C}$. Select $C(d)$ in $\mathbb{F}^{(p-m)\times p}[d]$ so that $\begin{bmatrix} G_b(d) \\ C(d) \end{bmatrix}$ is unimodular, and $D(d) \in \mathbb{F}[d]^{p\times m}$ and $S(d) \in \mathbb{F}[d]^{p\times(p-m)}$ so that*

$$\begin{bmatrix} G_b(d) \\ C(d) \end{bmatrix} [\, D(d) \mid S(d) \,] = I_p.$$

Then $S(d)$ is a basic (*i.e. right prime*) *syndrome former of $\mathscr{C}$, and its maximal order minors are equal, up to units, to the complementary maximal order minors of $G_b(d)$.*

The above lemma has several interesting consequences. First of all, the degree of $\mathscr{C}_\perp$ is equal to the degree of $\mathscr{C}$, and row degrees $\psi_1, \ldots, \psi_{p-m}$ of any canonical encoder of $\mathscr{C}_\perp$ satisfy

$$\sum_{i=1}^{p-m} \psi_i = \sum_{i=1}^{m} \phi_i = \deg \mathscr{C}.$$

The transpose of any canonical encoder $G_{c\perp}(d)$ of $\mathscr{C}_\perp$,

$$S_c(d) := G_{c\perp}(d)^{\mathrm{T}}$$

is a polynomial syndrome former of $\mathscr{C}$, right prime and column reduced with column degrees $\psi_1, \ldots, \psi_{p-m}$, that will be called *canonical*. Consequently, all *minimal syndrome formers* of $\mathscr{C}$ have McMillan degree $\sum_{i=1}^{p-m} \psi_i$, and are biuniquely parametrized by the MFDs $S_c(d)Q(d)^{-1}$, as $Q(d)$ sweeps all $(p-m)\times(p-m)$ polynomial matrices with $\deg \mathrm{col}_i(Q) \leqslant \deg \mathrm{col}_i(S_c)$, $i = 1, \ldots, p-m$, and $Q(0)$ invertible.

When considering the way of operating of a syndrome former $S(d)$, we may ask whether its finite support syndromes are all induced by sequences $\mathbf{v}$ that differ in a finite number of positions from some codeword $\mathbf{w}$ of $\mathscr{C}$. In other terms, is there any condition on $S(d)$ guaranteeing that finite support syndromes imply finite support errors?

This problem is quite similar to (non)catastrophic error generation, and the structural condition on the syndrome former is dual w.r.t the condition on (non) catastrophic encoders.

Proposition 9.2. *Let $P(d)Q(d)^{-1}$ be an irreducible right MFD of a causal syndrome former $S(d)$ of $\mathscr{C}$. The following are equivalent*:

(i) *for all $\hat{\mathbf{v}}(d)$ in $\mathbb{F}((d))^p$, if the syndrome $\hat{\mathbf{v}}(d)S(d)$ has finite support, then $\hat{\mathbf{v}}(d) - \hat{\mathbf{w}}(d)$ has finite support, for some codeword $\hat{\mathbf{w}}(d) \in \mathscr{C}$;*
(ii) *$P(d)$ factorizes into $P(d) = \bar{P}(d)\Delta(d)$, where $\bar{P}(d)$ is right prime and $\det \Delta(d) = \alpha d^k$, $0 \neq \alpha \in \mathbb{F}$, $k \in \mathbb{N}$.*

Proof. (ii) $\Rightarrow$ (i) Note that $\bar{P}(d)$ has a polynomial left inverse $L(d)$ and $\Delta(d)$ has a Laurent polynomial inverse. So, if $\hat{\mathbf{s}}(d) := \hat{\mathbf{v}}(d)S(d)$ has finite support, $\hat{\mathbf{s}}(d)Q(d)\Delta(d)^{-1}L(d)$ has finite support too, and

$$[\hat{\mathbf{v}}(d) - \hat{\mathbf{s}}(d)Q(d)\Delta(d)^{-1}L(d)]S(d) = 0$$

This implies that

$$\hat{\mathbf{w}}(d) := \hat{\mathbf{v}}(d) - \hat{\mathbf{s}}(d)Q(d)\Delta(d)^{-1}L(d) \tag{36}$$

is a codeword, and $\hat{\mathbf{v}}(d) - \hat{\mathbf{w}}(d)$ has finite support.

(i) $\Rightarrow$ (ii) Suppose that $P(d)$ factorizes into $P(d) = \bar{P}(d)M(d)$, with $\bar{P}(d)$ right prime and $M(d)$ nonsingular, with $\det M \neq \alpha d^k$.

The right MFD $M(d)Q(d)^{-1}$ is irreducible, as any right common factor of $M(d)$ and $Q(d)$ is also a right common factor of $P(d)$ and $Q(d)$. So, if $X(d)^{-1}Y(d)$ is an irreducible left MFD of $M(d)Q(d)^{-1}$, $\det Y = \det M$ implies that $\det Y \neq \alpha d^k$.

