Codes on Graphs: Fundamentals

G. David Forney, Jr.*

Abstract

This paper develops a fundamental theory of realizations of linear and group codes on general graphs using elementary group theory, including basic group duality theory. The properties of fragments of realizations are analyzed, particularly leaf (degree-1) and trellis (degree-2) fragments. A state space theorem for leaf fragments yields a simple proof of the minimality of trim and proper cycle-free realizations, as well as a decomposition of a general realization into a cyclic leafless 2-core and a number of cycle-free leaf fragments. Different types of observability and controllability are defined and analyzed. A general structure theory of trellis fragments is developed, and applied to edge cuts of cyclic realizations. It is shown that finite linear realizations may always be locally reduced to be state-trim, observable and controllable.

Index terms — Group codes, linear codes, graphical models.

1 Introduction

The subject of “codes on graphs” was founded by Tanner in [18], which showed how codes such as low-density parity-check (LDPC) codes may be defined by a graphical system of variables and constraints, and decoded by generic decoding algorithms (“sum-product,” “max-sum”) that are optimal in the cycle-free case.

After years of obscurity, Tanner’s results were rediscovered (largely independently) in Wiberg’s doctoral thesis [21, 22], which showed that capacity-approaching codes (e.g., turbo codes and LDPC codes) and their decoding algorithms could be viewed a common graphical framework. By introducing internal (“state”) variables, Wiberg made connections with topics such as convolutional codes, trellis codes, and tail-biting trellis codes, as well as with classical linear system theory.

The first paper in this series [3] introduced an algebraic approach to realizations of linear and group codes by systems of variables and constraints, as in the behavioral system theory of Willems [23]. Its key observation was that, without any loss of generality or essential increase of complexity, every realization may be assumed to be “normal;” i.e., to satisfy the “normal degree constraints” that all external variables (“symbols”) have degree 1, whereas all internal variables (“states”) have degree 2. Two important consequences are:

• (Normal graph) A normal realization is naturally represented by a “normal graph.”

• (Normal realization duality) A linear or group normal realization \( \mathcal{R} \) has a well-defined dual realization \( \mathcal{R}^\circ \), such that if \( \mathcal{R} \) realizes a linear or group code \( \mathcal{C} \), then \( \mathcal{R}^\circ \) realizes \( \mathcal{C}^\perp \).

*Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 (email: forneyd@comcast.net). Part of this paper was presented at the 2012 Allerton Conference [6].
In this paper, we study the fundamental system-theoretic properties of general normal linear and (finite abelian) group realizations, in many cases generalizing the results of [7] for linear tail-biting trellis realizations. We use only elementary group theory (including elementary duality theory of finite abelian groups), reviewed in Section 3, not because group codes are so important, but rather because in this setting we feel that a result that cannot be proved by elementary group theory is probably not fundamental.

Our results apply not just to codes, but to any linear or group system defined by a network of variables and constraints; e.g., classical linear systems, or various types of physical systems. However, we use the language of coding theory.

As in Willems [23] or Vontobel and Loeliger [20], we analyze a realization by cutting it into fragments, or by combining fragments into larger fragments. The smallest fragment is a single constraint; the largest fragment is the complete realization.

In Section 2, we begin a systematic study of fragments of realizations, which are more complex than constraints or realizations in that they have both internal and external state variables. We focus particularly on leaf fragments (fragments with a single external state variable), which are the key to analysis of cycle-free realizations, and on trellis fragments (fragments with two external state variables), which arise not only in classical and tail-biting trellis realizations, but also whenever a cycle is cut. Analysis of fragments of higher degree is left mostly for future research.

Section 4 studies the external properties of fragments, which resemble those of constraint codes. The results of [7] on trimness, properness, and local reducibility of constraint codes are generalized to fragments. Using a fundamental structure theorem for subdirect products, we develop an elementary but powerful state space theorem for leaf fragments, which leads not only to an improved proof for the fundamental “minimal $\Leftrightarrow$ trim + proper” theorem [7] for cycle-free realizations, but also to a general decomposition of a cyclic realization into a number of cycle-free leaf fragments plus a leafless “2-core,” which is the essential cyclic core of the realization.

Section 5 studies the internal properties of fragments, which resemble those of realizations. We consider several different notions of observability and controllability. We define internal observability and controllability similarly to observability and controllability of realizations, and characterize them in similar ways. In particular, a fragment is internally controllable if its constraints are independent. We define external observability and controllability in a way that turns out to be more closely related to classical notions of observability and controllability. We attempt to clarify the distinctions between these various definitions. Finally, as in [7], we observe that a realization can be internally unobservable or uncontrollable only if it is cyclic.

In Section 6, we develop a general structure theory of trellis fragments. Assuming trimness and properness, we develop dual decompositions of the transition space and the unobservable transition space of a trellis fragment, which govern its external controllability and observability properties, respectively. We show that when trellis fragments are combined, the combination is neither less controllable nor less observable than its components.

Finally, Section 7 develops results concerning non-state-trimness, internal unobservability and internal uncontrollability of cyclic realizations. For linear realizations, we find, as in [7] that local reductions can eliminate all of these defects. However, these reductions do not necessarily work in the general group case. This is the only place in this paper where we observe a difference between the linear and group cases.
2 Realizations and fragments

In this section, we review realizations, normal realizations, normal graphs, and behaviors. We introduce extended behaviors, which, while redundant, yield nice proofs. Finally, we formally introduce fragments.

2.1 Realizations

In this paper, as in [3], a realization $R$ of a code $C$ will be defined by a system of variables and constraints, in the style of behavioral system theory [23].

For a linear or group realization, each variable $V$ will take values $v$ in a finite-dimensional vector space $V$ over a base field $\mathbb{F}$, or in a finite abelian group $\mathcal{V}$, respectively. We will often identify a variable by its alphabet $\mathcal{V}$ or its value $v \in \mathcal{V}$. In the linear case, the “size” of $V$ will be measured by the dimension $\dim V$ of its alphabet, whereas in the group case it will be indicated by the size $|\mathcal{V}|$ of its alphabet. Otherwise, we will use common notation for linear and group realizations.

We distinguish two types of variables. External variables, or symbols, are the variables of the code $C$ that is being realized. Internal variables, or states, are additional auxiliary variables introduced by the designer of the realization for some purpose. The essential difference is that internal variables may be changed at will, whereas external variables are fixed a priori.

Constraints will usually be defined by linear or group constraint codes $C_i$ that involve subsets of the variables, which define which combinations of values of those variables are valid. Some simple constraints are:

- An equality constraint is defined by a repetition code $C_{\equiv \mathcal{V}}$ of length $n$ over a set of $n$ variables with a common alphabet $\mathcal{V}$. In a graphical representation, an equality constraint will be denoted by a box containing an equals ($\equiv$) sign.

- A zero-sum constraint is defined by a zero-sum (single-parity-check) code $C_{\pm \mathcal{V}}$ of length $n$ over a set of $n$ variables with a common linear or group alphabet $\mathcal{V}$. In a graphical representation, a zero-sum constraint will be denoted by a box containing a plus ($+$) sign.

- A sign-inversion constraint is defined by a zero-sum code $C_{\sim \mathcal{V}}$ of length 2; i.e., the constraint is $v_1 + v_2 = 0$, or equivalently $v_1 = -v_2$. In a graphical representation, a sign-inversion constraint will be denoted by a box containing a negation ($\sim$) sign, or by a small circle.

- An isomorphism constraint is defined by a code $C = \{(a, \varphi(a)) \in A \times B : a \in A\}$, where $A$ and $B$ are isomorphic groups and $\varphi : A \to B$ is an isomorphism between them. Equality constraints of length 2 and sign-inversion constraints are examples of isomorphism constraints. In a graphical representation, an isomorphism constraint will be denoted by a box containing a left-right arrow ($\leftrightarrow$) sign.

A finite realization $R$ is thus defined by a set of symbol alphabets $A_k, k \in I_A$, a set of state alphabets $S_j, j \in I_S$, and a set of constraint codes $C_i, i \in I_C$, for some finite index sets $I_A, I_S$ and $I_C$. We define the symbol configuration space as the Cartesian (external direct) product $A = \prod_{k \in I_A} A_k$, and the state configuration space as $S = \prod_{j \in I_S} S_j$. Each constraint code $C_i$ involves a subset of the symbol variables with indices $I_{A(i)} \subseteq I_A$, and a subset of the state variables with indices $I_{S(i)} \subseteq I_S$. Thus $C_i$ is a subgroup or subspace of $A^{(i)} \times S^{(i)}$, where $A^{(i)} = \prod_{k \in I_{A(i)}} A_k$ and $S^{(i)} = \prod_{j \in I_{S(i)}} S_j$. 


The \textit{(internal) behavior} \( \mathcal{B} \) of \( \mathcal{R} \) is the set of valid configurations \((a, s) \in \mathcal{A} \times \mathcal{S} \) for which all constraints are satisfied; \textit{i.e.}, such that \( (a(i), s(i)) \in \mathcal{C}_i \) for all \( i \in \mathcal{I}_\mathcal{C} \). If \( \mathcal{R} \) is a linear or group realization, then \( \mathcal{B} \) is a subspace or subgroup of \( \mathcal{A} \times \mathcal{S} \). The \textit{code} or \textit{external behavior} \( \mathcal{C} \) realized by \( \mathcal{R} \) is the projection \( \mathcal{B}|_\mathcal{A} \); \textit{i.e.}, the set of all symbol configurations \( a \in \mathcal{A} \) such that \( (a, s) \in \mathcal{B} \) for some \( s \in \mathcal{S} \). If \( \mathcal{R} \) is a linear or group realization, then \( \mathcal{C} \) is a subspace or subgroup of \( \mathcal{A} \); \textit{i.e.}, a linear or group code. Two realizations are \textit{equivalent} if they realize the same code \( \mathcal{C} \).

A common graphical representation of a realization \( \mathcal{R} \) is a \textit{Tanner graph} \cite{b6}, namely a bipartite graph in which one set of vertices represents variables, a second set represents constraints, and an edge is drawn between a variable and a constraint if that variable is involved in that constraint. Figure 1(a) shows a generic Tanner graph, with closed circles representing symbol variables, open circles representing state variables, and boxes representing constraints.

![Tanner Graph](image)

\textbf{Figure 1:} (a) Tanner graph; (b) corresponding normal graph.

\section{2.2 Normal realizations}

The \textit{degree} of a variable is the number of constraints in which it is involved. A \textit{realization} is \textit{normal} if all symbol variables have degree 1, and all nontrivial state variables have degree 2\cite{b6}.

A normal realization is naturally represented by a \textit{normal graph} \cite{b3}, in which constraints are represented by vertices, state variables by edges, and symbol variables by half-edges, with a half-edge or edge incident on the one or two vertices representing the variable(s) in which the corresponding variable is involved.

As shown in \cite{b3}, any realization may be “normalized” with essentially no change in realization complexity as follows. To convert the Tanner graph of any realization into the normal graph of a corresponding normal realization, simply convert the variable vertices to equality constraints, and add half-edges in the case of symbol variables. For example, the generic Tanner graph of Figure 1(a) may be converted to the normal graph of Figure 1(b), where symbol half-edges are represented by a special “dongle” symbol (\( \vdash \)).

In view of this simple conversion, we may and will assume that all realizations are normal.

\footnote{Degree-1 state variables do not impose any constraints on a realization, so their alphabets may be assumed to be trivial, or they may be deleted. In this section we will assume that degree-1 state variables have been deleted.}
If the graph of a normal realization is disconnected, then the realization realizes the Cartesian product of the codes realized by each component. Therefore we may and will consider only normal realizations with connected graphs.

2.3 Extended behaviors

We now introduce the extended behavior $\mathcal{B} \subseteq A \times S \times S$ of a normal realization. We will see that although the extended behavior $\mathcal{B}$ is a redundant version of the behavior $\mathcal{B} \subseteq A \times S$, it yields the simplest known proof of the normal realization duality theorem, and also a nice development of the controllability properties of linear or group realizations.

Again, a normal realization $R$ is defined by symbol alphabets $A_k$, $k \in I_A$, state alphabets $S_j$, $j \in I_S$, and constraint codes $C_i$, $i \in I_C$. By the normal degree restrictions, as we run through all constraint codes $C_i$, each symbol variable appears precisely once, and each nontrivial state variable precisely twice. We denote the two values of a given state variable $S_j$ by $s_j$ and $s_j'$; it does not matter which one is primed. Then each element of the Cartesian (external direct) product $\prod_{i \in I_C} C_i$ of all constraint codes is a triple $(a, s, s') \in A \times S \times S$. At some risk of notational ambiguity, we will identify this Cartesian product with $R$; i.e., $R = \prod_{i \in I_C} C_i$.

We then define the extended behavior $\mathcal{B}$ as the set of all valid configurations in $R$; i.e., $\mathcal{B} = \{(a, s, s) \in R\}$. Evidently $\mathcal{B}$ is isomorphic to the behavior $\mathcal{B}$ via projection onto $A \times S$, and the code $C$ realized by $R$ is the projection of $\mathcal{B}$ onto $A$.

2.4 Fragments

We now begin our formal study of fragments of normal realizations. A fragment may range from a single constraint code to the entire realization. As we will see, realizations may be analyzed by studying the effects of connecting or disconnecting fragments.

If $R$ is a normal realization with a connected normal graph $G$, then a fragment $F$ of $R$ is a part of $R$ that corresponds to a connected subgraph $G_F$ of $G$ obtained by “cutting” certain edges of $G$ into two half-edges. (Vontobel and Loeliger call this “drawing a box” [20].)

Thus whereas a normal graph $G$ has three kinds of elements, namely constraint vertices, state edges, and symbol half-edges, a fragment has a fourth kind: namely, a state half-edge. We call the corresponding state variable an external state variable, relative to the fragment $F$. The set of all external state variables of $F$ is called the boundary $\partial(F)$ of $F$.

A fragment $F$ thus contains a nonempty subset $\{C_i : i \in I_CF\}$ of the constraint codes of $R$ (vertices of $G$), and the corresponding subset $\{A_k : k \in I_AF\}$ of symbol variables of $R$ (half-edges of $G$) that are involved in these constraint codes. It further contains the subset $\{S_j : j \in I_SF,\text{int}\}$ of the state variables of $R$ (edges of $G$) that are involved in two of the constraint codes of $F$ as internal state variables, again relative to the fragment $F$, as well as the set $\partial(F) = \{S_j : j \in I_SF,\text{ext}\}$ of external state variables of $F$. The respective configuration spaces are denoted by $A^F = \prod_{k \in I_AF} A_k$, $S^{F,\text{int}} = \prod_{j \in I_SF,\text{int}} S_j$, and $S^{F,\text{ext}} = \prod_{j \in I_SF,\text{ext}} S_j$.

There are two important special cases. A fragment with no internal state variables is simply a single constraint code $C_i$. A fragment with no external state variables is the entire normal realization $R$. 

