Erasure decoding of convolutional codes with the help of linear systems

Julia Lieb

Institute of Mathematics
University of Zurich

Joint work with Joachim Rosenthal

Convolutional Codes

Definition

A **convolutional code** $\mathcal{C}$ of **rate** k/n is a free $\mathbb{F}[z]$-submodule of $\mathbb{F}[z]^n$ of rank k. There is $G(z) \in \mathbb{F}[z]^{n\times k}$ of full column rank such that

$$\mathcal{C} = \{v(z) \in \mathbb{F}[z]^n \mid v(z) = G(z)m(z) \text{ for some } m(z) \in \mathbb{F}[z]^k\}.$$

$G(z)$ is called **generator matrix** of the code. The **degree** δ of $\mathcal{C}$ is defined as the maximal degree of the $k \times k$-minors of $G(z)$. One calls $\mathcal{C}$ an (n, k, δ) convolutional code.

$m(z) = m_0 + m_1 z + \cdots + m_N z^N$ is encoded to $v(z) = G(z)m(z)$:

$v_0 = G_0 m_0$, $v_1 = G_1 m_0 + G_0 m_1$, $v_2 = G_2 m_0 + G_1 m_1 + G_0 m_2$, ...

Convolutional Codes

Definition

A **convolutional code** $\mathcal{C}$ of **rate** k/n is a free $\mathbb{F}[z]$-submodule of $\mathbb{F}[z]^n$ of rank k. There is $G(z) \in \mathbb{F}[z]^{n\times k}$ of full column rank such that

$$\mathcal{C} = \{v(z) \in \mathbb{F}[z]^n \mid v(z) = G(z)m(z) \text{ for some } m(z) \in \mathbb{F}[z]^k\}.$$

$G(z)$ is called **generator matrix** of the code. The **degree** δ of $\mathcal{C}$ is defined as the maximal degree of the $k \times k$-minors of $G(z)$. One calls $\mathcal{C}$ an (n, k, δ) convolutional code.

Definition

If there exists a **parity-check matrix** $H(z) \in \mathbb{F}[z]^{(n-k)\times n}$ of full rank, such that

$$\mathcal{C} = \{v(z) \in \mathbb{F}[z]^n \mid H(z)v(z) = 0 \in \mathbb{F}[z]^{n-k}\},$$

$\mathcal{C}$ is called **non-catastrophic**.

Linear Systems and Convolutional Codes

Consider discrete-time linear systems (A, B, C, D) of the form

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t \\ y_t &= Cx_t + Du_t \\ x_0 &= 0 \end{aligned}$$

with $s, t \in \mathbb{N}_0$, $A \in \mathbb{F}^{s\times s}$, $B \in \mathbb{F}^{s\times k}$, $C \in \mathbb{F}^{(n-k)\times s}$, $D \in \mathbb{F}^{(n-k)\times k}$, input sequence $(u_t)_{t\in\mathbb{N}_0} \subset \mathbb{F}^k$, state sequence $(x_t)_{t\in\mathbb{N}_0} \subset \mathbb{F}^s$, output sequence $(y_t)_{t\in\mathbb{N}_0} \subset \mathbb{F}^{n-k}$.

Linear Systems and Convolutional Codes

Consider discrete-time linear systems (A, B, C, D) of the form

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t \\ y_t &= Cx_t + Du_t \\ x_0 &= 0 \end{aligned}$$

with $s, t \in \mathbb{N}_0$, $A \in \mathbb{F}^{s\times s}$, $B \in \mathbb{F}^{s\times k}$, $C \in \mathbb{F}^{(n-k)\times s}$, $D \in \mathbb{F}^{(n-k)\times k}$, input sequence $(u_t)_{t\in\mathbb{N}_0} \subset \mathbb{F}^k$, state sequence $(x_t)_{t\in\mathbb{N}_0} \subset \mathbb{F}^s$, output sequence $(y_t)_{t\in\mathbb{N}_0} \subset \mathbb{F}^{n-k}$. The set of finite-weight trajectories $(u_t)_{t\in\mathbb{N}_0}, (x_t)_{t\in\mathbb{N}_0}, (y_t)_{t\in\mathbb{N}_0}$ of (A, B, C, D) corresponds to an (n, k, δ) convolutional code $\mathcal{C}(A, B, C, D)$ with $s \geq \delta$, where the codewords are defined via their coefficient vectors $v_t = \begin{pmatrix} y_t \\ u_t \end{pmatrix}$. If $s = \delta$, (A, B, C, D) is called **minimal ISO (input-state-output) representation** of $\mathcal{C}(A, B, C, D)$.

RY 1999: J. Rosenthal, E.V. York: BCH convolutional codes, IEEE Trans. Inform. Theory, Vol. 45, No. 6 (1999), 1833-1844.