The expansion of $Y(d)^{-1}$ includes some series with infinite support. So, there exists $\mathbf{c} \in \mathbb{F}^m$ such that $\hat{\mathbf{q}}(d) := \mathbf{c}Y(d)^{-1}$ has infinite support and $\hat{\mathbf{q}}(d)Y(d)$ is polynomial. On the other hand,

$$\hat{\mathbf{b}}(d) := \hat{\mathbf{q}}(d)X(d)$$

has infinite support, otherwise $\hat{\mathbf{q}}(d)[X(d) \mid Y(d)]$ would have finite support, which is unconsistent with the left primeness of $[X(d) \mid Y(d)]$.

Consider now a polynomial left inverse $L(d)$ of $\bar{P}(d)$, and the infinite support signal

$$\hat{\mathbf{v}}(d) := \hat{\mathbf{b}}(d)L(d) \in \mathbb{F}((d))^p.$$

The corresponding syndrome is given by

$$\hat{\mathbf{s}}(d) = \hat{\mathbf{v}}(d)S(d) = \hat{\mathbf{b}}(d)L(d)\bar{P}(d)X(d)^{-1}Y(d) = \hat{\mathbf{b}}X(d)^{-1}Y(d) = \hat{\mathbf{q}}(d)Y(d)$$

and therefore has finite support.

Finally, suppose that $\hat{\mathbf{w}}(d)$ is any codeword of $\mathscr{C}$, and consider a basic encoder $G(d)$ of $\mathscr{C}$, with polynomial right inverse $C(d)$. Then

$$\begin{bmatrix} L(d) \\ G(d) \end{bmatrix} [\bar{P}(d) \mid C(d)] = \begin{bmatrix} I_{p-m} & * \\ 0 & I_m \end{bmatrix}$$

implies that $\begin{bmatrix} L(d) \\ G(d) \end{bmatrix}$ is unimodular, and the difference

$$\hat{\mathbf{v}}(d) - \hat{\mathbf{w}}(d) = \hat{\mathbf{b}}(d)L(d) - \hat{\mathbf{u}}(d)G(d) = [\hat{\mathbf{b}}(d) \mid -\hat{\mathbf{u}}(d)] \begin{bmatrix} L(d) \\ G(d) \end{bmatrix}$$

cannot be finite support, as $[\hat{\mathbf{b}}(d) \mid -\hat{\mathbf{u}}(d)]$ is not. □

Remark. In case $S(d)$ satisfies the conditions of Proposition 9.2, let $[s_{\min}, s_{\max}]$ be an interval that includes the support of $Q(d)\Delta(d)^{-1}L(d)$ and suppose that $\mathscr{C}$ is ℓ-observable. If the syndrome $\hat{\mathbf{v}}(d)S(d)$ remains zero on an interval $[a, b]$ larger than $\ell + s_{\max} - s_{\min}$,

- the restriction of $\hat{\mathbf{v}}(d)$ to $[a + s_{\max}, b + s_{\min}]$ is the restriction of a legal codeword;
- the restriction of $\hat{\mathbf{v}}(d) - [(\mathrm{id} - \mathscr{P}_a)\hat{\mathbf{s}}(d)]Q(d)\Delta(d)^{-1}L(d)$ to $[a + s_{\max}, +\infty]$ is the "tail" of a legal codeword.

Consequently, if the syndrome remains zero on a suitably large time interval $[a, b]$, a correction procedure (36) on the received message $\hat{\mathbf{v}}(d)$ can be implemented, that uses only syndrome symbols from time a onwards. The resulting sequence is compatible with any correction that could have been introduced before time a.

Upon applying arbitrary output injection and static output compensation to a minimal state space realization of a canonical syndrome former $S_c(d)$ of $\mathscr{C}$, we obtain all minimal syndrome formers of $\mathscr{C}$.

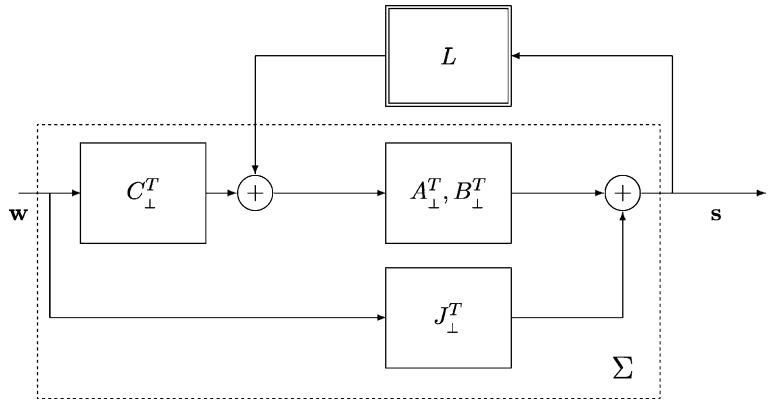

Fig. 3.

