Minimal-span bases, linear system theory, and the invariant factor theorem

G. David Forney, Jr.

MIT
Cambridge MA 02139 USA

DIMACS Workshop on
Algebraic Coding Theory and Information Theory

DIMACS Center, Rutgers University, Piscataway, NJ
15 December 2003

Background

1970: "Convolutional codes I: Algebraic structure"

- Key tool: the invariant factor theorem

1976: "Minimal bases of rational vector spaces, with applications to multivariable linear systems"

- Similar results, without the invariant factor theorem
- Minimal basis = set of shortest independent generators

1988-98: Trellis-oriented generator matrices for linear block codes

- Minimal state-space realizations of linear block codes
- Trellis-oriented basis = set of shortest-span independent generators
- Theory is elementary, once ordering of coordinates is specified

1993: "Dynamics of group codes: State spaces, trellis diagrams, and canonical encoders"

- Minimal state-space realizations depend only on group structure

Conclusions and speculations

- Theory of minimal realizations of linear systems is
  - elementary, more so than than the invariant factor theorem;
  - basically group-theoretic
- Can the IFT be proved using minimal realization theory?

Outline

Develop theory of minimal realizations of linear systems

- Key: Minimal-span bases
- Demonstrate that the theory is elementary

Easy proof that the ring of polynomials (resp. finite sequences) is a principal ideal domain

- Based on structure of linear time-invariant systems over $\mathbb{F}$

However, our proof of the IFT is still mainly algebraic

Open question:

- relation between minimal-span bases and invariant-factor bases

For algebraic coding theorists:

- A different kind of algebra

Definitions

Sequence space $(\mathbb{F}^n)^{\mathcal{I}}$

- $\mathbb{F}$: a field
- **time axis** $\mathcal{I} \subseteq \mathbb{Z}$: a discrete index set
- **sequence** $\mathbf{x} \in (\mathbb{F}^n)^{\mathcal{I}} = \{x_k \in \mathbb{F}^n, k \in \mathcal{I}\}$
  - D-transform $x(D) = \sum_k x_k D^k$
- $(\mathbb{F}^n)^{\mathcal{I}} \cong (\mathbb{F}^{\mathcal{I}})^n$ is a vector space over $\mathbb{F}$

Discrete-time linear system (code) $\mathcal{C}$ over $\mathbb{F}$

- $\mathcal{C}$: any subspace of $(\mathbb{F}^n)^{\mathcal{I}}$

Degree, delay, support, span of a sequence $\mathbf{x} \neq \mathbf{0}$

- **degree**: $\deg \mathbf{x} =$ greatest $k \in \mathcal{I}$ such that $x_k \neq 0$
- **delay**: $\mathrm{del}\ \mathbf{x} =$ least $k \in \mathcal{I}$ such that $x_k \neq 0$
- **support**: $\mathrm{supp}\ \mathbf{x} = [\deg \mathbf{x}, \mathrm{del}\ \mathbf{x}]$
- **span**: $\mathrm{span}\ \mathbf{x} = \deg \mathbf{x} - \mathrm{del}\ \mathbf{x} \geq 0$.
- if $\mathbf{x} = \mathbf{0}$, then $\deg \mathbf{x} = -\infty$, $\mathrm{del}\ \mathbf{x} = \infty$

Classification of sequences:

	$\mathrm{del}\ \mathbf{x} = -\infty$	$\mathrm{del}\ \mathbf{x} > -\infty$	$\mathrm{del}\ \mathbf{x} \geq 0$
$\deg \mathbf{x} = \infty$	bi-infinite	Laurent	causal
$\deg \mathbf{x} < \infty$	anti-Laurent	finite	polynomial
$\deg \mathbf{x} \leq 0$	anti-causal	anti-polynomial	scalar

Time-invariance of a system $\mathcal{C}$

- $\mathcal{C}$ is **time-invariant** if $\mathcal{I} = \mathbb{Z}$ and $D\mathcal{C} = \mathcal{C}$
- $\mathcal{C}$ is **semi-time-invariant** if $D\mathcal{C} \subset \mathcal{C}$