5
A fragment has two kinds of half-edges, corresponding to symbol variables and external state variables, respectively. Again, the essential difference is that symbol variables are specified a priori, whereas state variables may be altered at will. To keep this distinction, we will continue to represent symbol variables in figures by our usual “dongle” symbol (⊢), whereas we will represent external state variables by an unadorned half-edge (–).

**Example 1** (trellis fragment). For example, Figure 2 shows a fragment \( F[j,k] \) of a trellis realization \( R \), as in [10]. The constraint codes are \( \{C_i : i \in [j,k]\} \); the symbol variables are \( \{A_i : i \in [j,k]\} \); the internal state variables are \( \{S_i : i \in (j,k)\} \); and the external state variables are \( \{S_j, S_k\} \), comprising the fragment boundary \( \partial(F[j,k]) \).

![Figure 2: Fragment \( F[j,k] \) of a trellis realization \( R \).](image)

The internal behavior \( B^F \) of a fragment \( F \) is the set of all configurations \( (a^F, s^F_{\text{ext}}, s^F_{\text{int}}) \in A^F \times S^F_{\text{ext}} \times S^F_{\text{int}} \) that satisfy all constraints \( C_i \) for all \( i \in \mathcal{I}_C \). Its external behavior \( C^F \) is the projection of \( B^F \) onto \( A^F \times S^F_{\text{ext}} \).

If \( F \) is the entire normal realization, then \( F \) has no external state variables, and these definitions reduce to those for a normal realization. If \( F \) is a constraint code \( C_i \) with no internal state variables, then \( B^F = C^F = C_i \).

We observe that if we conflate symbol and external state half-edges, then a fragment \( F \) is a normal realization of its external behavior \( C^F \).

The degree \( \deg(F) \) of a fragment \( F \) will be defined as the number of its external state variables; i.e., the size \( |\partial(F)| \) of its boundary. As already noted, a fragment of degree 0 is simply a normal realization. Fragments of degrees 1, 2, 3, and \( \geq 4 \) will be called leaf, trellis, cubic and hypercubic fragments, respectively. Similarly, constraint codes of degrees 1, 2, 3, and \( \geq 4 \) will be called leaf, trellis, cubic and hypercubic constraint codes, respectively.

### 3 Groups, vector spaces and duality

In this section, we will develop in parallel the general principles of realizations of group codes over finite abelian groups and of linear codes over a field \( \mathbb{F} \). The two theories involve similar algebra, and indeed coincide when \( \mathbb{F} \) is a finite field. All proofs will be group-theoretic, and therefore we will generally use group-theoretic language, although from time to time we will translate our results into vector space (linear algebra) terminology, which is no doubt more familiar to most readers.

#### 3.1 Finite abelian groups and vector spaces

If \( G \) is an abelian group, then any subgroup \( H \subseteq G \) is automatically normal, and the cosets of \( H \) in \( G \) form a quotient group \( G/H \). If \( G \) is finite, then \( |G| = |H||G/H| \), and given any set \( [G/H] \) of coset representatives for \( G/H \), every element \( g \in G \) can be uniquely represented as a sum \( g = h + r \) with \( h \in H, r \in [G/H] \). However, in general \( [G/H] \) cannot be taken as a subgroup of \( G \); e.g., there is no subgroup of \( \mathbb{Z}_4 \) that can be taken as a set of coset representatives for \( \mathbb{Z}_4/2\mathbb{Z}_4 \).
More generally, a normal series is a chain of subgroups \( G_n = \{0\} \subseteq G_{n-1} \subseteq \cdots \subseteq G_0 = G \). The factor groups of the normal series are the quotient groups \( G_i/G_{i+1} \). (Alternatively, we may let \( G_n \neq \{0\} \), in which case we will regard \( G_n \) as a factor group.) If \( G \) is finite, then we have \( |G| = \prod_i |G_i/G_{i+1}| \). If \( |G_i/G_{i+1}|, 0 \leq i \leq n-1 \), are any sets of coset representatives for each quotient group, then the elements of \( G \) may be uniquely represented as sums of coset representatives, one from each set; this is called a chain coset decomposition.

Similarly, if \( V \) is a finite-dimensional vector space over \( \mathbb{F} \), then the cosets of any subspace \( W \) in \( V \) form a quotient space \( V/W \) such that \( \dim V = \dim W + \dim V/W \). For vector spaces, it is always possible to find a coset representative subspace \( [V/W] \subseteq V \) of dimension \( \dim V/W \) such that \( V \) is equal to the (internal) direct product \( W \times [V/W] \), so any basis for \( W \) and basis for \([V/W]\) together form a basis for \( V \).

More generally, a normal series of subspaces of a vector space \( V \) is a chain of subspaces \( V_n = \{0\} \subseteq V_{n-1} \subseteq \cdots \subseteq V_0 = V \). The factors of the normal series are the quotient spaces \( V_i/V_{i+1} \). (If we let \( V_n \neq \{0\} \), then \( V_n \) is also a factor.) If \( V \) is finite-dimensional, then we have \( \dim V = \sum_i \dim V_i/V_{i+1} \). If \( |V_i/V_{i+1}|, 0 \leq i \leq n-1 \), are coset representative subspaces for the quotient spaces, then \( V \) is the direct product of the subspaces \( [V_i/V_{i+1}] \), \( 0 \leq i \leq n-1 \), so the union of any set of bases for these subspaces is a basis for \( V \). Thus in the vector space case, unlike the group case, the order of factors does not matter.

The fundamental theorem of homomorphisms says that if \( f : G \to H \) is a homomorphism with image \( f(G) \) and kernel \( K \subseteq G \), then \( K \) is a normal subgroup of \( G \), and \( f(G) \cong G/K \). On the other hand, if \( H \) is any normal subgroup of \( G \), then the natural map \( \pi : G \to G/H \) defined by \( \pi(g) = H + g \) is a homomorphism with kernel \( H \) and image \( G/H \).

The correspondence theorem says that if \( H \) is a subgroup of \( G \), and \( f \) is a homomorphism such that \( f : G \to f(G) \) and \( f : H \to f(H) \) have the same kernel \( K \subseteq G \), then \( G/H \cong f(G)/f(H) \). Thus the natural map \( \pi : G \to G/G_n \) maps a normal series \( G_n \subseteq G_{n-1} \subseteq \cdots \subseteq G_0 = G \) to a normal series of quotients, \( \{0\} \subseteq G_{n-1}/G_n \subseteq \cdots \subseteq G/G_n \), whose factor groups are isomorphic to those of the original series, apart from \( G_n \).

### 3.2 Duality

A finite abelian group \( G \) has a dual group \( \hat{G} \) (namely, its character group) such that there exists a well-defined pairing \( \langle \hat{g}, h \rangle \in \mathbb{R}/\mathbb{Z} \) for all \( v \in V, \hat{v} \in \hat{V} \) that is bihomomorphic: i.e., \( \langle 0, v \rangle = \langle \hat{v}, 0 \rangle = 0, \langle \hat{v}_1 + \hat{v}_2, v \rangle = \langle \hat{v}_1, v \rangle + \langle \hat{v}_2, v \rangle \), and so forth. In the finite abelian case, \( G \) and \( \hat{G} \) are isomorphic.

If \( H \) is a subgroup of \( G \), then the orthogonal subgroup \( H^\perp \) is a subgroup of \( \hat{G} \), and \( \langle H^\perp \rangle^\perp = H \). The product \( |H||H^\perp| \) is equal to \( |G| = |\hat{G}| \). Moreover, \( H^\perp \) acts as the dual group to the quotient group \( G/H \), with the pairing \( \langle \hat{g}, H + g \rangle = \langle \hat{g}, g \rangle \) for \( \hat{g} \in H^\perp, H + g \in G/H \), so in the finite abelian case \( H^\perp \) is actually isomorphic to \( G/H \). More generally, we have quotient group duality: if \( J \subseteq H \subseteq G \), then the quotient group \( J^\perp/H^\perp \) acts as the dual group to \( H/J \).

Similarly, a finite-dimensional vector space \( V \) over \( \mathbb{F} \) has a dual space \( \hat{V} \) of the same dimension such that there exists a well-defined inner product \( \langle \hat{v}, v \rangle \in \mathbb{F} \) for all \( v \in V, \hat{v} \in \hat{V} \) that is bilinear: i.e., \( \langle 0, v \rangle = \langle \hat{v}, 0 \rangle = 0, \langle \hat{v}_1 + \hat{v}_2, v \rangle = \langle \hat{v}_1, v \rangle + \langle \hat{v}_2, v \rangle \), and so forth.

If \( W \) is a subspace of \( V \), then the orthogonal subspace \( W^\perp \) is a subspace of \( \hat{V} \), and \( \langle W^\perp \rangle^\perp = W \). The sum \( \dim W + \dim W^\perp \) is equal to \( \dim V = \dim \hat{V} \). Moreover, \( W^\perp \) acts as the dual space to \( V/W \), with the inner product \( \langle \hat{v}, W + v \rangle = \langle \hat{v}, v \rangle \) for \( \hat{v} \in W^\perp, W + v \in V/W \).
If $G = \Pi_k G_k$ is a finite Cartesian (external direct) product of a collection of groups or vector spaces $G_k$, then the dual group or space to $G$ is $\hat{G} = \Pi_k \hat{G}_k$, and the pairing or inner product between $g \in G$ and $\hat{g} \in \hat{G}$ is given by the componentwise sum

$$\langle \hat{g}, g \rangle = \sum_k \langle \hat{g}_k, g_k \rangle.$$ 

If $H = \Pi_k H_k \subseteq V$ is a Cartesian product of subgroups or subspaces $H_k \subseteq G_k$, then the orthogonal group or space is the Cartesian product $H^\perp = \Pi_k (H_k)^\perp \subseteq \hat{G}$.

### 3.3 Projection/cross-section duality

The most useful duality relationship for us will be projection/cross-section duality. If $A$ and $B$ are groups (or vector spaces), and $C$ is a subgroup (or subspace) of $A \times B$, then the projection of $C$ onto $A$ is defined as $C|_A = \{ a \in A : (a, b) \in C \text{ for some } b \in B \}$, and the cross-section of $C$ onto $A$ is defined as $C_A = \{ a \in A : (a, 0) \in C \}$. The projection/cross-section duality theorem says that if $C^\perp \subseteq \hat{A} \times \hat{B}$ is the orthogonal subgroup (or subspace) to $C$, then $(C|_A)^\perp = (C_A)^\perp \subseteq \hat{A}$. For a one-line proof, see [3].

We illustrate projection/cross-section duality in Figure 3. In Figure 3(a), the constraint $C$ is illustrated as a constraint code on the two variables $A$ and $B$. If we introduce another constraint on $B$, namely the dummy constraint $b \in B$ (triangular symbol), then $B$ becomes an internal state variable, and the resulting realization can be seen to realize the projection $C|_A = \{ a \in A : (a, b) \in C \text{ for some } b \in B \}$.

Similarly, in Figure 3(b), the dual constraint, defined by the orthogonal code $C^\perp$, is illustrated as a constraint on the two dual variables $\hat{A}$ and $\hat{B}$. If we introduce another constraint on $\hat{B}$, namely the zero constraint $\hat{b} = 0$ (triangular symbol), then $\hat{B}$ becomes an internal state variable, and the resulting realization evidently realizes the cross-section $(C^\perp)|_{\hat{A}} = \{ \hat{a} \in \hat{A} : (a, 0) \in C^\perp \}$. (After normal realization duality has been discussed in Section 3.5 it will become clear that Figure 3(b) is the dual realization to the normal realization of Figure 3(a), since $B^\perp = \{ 0 \}$.)

It will be convenient to introduce special symbols for the dummy (“free,” “open”) and zero (“grounded,” “pinned”) constraints. We illustrate our conventions in Figure 4.

### 3.4 Sum/intersection duality

Another useful duality relationship is sum/intersection duality: if $A$ and $B$ are subgroups of $G$, then $(A + B)^\perp = A^\perp \cap B^\perp$, a subgroup of the dual group $\hat{G}$. 

8
We illustrate sum/intersection duality in Figure 5. In Figure 5(a), the sum $A + B$ is realized by a symbol variable incident on a zero-sum constraint of length 3 over $G$, with the two other incident state variables constrained to $A \subseteq G$ and $B \subseteq G$, respectively. In Figure 5(b), the intersection $A^\perp \cap B^\perp$ is realized by a dual symbol variable incident on an equality constraint of length 3 over $\hat{G}$, with the two other incident state variables constrained to $A^\perp \subseteq \hat{G}$ and $B^\perp \subseteq \hat{G}$, respectively. (Again, after Section 3.5 it will become clear that Figures 5(a) and 5(b) depict dual normal realizations, since equality and zero-sum constraints are duals. No sign inversions are needed, because $-G = G$.)

$$\begin{array}{c}
A + B \\
\langle A \rangle \\
G \\
\langle B \rangle
\end{array}$$

$$\begin{array}{c}
A^\perp \cap B^\perp \\
\langle A^\perp \rangle \\
\hat{G} \\
\langle B^\perp \rangle
\end{array}$$

Figure 5: (a) sum $A + B$; (b) intersection $A^\perp \cap B^\perp$.

The sum $A + B$ is said to be an (internal) direct product $A \times B$, and $A$ and $B$ are said to be independent, if and only if $A \cap B = \{0\}$. Then and only then every element of $A + B$ may be expressed uniquely as a sum $a + b$, so $|A + B| = |A||B|$ in the group case, or $\dim(A + B) = \dim A + \dim B$ in the linear case. In particular, $a + b = 0$ implies $a = b = 0$. By sum/intersection duality, $A$ and $B$ are independent if and only if $A^\perp + B^\perp = G$.

3.5 Normal realization duality

We now define the dual realization $R^\circ$ to a normal linear or group realization $R$ as in [3], and give the simplest proof of the normal realization duality theorem that we know, after Koetter [14].

Again, $R$ is defined by symbol alphabets $A_k, k \in \mathcal{I}_A$, state alphabets $S_j, j \in \mathcal{I}_S$, and constraints $C_i, i \in \mathcal{I}_C$, with the normal degree restrictions that each symbol variable has degree 1 and each nontrivial state variable has degree 2. Denoting the two values of each state variable $S_j$ by $s_j$ and $s'_j$, we identify $R$ with the Cartesian product $R = \prod_{i \in \mathcal{I}_C} C_i \subseteq A \times S \times S$. The extended behavior of $R$ is $\mathcal{B} = \{(a, s, s) \in R\}$, its behavior is $\mathcal{B} = \mathcal{B}|_{A \times S} \cong \mathcal{B}$, and it realizes the code $C = \mathcal{B}|_A = \mathcal{B}|_A$.

The dual realization $R^\circ$ to $R$ is then defined as the normal realization with the dual symbol alphabets $\hat{A}_k, k \in \mathcal{I}_A$, the dual state alphabets $\hat{S}_j, j \in \mathcal{I}_S$, the orthogonal constraint codes $(C_i)^\perp, i \in \mathcal{I}_C$, and finally sign inversion constraints $\hat{s}_j = -s'_j$ on the two values of each dual state variable $\hat{S}_j$.