Actually, suppose that $\Sigma_\perp = (A_\perp, B_\perp, C_\perp, J_\perp)$ is a minimal realization of the canonical encoder $S_c(d)^{\mathrm{T}}$ of $\mathscr{C}_\perp$. Then the dual system $\Sigma = (A_\perp^{\mathrm{T}}, C_\perp^{\mathrm{T}}, B_\perp^{\mathrm{T}}, J_\perp^{\mathrm{T}})$

$$\begin{aligned}\bar{\mathbf{x}}_{t+1} &= \bar{\mathbf{x}}_t\, A_\perp^{\mathrm{T}} + \mathbf{w}_t\, C_\perp^{\mathrm{T}}\\ \mathbf{s}_t &= \bar{\mathbf{x}}_t\, B_\perp^{\mathrm{T}} + \mathbf{w}_t\, J_\perp^{\mathrm{T}}\end{aligned}$$

provides a minimal realization of $S_c(d)$. An output injection $\mathbf{s}_t L$, $L \in \mathbb{F}^{(p-m)\times n}$, modifies the above equations as follows (Fig. 3):

$$\begin{aligned}\bar{\mathbf{x}}_{t+1} &= \bar{\mathbf{x}}_t\, A_\perp^{\mathrm{T}} + \mathbf{w}_t\, C_\perp^{\mathrm{T}} + \mathbf{s}_t L\\ &= \bar{\mathbf{x}}_t (A_\perp^{\mathrm{T}} + B_\perp^{\mathrm{T}} L) + \mathbf{w}_t (C_\perp^{\mathrm{T}} + J_\perp^{\mathrm{T}} L)\\ \mathbf{s}_t &= \bar{\mathbf{x}}_t\, B_\perp^{\mathrm{T}} + \mathbf{w}_t\, J_\perp^{\mathrm{T}}\end{aligned}$$

and the transfer matrix of the resulting system $\Sigma^{(L)} = (A_\perp^{\mathrm{T}} + B_\perp^{\mathrm{T}} L,\ C_\perp^{\mathrm{T}} + J_\perp^{\mathrm{T}} L,\ B_\perp^{\mathrm{T}},\ J_\perp^{\mathrm{T}})$ is given by

$$\begin{aligned}S^{(L)}(d) &= [C_\perp^{\mathrm{T}} d(I_n - dA_\perp^{\mathrm{T}})^{-1} B_\perp^{\mathrm{T}} + J_\perp^{\mathrm{T}}][I_{p-m} - Ld(I_n - dA_\perp^{\mathrm{T}})^{-1} B_\perp^{\mathrm{T}}]^{-1}\\ &= S_c(d)[I_{p-m} - LX(d)]^{-1},\end{aligned}$$

where $X(d) := (I_n - dA_\perp^{\mathrm{T}})^{-1} B_\perp^{\mathrm{T}} d$. If the minimal realization of $G_{c\perp}(d)$ is obtained via the procedure of Section 5, we have that $X(d)^{\mathrm{T}}$ has the following structure:

$$X(d)^{\mathrm{T}} = \begin{bmatrix} d & d^2 & \dots & d^{\psi_1} & & & & & & & & \\ & & & & d & d^2 & \dots & d^{\psi_2} & & & & \\ & & & & & & & & \ddots & & & \\ & & & & & & & & & d & d^2 & \dots & d^{\psi_m} \end{bmatrix}$$

and, consequently, the matrix $I_{p-m} - LX(d)$ describes all $(p-m)\times(p-m)$ polynomial matrices with constant term I_{p-m} and ith -column degree not greater than ψ_i, $i = 1, \dots, p-m$, as L varies in $\mathbb{F}^{(p-m)\times n}$.

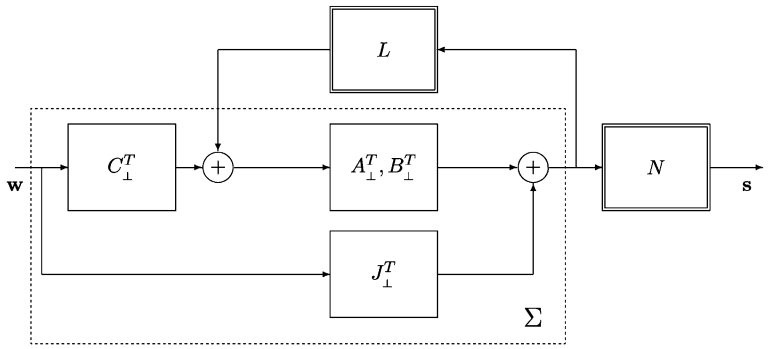

Fig. 4.