Finite and polynomial linear systems

Polynomial linear systems

- A sequence $\mathbf{x}$ is **polynomial** if its D-transform $x(D)$ is polynomial
  - $\mathbb{F}[D]$: ring of polynomial sequences
  - $\mathbb{F}^n[D] \cong (\mathbb{F}[D])^n$: module of n-tuples of polynomial sequences
  - $(\mathbb{F}[D])^n$ is a semi-time-invariant linear system
- **Polynomial linear system** $\mathcal{C}$ over $\mathbb{F}^n$:
  a subset $\mathcal{C} \subseteq (\mathbb{F}[D])^n$ that is closed under addition and multiplication by scalars
- Polynomial linear semi-time-invariant (LSTI) system $\mathcal{C}$ over $\mathbb{F}^n$:
  a subset $\mathcal{C} \subseteq (\mathbb{F}[D])^n$ that is closed under addition and multiplication by scalars or by D; *i.e.,* multiplication by polynomials

Finite linear systems

- A sequence $\mathbf{x}$ is **finite** if it has a finite number of nonzero coefficients
  - $\mathbb{F}[D, D^{-1}]$: ring of finite sequences
  - $\mathbb{F}^n[D, D^{-1}] \cong (\mathbb{F}[D, D^{-1}])^n$: module of n-tuples of finite sequences
  - $(\mathbb{F}[D, D^{-1}])^n$ is a time-invariant linear system
- **Finite linear system** $\mathcal{C}$ over $\mathbb{F}^n$:
  a subset $\mathcal{C} \subseteq (\mathbb{F}[D, D^{-1}])^n$ that is closed under addition and multiplication by scalars
- Finite linear time-invariant (LTI) system $\mathcal{C}$ over $\mathbb{F}^n$:
  a subset $\mathcal{C} \subseteq (\mathbb{F}[D, D^{-1}])^n$ that is closed under addition and multiplication by scalars, D or D^{-1}; *i.e.,* multiplication by finite sequences

We will focus on finite linear systems

- Finite and polynomial linear systems are almost identical
- Finite linear systems can be time-invariant

Minimal-span bases for finite linear systems

Basis for a finite linear system $\mathcal{C}$:
a linearly independent set $\mathcal{G}$ of finite **generators** $\mathbf{g}_i \in \mathcal{C}$ such that
$\mathcal{C}$ is the set of all finite $\mathbb{F}$-linear combinations of generators

Minimal-span basis for a finite linear system $\mathcal{C}$:
a basis $\mathcal{G}$ for $\mathcal{C}$ such that
no generator can be replaced by a shorter-span generator

Predictable support property for a set $\mathcal{G} = \{\mathbf{g}_i\}$ of finite generators:
if $\sum_{i \in \mathcal{J}} \alpha_i \mathbf{g}_i$ is any finite linear combination with $\alpha_i \neq 0, i \in \mathcal{J}$, then

$$\text{supp} \sum_{i \in \mathcal{J}} \alpha_i \mathbf{g}_i = [(\min_{\mathcal{J}} \text{del } \mathbf{g}_i), (\max_{\mathcal{J}} \deg \mathbf{g}_i)];$$

i.e., cancellation of minimum-delay or max-degree terms never occurs.

Theorem 1 (Minimal-span basis = PSP) *Given a finite linear system $\mathcal{C} \in ((\mathbb{F}^n)^{\mathcal{I}})_f$ and a basis $\mathcal{G}$ for $\mathcal{C}$, where $\mathcal{I} \subseteq \mathbb{Z}$, the following are equivalent:*

(a) $\mathcal{G}$ is a minimal-span basis for $\mathcal{C}$;

(b) $\mathcal{G}$ has the predictable support property.

Proof. There is a $\mathbf{x} \in \mathcal{C}$ that can be substituted for a longer-span generator $\mathbf{g}_i \in \mathcal{G}$ if and only if there is a linear combination of generators including $\mathbf{g}_i$ for which the predictable support property fails. □

Corollary 2 (Algebraic test for PSP) *A set $\mathcal{G}$ of generators $\mathbf{g}_i \in (\mathbb{F}^n)^{\mathcal{I}}$ has the predictable support property if and only if for each $k \in \mathcal{I}$, the set of time-k symbols g_{ik} of generators $\mathbf{g}_i \in \mathcal{G}$ that start at time k is linearly independent, and similarly the set of time-k symbols g_{ik} of generators $\mathbf{g}_i$ that stop at time k is linearly independent.*

Consequently the number of generators $\mathbf{g}_i \in \mathcal{G}$ that start or stop at any time $k \in \mathcal{I}$ is not greater than n.

Minimal state-space realizations and minimal-span bases

Elementary realization of a single generator $\mathbf{g}_i$

A single generator $\mathbf{g}_i$ with support $[\mathrm{del}\, \mathbf{g}_i, \deg \mathbf{g}_i]$ may be realized by an elementary state realization with a one-dimensional state space which is "active" during $[\mathrm{del}\, \mathbf{g}_i, \deg \mathbf{g}_i]$ and "inactive" otherwise.