The Cartesian product of all dual constraint codes is thus $\prod_{i \in \mathcal{I}_C} (C_i)^\perp = R^\perp$. The extended behavior of $R^\circ$ is the set $\hat{\mathcal{B}}^\circ = \{(\hat{a}, s, -s) \in R^\perp\}$ of all valid configurations $(\hat{a}, s, -s) \in R^\perp$. Its behavior is $\hat{\mathcal{B}}^\circ = (\hat{\mathcal{B}}^\circ)|_{\hat{A} \times \hat{S}} \cong \hat{\mathcal{B}}^\circ$, and it realizes the code $C^\circ = (\hat{\mathcal{B}}^\circ)|_A = (\hat{\mathcal{B}}^\circ)|_{\hat{A}}$.

We further define the constraint space of $R$ as the orthogonal space $\mathcal{B}^\perp$ to the extended behavior $\mathcal{B}$; thus $(a, s, s') \in \mathcal{B}$ if and only if $(a, s, s') \perp \mathcal{B}^\perp$. We note that the extended behavior $\hat{\mathcal{B}}$ may be expressed as $\hat{\mathcal{B}} = R \cap \mathcal{V}$, where $\mathcal{V}$ denotes the validity space $\mathcal{V} = \{(a, s, s) \in A \times S \times S\}$; i.e., the valid configurations $(a, s, s) \in A \times S \times S$ are those that satisfy both the constraints of $R$ and the constraints of $\mathcal{V}$. It follows by sum/intersection duality that the constraint space may be expressed as $\mathcal{B}^\perp = R^\perp + \mathcal{V}^\perp$, where $\mathcal{V}^\perp = \{(0, s, -s) \in A \times S \times S\}$ is the validity constraint space.
Lemma. The code $C^\circ$ realized by $\mathcal{R}^\circ$ is the constraint space cross-section $(\mathfrak{B}^\perp)_{\hat{A}}$.

Proof: For $(\hat{a}, \hat{s}, \hat{s}') \in R^\perp$, the coset $(\hat{a}, \hat{s}, \hat{s}') + \mathcal{V}^\perp$ contains the element $(\hat{a}, 0, 0)$ if and only if $\hat{s} = -\hat{s}'$. Thus $C^\circ = (R^\perp + \mathcal{V}^\perp)_{\hat{A}}$. $\square$

Theorem (normal realization duality). $C^\circ = C^\perp$.

Proof: By projection/cross-section duality, $C^\circ = (\mathfrak{B}^\perp)_{\hat{A}} = (\mathfrak{B}^\perp | A) = C^\perp$. $\square$

The normal realization duality theorem is illustrated by the dual normal realizations of Figure 6. Figure 6(a) realizes $C$ as the set of all $a \in A$ such that there exists some $(a, s, s') \in R$ with $s = s'$. Figure 6(b) realizes $C^\circ$ as the set of all $\hat{a} \in \hat{A}$ such that there exists some $(\hat{a}, \hat{s}, \hat{s}') \in R^\perp$ with $\hat{s} = -\hat{s}'$. Since these realizations are duals, by normal realization duality we have $C^\circ = C^\perp$.

\begin{figure}[h]
\centering
\begin{tabular}{cc}
\hspace{2cm} & \hspace{2cm} \\
$C$ & $\mathcal{R}$ \\
\hspace{1cm} $\mathfrak{s} \in S$ & $\mathfrak{s}' \in S'$ \\
\hspace{1cm} (a) & (b) \\
\end{tabular}
\caption{Dual normal realizations of $C$ and $C^\perp$.}
\end{figure}

More generally, since a fragment $F$ is a normal realization of its external behavior $C^F$, we will define the dual fragment $F^\circ$ as the dual normal realization to $F$. Then, by normal realization duality, $F^\circ$ realizes $(C^F)^\perp$.

3.6 Generalized normal realization duality

The normal realization duality theorem may be generalized so as to exhibit greater symmetry between primal and dual realizations as follows.

The primal generalized realization $R$ is defined as before, but with extended behavior $\mathfrak{B} = \{(a, s, s') \in R : s'_j = \varphi_j(s_j), \forall j \in I_S\}$, where, for each $j$, $\varphi_j : S_j \rightarrow S_j'$ is an isomorphism between state spaces $S_j$ and $S_j'$. The dual generalized realization $R^\circ$ is then defined with extended behavior $\mathfrak{B}^\circ = \{(\hat{a}, \hat{s}, \hat{s}') \in R^\circ : \hat{s}_j = -\hat{\varphi}_j(\hat{s}_j'), \forall j \in I_S\}$, where $\hat{\varphi}_j : \hat{S}_j' \rightarrow \hat{S}_j$ is the adjoint isomorphism, namely the unique homomorphism $\hat{\varphi}_j : \hat{S}_j' \rightarrow \hat{S}_j$ such that $(\hat{\varphi}_j(\hat{s}_j'), s_j) = (\hat{s}_j', \varphi_j(s_j))$ for all $s_j \in S_j, \hat{s}_j' \in \hat{S}_j'$. If $\varphi_j$ is the equality isomorphism on $S_j$, then its adjoint $\hat{\varphi}_j$ is the equality isomorphism on $\hat{S}_j$, so its negative adjoint is the sign-inversion isomorphism on $\hat{S}_j$.

The constraint codes $\{(s_j, \varphi_j(s_j)) : s_j \in S_j\}$ and $\{(-\hat{\varphi}_j(\hat{s}_j'), s_j') : \hat{s}_j' \in \hat{S}_j'\}$ are then orthogonal for each $j \in I_S$, so by normal realization duality the codes realized by $R$ and $R^\circ$ are orthogonal.

Figure 7 illustrates generalized dual normal realizations. The box labeled by $\leftrightarrow$ in Figure 7(a) represents the isomorphism constraints $\{s'_j = \varphi_j(s_j)\}$, whereas the box labeled by $\Leftrightarrow$ and a small circle (representing a sign inversion) in Figure 7(b) represents the negative adjoint isomorphism constraints $\{\hat{s}_j = -\hat{\varphi}_j(\hat{s}_j')\}$.

\begin{figure}[h]
\centering
\begin{tabular}{cc}
\hspace{2cm} & \hspace{2cm} \\
$C$ & $\mathcal{R}$ \\
\hspace{1cm} $\mathfrak{s} \in S$ & $\mathfrak{s}' \in S'$ \\
\hspace{1cm} (a) & (b) \\
\end{tabular}
\caption{Dual generalized normal realizations of $C$ and $C^\perp$.}
\end{figure}

2This argument is not circular, since the normal realization duality theorem has already been proved algebraically.
Moreover, we may correspondingly generalize normal graphs so as to exhibit greater symmetry between primal and dual graphs as follows. The primal generalized normal graph is defined as before, except that the ends of each generalized edge represent values \( s_j \) and \( s'_j \) of isomorphic state spaces \( S_j \) and \( S'_j \), subject to some isomorphism constraint \( s'_j = \varphi_j(s_j) \). In the dual generalized normal graph, each dual generalized edge represents the negative adjoint isomorphism constraint \( \hat{s}_j = -\hat{\varphi}_j(\hat{s}'_j) \) between the dual state spaces \( \hat{S}_j \) and \( \hat{S}'_j \). With such generalized edges, the primal and dual graphs then have the same graph topology.

### 3.7 Observability and controllability of realizations

We now discuss the properties of observability and controllability of realizations as defined in [7].

Using extended behaviors and normal realization duality, we obtain an elegant proof of observability/controllability duality, and a nice generalization of the controllability test of [7].

A realization \( \mathcal{R} \) is called observable, or one-to-one, if there is precisely one configuration \( (a, s) \in \mathcal{B} \) corresponding to each \( a \in \mathcal{C} = \mathcal{B}_{|A} \). Alternatively, \( \mathcal{R} \) is observable if there is precisely one configuration \( (a, s, s) \) in the extended behavior \( \hat{\mathcal{B}} \) for each \( a \in \mathcal{C} = \mathcal{B}_{|A} \).

Equivalently, \( \mathcal{R} \) is observable if the kernel \( \mathcal{B}^u = \{(0, s) \in \mathcal{B}\} \) of the projection of the behavior \( \mathcal{B} \) onto \( \mathcal{A} \) is trivial, or if the kernel \( \hat{\mathcal{B}}^u = \{(0, s, s) \in \mathcal{R}\} \) of the projection of the extended behavior \( \hat{\mathcal{B}} \) onto \( \hat{\mathcal{A}} \) is trivial. We call \( \mathcal{B}^u \) the unobservable behavior; it consists of all unobservable valid configurations \( (0, s) \in \mathcal{B} \). Similarly, the unobservable extended behavior \( \hat{\mathcal{B}}^u \) consists of all unobservable valid extended configurations \( (0, s, s) \in \hat{\mathcal{B}} \). We define the unobservable state configuration space \( \mathcal{S}^u \) as \( \{(0, s) \in \mathcal{R}\} \); thus \( \mathcal{S}^u \cong \mathcal{B}^u \cong \mathcal{B}^u \), and \( \mathcal{R} \) is observable if and only if \( \mathcal{S}^u \) is trivial.

By the fundamental theorem of homomorphisms, \( \mathcal{B}/\mathcal{B}^u \cong \mathcal{C} \); alternatively, \( \hat{\mathcal{B}}/\hat{\mathcal{B}}^u \cong \hat{\mathcal{C}} \). Thus \( |\mathcal{B}|/|\mathcal{S}^u| = |\mathcal{C}| \), or in the linear case \( \dim \mathcal{B} - \dim \mathcal{S}^u = \dim \mathcal{C} \). Hence \( |\mathcal{C}| \leq |\mathcal{B}| \), or in the linear case \( \dim \mathcal{C} \leq \dim \mathcal{B} \), with equality if and only if \( \mathcal{R} \) is observable; this is the basic observability test for \( \mathcal{R} \).

In other words, by definition, a realization is unobservable if and only if its internal behavior is redundant, in the sense that \( |\mathcal{B}| > |\mathcal{C}| \).

Figure 8(a) shows a normal realization of the unobservable state configuration space \( \mathcal{S}^u \) as the set of all \( s \in \mathcal{S} \) such that there exists an unobservable valid extended configuration \( (0, s, s) \in \mathcal{R} \).

![Figure 8: Dual normal realizations of \( \mathcal{S}^u \) and \( \hat{\mathcal{S}}^c \).](image)

Dually, we will say that a realization \( \mathcal{R} \) is controllable, or has independent constraints, if its realization constraint space \( \mathcal{R}^\perp = \prod_{i \in \mathcal{C}} (\mathcal{C}_i)^\perp \) and the validity constraint space \( \mathcal{V}^\perp = \{(0, \hat{s}, -\hat{s}) \in \hat{\mathcal{A}} \times \hat{\mathcal{S}} \times \hat{\mathcal{S}}\} \) are independent.

Thus \( \mathcal{R} \) is controllable if and only if the constraint space \( \mathcal{B}^\perp = \mathcal{R}^\perp + \mathcal{V}^\perp \) is the internal direct product \( \mathcal{B}^\perp \times \mathcal{V}^\perp \), or equivalently if and only if \( \mathcal{R}^\perp \cap \mathcal{V}^\perp = \{0\} \). But \( \mathcal{R}^\perp \cap \mathcal{V}^\perp = \{(0, \hat{s}, -\hat{s}) \in \mathcal{R}^\perp\} = (\mathcal{B}^\perp)^\perp \), the unobservable extended behavior of the dual realization \( \mathcal{R}^\circ \). This establishes observability/controllability duality [7] Theorem 4]: \( \mathcal{R} \) is controllable if and only if \( \mathcal{R}^\circ \) is observable.

We note that, by sum/intersection duality, \( \mathcal{R}^\perp \cap \mathcal{V}^\perp = \{0\} \) if and only if \( \mathcal{R} + \mathcal{V} = \mathcal{A} \times \mathcal{S} \times \mathcal{S} \); thus \( \mathcal{R} \) is controllable if and only if \( \mathcal{R} + \mathcal{V} = \mathcal{A} \times \mathcal{S} \times \mathcal{S} \).
The dual normal realization to that of Figure 8(a) is that of Figure 8(b), because (a) the orthogonal constraint to $\mathcal{R} = \prod_i C_i$ is $\mathcal{R}^\perp = \prod_i (C_i)^\perp$; (b) the orthogonal constraint to the equality constraint of length 3 over $\mathcal{S}$ is the zero-sum constraint of length 3 over $\hat{\mathcal{S}}$; (c) the internal state variables $\hat{s}, \hat{s}'$ are negated in this zero-sum constraint (the sign inversions are represented by small circles), so the zero-sum constraint is $-\hat{s} - \hat{s}' + \hat{s}^c = 0$, or equivalently $\hat{s}^c = \hat{s} + \hat{s}'$. Thus this dual realization realizes $\{\hat{s} + \hat{s}' : (\hat{a}, \hat{s}, \hat{s}') \in \mathcal{R}^\perp\}$, which we define as the controllable subspace $\hat{\mathcal{S}}^c \subseteq \hat{\mathcal{S}}$ of the dual realization $\mathcal{R}^\circ$.

By normal realization duality, we have immediately:

**Theorem (unobservable state configuration space/controllable subspace duality)** The unobservable state configuration space $\mathcal{S}^u$ of $\mathcal{R}$ and the controllable subspace $\hat{\mathcal{S}}^c$ of $\mathcal{R}^\circ$ are orthogonal; i.e., $\hat{\mathcal{S}}^c = (\mathcal{S}^u)^\perp$. Thus $\mathcal{R}^\circ$ is controllable if and only if $\hat{\mathcal{S}}^c = \hat{\mathcal{S}}$.

A primal configuration $(a, s, s') \in \mathcal{R}$ is valid if $s = s'$, whereas a dual configuration $(\hat{a}, \hat{s}, \hat{s}') \in \mathcal{R}^\perp$ is valid if $\hat{s} = -\hat{s}'$. The controllable subspace of a primal realization $\mathcal{R}$ must therefore be defined with the inverse sign, namely $\mathcal{S}^c = \{s - s' : (a, s, s') \in \mathcal{R}\}$. Then, by normal realization duality, we have $\mathcal{S}^c = (\hat{\mathcal{S}}^u)^\perp$.

It follows that $\hat{\mathcal{S}}^u$ acts as the dual group or space to $\mathcal{S}/\mathcal{S}^c$, which in our setting implies that they are isomorphic. Thus $|\hat{\mathcal{S}}^u| = |\mathcal{S}|/|\mathcal{S}^c|$; or, in the linear case, $\dim \hat{\mathcal{S}}^u = \dim \mathcal{S} - \dim \mathcal{S}^c$. If we take $|\mathcal{S}|/|\mathcal{S}^c|$ or $\dim \mathcal{S} - \dim \mathcal{S}^c$ as a measure of the uncontrollability of $\mathcal{R}$, and $|\hat{\mathcal{S}}^u|$ or $\dim \hat{\mathcal{S}}^u$ as a measure of the unobservability of $\mathcal{R}^\circ$, then this says that they are both “the same size.”