Finally, if the output of $\Sigma^{(L)}$ is filtered through an invertible nondynamical system $N \in \mathbb{F}^{(p-m)\times(p-m)}$, we end up with a state space model $\Sigma^{(L,N)} = (A_\perp^{\mathrm{T}} + B_\perp^{\mathrm{T}} L,\ C_\perp^{\mathrm{T}} + J_\perp^{\mathrm{T}} L,\ B_\perp^{\mathrm{T}} N,\ J_\perp^{\mathrm{T}} N)$ of a new syndrome former, with equations (Fig. 4)

$$\begin{aligned} \bar{\mathbf{x}}_{t+1} &= \bar{\mathbf{x}}_t (A_\perp^{\mathrm{T}} + B_\perp^{\mathrm{T}} L) + \mathbf{w}_t (C_\perp^{\mathrm{T}} + J_\perp^{\mathrm{T}} L) \\ \mathbf{s}_t &= \bar{\mathbf{x}}_t\, B_\perp^{\mathrm{T}} N + \mathbf{w}_t\, J_\perp^{\mathrm{T}} N. \end{aligned}$$

and transfer matrix

$$\begin{aligned} S^{(L,N)}(d) &= [C_\perp^{\mathrm{T}} d(I_n - dA_\perp^{\mathrm{T}})^{-1} B_\perp^{\mathrm{T}} + J_\perp^{\mathrm{T}}] \\ &\quad \times [N^{-1} - N^{-1} L d (I_n - dA_\perp^{\mathrm{T}})^{-1} B_\perp^{\mathrm{T}}]^{-1} \\ &= S_c(d)[N^{-1} - N^{-1} L X(d)]^{-1}. \end{aligned} \tag{37}$$

Varying N in $Gl(p-m, \mathbb{F})$ and L in $\mathbb{F}^{(p-m)\times n}$, the denominator matrices $N^{-1} - N^{-1}LX(d)$ in (37) biuniquely represent all $(p-m)\times(p-m)$ matrices $Q(d)$ with invertible constant term $Q(0)$ and column degrees not greater than the corresponding ones in $S_c(d)$. Hence (37) provides all minimal syndrome formers of $\mathscr{C}$.

10. Decoupled encoders and code decomposition

Consider a $[p, m]$-convolutional code $\mathscr{C}$, and let $p_1, \ldots, p_k$ be nonzero integers such that $\sum_{i=1}^k p_i = p$. An encoder $G(d)$ of the code is $(p_1, \ldots, p_k)$-*decoupled* if there exist positive integers $m_1, \ldots, m_k$ with $\sum_{i=1}^k m_i = m$ such that, possibly up to a column permutation,

$$G(d) = \mathrm{diag}\{G_1(d), \ldots, G_k(d)\}, \quad G_i(d) \in \mathbb{F}(d)^{m_i \times p_i}, \ \ i = 1, \ldots, k.$$

Upon partitioning an information sequence $\hat{\mathbf{u}}(d) \in \mathbb{F}((d))^m$ into $[\hat{\mathbf{u}}_1(d) \cdots \hat{\mathbf{u}}_k(d)]$, $\hat{\mathbf{u}}_i(d) \in \mathbb{F}((d))^{m_i}$, we have

$$\hat{\mathbf{u}}(d)G(d) = [\hat{\mathbf{w}}_1(d) \cdots \hat{\mathbf{w}}_k(d)], \quad \hat{\mathbf{w}}_i(d) = \hat{\mathbf{u}}_i(d)G_i(d), \ \ i = 1, \ldots, k,$$

and therefore

$$\mathscr{C} = \mathscr{C}_1 \oplus \cdots \oplus \mathscr{C}_k, \tag{38}$$

where $\mathscr{C}_i$ is the $[p_i, m_i]$-convolutional code generated by $G_i(d)$. As a consequence, the existence of a decoupled encoder for a $[p, m]$-convolutional code $\mathscr{C}$ is equivalent to the possibility of representing $\mathscr{C}$ as an "external" direct sum of $k \geqslant 2$ smaller $[p_i, m_i]$-convolutional codes. This in particular implies that in all codewords of $\mathscr{C}$ the values taken by a suitable set of p_i components are completely independent of the values of the remaining set of $p - p_i$ components, and no cross information from either set can be retrieved when implementing an error correcting procedure.

The purpose of this section is to investigate the existence and the structure of the decoupled encoders of $\mathscr{C}$ and, in particular, of the minimal ones, and to develop appropriate algorithms to compute direct summands appearing in (38).

For sake of simplicity, we shall assume that all columns of the encoders are different from zero (if not, we just consider codewords with a smaller number of components).

As any encoder of $\mathscr{C}$ is full rank, its columns constitute a generator set of $\mathbb{F}((d))^m$. The determination of decoupled encoders of $\mathscr{C}$ is straightly connected with the partition of the columns of its encoders into sets $\mathscr{G}_1, \ldots, \mathscr{G}_k$ such that

$$\mathbb{F}((d))^m = \operatorname{span} \mathscr{G}_1 \oplus \cdots \oplus \operatorname{span} \mathscr{G}_k.$$

Definition 10.1. A set of nonzero generators of $\mathbb{F}((d))^m$, $\mathscr{G} = \{\hat{\mathbf{v}}_1(d), \hat{\mathbf{v}}_2(d), \ldots, \hat{\mathbf{v}}_p(d)\}$ and a decomposition of $\mathbb{F}((d))^m$ in direct sum

$$\mathbb{F}((d))^m = V_1 \oplus V_2 \oplus \cdots \oplus V_k, \tag{39}$$

are compatible if every vector of $\mathscr{G}$ belongs to a summand of (39) (and, obviously, to only one).

