on minimality of ISO representation of basic 2D convolutional codes

rita simões
raquel pinto
CIDMA
CIDMA, university of aveiro

5ICMCTA
28-31 August 2017
International Castle Meeting on Coding Theory and Applications

vihula manor

outline of the talk

- motivation
- polynomial matrices
- 1D convolutional codes
- 2D convolutional codes
- minimality of ISO representations of basic 2D convolutional codes
- conclusions and future work

motivation

- 2D convolutional code are naturally suitable to deal with data recorded in two dimensions.

- one of the most important problems studied in the theory of convolutional codes is the minimality of the representations of these codes.

- unlike the 1D case, it does not exist necessary and sufficient conditions for the minimality of a realization of 2D convolutional codes.

- let $\mathbb{F}$ be a finite field and let $\overline{\mathbb{F}}$ denote the algebraic closure of $\mathbb{F}$; denote by $\mathbb{F}[z]$ the ring of polynomials in one variable with coefficients in the $\mathbb{F}$ and $\mathbb{F}(z)$ the field of fractions of $\mathbb{F}[z]$. let $U(z) \in \mathbb{F}[z]^{n\times k}$, with $n \geq k$:

definition: $U(z)$ is unimodular if $n = k$ and $\det(U(z)) \in \mathbb{F}\backslash\{0\}$;

proposition: the following are equivalent:

(a) $U(z)$ is right prime (rP);

(b) $\forall \hat{v}(z) \in \mathbb{F}(z)^k$, $U(z)\hat{v}(z) \in \mathbb{F}[z]^n \Rightarrow \hat{v}(z) \in \mathbb{F}[z]^k$;

(c) the $k \times k$ minors of $U(z)$ have no common factor;

(d) $U(z)$ admits a polynomial left inverse;

(e) $U(\lambda)$ is full column rank, for all $\lambda \in \overline{\mathbb{F}}$.

- a matrix is left prime (ℓP) if its transpose is rP.

- denote by $\mathbb{F}[z_1, z_2]$ the ring of polynomials in two variables, z_1 and z_2, with coefficients in the $\mathbb{F}$ and $\mathbb{F}(z_1, z_2)$ the field of fractions of $\mathbb{F}[z_1, z_2]$.

let $G(z_1, z_2) \in \mathbb{F}[z_1, z_2]^{n \times k}$, with $n \geq k$:

definition: $G(z_1, z_2)$ is **unimodular** if $n = k$ and $\det(G(z_1, z_2)) \in \mathbb{F} \backslash \{0\}$;

proposition: the following are equivalent:

(a) $G(z_1, z_2)$ is **right factor prime** (rFP);

(b) $\forall \hat{u}(z_1, z_2) \in \mathbb{F}(z_1, z_2)^k$,

$$G(z_1, z_2)\hat{u}(z_1, z_2) \in \mathbb{F}[z_1, z_2]^n \Rightarrow \hat{u}(z_1, z_2) \in \mathbb{F}[z_1, z_2]^k;$$

(c) the $k \times k$ minors of $G(z_1, z_2)$ have no common factor;

- a matrix is **left factor prime** (ℓFP) if its transpose is rFP.

polynomial matrices in two indeterminates

- let $G(z_1, z_2) \in \mathbb{F}[z_1, z_2]^{n\times k}$, with $n \geq k$:

proposition: the following are equivalent:

(a) $G(z_1, z_2)$ is right zero prime (rZP);

(b) $G(z_1, z_2)$ admits a polynomial left inverse;

(c) $G(\lambda_1, \lambda_2)$ is full column rank, for all $\lambda_1, \lambda_2 \in \overline{\mathbb{F}}$.

- a matrix is left zero prime (ℓZP) if its transpose is rZP.

note: zero primeness implies factor primeness.

definition: a 1D (finite support) convolutional code $\mathcal{C}$ of rate k/n is a (free) $\mathbb{F}[z]-$submodule of $\mathbb{F}^n[z]$ with rank k; the elements of $\mathcal{C}$ are called codewords.

- a full column rank matrix $G(z) \in \mathbb{F}[z]^{n\times k}$ is an encoder of $\mathcal{C}$ if

$$\mathcal{C} = \left\{\widehat{v}(z) \in \mathbb{F}[z]^n \mid \widehat{v}(z) = G(z)\widehat{u}(z), \text{ with } \widehat{u}(z) \in \mathbb{F}[z]^k\right\}$$

- the complexity δ of $\mathcal{C}$ is the maximum of the degree of the $k \times k$ minors of any encoder of $\mathcal{C}$; we say that $\mathcal{C}$ is an (n, k, δ) convolutional code.