Linear Systems and Convolutional Codes

Definition

A linear system is called **reachable** if the **reachability matrix** $\mathcal{R}(A,B) := [B, AB, \ldots, A^{s-1}B] \in \mathbb{F}^{s \times sk}$ satisfies $\mathrm{rk}(\mathcal{R}(A,B)) = s$ and **observable** if the **observability matrix**

$$\mathcal{O}(A,C) = \begin{bmatrix} C \\ \vdots \\ CA^{s-1} \end{bmatrix} \in \mathbb{F}^{(n-k)s \times s} \text{ satiesfies } \mathrm{rk}(\mathcal{O}(A,B)) = s.$$

Linear Systems and Convolutional Codes

Definition

A linear system is called **reachable** if the **reachability matrix** $\mathcal{R}(A,B) := [B, AB, \ldots, A^{s-1}B] \in \mathbb{F}^{s \times sk}$ satisfies $\mathrm{rk}(\mathcal{R}(A,B)) = s$ and **observable** if the **observability matrix**

$$\mathcal{O}(A,C) = \begin{bmatrix} C \\ \vdots \\ CA^{s-1} \end{bmatrix} \in \mathbb{F}^{(n-k)s \times s} \text{ satiesfies } \mathrm{rk}(\mathcal{O}(A,B)) = s.$$

Theorem (RY1999)

(A,B,C,D) is a minimal ISO representation of $\mathcal{C}(A,B,C,D)$ if and only if it is reachable.

Linear Systems and Convolutional Codes

Definition

A linear system is called **reachable** if the **reachability matrix** $\mathcal{R}(A,B) := [B, AB, \ldots, A^{s-1}B] \in \mathbb{F}^{s \times sk}$ satisfies $\mathrm{rk}(\mathcal{R}(A,B)) = s$ and **observable** if the **observability matrix**

$$\mathcal{O}(A,C) = \begin{bmatrix} C \\ \vdots \\ CA^{s-1} \end{bmatrix} \in \mathbb{F}^{(n-k)s \times s} \text{ satiesfies } \mathrm{rk}(\mathcal{O}(A,B)) = s.$$

Theorem (RY1999)

(A,B,C,D) is a minimal ISO representation of $\mathcal{C}(A,B,C,D)$ if and only if it is reachable.

Theorem (RY1999)

Assume that (A,B,C,D) is reachable. Then $\mathcal{C}(A,B,C,D)$ is non-catastrophic if and only if (A,B,C,D) is observable.

Low-delay erasure decoding

- Lieb, J.; Rosenthal, J.: Erasure decoding of convolutional codes using first-order representations, Math. Control Signals Syst. 33:3 (2021), p. 499-513, doi: 10.1007/s00498-021-00289-9.

Low-delay erasure decoding

- Lieb, J.; Rosenthal, J.: Erasure decoding of convolutional codes using first-order representations, Math. Control Signals Syst. 33:3 (2021), p. 499-513, doi: 10.1007/s00498-021-00289-9.
- Tomas, V. T.: Complete-MDP Convolutional Codes over the Erasure Channel, PhD thesis, Universidad de Alicante, 2010.

Low-delay erasure decoding

- Lieb, J.; Rosenthal, J.: Erasure decoding of convolutional codes using first-order representations, Math. Control Signals Syst. 33:3 (2021), p. 499-513, doi: 10.1007/s00498-021-00289-9.
- Tomas, V. T.: Complete-MDP Convolutional Codes over the Erasure Channel, PhD thesis, Universidad de Alicante, 2010.

$(y_0, u_0), \ldots, (y_j, u_j)$ are the first $j+1$ coefficients of a codeword in $\mathcal{C}(A, B, C, D)$ if and only if

$$\left[\begin{array}{c|cccc} & D & 0 & \ldots & 0 \\ -I & CB & \ddots & \ddots & \vdots \\ & \vdots & \ddots & \ddots & 0 \\ & CA^{j-1}B & \ldots & CB & D \end{array}\right] \left(\begin{array}{c} y_0 \\ \vdots \\ y_j \\ \hline u_0 \\ \vdots \\ u_j \end{array}\right) = 0.$$

$$y_0 = Cx_0 + Du_0 = Du_0$$

$$y_1 = Cx_1 + Du_1 = C(Ax_0 + Bu_0) + Du_1 = CBu_0 + Du_1,$$

$$\vdots$$

Low-delay erasure decoding

Assume that $v_0, \ldots, v_{i-1}$ are known and v_i contains erasures:

$$\left[\begin{array}{c|cccc} & D & 0 & \ldots & 0 \\ & CB & \ddots & \ddots & \vdots \\ -I & \vdots & \ddots & \ddots & 0 \\ & CA^{j-1}B & \ldots & CB & D \end{array}\right] \left(\begin{array}{c} y_i \\ \vdots \\ y_{i+j} \\ \hline u_i \\ \vdots \\ u_{i+j} \end{array}\right) = \beta$$

where β is a known vector depending on $v_0, \ldots, v_{i-1}$.