If a generator set $\mathscr{G}$ is compatible with (39), it is clear that

(i) $\mathscr{G}_i := V_i \cap \mathscr{G}$, $i = 1, \ldots, k$, provide a partition of $\mathscr{G}$

$$\mathscr{G} = \mathscr{G}_1 \dot{\cup} \mathscr{G}_2 \dot{\cup} \cdots \dot{\cup} \mathscr{G}_k$$

and $V_i = \operatorname{span}(\mathscr{G}_i)$, $i = 1, \ldots, k$.

(ii) if $\mathscr{B} := \{\hat{\mathbf{v}}_{i_1}(d), \ldots, \hat{\mathbf{v}}_{i_m}(d)\} \subset \mathscr{G}$ is a basis of $\mathbb{F}((d))^m$, the vectors of $\mathscr{G}_i$ are linearly dependent on $\mathscr{B}_i := \mathscr{G}_i \cap \mathscr{B}$.

(iii) there exists a unique finest direct sum decomposition

$$\mathbb{F}((d))^m = \bar{V}_1 \oplus \bar{V}_2 \oplus \cdots \oplus \bar{V}_h \tag{40}$$

compatible with $\mathscr{G}$. Each summand of any other compatible decomposition of $\mathbb{F}((d))^m$ can be expressed as a suitable sum of some $\bar{V}_i$s in (40).

In order to obtain the partition of $\mathscr{G} = \{\hat{\mathbf{v}}_1(d), \ldots, \hat{\mathbf{v}}_p(d)\}$ associated with the finest decomposition (40), we select a basis $\mathscr{B} \subset \mathscr{G}$ and introduce on $\mathscr{G}$ an equivalence relation as follows.

For $\nu = 1, \ldots, p$, denote by $\mathscr{M}_\nu$ the smallest subset of $\mathscr{B}$ such that $\hat{\mathbf{v}}_\nu(d) \in \operatorname{span} \mathscr{M}_\nu$ and let $\hat{\mathbf{v}}_i(d) \sim \hat{\mathbf{v}}_j(d)$ if there exists a chain $\mathscr{M}_i = \mathscr{M}_{\nu_1}, \mathscr{M}_{\nu_2}, \ldots, \mathscr{M}_{\nu_q} = \mathscr{M}_j$ such that $\mathscr{M}_{\nu_\ell} \cap \mathscr{M}_{\nu_{\ell+1}} \neq \emptyset,\ \ell = 1, \ldots, q-1$. It is easy to check that $\hat{\mathbf{v}}_i(d) \sim \hat{\mathbf{v}}_j(d)$ if and only if $\hat{\mathbf{v}}_i(d)$ and $\hat{\mathbf{v}}_j(d)$ belong to the same subspace in the finest direct sum decomposition (40) compatible with $\mathscr{G}$. From a computational point of view, we arrive at decomposition (40) through the following steps:

Step 1: Select an $m \times m$ nonsingular submatrix $B(d)$ of $[\hat{\mathbf{v}}_1(d) \cdots \hat{\mathbf{v}}_p(d)]$ and put

$$V(d) = B(d)^{-1}[\hat{\mathbf{v}}_1(d) \cdots \hat{\mathbf{v}}_p(d)].$$

Step 2: Construct the $m \times p$ Boolean matrix A defined by

$$A_{ij} = \begin{cases} 1 & \text{if } V_{ij} \neq 0, \\ 0 & \text{if } V_{ij} = 0. \end{cases}$$

Step 3: Compute $(A^{\mathrm{T}}A)^{p-1}$ and determine a permutation matrix $P \in \mathbb{F}^{p\times p}$ such that

$$P^{\mathrm{T}}(A^{\mathrm{T}}A)^{p-1}P = \operatorname{diag}\{N_1, \ldots, N_h\},$$

where $N_i = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} [\,1 \quad \cdots \quad 1\,] \in \mathbb{F}^{p_i \times p_i},\ i = 1, \ldots, h.$

Step 4: Partitionate $[\hat{\mathbf{v}}_1(d) \cdots \hat{\mathbf{v}}_p(d)]P$ into

$$[L_1(d)|\cdots|L_h(d)],\ L_i(d) \in \mathbb{F}((d))^{m\times p_i},\ i = 1, \ldots, h,$$

and denote by $\mathscr{G}_i$ the subset of $\mathscr{G}$ whose vectors are the columns of $L_i(d)$, $i = 1, \ldots, h$.

Proposition 10.2. *Let $\mathscr{G} = \{\hat{\mathbf{v}}_1(d), \ldots, \hat{\mathbf{v}}_p(d)\}$ be a set of nonzero generators of $\mathbb{F}((d))^m$. The above algorithm provides the partition of $\mathscr{G}$ associated with the finest compatible decomposition of $\mathbb{F}((d))^m$.*

Proof. We prove first that

$$\hat{\mathbf{v}}_i(d) \sim \hat{\mathbf{v}}_j(d) \iff (A^{\mathrm{T}}A)_{ij}^{p-1} = 1. \tag{41}$$

Observe that $A_{ij} = 1 \iff \hat{\mathbf{v}}_i(d) \in \mathscr{M}_j$.