- two encoders are equivalent if they generate the same code

lemma: if $\mathcal{C}$ admits a rP encoder then all its encoders are rP;

definition: a 1D convolutional code that admits rP encoders is called basic (or noncatastrophic).

- a **1D linear system** $\Sigma = (A, B, C, D)$ is given by the updating equations

$$x(t+1) = Ax(t) + Bu(t)$$

$$y(t) = Cx(t) + Du(t)$$

where $A \in \mathbb{F}^{s\times s}$, $B \in \mathbb{F}^{s\times k}$, $C \in \mathbb{F}^{(n-k)\times s}$, $D \in \mathbb{F}^{(n-k)\times k}$, $s, n, k \in \mathbb{N}$, $n > k$ and $x(0) = 0$; we say that Σ has dimension s.

state	input	output
$\{x(t)\}_{t\in\mathbb{N}}$	$\{u(t)\}_{t\in\mathbb{N}}$	$\{y(t)\}_{(t\in\mathbb{N}}$
$\downarrow$	$\downarrow$	$\downarrow$
$\hat{x}(z) = \sum_{t\in\mathbb{N}} x(t)z^t$	$\hat{u}(z) = \sum_{t\in\mathbb{N}} u(t)z^t$	$\hat{y}(z) = \sum_{t\in\mathbb{N}} y(t)z^t$

- define $\hat{v}(z) = \begin{bmatrix} \hat{u}(z) & \hat{y}(z) \end{bmatrix}^T \in \mathbb{F}[z]^n$ to be the code vector.

definition: the finite support input-output trajectories $(\hat{u}(z), \hat{y}(z))$ of Σ with corresponding state $\hat{x}(z)$ also having finite support are called finite-weight input-output trajectories of Σ.

theorem [Rosenthal&York,1999] the set of finite-weight input-output trajectories of $\Sigma = (A, B, C, D)$ is a 1D convolutional code of rate k/n, denoted by $\mathcal{C}(A, B, C, D)$.

- Σ is called an input-state-output (ISO) representation of $\mathcal{C}(A, B, C, D)$

note: all the 1D convolutional codes admit (many) ISO representations definition: let $\Sigma = (A, B, C, D)$ be a 1D linear system with dimension s; Σ is reachable if the matrix $\begin{bmatrix} I_s - Az & Bz \end{bmatrix}$ is ℓP.

definition: an ISO representation of a 1D convolutional code is minimal if it has minimal dimension among all the ISO representations of the code.

theorem [Rosenthal&York&Schumacher,1996] let Σ be an ISO representation of an (n, k, δ) convolutional code $\mathcal{C}$; then Σ is a minimal ISO representation of $\mathcal{C}$ iff Σ is reachable.

- moreover, a minimal ISO representation $\Sigma = (A, B, C, D)$ of $\mathcal{C}$ has dimension δ and any other minimal ISO representation of $\mathcal{C}$ is of the form $\tilde{\Sigma} = \left(SAS^{-1}, SB, CS^{-1}, D\right)$, where S is a $\delta \times \delta$ invertible constant matrix.

definition: a **2D (finite support) convolutional code $\mathcal{C}$ of rate k/n** is a free $\mathbb{F}[z_1, z_2]-$submodule of $\mathbb{F}^n[z_1, z_2]$ with rank k; the elements of $\mathcal{C}$ are called **codewords**.

- a full column rank matrix $\widehat{G}(z_1, z_2) \in \mathbb{F}[z_1, z_2]^{n\times k}$ is an **encoder of $\mathcal{C}$** if

$$\mathcal{C} = \left\{ \widehat{v}(z_1, z_2) \in \mathbb{F}[z_1, z_2]^n \mid \widehat{v}(z_1, z_2) = \widehat{G}(z_1, z_2)\widehat{u}(z_1, z_2) \right.$$
$$\left. \text{with } \widehat{u}(z_1, z_2) \in \mathbb{F}[z_1, z_2]^k \right\}$$

- the **complexity** δ of $\mathcal{C}$ is the maximum of the degree of the $k \times k$ minors of any encoder of $\mathcal{C}$

- if $\mathcal{C}$ admits a rZP (rFP) encoder then all its encoders are rZP (rFP)

definition: a 2D convolutional code that admits rZP (respectively, rFP) encoders is called **basic** (respectively, **noncatastrophic**).