Finally, we observe that $\mathcal{S}^c$ is the image of the validity check homomorphism $\mathcal{R} \rightarrow \mathcal{S}$ defined by $(a, s, s') \rightarrow s - s'$, whose kernel is the extended behavior $\mathcal{B}$. By the fundamental theorem of homomorphisms, we thus have $\mathcal{R}/\mathcal{B} \cong \mathcal{S}^c$. We therefore obtain the following generalization of the controllability test of [7, Theorem 6]:

**Theorem (controllability test)** For a linear or group realization $\mathcal{R}$ with behavior $\mathcal{B}$ and controllable subspace $\mathcal{S}^c \subseteq \mathcal{S}$, we have $|\mathcal{R}|/|\mathcal{B}| = |\mathcal{S}^c| \leq |\mathcal{S}|$, or in the linear case $\dim \mathcal{R} - \dim \mathcal{B} = \dim \mathcal{S}^c \leq \dim \mathcal{S}$, with equality if and only if $\mathcal{R}$ is controllable.

Alternatively, this controllability test follows directly from the fact that $|\mathcal{B}^\perp| \leq |\mathcal{R}^\perp|$, with equality if and only if $\mathcal{R}$ is controllable, since $|\mathcal{B}^\perp| = |A|\mathcal{S}^2/|\mathcal{B}|$, $|\mathcal{R}^\perp| = |A|\mathcal{S}^2/|\mathcal{R}|$, and $|\mathcal{V}^\perp| = |\mathcal{S}|$. In other words, by definition, a realization is uncontrollable if and only if its constraints are redundant, in the sense that $|\mathcal{R}^\perp + \mathcal{V}^\perp| < |\mathcal{R}^\perp|$. |

4 External properties of fragments

In this section we begin our analysis of realizations via fragments. Our main tool will be a simple but fundamental structure theorem for subdirect products [13] (subgroups of a direct product $A \times B$—i.e., length-2 group codes). We define trimness and properness for fragments, and generalize various results of [7] from constraint codes to fragments. We show how symbol variables may be reduced.

We give a basic structure theorem for leaf fragments, which characterizes the external state space for any trim and proper leaf fragment. Using this result, we obtain an improved proof of the “minimal $\Leftrightarrow$ trim and proper” theorem, which is the key result for cycle-free realizations. For trim and proper cyclic realizations, this result allows a decomposition into cycle-free leaf fragments and a leafless “2-core,” which is the cyclic essence of the realization.
4.1 Fundamental theorem of subdirect products

We will use repeatedly the following fundamental result, which establishes the structure of any linear or group constraint on two variables. For further discussion, see Section VIII-D.

**Fundamental theorem of subdirect products.** Given groups $A$ and $B$ and a subgroup $C \subseteq A \times B$, let $C_A$ and $C_B$ be the projections of $C$ onto $A$ and $B$, respectively, and let $C_A$ and $C_B$ be the cross-sections of $C$ on $A$ and $B$, respectively. Then $C_A/C_B \cong C/B/C_B \cong C/(C_A \times C_B)$.

**Proof:** Evidently $C_A \times C_B \subseteq C$. Since the kernels of the projections of $C$ and $C_A \times C_B$ onto $B$ are both equal to $C_A$, and their images are equal to $C_B$ and $C_B$, respectively, by the correspondence theorem we have $C/(C_A \times C_B) \cong C_B/C_B$. Similarly, $C/(C_A \times C_B) \cong C_A/C_A$.

This theorem and its proof are illustrated in Figure 9 where $\pi_A$ and $\pi_B$ denote projections onto $A$ and $B$, respectively. The theorem simply says that the three normal series $C_A \times C_B \subseteq C$, $C_A \subseteq C_B$, and $C_B \subseteq C_B$ have isomorphic factor groups.

![Fundamental theorem of subdirect products](image)

Figure 9: Fundamental theorem of subdirect products.

Figure 10(a) shows a generic normal realization of a subdirect product $C \subseteq A \times B$ according to this theorem.

![Realizations of dual subdirect products](image)

Figure 10: Realizations of dual subdirect products.

The first constraint in Figure 10(a) is $\{(a, a + C_A) \in A \times C_A \mid a \in C_A\}$. This may be viewed as the combination of constraints based on the inclusion map $A \hookrightarrow C_A$ and the natural map $C_A \twoheadrightarrow C_A/C_A$, or equivalently constraints based on the natural map $A \twoheadrightarrow A/C_A$ and the inclusion map $A/C_A \hookrightarrow C_A/C_A$. The isosceles trapezoids denote such inclusion/natural-map constraints on two variables, and indicate which variable alphabet is smaller.

The central constraint in Figure 10(a), denoted by a rectangular box labeled by $\leftrightarrow$, enforces the isomorphism $C_A/C_A \leftrightarrow C_B/C_B$. Note that this isomorphism constraint could be represented by a generalized edge, as in Section VIII-D. The last constraint is the inclusion/natural-map constraint $\{(b, b + C_B) \in B \times C_B \mid b \in C_B\}$.

Notice that, as shown in Figures 9 and 10 if we impose a zero constraint on $B$, then Figure 10(a) realizes $C_A$, whereas if we impose a dummy constraint on $B$, then Figure 10(a) realizes $C_A$.

Figure 10(b) shows the dual normal realization of the orthogonal subdirect product $C^\perp \subseteq A^\perp \times B^\perp$. Note that by projection/cross-section duality $(C^\perp)_A = (C_A)^\perp$ and $(C^\perp)_A = (C_A)^\perp$; therefore, by quotient group duality, $(C^\perp)_A/(C^\perp)_A$ acts as the dual group to $C_A/C_A$. Similarly, $(C^\perp)_B/(C^\perp)_B$.

---

Footnote: More generally, by imposing an appropriate constraint on $B$, we can make Figure 10(a) realize the quotient group $D/C_A$ for any $D$ such that $C_A \subseteq D \subseteq C_A$ is a normal series. This is the essence of the “most beautiful behavioral control theorem” [19].
acts as the dual group to $C_B/C_B$. The dual isomorphism in Figure 10(b) is the negative adjoint isomorphism to that in Figure 10(a); see Section 3.6.

**Remark.** We remark that any homomorphism $\varphi : A \to B$ may be viewed a special case of Figure 10(a), in which the first constraint is based on the natural map $A \to A/(\ker \varphi)$, and the last is based on the inclusion map $\im(\varphi) \hookrightarrow B$; the fundamental theorem of subdirect products then reduces to the fundamental theorem of homomorphisms, namely $\im(\varphi) \cong A/(\ker \varphi)$. Note that a homomorphism $\phi : B \to A$ in the reverse direction is also a special case. In this sense, a subdirect product $C \subseteq A \times B$ may be seen as a bidirectional (or behavioral) generalization of a homomorphism.

Moreover, if Figure 10(a) represents any homomorphism $\varphi : A \to B$, then it is easily seen that the dual realization of Figure 10(b) represents the negative adjoint homomorphism $-\varphi : B \to A$ (see Section 3.6 and [3, Section VIII-C]). Thus an orthogonal subdirect product $C^\perp \subseteq \hat{A} \times \hat{B}$ may be seen as a bidirectional generalization of a negative adjoint homomorphism.

### 4.2 Trimness and properness for fragments

In [7] we defined trimness and properness for constraint codes, showed that these were dual properties, and showed that lack of either of these properties implies local reducibility. In this section we will straightforwardly generalize these results to fragments.

A fragment $\mathcal{F}$ with external behavior $\mathcal{C}_F \subseteq \mathcal{A}_F \times \mathcal{S}_{F,\text{ext}}$ will be called *trim* if the projection of $\mathcal{C}_F$ onto every external state variable $\mathcal{S}_j$ of $\mathcal{F}$ is $\mathcal{S}_j$; i.e., if every such projection is surjective (onto). $\mathcal{F}$ will be called *proper* if there is no element of $\mathcal{C}_F$ whose support is a single external state variable $\mathcal{S}_j$ of $\mathcal{F}$; i.e., if for all $\mathcal{S}_j$ the cross-section $(\mathcal{C}_F)_{|\mathcal{S}_j}$ is trivial, so the projection of $\mathcal{C}_F$ onto $\mathcal{S}_j$ is injective (one-to-one). These generalize the corresponding definitions for constraint codes [7].

By projection/cross-section duality, $(\mathcal{C}_F)_{|\mathcal{S}_j} = \mathcal{S}_j$ if and only if $((\mathcal{C}_F)_{|\mathcal{S}_j}) = \{0\}$. Thus we have *trim/proper duality*: a fragment $\mathcal{F}$ is trim at $\mathcal{S}_j$ if and only if its dual fragment $\mathcal{F}^\circ$ is proper at $\mathcal{S}_j$. This is a straightforward generalization of trim/proper duality for constraint codes [7, Theorem 1].

We now show that if $\mathcal{F}$ is not trim or proper at some external state $\mathcal{S}_j$, then $\mathcal{R}$ is locally reducible. This is a straightforward generalization of [7, Theorem 2].

We partition the variables involved in $\mathcal{C}_F$ into two subsets, one consisting of $\mathcal{S}_j$, and the other consisting of all other variables involved in $\mathcal{C}_F \subseteq \mathcal{A}_F \times \mathcal{S}_{F,\text{ext}}$. Thus $\mathcal{C}_F \subseteq (\mathcal{A}_F \times \mathcal{S}_{F,j}) \times \mathcal{S}_j$, where $\mathcal{S}_{F,j} = \prod_{j' \in \mathcal{S}_{F,\text{ext}} \setminus \{j\}} \mathcal{S}_{j'}$. Then, by the fundamental theorem of subdirect products, $\mathcal{C}_F$ has the realization labelled by $\mathcal{F}$ in Figure 11.

![Figure 11: Fragment $\mathcal{F}$ connected to constraint $\hat{\mathcal{C}}_i$ by state $\mathcal{S}_j$, and reduced $\tilde{\mathcal{F}}$, $\tilde{\mathcal{C}}_i$, and $\tilde{\mathcal{S}}_j$.](image)

Here we have introduced the reduced state space $\tilde{\mathcal{S}}_j = (\mathcal{C}_F)_{|\mathcal{S}_j}/(\mathcal{C}_F)_{|\mathcal{S}_j}$. In view of the normal series $\{0\} \subseteq (\mathcal{C}_F)_{|\mathcal{S}_j} \subseteq (\mathcal{C}_F)_{|\mathcal{S}_j} \subseteq \mathcal{S}_j$, we see that $|\tilde{\mathcal{S}}_j| \leq |\mathcal{S}_j|$, with equality if and only if $\{0\} = (\mathcal{C}_F)_{|\mathcal{S}_j}$ and $(\mathcal{C}_F)_{|\mathcal{S}_j} = \mathcal{S}_j$; i.e., if and only if $\mathcal{C}_F$ is trim and proper at $\mathcal{S}_j$.

We have also introduced a reduced fragment $\tilde{\mathcal{F}}$ involving $\tilde{\mathcal{S}}_j$ and the variables other than $\mathcal{S}_j$ that are involved in $\mathcal{F}$. We see that in any realization $\mathcal{R}$ that includes $\mathcal{F}$, we may replace $\mathcal{F}$ by $\tilde{\mathcal{F}}$.
and an inclusion/natural map constraint \( \{(s_j, s_j + (C^F)_{S_j}) \in S_j \times \tilde{S}_j : s_j \in (C^F)_{S_j}\} \). Moreover, we can then combine the latter constraint with the neighboring constraint code \( C_i \) that also involves \( S_j \) to obtain a reduced constraint code \( \tilde{C}_i \); this amounts to restricting \( S_j \) to \( (C^F)_{S_j} \) in \( C_i \), and merging states \( s_j \in (C^F)_{S_j} \) to their cosets \( s_j + (C^F)_{S_j} \in \tilde{S}_j \). As a result, we obtain an equivalent realization \( \tilde{R} \) with the same graph topology, but with \( F, S_j \) and \( C_i \) reduced to \( \tilde{F}, \tilde{S}_j \) and \( \tilde{C}_i \). We call this a local reduction of \( R \).

Since state variables may be specified as we like, we may and will assume the following:

**Standing assumption:** All fragments are trim and proper at all external state variables. □

### 4.3 Reduced symbol alphabets

Let \( F \) be a linear or group fragment with external behavior \( C^F \subseteq A^F \times S^{F,\text{ext}} \). It is natural to partition the variables involved in \( C^F \) into the two subsets \( A^F \) and \( S^{F,\text{ext}} \), namely the symbol and external state variables, respectively.

By the fundamental theorem of subdirect products, we obtain the realization of \( F \) shown in Figure 12. Here we have introduced the reduced symbol alphabet \( \bar{A}^F = A^F / A^F \), where \( \bar{A}^F = (C^F)_{|\bar{A}^F} \) is the free symbol alphabet of \( F \), and \( \bar{A}^F = (C^F)_{\bar{A}^F} \) is the pinned symbol alphabet of \( F \).

![Figure 12: Realization of constraint code \( C^F \subseteq A^F \times S^{F,\text{ext}} \).](image)

The pinned symbol alphabet \( \bar{A}^F = (C^F)_{\bar{A}^F} \) of \( F \) comprises all symbol configurations \( a^F \in A^F \) that can occur with \( s^{F,\text{ext}} = 0 \). Figures 13(a)-(b) show realizations of \( \bar{A}^F \).

![Figure 13: Realizations of (a) \( \bar{A}^F \); (b) \( \tilde{A}^F \); (c) \( \tilde{A}^F \); (d) \( \tilde{A}^F \).](image)

The free symbol alphabet \( \bar{A}^F = (C^F)_{|\bar{A}^F} \) of \( F \) comprises all symbol configurations \( a^F \in A^F \) that can occur with any \( s^{F,\text{ext}} \in S^{F,\text{ext}} \). The dual realizations of Figures 13(c)-(d) realize the free symbol alphabet \( \bar{A}^F = ((C^F)^\perp)_{\bar{A}^F} \) of the dual fragment \( F^\circ \), which by normal realization duality is the orthogonal alphabet to \( \bar{A}^F \).

If \( F \) is a normal realization \( R \) (i.e., \( F \) has no external state variables) that realizes \( C \), then \( \bar{A}^R = \bar{A}^R = C \).

By the normal degree restrictions, the symbol variables \( A^F \) that are involved in \( F \) influence the rest of the realization only via the reduced symbol variables \( \bar{A}^F \). Consequently, the external behavior \( C^F \) of \( F \) and the code \( C \) realized by \( R \) should be regarded as “coset codes” over the cosets of \( A^F \) in \( \bar{A}^F \), rather than over \( A^F \) itself. (For this reason, in earlier work such as [8], we have called the pinned symbol alphabet \( \bar{A}^F \) “nondynamical.”)
If $\mathcal{A}_e$ is nontrivial, then the minimum distance between symbol configurations in $C$ cannot exceed the minimum distance within $\mathcal{A}_e$, for any notion of distance, since every $a^f \in \mathcal{A}_e$, combined with zeros elsewhere, is a codeword in $C$.

Moreover, in decoding $C$, no information based on other symbols can be given about the symbol value $a \in \mathcal{A}_e$, beyond the relative likelihood of lying in each coset of $\mathcal{A}_e$.