On the other hand, as $(A^{\mathrm{T}}A)_{ij} = 1$ if and only if there exists $s \in \{1, \ldots, p\}$ such that $A_{si} = A_{sj} = 1$, we have

$$\begin{aligned}(A^{\mathrm{T}}A)_{ij} = 1 &\iff \exists\, \hat{\mathbf{v}}_s(d) \in \mathscr{G} :\ \hat{\mathbf{v}}_s(d) \in \mathscr{M}_i \cap \mathscr{M}_j \\ &\iff \mathscr{M}_i \cap \mathscr{M}_j \neq \emptyset,\end{aligned}$$

and, more generally, for all $n \in \mathbb{N}$

$$\begin{aligned}&(A^{\mathrm{T}}A)^n_{ij} = 1\\ &\quad\Longleftrightarrow\ \exists \nu_2, \dots, \nu_n : (A^{\mathrm{T}}A)_{i\nu_2} = (A^{\mathrm{T}}A)_{\nu_2\nu_3} = \cdots = (A^{\mathrm{T}}A)_{\nu_n j} = 1\\ &\quad\Longleftrightarrow\ \exists \nu_1 = i, \nu_2, \dots, \nu_n, \nu_{n+1} = j : \mathscr{M}_{\nu_\ell} \cap \mathscr{M}_{\nu_{\ell+1}} \neq \emptyset,\ \ell = 1, \dots, n.\end{aligned}$$

Consequently,

$$\hat{\mathbf{v}}_i(d) \sim \hat{\mathbf{v}}_j(d) \iff (A^{\mathrm{T}}A)^k_{ij} = 1,\ \exists k. \tag{42}$$

Since $(A^{\mathrm{T}}A)_{ii} = 1,\ i = 1, \dots, p$, we have also

$$(A^{\mathrm{T}}A)^n_{ij} = 1 \implies (A^{\mathrm{T}}A)^{n+1}_{ij} = 1,\ \forall n \in \mathbb{N},\ \forall i, j. \tag{43}$$

On the other hand

$$(A^{\mathrm{T}}A)^n_{ij} = 1 \Longrightarrow (A^{\mathrm{T}}A)^{n-1}_{ij} = 1,\ \forall\, i, j \in \{1, \dots, p\},\ \forall n \geqslant p. \tag{44}$$

In fact, if $(A^{\mathrm{T}}A)^n_{ij} = 1$, there exist $\mathscr{M}_i = \mathscr{M}_{\nu_1}, \mathscr{M}_{\nu_2}, \dots, \mathscr{M}_{\nu_{n+1}} = \mathscr{M}_j$ with $\mathscr{M}_{\nu_\ell} \cap \mathscr{M}_{\nu_{\ell+1}} \neq \emptyset$, $\ell = 1, \dots, n$. As $|\mathscr{G}| = p$, there exist $k_1 < k_2$ such that $\nu_{k_1} = \nu_{k_2}$, and $\mathscr{M}_i = \mathscr{M}_{\nu_1}, \mathscr{M}_{\nu_2}, \dots, \mathscr{M}_{\nu_{k_1}} = \mathscr{M}_{\nu_{k_2}}, \dots, \mathscr{M}_{\nu_{n+1}} = \mathscr{M}_j$ satisfies $\mathscr{M}_{\nu_\ell} \cap \mathscr{M}_{\nu_{\ell+1}} \neq \emptyset,\ \ell = 1, \dots, k_1 - 1,\ \ell = k_2, \dots, n$. This, together with (43) imply $(A^{\mathrm{T}}A)^{n-1}_{ij} = 1$. (41) follows immediately from (42) and (44).

It is clear now that a permutation matrix $P \in \mathbb{F}^{p\times p}$ sorts the columns of $[\hat{\mathbf{v}}_1(d) \cdots \hat{\mathbf{v}}_p(d)]$ according to the equivalence classes of $\sim$ if and only if

$$P^{\mathrm{T}}(A^{\mathrm{T}}A)^{p-1}P = \operatorname{diag}\{N_1, \dots, N_h\},$$

and the equivalence classes of $\sim$ are constituted by the columns of $L_i(d) \in \mathbb{F}(d)^{m\times p_i}$, $i = 1, \dots, h$, in

$$[L_1(d) \mid \cdots \mid L_h(d)] = [\hat{\mathbf{v}}_1(d) \cdots \hat{\mathbf{v}}_p(d)]P. \qquad \square$$

The partition of the columns of an encoder of $\mathscr{C}$ associated with the finest decomposition (40) of $\mathbb{F}((d))^m$, does not depend on the particular encoder and therefore is a code property. In fact, let $G(d)$ and $\tilde{G}(d)$ be two encoders of $\mathscr{C}$, $P \in \mathbb{F}^{p\times p}$ a permutation matrix, and consider the column partitions

$$G(d)P = [G_1(d)|\cdots|G_h(d)], \quad G_i(d) \in \mathbb{F}(d)^{m\times p_i},\ i = 1, \dots, h,$$

$$\tilde{G}(d)P = [\tilde{G}_1(d)|\cdots|\tilde{G}_h(d)], \quad \tilde{G}_i(d) \in \mathbb{F}(d)^{m\times p_i},\ i = 1, \dots, h.$$