- a (first quarter plane) 2D linear system $\Sigma = (A_1, A_2, B_1, B_2, C, D)$, with dimension s, is given by the updating equations

$$\begin{aligned} x(i+1, j+1) &= A_1 x(i, j+1) + A_2 x(i+1, j) + \\ &\quad + B_1 u(i, j+1) \; + \; B_2 u(i+1, j) \\ y(i,j) &= Cx(i,j) + Du(i,j), \end{aligned}$$

where $A_1, A_2 \in \mathbb{F}^{s\times s}$, $B_1, B_2 \in \mathbb{F}^{s\times k}$, $C \in \mathbb{F}^{(n-k)\times s}$, $D \in \mathbb{F}^{(n-k)\times k}$, $s, n, k \in \mathbb{N}$, $n > k$ and $u(i,j) = x(i,j) = 0$ for $i < 0$ or $j < 0$ and $x(0,0) = 0$.

state

$$\{x(i,j)\}_{(i,j)\in\mathbb{N}^2}$$

$$\downarrow$$

$$\hat{x}(z_1,z_2) = \sum_{(i,j)\in\mathbb{N}^2} x(i,j) z_1^i z_2^j$$

input

$$\{u(i,j)\}_{(i,j)\in\mathbb{N}^2}$$

$$\downarrow$$

$$\hat{u}(z_1,z_2) = \sum_{(i,j)\in\mathbb{N}^2} u(i,j) z_1^i z_2^j$$

output

$$\{y(i,j)\}_{(i,j)\in\mathbb{N}^2}$$

$$\downarrow$$

$$\hat{y}(z_1,z_2) = \sum_{(i,j)\in\mathbb{N}^2} y(i,j) z_1^i z_2^j$$

- define $\hat{v}(z_1, z_2) = \begin{bmatrix} \hat{u}(z_1, z_2) & \hat{y}(z_1, z_2) \end{bmatrix}^T \in \mathbb{F}[z_1, z_2]^n$ to be the code vector.

definition: the finite support input-output trajectories $(\hat{u}(z_1, z_2), \hat{y}(z_1, z_2))$ of Σ with corresponding state $\hat{x}(z_1, z_2)$ also having finite support are called finite-weight input-output trajectories of Σ.

theorem: [Pinto&Napp&Perea,2010] the set of finite-weight input-output trajectories of Σ is a 2D convolutional code of rate k/n, denoted by $\mathcal{C}(A_1, A_2, B_1, B_2, C, D)$.

- Σ is called an input-state-output (ISO) representation of $\mathcal{C}(A_1, A_2, B_1, B_2, C, D)$

note: all the 2D convolutional codes admit (many) ISO representations definition: let $\Sigma = (A_1, A_2, B_1, B_2, C, D)$ be a 2D linear system with dimension s;

– Σ is modally reachable if $\begin{bmatrix} I_s - A_1 z_1 - A_2 z_2 & B_1 z_1 + B_2 z_2 \end{bmatrix}$ is ℓFP.

– Σ is locally reachable if $\mathcal{R} = \begin{bmatrix} R_1 & R_2 & R_3 & \cdots \end{bmatrix}$ is full row rank, where R_k represents the block matrix including all columns defined by

$$\left(A_1 \,{}^{i-1}\sqcup^{j}\, A_2\right) B_1 + \left(A_1 \,{}^{i}\sqcup^{j-1}\, A_2\right) B_2$$

with $i + j = k$, for $i, j \geq 0$ and

$$A_1 \,{}^{r}\sqcup^{t}\, A_2 = 0, \text{ when either } r \text{ or } t \text{ is negative,}$$

$$A_1 \,{}^{r}\sqcup^{0}\, A_2 = A_1^r, \quad A_1 \,{}^{0}\sqcup^{t}\, A_2 = A_2^t, \text{ for } r, t \geq 0,$$

$$A_1 \,{}^{r}\sqcup^{t}\, A_2 = A_1 \left(A_1 \,{}^{r-1}\sqcup^{t}\, A_2\right) + A_2 \left(A_1 \,{}^{r}\sqcup^{t-1}\, A_2\right), \text{ for } r, t \geq 1.$$

Kalman reachability canonical form

proposition [Pinto&Napp&Perea,2010] a 2D linear system $\Sigma = (A_1, A_2, B_1, B_2, C, D)$ with dimension s, k inputs and $n-k$ outputs, is algebraically equivalent to a system in the Kalman reachability canonical form, i.e., there exist an invertible constant matrix S such that

$$SA_1S^{-1} = \begin{bmatrix} A_{11}^{(1)} & A_{12}^{(1)} \\ 0 & A_{22}^{(1)} \end{bmatrix}, \quad SA_2S^{-1} = \begin{bmatrix} A_{11}^{(2)} & A_{12}^{(2)} \\ 0 & A_{22}^{(2)} \end{bmatrix},$$

$$SB_1 = \begin{bmatrix} B_1^{(1)} \\ 0 \end{bmatrix}, \quad SB_2 = \begin{bmatrix} B_1^{(2)} \\ 0 \end{bmatrix}, \quad CS^{-1} = [C_1 \;\; C_2]$$

where $A_{11}^{(1)}, A_{11}^{(2)} \in \mathbb{F}^{\delta\times\delta}$, $B_1^{(1)}, B_1^{(2)} \in \mathbb{F}^{\delta\times k}$, $C_1 \in \mathbb{F}^{(n-k)\times\delta}$, with $s \geq \delta$ and the remaining matrices of suitable dimensions.