The erased components of $y_i, \ldots, y_{i+j}, u_i, \ldots, u_{i+j}$ are are the unknown variables of the system of linear equations.

The parameter j represents the time-delay, with which v_i is recovered, and should be kept as small as possible.

Restart of recovery after too many erasures

To restart the decoding in case of too many erasures for the recovery of v_i, use that for $i, j, l \in \mathbb{N}_0$,

$$\begin{bmatrix} C \\ \vdots \\ CA^j \end{bmatrix} x_{i+l} + \left[\begin{array}{c|cccc} -I & D & 0 & \dots & 0 \\ & CB & \ddots & \ddots & \vdots \\ & \vdots & \ddots & \ddots & 0 \\ & CA^{j-1}B & \dots & CB & D \end{array}\right] \left(\begin{array}{c} y_{i+l} \\ \vdots \\ y_{i+l+j} \\ \hline u_{i+l} \\ \vdots \\ u_{i+l+j} \end{array}\right) = 0$$

where $$x_{i+l} = A^{i+l-1}Bu_0 + \cdots + Bu_{i+l-1}.$$

As $u_0, \dots, u_{i-1}$ are known, $x_0, \dots, x_i$ are known as well and we first try to recover x_{i+1} with minimum possible time delay j.
If this is not possible (within a certain delay), we try to recover x_{i+l} for $l > 1$. Unknows for this system of equations are x_{i+l} and the erased components of $v_{i+l}, \dots, v_{i+l+j}$.

Restart of recovery after too many erasures

$$\begin{bmatrix} C \\ \vdots \\ CA^j \end{bmatrix} x_{i+l} + \left[\begin{array}{c|ccc} & D & 0 & \dots & 0 \\ -I & CB & \ddots & \ddots & \vdots \\ & \vdots & \ddots & \ddots & 0 \\ & CA^{j-1}B & \dots & CB & D \end{array}\right] \left(\begin{array}{c} y_{i+l} \\ \vdots \\ y_{i+l+j} \\ \hline u_{i+l} \\ \vdots \\ u_{i+l+j} \end{array}\right) = 0$$

where

$$x_{i+l} = A^{i+l-1}Bu_0 + \cdots + Bu_{i+l-1}.$$

If x_{i+l} can be recovered, the low-delay decoding can be restarted (even if parts of $v_i, \dots, v_{i+l-1}$ have to be declared as lost for the moment) and the recovery of $u_i, \dots, u_{i+l-1}$ is not necessary for the recovery of further symbols.

Termination

Proposition

Let $\gamma := \deg(m(z))$ *and* $\mu := \deg(G(z))$*. For* $w \in \mathbb{N}_0$*, define*

$$E_w := \begin{bmatrix} CA^{\gamma+\mu}B & \cdots & CB \\ \vdots & & \vdots \\ CA^{\gamma+\mu+w}B & \cdots & CA^wB \end{bmatrix}.$$

One has, $E_w \cdot [u_0^\top, \ldots, u_{\gamma+\mu}^\top]^\top = 0$ *for all* $w \in \mathbb{N}_0$*.*

Let $\tilde{E}_w$ denote the submatrix of E_w consisting of the columns corresponding to the unknown components of $[u_0^\top, \ldots, u_{\gamma+\mu}^\top]$.

If there exists $w \in \{0, \ldots, \delta - 1\}$ such that $\tilde{E}_w$ has full column rank, we can use the system of linear equations induced by $E_w \cdot [u_0^\top, \ldots, u_{\gamma+\mu}^\top]^\top = 0$ to recover the unknown components of $[u_0^\top, \ldots, u_{\gamma+\mu}^\top]^\top$ and then also obtain $[y_0^\top, \ldots, y_{\gamma+\mu}^\top]^\top$.

Advantages of presented decoding algorithm

- **Reduced complexity due to echelon form of systems of equations**

  For decoding with delay at most T, the complexity is determined by solving a linear system with $(n-k)(T+1)$ equations and $e \leq (n-k)(T+1)$ unknowns, i.e. $O(((n-k)(T+1))^{0.8} \cdot e^2)$.

  Our systems of equations are parts of systems in echelon form. Hence, the erasures in $y_i, \ldots, y_{i+T}$ can be neglected for the complexity analysis and only the erasures in $u_i, \ldots, u_{i+T}$ matter.

Advantages of presented decoding algorithm

- **Reduced complexity due to echelon form of systems of equations**

  For decoding with delay at most T, the complexity is determined by solving a linear system with $(n-k)(T+1)$ equations and $e \leq (n-k)(T+1)$ unknowns, i.e. $O(((n-k)(T+1))^{0.8} \cdot e^2)$.

  Our systems of equations are parts of systems in echelon form. Hence, the erasures in $y_i, \ldots, y_{i+T}$ can be neglected for the complexity analysis and only the erasures in $u_i, \ldots, u_{i+T}$ matter.

- **Reduced decoding delay due to termination algorithm**

  The termination algorithm can allow for an early determination of the whole decoding process.