As

$$\tilde{G}(d) = T(d)G(d)$$

for some nonsingular matrix $T(d) \in \mathbb{F}(d)^{m\times m}$, it follows that rank $G_i(d) = \operatorname{rank} \tilde{G}_i(d)$, $i = 1, \dots, h$, and

$$\mathbb{F}((d))^m = \operatorname{span} G_1(d) \oplus \cdots \oplus \operatorname{span} G_h(d)$$

if and only if

$$\mathbb{F}((d))^m = \text{span } \tilde{G}_1(d) \oplus \cdots \oplus \text{span } \tilde{G}_h(d).$$

Therefore, equivalent encoders of $\mathscr{C}$ exhibit the same column partitions, compatible with the finest sum decomposition of $\mathbb{F}((d))^m$.

Remark. As Step 1 in the above algorithm produces a systematic encoder, in order to find a column partition associated with (40) we can always assume that the encoder is systematic, and apply the algorithm, starting on Step 2.

Keeping in the spirit of the previous sections of the paper, we provide now a parametrization of all minimal decoupled encoders of $\mathscr{C}$. To that purpose we construct first a canonical decoupled one, starting from a canonical encoder $G_c(d)$, and considering the partition $G_c(d)P = [G_1(d)|\cdots|G_h(d)]$, $G_i(d) \in \mathbb{F}[d]^{m\times p_i}$, with rank $G_i(d) = m_i$, $i = 1,\ldots,h$, compatible with the finest sum decomposition of $\mathbb{F}((d))^m$. We select an $m_i \times p_i$ full rank submatrix of $G_i(d)$, $\tilde{G}_i(d)$, $i = 1,\ldots,h$, and factorize it into

$$\tilde{G}_i(d) = M_i(d)\bar{G}_i(d),$$

where $\bar{G}_i(d) \in \mathbb{F}[d]^{m_i\times p_i}$ is left prime, and $M_i(d) \in \mathbb{F}[d]^{m_i\times m_i}$ is a left maximal divisor of $\tilde{G}_i(d)$.

If $\hat{\mathbf{r}}(d) \in \mathbb{F}[d]^{1\times p_i}$ is any row of $G_i(d)$, there exists a rational row vector $\hat{\mathbf{x}}(d)$ such that $\hat{\mathbf{r}}(d) = \hat{\mathbf{x}}(d)\bar{G}_i(d)$ and therefore $\hat{\mathbf{r}}(d)\bar{G}_i(d)^{-1} = \hat{\mathbf{x}}(d)$. As $\bar{G}_i(d)^{-1}$ is polynomial and right prime, $\hat{\mathbf{x}}(d)$ is polynomial too. Consequently,

$$G_i(d) = X_i(d)\bar{G}_i(d), \quad X_i(d) \in \mathbb{F}[d]^{m\times m_i},$$

and we have

$$G_c(d)P = [X_1(d)|\cdots|X_h(d)]\text{diag}\{\bar{G}_1(d),\ldots,\bar{G}_h(d)\}.$$

As $\bar{G}_i(d)$, $i = 1,\ldots,h$, are left prime, so is $\text{diag}\{\bar{G}_1(d),\ldots,\bar{G}_h(d)\}$, which implies, in particular, that $[X_1(d)|\cdots|X_h(d)]$ is unimodular.

For a suitable choice of $X_i(d)$, the submatrices $\bar{G}_i(d)$, $i = 1,\ldots,h$, and therefore also $\text{diag}\{\bar{G}_1(d),\ldots,\bar{G}_h(d)\}$, are row reduced. Thus, $\text{diag}\{\bar{G}_1(d),\ldots,\bar{G}_h(d)\} = [X_1(d)|\cdots|X_h(d)]^{-1}G_c(d)P$ is a canonical decoupled encoder of $\mathscr{C}$.

Any other minimal decoupled encoder realizing the finest decomposition of $\mathscr{C}$ is given by

$$\begin{bmatrix} D_1(d) & & \\ & \ddots & \\ & & D_h(d) \end{bmatrix}^{-1} [X_1(d)|\cdots|X_h(d)]^{-1}G_c(d)P$$
$$= [X_1(d)D_1(d)|\cdots|X_h(d)D_h(d)]^{-1}G_c(d)P,$$

where $D_i(d) \in \mathbb{F}[d]^{m_i\times m_i}$ is an invertible polynomial matrix, whose row degrees do not exceed the corresponding row degrees in $\bar{G}_i(d)$ and $D_i(0)$ is nonsingular, $i = 1,\ldots,h$.

11. Concluding remarks

In this paper, several applications of MFDs techniques to analysis, realization and parametrization of encoders and syndrome formers of convolutional codes over an arbitrary field have been discussed.