Finally, we observe that the inclusion/natural-map constraint $\{(a^f, a^f + \mathcal{A}_e) \in \mathcal{A}_e \times \mathcal{A}_e : a^f \in \mathcal{A}_e\}$ at the left of Figure 12 (called an “interface node” in 4) will naturally appear in any minimal realization.

Dually, in any realization $R$ that includes $C^\mathcal{F}$, the symbol configuration $a^f \in \mathcal{A}_e$ must lie in the free (“trimmed”) symbol alphabet $\mathcal{A}_e$. Because symbol variable alphabets are determined externally, we do not restrict $\mathcal{A}_e$ to $\mathcal{A}_e$, but rather let the inclusion constraint do the trimming.

In Figure 12, we have also introduced a reduced state alphabet $\tilde{\mathcal{S}}_{\mathcal{F},\text{ext}} = (\mathcal{C}_{\mathcal{F}})_{|S_{\mathcal{F},\text{ext}}}/(\mathcal{C}_{\mathcal{F}})_{|S_{\mathcal{F},\text{ext}}}$, which is isomorphic to the reduced symbol alphabet $\tilde{\mathcal{A}}_e$ by the fundamental theorem of subdirect products. However, since the component external state variables $S_j$ of $\tilde{\mathcal{S}}_{\mathcal{F},\text{ext}}$ are in general incident on different neighboring constraint codes, we will not usually want to aggregate them.

However, suppose that there is only one such external state variable $S_j$—i.e., $\mathcal{F}$ is a leaf fragment. If $\mathcal{F}$ is trim and proper, so that $S_j$ is reduced, then the realization of Figure 12 reduces to that of Figure 14, which illustrates the following simple but important theorem:

**Theorem (state space theorem for leaf fragments)** If $\mathcal{F}$ is a trim and proper linear or group leaf fragment with external behavior $C^\mathcal{F} \subseteq \mathcal{A}_e \times S_j$, then the external state space $S_j$ is isomorphic to the reduced symbol alphabet $\tilde{\mathcal{A}}_e = \tilde{\mathcal{A}}_e / \mathcal{A}_e = (\mathcal{C}_{\mathcal{F}})_{|\mathcal{A}_e} / (\mathcal{C}_{\mathcal{F}})_{|\mathcal{A}_e}$.

**Proof:** We have $(\mathcal{C}_{\mathcal{F}})_{|\mathcal{A}_e} / (\mathcal{C}_{\mathcal{F}})_{|\mathcal{A}_e} \cong (\mathcal{C}_{\mathcal{F}})_{|S_j} / (\mathcal{C}_{\mathcal{F}})_{|S_j}$, by the fundamental theorem of subdirect products. But by trimness $(\mathcal{C}_{\mathcal{F}})_{|S_j} = S_j$, and by properness $(\mathcal{C}_{\mathcal{F}})_{|S_j} = \{0\}$.

![Figure 14: State space theorem for a trim and proper leaf fragment.](image)

**4.4 Connecting fragments**

We will now consider taking a pair of fragments, $\mathcal{F}_1$ and $\mathcal{F}_2$, with external behaviors $C_1 \subseteq \mathcal{A}(1) \times S(1)$ and $C_2 \subseteq \mathcal{A}(2) \times S(2)$, and connecting external state variables with isomorphic state spaces $S_j$ and $S_j'$ via a generalized edge—i.e., an isomorphism constraint based on an isomorphism $S_j \leftrightarrow S_j'$—to form a combined fragment with external behavior $C_{12}$, as shown in Figure 15. The isomorphism could be an equality constraint or a sign-inversion constraint. (This operation is called “closing the box” by Vontobel and Loeliger 20.) Here $C_1 \subseteq \mathcal{A}(1) \times S(1 \cup) \times S_j$, where $S(1 \cup)$ is the Cartesian product of all external state spaces of $C_1$ except for $S_j$, and similarly $C_2 \subseteq \mathcal{A}(2) \times S(2 \cup) \times S_j'$.

Finally, $C_{12} \subseteq \mathcal{A}(12) \times S(12)$, where $\mathcal{A}(12) = \mathcal{A}(1) \times \mathcal{A}(2)$ and $\mathcal{A}(12) = \mathcal{A}(1) \times S(1 \cup) \times S(2 \cup)$.

![Figure 15: Connecting two fragments via an isomorphism $S_j \leftrightarrow S_j'$.](image)
The external behavior of the combined fragment is
\[ C_{12} = \{(a^{(1)}, s^{(1)}, s_j), (a^{(2)}, s^{(2)}, s'_j) \in C_1 \times C_2 : s_j \leftrightarrow s'_j \}_{A^{(12)} \times S^{(12)}}, \]
where \( \leftrightarrow \) denotes correspondence under the given isomorphism.

We then have the following simple but important lemma:

**Lemma (connected fragments)** If two linear or group fragments \( F_1, F_2 \) are connected via an isomorphism between external state spaces \( S_j \) and \( S'_j \), and \( F_{12} \) is the combined fragment, then:

(a) If \( F_1 \) and \( F_2 \) are trim, then \( F_{12} \) is trim.

(b) If \( F_1 \) and \( F_2 \) are proper, then \( F_{12} \) is proper.

**Proof:** (a) If \( F_1 \) is trim, then every value \( \sigma \) of every external state variable of \( F_1 \) appears in some valid configuration of \( C_1 \) in combination with some value \( s_j \in S_j \); and if \( F_2 \) is trim, then the corresponding value \( s'_j \in S'_j \) under the given isomorphism appears in some valid configuration of \( C_2 \), so \( \sigma \) must appear in some valid configuration in \( C_{12} \); and similarly for the state variables of \( C_2 \).

(b) follows by trim/proper duality from (a).

Since any connected cycle-free graph may be constructed by starting with its vertices and iteratively connecting vertices via edges as above, we have the following corollary:

**Theorem (cycle-free fragments)** If a fragment \( F \) is cycle-free and all of its constraint codes are trim and proper, then \( F \) is trim and proper.

A **cycle-free leaf fragment** \( F \) is a fragment of degree 1 whose graph is cycle-free. In graph theory, cycle-free leaf fragments are called **rooted trees**, with the **root** being the unique external state variable \( S_j \) of \( F \).

The state-space theorem for leaf fragments and the cycle-free fragment theorem then combine to yield the following important result:

**Theorem (cycle-free leaf fragments)** If a linear or group cycle-free leaf fragment \( F \) has external behavior \( C^F \subseteq A^F \times S_j \) and all of its constraint codes are trim and proper, then its external state space \( S_j \) is isomorphic to \( \tilde{A}^F = \overline{A^F} / A^F \).

### 4.5 Minimal cycle-free realizations

In the next two subsections, we will apply the cycle-free leaf fragment theorem to the two mutually exclusive cases of cycle-free and cyclic realizations, respectively. In this subsection, we will consider cycle-free realizations, and show that we can obtain an improved proof of one direction of the **“minimal = trim + proper” theorem** of [7, Theorem 3], which says that a cycle-free realization is minimal if and only if every constraint code is trim and proper, and furthermore that the state spaces of a minimal cycle-free realization are uniquely defined, up to isomorphism.

It is well known in graph theory that a connected graph \( G \) is cycle-free if and only if every edge is a cut set; that is, cutting any edge \( S_j \) disconnects the graph into two fragments, which we denote by \( F_j \) and \( P_j \) (for “future” and “past”). If \( G \) is cycle-free, then both fragments are cycle-free leaf fragments (rooted trees), with the edge \( S_j \) being their common root.
Lemma (trim + proper ⇒ minimal). Given a finite, connected, cycle-free linear or group normal realization $\mathcal{R}$ of a code $\mathcal{C}$, if every constraint code $\mathcal{C}_i$ is both trim and proper, then every state space $S_j$ is isomorphic to $\mathcal{C}_{|A^F_j} / \mathcal{C}_{|A^P_j}$ and also to $\mathcal{C}_{|A^P_j} / \mathcal{C}_{|A^P_j}$, where $F_j$ and $P_j$ are the two cycle-free leaf fragments of $\mathcal{R}$ created by a cut of the edge $S_j$. Moreover, $S_j$ is minimal.

Proof: If $\mathcal{R}$ is cycle-free, then both $F_j$ and $P_j$ are cycle-free leaf fragments with external state space $S_j$, so from the cycle-free leaf fragment theorem, we have $S_j \cong (C^F_j)_{|A^F_j} / (C^F_j)_{|A^F_j}$ and $S_j \cong (C^P_j)_{|A^P_j} / (C^P_j)_{|A^P_j}$. Thus, as shown in Figure 16, $\mathcal{C}$ is the subdirect product

$$\mathcal{C} = \{(a^{F_j}, a^{P_j}) : a^{F_j} + (C^F_j)_{|A^F_j} \leftrightarrow a^{P_j} + (C^P_j)_{|A^P_j}\} \subseteq A^{F_j} \times A^{P_j},$$

where $\leftrightarrow$ denotes correspondence under the isomorphisms

$$A^{F_j} \cong (C^F_j)_{|A^F_j} \cong (C^P_j)_{|A^P_j} = A^{P_j}$$

This implies $C_{|A^F_j} = (C^F_j)_{|A^F_j}$, $C_{|A^P_j} = (C^P_j)_{|A^P_j}$, $C_{|A^F_j} = (C^F_j)_{|A^F_j}$, and $C_{|A^P_j} = (C^P_j)_{|A^P_j}$.

![Figure 16: Realization of $\mathcal{C}$ as a subdirect product.](image)

Moreover, at least $|S_j|$ different states are required, since if $a^{F_j}$ and $a^{P_j}$ are not in corresponding cosets of $(C^F_j)_{|A^F_j}$ and $(C^P_j)_{|A^P_j}$, then $(a^{F_j}, a^{P_j}) \notin \mathcal{C}$. Thus $S_j$ is minimal.

The proof of the converse (minimal ⇒ trim + proper) is easy: if any constraint code $\mathcal{C}_i$ is not trim or proper at some state space $S_j$, then $S_j$ may be locally reduced by trimming or merging as in Figure 11. Thus we have simplified the proof of the following fundamental theorem:

Theorem (minimal = trim + proper [7]). Given a normal linear or group realization of a code $\mathcal{C}$ on a finite connected cycle-free graph $\mathcal{G}$, the following are equivalent:

1. Every constraint code $\mathcal{C}_i$ is both trim and proper.
2. Every state space $S_j$ is isomorphic to $\mathcal{C}_{|A^F_j} / \mathcal{C}_{|A^F_j}$ and to $\mathcal{C}_{|A^P_j} / \mathcal{C}_{|A^P_j}$.
3. The realization is minimal.

This theorem is closely related to the state space theorem [8, 23], which says that if a linear or group realization $\mathcal{R}$ of a code $\mathcal{C}$ has a state space $S_j$ such that a cut of the edge associated with $S_j$ partitions $\mathcal{R}$ into two disconnected fragments $F_j$ and $P_j$, then $(C^F_j)_{|A^F_j} / (C^F_j)_{|A^F_j} \cong (C^P_j)_{|A^P_j} / (C^P_j)_{|A^P_j} \cong \mathcal{C} / ((C^F_j)_{|A^F_j} \times (C^P_j)_{|A^P_j})$ (by the fundamental theorem of subdirect products [8]), and the minimal state space $S_j$ is isomorphic to any of these quotient groups.

In turn, the state space theorem is a special case of the cut-set bound [21, 22], which says that if $T_x \subseteq T_S$ is a minimal state index subset (cut set) such that a cut of the edges $S_j$ indexed by $T_x$ partitions $\mathcal{R}$ into two disconnected fragments $F_j$ and $P_j$, then the size $|S^x|$ of the composite state configuration space $S^x = \prod_{j \in T_x} S_j$ is at least as large as $|\mathcal{C} / ((C^F_j)_{|A^F_j} \times (C^P_j)_{|A^P_j})|$, with equality if and only if $S^x \cong \mathcal{C} / ((C^F_j)_{|A^F_j} \times (C^P_j)_{|A^P_j})$.
4.6 Cycle-free leaf fragments and 2-cores

We now consider cyclic realizations.

The cyclomatic number of a connected graph $G$ is the number of (ordinary) edges minus the number of vertices plus one. By elementary graph theory, the cyclomatic number is non-negative, and is equal to zero if and only if $G$ is cycle-free. We call $G$ cyclic if it is not cycle-free; i.e., if its cyclomatic number is nonzero. The cyclomatic number of $G$ is equal to the minimum number of edge cuts required to make $G$ cycle-free, and also to the maximum number of edge cuts that can be made without disconnecting $G$. The cyclomatic number of $G$ may thus be regarded as the “loopiness index” of $G$.

In graph theory, the 2-core of a connected graph $G$ is its maximal connected subgraph such that all vertices have degree 2 or more; i.e., the 2-core is its maximal leafless subgraph. (As in [7], we call a connected leafless graph a “generalized cycle.”) The 2-core of $G$ may be found by repeatedly deleting leaf vertices until none remain; thus an edge is in the 2-core if and only if it is part of a cycle. The 2-core is empty if and only if $G$ is cycle-free.

Since the 2-core may be obtained from $G$ by deleting leaf vertices and their associated edges, the cyclomatic number of the 2-core of $G$ is the same as that of $G$. The 2-core thus comprises the essential cyclic skeleton of $G$ after all leaves have been stripped away.

In our context, we will define the 2-core $\bar{R}$ of a normal realization $R$ whose normal graph $\bar{G}$ is cyclic as the part of $R$ that remains after repeatedly deleting leaf constraints. The normal graph $\bar{G}$ of $\bar{R}$ is thus a generalized cycle.

The part of the realization that is not in the 2-core $\bar{R}$ consists of a number of cycle-free leaf fragments (rooted trees), shown schematically in Figure 17. Each such leaf fragment $F$ is connected to a constraint code of $\bar{R}$ via a single external state space (root) $S_F$. (Notice that Figure 16 is the special case of Figure 17 in which the 2-core is empty.)

Figure 17: Schematic representation of 2-core with two cycle-free leaf fragments.

Under our standing assumption that all constraint codes are trim and proper, the cycle-free leaf fragment theorem applies to each such leaf fragment $F$, so $S_F \cong \tilde{A}_F = (C_F)_{A_F}/(C_F)_{\tilde{A}_F}$. From the point of view of the 2-core, nothing changes if we replace each state space $S_F$ by the isomorphic reduced symbol alphabet $\tilde{A}_F$. With these substitutions, $C$ is realized by its 2-core $\bar{R}$.

In this way, we may partition any realization $R$ into two parts: a cycle-free part consisting of a number of cycle-free leaf fragments, and a cyclic part consisting of the 2-core $\bar{R}$ of $R$, whose graph $\bar{G}$ is a leafless generalized cycle.