Product realization of a generator set $\mathcal{G}$

A set $\mathcal{G} = \{\mathbf{g}_i\}$ of generators may be realized by summing the outputs of elementary realizations of each generator individually.

Theorem 3 *Given a linear system $\mathcal{C}$ and a minimal-span basis $\mathcal{G}$ for $\mathcal{C}$, the product realization of $\mathcal{G}$ is a minimal state-space realization of $\mathcal{C}$.*

Proof. Based on:

Theorem 4 (State space theorem) *Given a linear system $\mathcal{C}$ defined on a time axis $\mathcal{I}$ and a cut time j of $\mathcal{I}$, the minimal dimension of the state space Σ_j in any linear realization is $\dim \mathcal{C}/(\mathcal{C}_{:\mathcal{P}_j} \times \mathcal{C}_{:\mathcal{F}_j})$, where*

- *$\mathcal{C}_{:\mathcal{P}_j}$ is the subsystem of $\mathcal{C}$ with support in $\mathcal{P}_j = \{k \in \mathcal{I} \mid k < j\}$*
- *$\mathcal{C}_{:\mathcal{F}_j}$ is the subsystem of $\mathcal{C}$ with support in $\mathcal{F}_j = \{k \in \mathcal{I} \mid k > j\}$.*

Theorem 5 (Bases of subsystems) *Let $\mathcal{C} \subseteq ((\mathbb{F}^n)^{\mathcal{I}})_f$ be a finite linear system with minimal-span basis $\mathcal{G}$, and let $\mathcal{J} \subseteq \mathcal{I}$ be any subinterval of the time axis $\mathcal{I}$. Then the subsystem $\mathcal{C}_{:\mathcal{J}}$ is generated by the subset $\mathcal{G}_{\mathcal{J}} \subseteq \mathcal{G}$ of generators whose support is contained in $\mathcal{J}$.*

Proof. By the predictable support property, a sequence generated by $\mathcal{G}$ has support in $\mathcal{J}$ if and only if it is a linear combination of generators with support in $\mathcal{J}$. □

It follows that the minimal dimension of the state space Σ_j at cut time j in any state realization of $\mathcal{C}$ is the number of generators in a minimal-span basis $\mathcal{G}$ whose support covers j; *i.e.*, which are "active" at time j. □

Finite LTI systems over $\mathbb{F}$

Theorem 6 (Minimal-span bases for finite LTI systems over $\mathbb{F}$) *A nontrivial LTI system $\mathcal{C}$ over $\mathbb{F}$ has a minimal-span basis consisting of all time shifts $\{D^d\mathbf{g}, d \in \mathbb{Z}\}$ of a single polynomial generator $\mathbf{g}$ with* $\operatorname{del} \mathbf{g} = 0$.

Proof. By time-invariance, the shortest-span generator starting at time d is a time shift by D^d of the shortest-span generator starting at time 0. By Corollary 2, no more than one generator can start at any time. $\square$

An $\mathbb{F}[D, D^{-1}]$**-ideal** is a set of finite sequences that is closed under $\mathbb{F}[D, D^{-1}]$-linear combinations.

Lemma 7 $\mathbb{F}[D, D^{-1}]$*-ideal = finite LTI system over* $\mathbb{F}$*.*

A **principal ideal** is the set $(g(D)) = \{a(D)g(D) \mid a(D) \in \mathbb{F}[D, D^{-1}]\}$ of $\mathbb{F}[D, D^{-1}]$-multiples of a single finite sequence $g(D)$.

Theorem 8 (The finite sequences form a PID) *Every ideal in the ring $\mathbb{F}[D, D^{-1}]$ of finite sequences in D over a field $\mathbb{F}$ is a principal ideal; i.e., $\mathbb{F}[D, D^{-1}]$ is a principal ideal domain (PID).*

Proof. Theorem 6 and Lemma 7.

$p(D)$ is the **greatest common divisor** (gcd) of two finite sequences $g(D)$ and $h(D)$ if every common divisor of $g(D)$ and $h(D)$ divides $p(D)$.