Related to the basic issue of parametrizing all minimal encoders and syndrome formers of a code $\mathscr{C}$, some problems naturally arise, that could provide a first natural avenue for future investigations on the structural side. We mention here the performance evaluation of different $G(d)$ and $S(d)$, one obtains by varying matrices K, M, L and N, and, in particular, of encoders and syndrome formers having block diagonal or block triangular form, and therefore exhibiting some degree of decoupling between inputs and outputs.

A different investigation perspective, that somehow exhibits a stronger coding theoretic flavor, leads to asking whether the above results can eventually provide good codes and/or efficient decoding algorithms.

Perhaps more interesting, but definitely more difficult, is the extension of the above point of view to multidimensional coding theory [23,24], possibly applying the recent framework of codes defined on graphs [25,26]. As a matter of fact, the results of this paper are based on effective algorithms for polynomial matrices in one indeterminate, that only partially hold in a more general setting. In particular, a multidimensional counterpart of the minimality characterization via McMillan degree, considered in Section 5, and the subsequent parametrization procedure, are unavailable yet. Perhaps a different concept of minimality should be devised, but it seems there is still a long way to go along, despite several efforts spent by many researchers in the last few years.

References

[1] G.D. Forney Jr., Algebraic Structure of Convolutional Codes, and Algebraic System Theory, Springer-Verlag, Berlin-Heidelberg, 1991, pp. 527–557.
[2] G.D. Forney Jr., Convolutional codes I: algebraic structure, IEEE Trans. Inform. Theory 16 (1970) 720–738, Correction, Ibid., 17 (1971) 360.
[3] G.D. Forney Jr., Structural analysis of convolutional codes via dual codes, IEEE Trans. Inform. Theory 19 (1973) 512–518.
[4] G.D. Forney Jr., Minimal bases of rational vector spaces, with applications to multivariable systems, SIAM J.Control 13 (3) (1975) 493–520.
[5] R. Johannesson, K.S. Zigangirov, Fundamentals of convolutional coding, IEEE Press Series in Digital and Mobile Comm. (1999).
[6] R.J. McEliece, The Algebraic Theory of Convolutional Codes, in: V.S.Pless, W.C.Huffman, R.A.Brualdi (Eds.), Handbook of Coding Theory, vol. 1, North-Holland, Amsterdam, 1998.
[7] P. Piret, Convolutional Codes: An Algebraic Approach, MIT Press, Cambridge, MA, 1988.
[8] J.C. Willems, Models for dynamics, Dynamics Rep. 2 (1988) 171–269.
[9] J.C. Willems, Paradigms and puzzles in the theory of dynamical systems, IEEE Trans. Automat. Control 36 (1991) 259–294.
[10] T. Kailath, Linear Systems, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[11] H.H. Rosenbrock, State Space and Multivariable Theory, John Wiley, New York, 1970.
[12] S. MacLane, G. Birkhoff, Algebra, MacMillan, London, 1967.
[13] F.R. Gantmacher, The Theory of Matrices, vol. I, Chelsea, New York, 1959.
[14] M. Vidyasagar, Control System Synthesis: A Factorization Approach, MIT Press, 1985.
[15] H.A. Loeliger, T. Mittelholzer, Convolutional codes over groups, IEEE Trans. Inform. Theory 42 (1996) 1660–1686.
[16] J. Rosenthal, Connections between linear systems and convolutional codes, in: J.R.B. Marcus (Ed.), Codes, Systems and Graphical Models, Springer-Verlag, New York, 2001, pp. 39–66.
[17] P. Elias, Coding for Noisy Channels, IRE Conv. Record, part 4, pp. 37–47, 1955.
[18] G.D. Forney Jr., R. Johannesson, Z. Wan, Minimal and canonical rational generator matrices for convolutional codes, IEEE Trans. Inform. Theory 42 (6) (1996) 1865–1880.
[19] J.L. Massey, M.K. Sain, Codes, automata and continuous systems: explicit interconnections, IEEE Trans. Autom. Control 12 (6) (1967) 644–650.
[20] J. Rosenthal, J.M. Schumacher, E.V. York, On behaviors and convolutional codes, IEEE Trans. Inform. Theory part 1 42 (6) (1996) 1881–1891.
[21] J. Rosenthal, J.M. Schumacher, Realization by inspection, IEEE Trans. Autom. Control 42 (9) (1997) 1257–1263.
[22] R.E. Kalman, P.L. Falb, M.A. Arbib, Topics in mathematical system theory, McGraw-Hill, New York, 1969.
[23] E. Fornasini, M.E. Valcher, Algebraic aspects of 2d convolutional codes, IEEE Trans. Inform. Theory 40 (1994) 1068–1082.
[24] P. Weiner, Multidimensional convolutional codes, Ph.D. dissertation, University of Notre Dame, 1998.
[25] F.K. Kschischang, B.J. Frey, H.A. Loeliger, Factor graphs and the sum–product algorithm, IEEE Trans. Inform. Theory 47 (2) (2001) 498–519.
[26] G.D. Forney Jr., Codes on graphs: normal realizations, IEEE Trans. Inform. Theory 47 (2) (2001) 520–548.