We expect that the main difficulties in analysis and in decoding will be associated with the 2-core. For example, in iterative sum-product decoding (see, e.g., [3]), it is easy to see that the cycle-free leaf fragments need to be decoded only once, with the result being a “message” of relative weights of the elements of the reduced symbol alphabet $\tilde{A}_F = (C_F)_{A_F}/(C_F)_{\tilde{A}_F}$ (or, for max-sum decoding, the “message” may consist of the “best” element and its weight). Iterative decoding may then be performed on the 2-core graph $\bar{G}$ with these incoming messages held constant. For example, in iterative decoding of a tail-biting trellis code with parallel transitions, the relative weights of the parallel transitions need to be computed only once.
5 Internal properties of fragments

Heretofore we have mainly considered the external behavior $C^F \subseteq A^F \times S^F,\text{ext}$ of a fragment $F$, which is a generalization of a constraint code. Now we consider also the internal behavior $B^F \subseteq B^F \times S^F,\text{ext} \times S^F,\text{int}$, which is a generalization of the internal behavior of a realization.

We define various notions of observability and controllability for fragments, and give dual tests for each. Finally, we give an improved proof of the result of [7] that a proper cycle-free realization is observable, and, dually, that a trim cycle-free realization is controllable.

5.1 Internal and external behavior of fragments

Again, a fragment $F$ is specified by a set $\{C_i : i \in I_C^F\}$ of constraint codes, a set $\{A_k : k \in I_A^F\}$ of symbol variables, a set $\{S_j : j \in I_S^F,\text{int}\}$ of internal state variables, and a boundary set $\partial(F) = \{S_j : j \in I_S^F,\text{ext}\}$ of external state variables. The configuration spaces $A^F$, $S^F,\text{int}$ and $S^F,\text{ext}$ are the Cartesian products of the respective variable alphabets.

A fragment $F$ is a normal realization of its external behavior $C^F \subseteq A^F \times S^F,\text{ext}$. As with normal realizations, we will denote the two values of each internal state variable $S_j$ by $s_j$ and $s_j'$. Then each element of the Cartesian (external direct) product of all constraint codes in $F$ is a quadruple $(a, s^\text{ext}, s, s') \in A^F \times S^F,\text{ext} \times S^F,\text{int} \times S^F,\text{int}$. Again, at the risk of some notational ambiguity, we will identify this Cartesian product with $F$; i.e., $F = \prod_{i \in I_C^F} C_i$.

The extended internal behavior is $\tilde{B}^F = \{(a, s^\text{ext}, s, s') \in F = F \cap \mathcal{V}\}$, where $\mathcal{V}$ now denotes the validity space $\mathcal{V} = \{(a, s^\text{ext}, s, s') \in A \times S^F,\text{ext} \times S^F,\text{int} \times S^F,\text{int}\}$. The internal behavior $B^F = (\tilde{B}^F)\lceil_A \times S^F,\text{ext} \times S^F,\text{int}$ then has the normal realization shown in Figure 18(a). The external behavior $C^F$ is the projection of $B^F$ or $\tilde{B}^F$ onto $A^F \times S^F,\text{ext}$, which is obtained by imposing a dummy (free) constraint on $S^F,\text{int}$, as shown in Figure 18(b).

Figure 18: Normal realizations of (a) internal behavior $B^F$; (b) external behavior $C^F$; (c) orthogonal internal behavior $B^F$; (d) dual external behavior $C^F$.

If $F$ is a linear or group fragment, then by normal realization duality we have the normal realizations of the orthogonal internal behavior $B^F$ and the dual external behavior $C^F$ shown in Figures 18(c) and (d). (In Figure 18(d), the two-variable zero-sum constraint $-s - \hat{s} = 0$ is equivalent to the sign inversion constraint $\hat{s} = -\hat{s}'$.)

5.2 Observability and controllability of fragments

We now discuss observability and controllability for fragments. We will see that there are various possible notions of observability, and corresponding dual notions of controllability.
In general, the term “observable” applies to systems that have a set of internal (“state”) variables $S$ and a set of external variables $A$. A system is called “observable” if observation of an external configuration $a \in A$ determines the internal configuration $s \in S$.

A fragment $F$ has internal behavior $B^F \subseteq A^F \times S^F,\text{ext} \times S^F,\text{int}$ and external behavior $C^F = (B^F)|_{A^F \times S^F,\text{ext}}$. Thus the definition of observability depends on whether the behavior of the system is regarded as $B^F$ or $C^F$, and on which variables are regarded as internal and which as external. Consequently, we define three notions of observability, as follows:

- A fragment $F$ is externally observable if the projection $C^F \to A^F$ is one-to-one; i.e., if the symbol configuration $a$ determines the external state configuration $s^{\text{ext}}$.

- A fragment $F$ is internally observable if the projection $B^F \to C^F$ is one-to-one; i.e., if the symbol configuration $a$ and the external state configuration $s^{\text{ext}}$ together determine the internal state configuration $s^{\text{int}}$.

- A fragment $F$ is totally observable if the projection $B^F \to A^F$ is one-to-one; i.e., if the symbol configuration $a$ determines both $s^{\text{ext}}$ and $s^{\text{int}}$.

Evidently $F$ is totally observable if and only if it is both externally and internally observable.

If $F$ is a normal realization—i.e., if $F$ has no external state variables—then $C^F = (C^F)|_{A^F}$, so $F$ is trivially externally observable. Thus for normal realizations, the notions of internal observability and total observability coincide, and are equivalent to the notion of observability that we have used for normal realizations in [7] and in Section 3.7.

If $F$ is a constraint code—i.e., if $F$ has no internal state variables—then $B^F = C^F$, so $F$ is trivially internally observable. Thus for constraint codes, the notions of external observability and total observability coincide.

If $F$ is a leaf fragment with a single external state $S_j$, then our definition of external observability reduces to that of properness; i.e., $F$ is externally observable if and only if $F$ is proper.

More generally, for any linear or group fragment $F$, since the kernel of the projection of $C^F$ onto $A^F$ is isomorphic to the cross-section $(C^F)|_{S^F,\text{ext}}$, it follows that $F$ is externally observable if and only if $(C^F)|_{S^F,\text{ext}}$ is trivial.

Similarly, since the kernel of the projection $B^F \to C^F$ is isomorphic to $(B^F)|_{S^F,\text{int}}$, $F$ is internally observable if and only if $(B^F)|_{S^F,\text{int}}$ is trivial, and since the kernel of the projection $B^F \to A^F$ is isomorphic to $(B^F)|_{S^F,\text{ext} \times S^F,\text{int}}$, $F$ is totally observable if and only if $(B^F)|_{S^F,\text{ext} \times S^F,\text{int}}$ is trivial.

In general, for any of these definitions of observability, we will define a corresponding notion of controllability such that a linear or group fragment $F$ is controllable if and only if the dual fragment $F^\circ$ is observable.

In particular, we will say that a linear or group fragment $F$ is externally controllable if the projection $(C^F)|_{S^F,\text{ext}}$ is equal to $S^F,\text{ext}$ (i.e., the projection is surjective).

As in Section 3.7, we will say that a linear or group fragment $F$ is internally controllable if the constraint spaces $F^\perp = \prod_{i \in I^\perp} (C_i)^\perp$ and $V^\perp = \{ (0, 0, \hat{s}, s) \in A \times S^F,\text{ext} \times S^F,\text{int} \times S^F,\text{int} \}$ are independent, because then and only then the internally unobservable extended dual behavior $(\tilde{B}^\circ)^u = F^\perp \cap V^\perp$ is trivial.

Finally, a linear or group fragment $F$ will be called totally controllable if it is both internally and externally controllable.
Example 1 (trellis fragments, cont.) Let us see how these definitions apply to a fragment $F_{j,k}$ of a conventional state-space (trellis) realization, as shown in Figure 2.

$F_{j,k}$ is externally observable if the symbol sequence $a_{j,k}$ determines the “state transition” $(s_j, s_k)$. (In [10], this property is called “[j, k]-observability.”) $F_{j,k}$ is internally observable if the symbol sequence $a_{j,k}$ and $(s_j, s_k)$ determine the remaining state sequence $s_{j,k}$. As we shall show shortly (see Section 5.4), if all constraint codes are proper (“instantaneously invertible”), then $F_{j,k}$ is externally observable if $a_{j,k}$ determines the entire state sequence $s_{j,k}$, thus if $F_{j,k}$ is proper, as we generally assume, then $F_{j,k}$ is totally observable if and only if it is externally observable. In classical linear system theory, “observability” usually means what we call total observability, which coincides with external observability under the assumption of properness.

Dually, $F_{j,k}$ is externally controllable (or “[j, k]-controllable” [10]) if all state transitions $(s_j, s_k)$ can occur. This is what is usually called “controllability” in classical linear system theory. By trim-proper duality, $F_{j,k}$ is internally controllable if all constraint codes are trim. Thus if $F_{j,k}$ is trim, as we generally assume, then external, total and classical controllability coincide.

We conclude that for trim and proper trellis fragments, the classical linear-system-theoretic notions of observability and controllability coincide with what we call external or total observability and controllability.

5.3 Observability and controllability tests

We now develop dual tests for observability and controllability of linear or group fragments along the lines of those developed for realizations in Section 3.7.

As already noted, a linear or group fragment $F$ is externally observable if and only if the cross-section $(C^F)_{S^F, ext}$ is trivial. We now call this cross-section the externally unobservable state configuration space

$$S^F, ext, u = (C^F)_{S^F, ext} = \{ s^{ext} : (0, s^{ext}) \in C^F \}.$$  

Recalling that the image of the projection $C^F \to A^F$ is the free symbol alphabet $A^F$, we have the following external observability test: $|C^F| \geq |A^F|$, or in the linear case $\dim C^F \geq \dim A^F$, with equality if and only if $F$ is externally observable.

A linear or group fragment $F$ is internally observable if and only if the cross-section $(B^F)_{S^F, int}$ is trivial. As in Section 3.7, we define this cross-section as the internally unobservable state configuration space

$$S^F, int, u = \{ s^{int} : (0, s^{int}) \in B^F \}.$$  

We then have the following internal observability test: $|C^F| \leq |B^F|$, or in the linear case $\dim C^F \leq \dim B^F$, with equality if and only if $F$ is internally observable.

A linear or group fragment $F$ is totally observable if and only if the cross-section $(B^F)_{S^F, ext \times S^F, int}$, which is isomorphic to the kernel of the projection $B^F \to A^F$, is trivial. We define this cross-section as the totally unobservable state configuration space

$$S^F, ext/int, u = \{ (s^{ext}, s^{int}) : (0, s^{ext}, s^{int}) \in B^F \}.$$  

We then have the following total observability test: $|B^F| \geq |A^F|$, or in the linear case $\dim B^F \geq \dim A^F$, with equality if and only if $F$ is totally observable.
A normal realization of $S^F,\text{ext},u$ may be obtained by grounding $A^F$ in the Figure 18(b) realization of $C^F$, as shown in Figure 19(a). A normal realization of $S^F,\text{int},u$ may be obtained by grounding $A^F$ and $S^F,\text{ext}$ in the Figure 18(a) realization of $B^F$, as shown in Figure 20(a). A normal realization of $S^F,\text{ext/int},u$ may be obtained by grounding $A^F$ in the Figure 18(a) realization of $B^F$, as shown in Figure 21(a).

![Figure 19: Normal realizations of (a) $S^F,\text{ext},u$; (b) $\hat{S}^F,\text{ext}$.](image)

![Figure 20: Normal realizations of (a) $S^F,\text{int},u$; (b) $\hat{S}^F,\text{int}$.](image)

![Figure 21: Normal realizations of (a) $S^F,\text{ext/int},u$; (b) $\hat{S}^F,\text{ext/int}$.](image)

The dual normal realization to that of Figure 19(a) is that of Figure 19(b), which realizes the dual externally controllable subspace $\hat{S}^F,\text{ext} = ((C^F)^\perp)|_{S^F,\text{ext}}$ of the dual fragment $F^\circ$. By projection/cross-section duality, $F^\circ$ is externally controllable if and only if $\hat{S}^F,\text{ext} = S^F,\text{ext}$.

Dually, we define the externally controllable subspace of $F$ as

$$S^F,\text{ext} = (C^F)|_{S^F,\text{ext}},$$

the image of the projection $C^F \rightarrow S^F,\text{ext}$. Since the kernel of this projection is the pinned symbol alphabet $A^F = (C^F)\cdot A^F$, we have the following external controllability test: $|C^F|/|A^F| = |S^F,\text{ext}| \leq |S^F,\text{ext}|$, or in the linear case $\dim C^F - \dim A^F = \dim S^F,\text{ext} \leq \dim S^F,\text{ext}$, with equality if and only if $F$ is externally controllable.

The dual normal realization to that of Figure 20(a) is that of Figure 20(b), which realizes the dual internally controllable subspace $\hat{S}^F,\text{int} = \{\hat{s} + \hat{s}' : (\hat{a}, s, \hat{s}, s') \in F^\perp\}$ of the dual fragment $F^\circ$. By comparison with Figure 18(c), we see that $\hat{S}^F,\text{int} = ((B^F)^\perp)|_{\hat{S}^F,\text{int}}$, so by projection/cross-section duality, $\hat{S}^F,\text{int} = (S^F,\text{int},u)^\perp$. Thus $F^\circ$ is internally controllable if and only if $\hat{S}^F,\text{int} = \hat{S}^F,\text{int}$.
Dually, we define the *internally controllable subspace* of $\mathcal{F}$ as

$$S^{F,\text{int},c} = (\mathfrak{B}^{F})|_{S^{F,\text{int}}},$$

the image of the validity check homomorphism $\mathcal{F} \to S^{F,\text{int}}$ defined by $(a, s^{\text{ext}}, s, s') \mapsto s - s'$. Since the kernel is isomorphic to $\mathfrak{B}^{F}$, we have as in Section 3.7 the following internal controllability test: $|\mathcal{F}|/|\mathfrak{B}^{F}| = |S^{F,\text{int},c}| \leq |S^{F,\text{int}}|$, or in the linear case $\dim \mathcal{F} - \dim \mathfrak{B}^{F} = \dim S^{F,\text{int},c} \leq \dim S^{F,\text{int}}$, with equality if and only if $\mathcal{F}$ is internally controllable.

The dual normal realization to that of Figure 21(a) is that of Figure 21(b), which realizes the *dual totally controllable subspace* $\hat{S}^{F,\text{ext}/\text{int},c} = ((\mathfrak{B}^{F})^\perp)|_{\hat{S}^{F,\text{ext}} \times \hat{S}^{F,\text{int}}}$. By projection/cross-section duality, $\mathcal{F}^{c}$ is totally controllable if and only if $\hat{S}^{F,\text{ext}/\text{int},c} = \hat{S}^{F,\text{ext}} \times \hat{S}^{F,\text{int}}$.