Kalman reachability canonical form

- moreover,

$$\Sigma_1 = \left(A_{11}^{(1)}, A_{11}^{(2)}, B_1^{(1)}, B_1^{(2)}, C_1, D\right)$$

is locally reachable 2D linear system and

$$\mathcal{C}(A_1, A_2, B_1, B_2, C, D) = \mathcal{C}\left(A_{11}^{(1)}, A_{11}^{(2)}, B_1^{(1)}, B_1^{(2)}, C_1, D\right).$$

definition: an ISO representation of a 2D convolutional code is **minimal** if it has minimal dimension among all the ISO representations of the code.

corollary: minimal ISO representations of a 2D convolutional code must be locally reachable.

strong modal reachabillity

definition: let $\Sigma = (A_1, A_2, B_1, B_2, C, D)$ be a 2D linear system with dimension s; Σ is **strongly modally reachable** if

$$\begin{bmatrix} I_s - A_1 z_1 - A_2 z_2 & B_1 z_1 + B_2 z_2 \end{bmatrix} \text{ is } \ell ZP$$

note: strongly modally reachable systems are also modally reachable; but the converse is not true.

lemma: let $\mathcal{C}$ be a basic 2D convolutional code and let $\Sigma = (A_1, A_2, B_1, B_2, C, D)$ be a modally reachable ISO representation of $\mathcal{C}$; then Σ is strongly modally reachable.

projections of a 2D convolution code

definition: the projections of a 2D convolutional code $\mathcal{C}$ onto the two semi-axis $\{\ell e_i | \ell \in \mathbb{N}\}$, for $i = 1, 2$, with $e_1 = (1, 0)$ and $e_2 = (0, 1)$, respectively, are defined by

– $\mathcal{C}_1 = \mathrm{proj}_{z_1} \mathcal{C} = \{\hat{v}(z_1, 0) : \hat{v}(z_1, z_2) \in \mathcal{C}\}$

and

– $\mathcal{C}_2 = \mathrm{proj}_{z_2} \mathcal{C} = \{\hat{v}(0, z_2) : \hat{v}(z_1, z_2) \in \mathcal{C}\}$

note: $\mathcal{C}_1$ and $\mathcal{C}_2$ are 1D convolutional codes

- if $G(z_1, z_2) \in \mathbb{F}[z_1, z_2]^{n \times k}$ is an encoder of $\mathcal{C}$ then

$$\mathcal{C}_1 = \mathrm{Im}_{\mathbb{F}[z_1]} G(z_1, 0) \qquad \text{and} \qquad \mathcal{C}_2 = \mathrm{Im}_{\mathbb{F}[z_2]} G(0, z_2)$$

but they may not be encoders of $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively.

lemma: if $\mathcal{C}$ is a basic 2D convolution code then $\mathcal{C}_1$ and $\mathcal{C}_2$ are basic 1D convolution codes.

note: the noncatastrophicity of $\mathcal{C}$ does not imply the noncatastrophicity of $\mathcal{C}_1$ and $\mathcal{C}_2$.

- let $\Sigma = (A_1, A_2, B_1, B_2, C, D)$ be an ISO representation of a 2D convolutional code $\mathcal{C}$; then:

  – $\Sigma_1 = (A_1, B_1, C, D)$ is an ISO representation of $\mathcal{C}_1$

and

  – $\Sigma_2 = (A_2, B_2, C, D)$ is an ISO representation of $\mathcal{C}_2$.

minimality and strong modal reachability

theorem: let $\Sigma = (A_1, A_2, B_1, B_2, C, D)$ be a strongly modally reachable ISO representation of a 2D convolutional code; then Σ is a minimal ISO representation.

corollary: let $\mathcal{C}$ be a basic 2D convolutional code of rate k/n with a strongly modally reachable ISO representation Σ of dimension s; then $\mathcal{C}$ has complexity s and the projections $\mathcal{C}_1$ and $\mathcal{C}_2$ of $\mathcal{C}$ have also rate k/n and complexity s.

conclusion:

— if a basic 2D convolutional code admits a strongly modally reachable ISO representation then this ISO representation is minimal with dimension equal to the complexity of the code

future work:

— prove that all basic 2D convolutional codes admit a strongly modally reachable ISO representation;

— show that all minimal ISO representations of a basic 2D convolutional code are algebraically equivalent.

thank you for your attention!