Lemma 9 (GCDs) *The gcd of two finite sequences $g(D)$ and $h(D)$ is the generator of the ideal of all their $\mathbb{F}[D, D^{-1}]$-linear combinations:*

$$(g(D)) + (h(D)) = \{a(D)g(D) + b(D)h(D) \mid a(D), b(D) \in \mathbb{F}[D, D^{-1}]\}.$$

<u>Corollary</u>: there exist $a(D), b(D)$ such that

$$\gcd(g(D), h(D)) = a(D)g(D) + b(D)h(D).$$

Finite LTI systems over $\mathbb{F}^n$

Theorem 10 (Minimal-span bases for finite LTI systems over $\mathbb{F}^n$) *A finite LTI system $\mathcal{C}$ over $\mathbb{F}^n$ has a minimal-span basis consisting of all time shifts of $k \leq n$ finite generators $\{\mathbf{g}_i, 1 \leq i \leq k\}$ with* $\text{del}\, \mathbf{g}_i = 0$.

Proof. Choose a set of shortest-span linearly independent generators that start at time 0, and all their time shifts. By Corollary 2, there can be at most n of them. □

Notes: The integer k is the *rank* of $\mathcal{C}$ as a free $\mathbb{F}[D, D^{-1}]$-module. The fraction k/n is the *rate* of $\mathcal{C}$ as a code.

Theorem 11 (Invariant factor theorem for finite LTI systems) *If $\mathcal{C}$ is a finite LTI system, then there exists*

- *a basis $\{\mathbf{a}_j(D), 1 \leq j \leq n\}$ for $\mathbb{F}[D, D^{-1}]^n$, and*
- *a set of $k \leq n$ monic delay-zero finite sequences $\{\gamma_i(D), 1 \leq i \leq k\}$, called the invariant factors of $\mathcal{C}$,*

such that

- *$\gamma_i(D)$ divides $\gamma_{i+1}(D)$ for $1 \leq i < k$, and*
- *$\{\gamma_i(D)\mathbf{a}_j(D), 1 \leq i \leq k\}$ is an $\mathbb{F}[D, D^{-1}]$-basis for $\mathcal{C}$.*

Proof. Theorem 8 shows that $\mathbb{F}[D, D^{-1}]$ is a PID, and Theorem 10 shows that the rank of $\mathcal{C}$ as an $\mathbb{F}[D, D^{-1}]$-module is $k \leq n$. The rest of the argument follows standard module-theoretic proofs. □

Question: What is the relation (if any) between an invariant-factor basis of $\mathcal{C}$ and a minimal-span basis for $\mathcal{C}$?

Invariant-factor and minimal-span bases

An invariant-factor basis is not necessarily a minimal-span basis; *e.g.,*

$$\begin{bmatrix} 1+D & D \\ -D & 1-D \end{bmatrix}$$

is an invariant-factor basis for $\mathbb{F}[D, D^{-1}]^2$, which has minimal-span basis

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

However, the latter is also an invariant-factor basis for $\mathbb{F}[D, D^{-1}]^2$.

A minimal-span basis is not necessarily an invariant-factor basis; *e.g.,*

$$\begin{bmatrix} 1 & 1-D & 1-D \\ 1 & 1-D & 0 \end{bmatrix}$$

is a minimum-span basis for a rate-2/3 system $\mathcal{C}$ whose invariant factors are $\gamma_1(D) = 1, \gamma_2(D) = 1 - D$, and which has an invariant-factor basis

$$\begin{bmatrix} 1 & 1-D & 0 \\ 0 & 0 & 1-D \end{bmatrix}.$$

However, the latter is also a minimal-span basis for $\mathcal{C}$.

Theorem 12 (Canonical bases) *Every finite LTI system $\mathcal{C}$ has a basis which is both a minimal-span basis and an invariant-factor basis.*

Proof. Start with an invariant-factor basis $\{\gamma_i(D)\mathbf{a}_j(D), 1 \leq i \leq k\}$ for $\mathcal{C}$. If the starting-time coefficient matrix and the stopping-time coefficient matrix are full-rank over $\mathbb{F}$, then by Corollary 2 we are done. Otherwise, if the starting-time coefficient matrix is not full-rank, then there is an $\mathbb{F}$-linear combination $\mathbf{g}(D)$ of the basis n-tuples whose delay is greater than zero; substitute a time shift of $\mathbf{a}(D)$ for $\gamma_m(D)\mathbf{a}_m(D)$, where m is the least index of the n-tuples involved in the combination. We proceed similarly if the stopping-time coefficient matrix is not full-rank. The process must terminate in a finite number of steps with the desired basis. $\square$

Conclusion

Conclusion: another direction for algebraic coding theory

Invariant-factor decomposition depends on linearity, time-invariance

Minimal-realization theory may be extended further

- Group systems and codes
- Non-time-invariant systems and codes (including block)
- Systems and codes on cycle-free graphs

Minimal realizations not well-defined on graphs with cycles

- Even so, a duality theory still applies