Dually, we define the totally controllable subspace of $\mathcal{F}$ as

$$S^{F,\text{ext}/\text{int},c} = (\mathfrak{B}^{F})|_{S^{F,\text{ext}} \times S^{F,\text{int}}},$$

the image of the homomorphism $\mathcal{F} \to S^{F,\text{ext}} \times S^{F,\text{int}}$ defined by $(a, s^{\text{ext}}, s, s') \mapsto (s^{\text{ext}}, s - s')$. Since the kernel $\{(a, 0, s^{\text{int}}, s^{\text{int}}) \in \mathcal{F}\}$ is isomorphic to the pinned internal behavior $\mathfrak{B}^{F} = \{(a, 0, s^{\text{int}}) \in \mathfrak{B}^{F}\}$, we have the following total controllability test: $|\mathcal{F}|/|\mathfrak{B}^{F}| = |S^{F,\text{ext}/\text{int},c}| \leq |S^{F,\text{ext}}||S^{F,\text{int}}|$, or in the linear case $\dim \mathcal{F} - \dim \mathfrak{B}^{F} = \dim S^{F,\text{ext}/\text{int},c} \leq \dim S^{F,\text{ext}} + \dim S^{F,\text{int}}$, with equality if and only if $\mathcal{F}$ is totally controllable.

### 5.4 Connecting fragments, continued

As in Section 4.4, we again consider connecting a pair of fragments $\mathcal{F}_1$ and $\mathcal{F}_2$ with isomorphic external state spaces $\mathcal{S}_j$ and $\mathcal{S}_j'$ via an isomorphism $\mathcal{S}_j \leftrightarrow \mathcal{S}_j'$ to form a combined fragment $\mathcal{F}_{12}$, as shown in Figure 15. Again we have $\mathcal{C}_1 \subseteq \mathcal{A}^{(1)} \times \mathcal{S}^{(1\backslash j)} \times \mathcal{S}_j$, $\mathcal{C}_2 \subseteq \mathcal{A}^{(2)} \times \mathcal{S}^{(2\backslash j')} \times \mathcal{S}_j'$, and $\mathcal{C}_{12} \subseteq \mathcal{A}^{(12)} \times \mathcal{S}^{(12)}$, where $\mathcal{A}^{(12)} = \mathcal{A}^{(1)} \times \mathcal{A}^{(2)}$ and $\mathcal{S}^{(12)} = \mathcal{S}^{(1\backslash j)} \times \mathcal{S}^{(2\backslash j')}$. The internal behavior of $\mathcal{F}_1$ and $\mathcal{F}_2$ will be denoted by $\mathfrak{B}_1 \subseteq \mathcal{A}^{(1)} \times \mathcal{S}^{(1\backslash j)} \times \mathcal{S}_j \times \mathcal{S}^{(1),\text{int}}$ and $\mathfrak{B}_2 \subseteq \mathcal{A}^{(2)} \times \mathcal{S}^{(2\backslash j')} \times \mathcal{S}_j' \times \mathcal{S}^{(2),\text{int}}$. The internal behavior $\mathfrak{B}_{12}$ of the combined fragment $\mathcal{F}_{12}$ is

$$\mathfrak{B}_{12} = \left\{ ((a^{(1)}_j, s^{(1\backslash j)}_j, s_j, s^{(1),\text{int}}_j), (a^{(2)}_j, s^{(2\backslash j')}_j, s'_j, s^{(2),\text{int}}_j)) \in \mathfrak{B}_1 \times \mathfrak{B}_2 : s_j \leftrightarrow s'_j \right\},$$

where $\leftrightarrow$ again denotes correspondence under the given isomorphism, and its external behavior is $\mathcal{C}_{12} = (\mathfrak{B}_{12})|_{\mathcal{A}^{(12)} \times \mathcal{S}^{(12)}}$.

We may now extend our previous connected fragment lemma as follows:

**Lemma (connected fragments, cont.)** If two linear or group fragments $\mathcal{F}_1, \mathcal{F}_2$ are connected via an isomorphism between state spaces $\mathcal{S}_j$ and $\mathcal{S}_j'$, and $\mathcal{F}_{12}$ is the combined fragment, then:

(c) If $\mathcal{F}_1$ and $\mathcal{F}_2$ are externally observable, then $\mathcal{F}_{12}$ is externally observable.

(d) If $\mathcal{F}_1$ and $\mathcal{F}_2$ are externally controllable, then $\mathcal{F}_{12}$ is externally controllable.

(e) If $\mathcal{F}_1$ and $\mathcal{F}_2$ are proper and internally observable, then $\mathcal{F}_{12}$ is internally observable.

(f) If $\mathcal{F}_1$ and $\mathcal{F}_2$ are trim and internally controllable, then $\mathcal{F}_{12}$ is internally controllable.

(g) If $\mathcal{F}_1$ and $\mathcal{F}_2$ are proper and totally observable, then $\mathcal{F}_{12}$ is totally observable.

(h) If $\mathcal{F}_1$ and $\mathcal{F}_2$ are trim and totally controllable, then $\mathcal{F}_{12}$ is totally controllable.
Proof: (c) If $F_1$ is externally observable, then $a^{(1)} = 0$ implies $s^{(1\setminus j)} = 0$, and similarly for $F_2$. Hence $a^{(12)} = (a^{(1)}, a^{(2)}) = (0, 0)$ implies $s^{(12)} = (s^{(1\setminus j)}, s^{(2\setminus j')}) = (0, 0)$, so $F_{12}$ is externally observable. (d) follows by observability/controllability duality from (c).

(e) If $C_1$ is proper, then $a^{(1)} = 0$ and $s^{(1\setminus j)} = 0$ imply $s_j = 0$, and similarly for $F_2$. If $F_1$ is internally observable, then this implies $s^{(1), \text{int}} = 0$, and similarly for $F_2$. Hence $(a^{(12)}, s^{(12)}) = (0, 0)$ implies $s^{(12), \text{int}} = 0$, so $F_{12}$ is internally observable. (f) follows by trim/proper duality and observability/controllability duality from (e).

(g) Since a fragment is totally observable if and only if it is internally and externally observable, it follows that if $F_1$ and $F_2$ are proper and totally observable, then $F_{12}$ is totally observable. (h) follows by trim/proper duality and observability/controllability duality from (g).

Example 1 (trellis fragments, cont.) As already noted, if all constraint codes in a trellis fragment $F[j,k]$ are proper, then $F[j,k]$ is internally observable, and if all are trim, then $F[j,k]$ is internally controllable.

We now obtain a more elegant proof of the result [7, Theorem 11] that a proper cycle-free normal realization $R$ must be (internally) observable, and the dual result.

For brevity, as in [7], we will call a realization $R$ proper if all of its constraint codes are proper, and trim if all of its constraint codes are trim. Since a realization has no external state variables, this usage does not conflict with our usage for fragments. We will also use these terms in the same way for a 2-core $\bar{R}$.

Theorem (unobservable/uncontrollable realizations and cycles) A proper cycle-free linear or group realization $R$ is internally observable, and a trim cycle-free linear or group realization $R$ is internally controllable. A proper cyclic linear or group realization $R$ is internally observable if and only if its 2-core $\bar{R}$ is internally observable, and a trim cyclic linear or group realization $R$ is internally controllable if and only if its 2-core $\bar{R}$ is internally controllable.

Proof: The first statement follows from the cycle-free fragment theorem above. For the second, we observe that since a proper cycle-free leaf fragment is internally observable, the combination of such a fragment with a proper 2-core is internally observable if the 2-core is internally observable, from part (e) of the connected fragment theorem; on the other hand, if the 2-core is not internally observable, then it supports an unobservable sequence, so the combination supports an unobservable sequence. The last part follows from trim/proper and controllability/observability duality.
6 Trellis fragments

In this section, we develop a structure theory for trellis fragments, namely fragments with two external state variables. The most elementary trellis fragment is a single constraint code with two state variables, which in coding theory is called a trellis section. Trellis sections may be combined into a trellis chain, such as the trellis fragment $R^{(j,k)}$ shown in Figure 2.

If the two ends of a trellis chain are connected, then the result is a cyclic tail-biting trellis realization, which is evidently the only kind of finite realization that can be constructed solely from trellis sections. If the two ends are terminated with leaf constraint codes, or leaf trellis sections, then the result is a cycle-free conventional trellis realization, which is evidently the only other kind of finite realization that can be constructed from trellis sections and leaf trellis sections.

We develop a generic decomposition of the transition space of a trim and proper trellis fragment, which governs its external controllability properties, and a dual decomposition of its unobservable transition space, which governs its observability properties. Finally, we analyze combined trellis fragments, and show that observability and controllability do not decrease under combination.

6.1 Preliminaries

A trellis fragment has two external state variables, whose alphabets we will denote as $S_j$ and $S_k$, as in Figure 2. Correspondingly, we will denote the fragment as $\mathcal{F}^{(j,k)}$, its symbol configuration space as $A^{(j,k)}$, and its external behavior as $\mathcal{C}^{(j,k)}$. An element of the external behavior will be denoted by $(a^{(j,k)}, s_j, s_k) \in A^{(j,k)} \times S_j \times S_k$. Although we will not discuss the internal behavior $\mathcal{B}^{(j,k)}$ in this section, this notation is consistent with a standard internal trellis structure as shown in Figure 2.

We assume that $\mathcal{C}^{(j,k)}$ is trim and proper, so there are no elements of $\mathcal{C}^{(j,k)}$ of the form $(\vec{0}^{(j,k)}, s_j, 0)$ or $(\vec{0}^{(j,k)}, 0, s_k)$, and similarly for the dual external behavior $(\mathcal{C}^{(j,k)})^\perp$.

By the fundamental theorem of subdirect products, we obtain the realization of $\mathcal{C}^{(j,k)}$ shown in Figure 22(a). Here we employ the reduced symbol alphabet $\bar{A}^{(j,k)} = A^{(j,k)}/\mathcal{A}^{(j,k)}$, where $\mathcal{A}^{(j,k)} = (\mathcal{C}^\perp)|_{S_j \times S_k}$ is the free symbol alphabet of $\mathcal{F}^{(j,k)}$, and $A^\perp = (\mathcal{C}^\perp)_{S_j \times S_k}$ is its pinned symbol alphabet. Furthermore, $\bar{\mathcal{C}}^{(j,k)}$ denotes the reduced external behavior of $\mathcal{F}^{(j,k)}$, which in combination with the inclusion/natural map constraint between $A^{(j,k)}$ and $\bar{A}^{(j,k)}$ is equivalent to $\mathcal{C}^{(j,k)}$.

![Figure 22: Realizations of (a) external behavior $\mathcal{C}^{(j,k)} \subseteq A^{(j,k)} \times S_j \times S_k$ of trellis fragment $\mathcal{F}^{(j,k)}$; (b) external behavior $(\mathcal{C}^{(j,k)})^\perp \subseteq \bar{A}^{(j,k)} \times \hat{S}_j \times \hat{S}_k$ of dual trellis fragment $(\mathcal{F}^\perp)^{(j,k)}$.](image)

The dual realization of the external behavior $(\mathcal{C}^{(j,k)})^\perp \subseteq \bar{A}^{(j,k)} \times \hat{S}_j \times \hat{S}_k$ of the dual trellis fragment $(\mathcal{F}^\perp)^{(j,k)}$ is shown in Figure 22(b).

By the definition of the reduced external behavior $\bar{\mathcal{C}}^{(j,k)}$, the cross-section $(\bar{\mathcal{C}}^{(j,k)})_{\bar{A}^{(j,k)}}$ is trivial and $(\bar{\mathcal{C}}^{(j,k)})_{\bar{A}^{(j,k)}} = \bar{A}^{(j,k)}$, and similarly for $(\mathcal{C}^{(j,k)})^\perp$. Thus there are no nonzero elements of the form $(\bar{a}^{(j,k)}, 0, 0)$ in $\mathcal{C}^{(j,k)}$, and similarly for $(\mathcal{C}^{(j,k)})^\perp$. In summary, neither $\bar{\mathcal{C}}^{(j,k)}$ nor $(\bar{\mathcal{C}}^{(j,k)})^\perp$ has any elements with Hamming weight 1.
6.2 Transition spaces

As in [10], we define the transition space \( T^{(j,k)} \) of \( F^{(j,k)} \) as the projection of \( C^{(j,k)} \) or \( \tilde{C}^{(j,k)} \) onto \( S_j \times S_k \), and the unobservable transition space \( U^{(j,k)} \) as the cross-section of \( C^{(j,k)} \) or \( \tilde{C}^{(j,k)} \) on \( S_j \times S_k \). In other words, \( T^{(j,k)} \) is the set of all state transitions \((s_j, s_k) \in S_j \times S_k\) that can occur, and \( U^{(j,k)} \) is the subset of such transitions that can occur with \( a^{(j,k)} = 0 \).

By definition, \( T^{(j,k)} = (C^{(j,k)})_{|S_j \times S_k} \) is the externally controllable subspace of \( C^{(j,k)} \), and \( U^{(j,k)} = (C^{(j,k)})_{|S_j \times S_k} \) is the externally unobservable state configuration space of \( C^{(j,k)} \), so \( F^{(j,k)} \) is externally controllable if and only if \( T^{(j,k)} = S_j \times S_k \), and externally observable if and only if \( U^{(j,k)} \) is trivial.

**Theorem (transition spaces).** Let \( T^{(j,k)} \subseteq S_j \times S_k \) and \( U^{(j,k)} \subseteq T^{(j,k)} \) be respectively the transition space and unobservable transition space of a trim and proper linear or group trellis fragment \( F^{(j,k)} \), and let \( (T^\circ)^{(j,k)} \) and \( (U^\circ)^{(j,k)} \) be the transition space and unobservable transition space of the dual fragment \( (F^\circ)^{(j,k)} \). Then:

(a) (transition space duality [10]) \( (T^{(j,k)})^\perp = (U^\circ)^{(j,k)} \) and \( (U^{(j,k)})^\perp = (T^\circ)^{(j,k)} \).

(b) \( T^{(j,k)} \) and \( (T^\perp)^{(j,k)} \) are trim (e.g., \( (T^{(j,k)})_{|S_j} = S_j \)), and \( U^{(j,k)} \) and \( (U^\perp)^{(j,k)} \) are proper (e.g., \( (U^{(j,k)})_{|S_j} = \{0\} \)).

(c) \( T^{(j,k)} \cong \tilde{C}^{(j,k)} \), the reduced external behavior of \( F^{(j,k)} \).

(d) \( T^{(j,k)}/U^{(j,k)} \cong \tilde{A}^{(j,k)} \), the reduced symbol alphabet of \( F^{(j,k)} \).

**Proof:**

(a) Projection/cross-section duality.

(b) \( (T^{(j,k)})_{|S_j} = (C^{(j,k)})_{|S_j} \), so \( T^{(j,k)} \) is trim if \( C^{(j,k)} \) is trim; the rest follows from trim/proper duality.

(c) Since the kernel of the projection of \( \tilde{C}^{(j,k)} \) onto \( S_j \times S_k \) is \( \{(0,0,0)\} \) and its image is \( T^{(j,k)} \), we have \( T^{(j,k)} \cong \tilde{C}^{(j,k)} \), by the fundamental theorem of homomorphisms.

(d) By the fundamental structure theorem for subdirect products, \( (C^{(j,k)})_{|A^{(j,k)}} \cong (C^{(j,k)})_{|S_j \times S_k}/(C^{(j,k)})_{|S_j \times S_k} \).

Part (c) implies a one-to-one correspondence between transitions in \( T^{(j,k)} \) and elements of the reduced internal behavior \( \tilde{C}^{(j,k)} \). For each \( (s_j, s_k) \in T^{(j,k)} \), there thus exists a unique \( (s_j, \tilde{a}^{(j,k)}, s_k) \in \tilde{C}^{(j,k)} \), which corresponds to \( \bigcup (T^{(j,k)}) \) “parallel transitions” in \( C^{(j,k)} \).

Part (d) implies that there is a unique coset \( \tilde{a}^{(j,k)} \in \tilde{A}^{(j,k)} \) corresponding to each transition \( (s_j, s_k) \in T^{(j,k)} \) if and only if \( U^{(j,k)} \) is trivial; i.e., if and only if \( F^{(j,k)} \) is externally observable. Otherwise, each coset \( \tilde{a}^{(j,k)} \in \tilde{A}^{(j,k)} \) is associated with \( |U^{(j,k)}| \) different transitions \( (s_j, s_k) \in T^{(j,k)} \).

6.3 Reachable and unreachable spaces

We define the reachable subspaces \( V^{(j,k)} \subseteq S_j \) and \( W^{(j,k)} \subseteq S_k \) as the cross-sections of the transition space \( T^{(j,k)} \) on \( S_j \) and \( S_k \), respectively. Thus \( V^{(j,k)} \) is the subspace consisting of all states \( s_j \in S_j \) such that \((s_j, 0) \in T^{(j,k)} \), or, in other words, the set of all states \( s_j \) from which a transition to \( s_k = 0 \) is possible. Similarly, \( W^{(j,k)} \) is the set of all \( s_k \in S_k \) that can be reached from \( s_j = 0 \).
From the fundamental structure theorem for subdirect products, it follows immediately that:

**Theorem (reachable subspaces).** Given a trim transition space $\mathcal{T}^{[j,k]} \subseteq S_j \times S_k$, and its reachable subspaces $\mathcal{V}^{[j,k]} = (\mathcal{T}^{[j,k]} \backslash S_j)$ and $\mathcal{W}^{[j,k]} = (\mathcal{T}^{[j,k]} \backslash S_k)$, then

$$\frac{S_j}{\mathcal{V}^{[j,k]}} \cong \frac{S_k}{\mathcal{W}^{[j,k]}} \cong \mathcal{T}^{[j,k]}.$$  

In particular, the following are equivalent:  

- $\mathcal{V}^{[j,k]} = S_j \iff \mathcal{W}^{[j,k]} = S_k \iff \mathcal{T}^{[j,k]} = \mathcal{V}^{[j,k]} \times \mathcal{W}^{[j,k]} \iff$ the fragment $\mathcal{F}^{[j,k]}$ is externally controllable. 

In view of this theorem, we will call $S_j / \mathcal{V}^{[j,k]}$ and $S_k / \mathcal{W}^{[j,k]}$ the *unreachable quotients* of the transition space $\mathcal{T}^{[j,k]}$. The theorem says that the unreachable quotients are isomorphic, and trivial if and only if the fragment $\mathcal{F}^{[j,k]}$ is externally controllable.

This theorem shows that in any trim trellis fragment, the two external state spaces $S_j$ and $S_k$ partition into corresponding cosets of the reachable subspaces $\mathcal{V}^{[j,k]} \subseteq S_j$ and $\mathcal{W}^{[j,k]} \subseteq S_k$, such that there are transitions between any pair $(s_j, s_k)$ that lie in corresponding cosets, but no other transitions. Moreover, as shown in Figure 23, a trim transition space $\mathcal{T}^{[j,k]}$ partitions into disconnected cosets of $\mathcal{V}^{[j,k]} \times \mathcal{W}^{[j,k]}$, which we will call its reachable subspace. (This leads to well-known controllability results for classical linear systems.)

![Diagram](image)

Figure 23: A trim trellis transition space $\mathcal{T}^{[j,k]} \subseteq S_j \times S_k$ is the union of $|\mathcal{T}^{[j,k]} / (\mathcal{V}^{[j,k]} \times \mathcal{W}^{[j,k]})|$ disconnected cosets of its reachable subspace $\mathcal{V}^{[j,k]} \times \mathcal{W}^{[j,k]}$.

We conclude that a trim transition space $\mathcal{T}^{[j,k]}$ is the “product” of three factor groups:

1. A *start space* $\{0\} \times \mathcal{W}^{[j,k]}$, consisting of all elements of $\mathcal{T}^{[j,k]}$ of the form $(0, s_k)$.
2. A *stop space* $\mathcal{V}^{[j,k]} \times \{0\}$, consisting of all elements of $\mathcal{T}^{[j,k]}$ of the form $(s_j, 0)$.
3. An *unreachable quotient space* $\mathcal{T}^{[j,k]} / (\mathcal{V}^{[j,k]} \times \mathcal{W}^{[j,k]}) \cong S_j / \mathcal{V}^{[j,k]} \cong S_k / \mathcal{W}^{[j,k]}$.

Dually, we define the *unobservable projections* $\mathcal{X}^{[j,k]}$ and $\mathcal{Y}^{[j,k]}$ as the projections of $\mathcal{U}^{[j,k]}$ on $S_j$ and $S_k$, respectively. Using projection/cross-section duality, we see that $\mathcal{X}^{[j,k]} = ((\mathcal{V}^{[j,k]})^\perp)$ and $\mathcal{Y}^{[j,k]} = ((\mathcal{W}^{[j,k]})^\perp)$, where $(\mathcal{V}^{[j,k]})^\perp$ and $(\mathcal{W}^{[j,k]})^\perp$ are the reachable spaces of the dual realization. Thus we obtain the following dual to the reachable spaces theorem:
Theorem (unobservable projections). Given a proper unobservable transition space $U^{(j,k)} \subseteq S_j \times S_k$ and its projections $X^{(j,k)} = (U^{(j,k)})_{|S_j}$ and $Y^{(j,k)} = (U^{(j,k)})_{|S_k}$, then $X^{(j,k)} \cong Y^{(j,k)} \cong U^{(j,k)}$. In particular, $X^{(j,k)} = \{0\} \iff Y^{(j,k)} = \{0\} \iff U^{(j,k)} = \{0\} \times \{0\} \iff$ the fragment $F^{(j,k)}$ is externally observable. The unobservable projections $X^{(j,k)} = ((\mathcal{Y}^o)^{\perp})^{(j,k)}$ and $Y^{(j,k)} = ((\mathcal{W}^o)^{\perp})^{(j,k)}$ act as the dual groups to, and thus in our setting are isomorphic to, the unreachable quotients $\hat{S}_j/(\mathcal{Y}^o)^{\perp}$ and $\hat{S}_k/(\mathcal{W}^o)^{\perp}$ of the dual realization.

This theorem shows that the unobservable trajectories in any proper trellis form a set of parallel transitions between isomorphic subspaces $X^{(j,k)} \subseteq S_j$ and $Y^{(j,k)} \subseteq S_k$, with the transitions corresponding to those of the isomorphism $X^{(j,k)} \leftrightarrow Y^{(j,k)}$.

Example 2 (self-dual trellis fragment). Figure 24 depicts the external behavior $C^{(j,k)} = \langle (01,00,01), (10,01,00), (00,10,10) \rangle$ of a trellis fragment $F^{(j,k)}$ with $A^{(j,k)} = S_j = S_k = (\mathbb{Z}_2)^2$. Its reachable subspaces $\mathcal{Y}^{(j,k)}$ and $\mathcal{W}^{(j,k)}$ are strict subspaces of $(\mathbb{Z}_2)^2$, so it is externally uncontrollable; indeed, $C^{(j,k)}$ is the union of two disconnected cosets of $\langle (01,00,01), (10,01,00) \rangle$. The start space is generated by $(01,00,01)$, the stop space by $(10,01,00)$, and the unreachable space by $(00,10,10)$.

![Figure 24: Self-dual externally unobservable and uncontrollable trellis fragment.](image)

Notice that $C^{(j,k)}$ is self-dual; i.e., $C^{(j,k)} = (C^{(j,k)})^\perp$. Its unobservable transition space is $\langle (10,10) \rangle$, so it is externally unobservable. The projections $X^{(j,k)}$ and $Y^{(j,k)}$ are both equal to the orthogonal subspace $\langle 10 \rangle$ to the reachable subspace $\langle 01 \rangle$, and the two unobservable trajectories form a set of parallel transitions corresponding to the isomorphism between these projections.

6.4 Connecting trellis fragments

We now connect two trellis fragments with external behaviors $C^{(j,k)} \subseteq A^{(j,k)} \times S_j \times S_k$ and $C^{(k,\ell)} \subseteq A^{(k,\ell)} \times S_k \times S_\ell$ via their common state space $S_k$. The combination is a trellis fragment with symbol alphabet $A^{(j,\ell)} = A^{(j,k)} \times A^{(k,\ell)}$, external state spaces $S_j$ and $S_\ell$, and internal state space $S_k$.

The observability properties of the combined behavior are as follows:

Theorem (observability under connection). A connection of two proper trellis fragments is proper. Its unobservable projections $X^{(j,\ell)} \subseteq S_j$ and $Y^{(j,\ell)} \subseteq S_\ell$ are isomorphic to $Y^{(j,k)} \cap X^{(k,\ell)} \subseteq S_k$.

Proof: Properness has already been shown in the connected fragments lemma.

A nontrivial unobservable configuration $(0^{(j,\ell)}, s_j, s_\ell) \in C^{(j,\ell)}$ must be the concatenation of non-trivial unobservable configurations $(0^{(j,k)}, s_j, s_k) \in C^{(j,k)}$ and $(0^{(k,\ell)}, s_k, s_\ell) \in C^{(k,\ell)}$, with $s_j$, $s_k$ and $s_\ell$ all nonzero by properness. Moreover, the state $s_k$ must lie in the intersection of the unobservable projections $Y^{(j,k)} = (U^{(j,k)})_{|S_k}$ and $X^{(k,\ell)} = (U^{(k,\ell)})_{|S_k}$. In turn, the state $s_j$ must lie in the image of $Y^{(j,k)} \cap X^{(k,\ell)}$ under the unobservability isomorphism from $S_k$ to $S_j$, and similarly $s_\ell$ must lie in the image of $Y^{(j,k)} \cap X^{(k,\ell)}$ under the unobservability isomorphism from $S_k$ to $S_\ell$. 

29
The unobservable projections $X^{(j,\ell)}$ and $Y^{(j,\ell)}$ of the concatenated behavior are thus both isomorphic to $Y^{(j,k)} \cap X^{(k,\ell)}$, which is not larger than $Y^{(j,k)} \cong X^{(j,k)}$ or $X^{(k,\ell)} \cong Y^{(k,\ell)}$. Unobservable projections are thus nonincreasing under combination. In this sense, a combination is “no less observable” than its parts. In particular, if $R^{(j,k)}$ is externally observable or $R^{(k,\ell)}$ is externally observable, then their combination is externally observable.

Dually, the reachability properties of the combined behavior are as follows:

**Theorem (reachability under combination).** A combination of two trim trellis fragments is trim. Its unreachable quotients $S_j/Y^{(j,\ell)}$ and $S_\ell/W^{(j,\ell)}$ are isomorphic to $S_k/(W^{(j,k)} + Y^{(k,\ell)})$.

**Proof:** Trimness has already been shown in the connected fragments lemma.

**Direct Proof:** $s_\ell \in W^{(j,\ell)} \Leftrightarrow$ there is a transition $(0,s_\ell) \in T^{(j,\ell)} \Leftrightarrow$ there is an $s_k \in S_k$ such that $(0,s_k) \in T^{(j,k)}$ (i.e., $s_k \in W^{(j,k)}$) and $(s_k,s_\ell) \in T^{(k,\ell)}$. But by the reachable spaces theorem, $(s_k,s_\ell) \in T^{(k,\ell)}$ if and only if the coset $s_k + W^{(k,\ell)}$ corresponds to the coset $s_\ell + W^{(k,\ell)}$ under the reachability isomorphism between $S_k/Y^{(k,\ell)}$ and $S_\ell/W^{(k,\ell)}$. Thus $W^{(j,\ell)} \subseteq S_\ell$ corresponds to $W^{(j,k)} + Y^{(k,\ell)} \subseteq S_k$ under the reachability isomorphism, implying $S_\ell/W^{(j,\ell)} \cong S_k/(W^{(j,k)} + Y^{(k,\ell)})$. Similarly, we have $S_j/Y^{(j,\ell)} \cong S_k/(W^{(j,k)} + Y^{(k,\ell)})$.

**Duality Proof:** By the unobservable projections theorem, the unreachable quotients $S_j/Y^{(j,\ell)}$ and $S_\ell/W^{(j,\ell)}$ are isomorphic to the dual unreachable quotients $(X^{(\ell)})^{j,\ell}$ and $(Y^{(\ell)})^{j,\ell}$, respectively. By the observability under combination theorem, $(X^{(\ell)})^{j,\ell} \subseteq S_j$ and $(Y^{(\ell)})^{j,\ell} \subseteq S_\ell$ are isomorphic to $(Y^{(\ell)})^{j,k} \cap (X^{(\ell)})^{k,\ell} \subseteq S_k$. This implies

$$\frac{S_j}{((Y^{(\ell)})^{j,\ell})^\perp} \cong \frac{S_\ell}{((X^{(\ell)})^{j,\ell})^\perp} \cong \frac{S_k}{((Y^{(\ell)})^{j,k} \cap (X^{(\ell)})^{k,\ell})^\perp}.$$ 

But $((Y^{(\ell)})^{j,\ell})^\perp = Y^{(j,\ell)}$, $((X^{(\ell)})^{j,\ell})^\perp = W^{(j,\ell)}$, and, by sum/intersection duality,

$$((Y^{(\ell)})^{j,k} \cap (X^{(\ell)})^{k,\ell})^\perp = ((Y^{(\ell)})^{j,k})^\perp + ((X^{(\ell)})^{k,\ell})^\perp = W^{(j,k)} + Y^{(j,\ell)}.$$ 

The unreachable quotients $S_j/Y^{(j,\ell)}$ and $S_\ell/W^{(j,\ell)}$ of the combined behavior are thus both isomorphic to $S_k/(W^{(j,k)} + Y^{(k,\ell)})$, which is not larger than $S_k/Y^{(j,k)} \cong S_j/Y^{(j,\ell)}$ or $S_k/W^{(k,\ell)} \cong S_\ell/W^{(k,\ell)}$. Unreachability quotients are thus nonincreasing under combination. In this sense, a combination is “no less controllable” than its parts. In particular, if $R^{(j,k)}$ is externally controllable or $R^{(k,\ell)}$ is externally controllable, then their combination is externally controllable.

### 7 Cyclic realizations

Given a normal realization $R$ with graph $G$, if one edge of $G$ is removed, then the resulting graph is either connected or disconnected. We have already analyzed the latter case in Section 4.4 *et seq.* In this section, we will analyze the case where $G$ remains connected, which can occur only when $G$ is cyclic. The cyclomatic number (“loopiness index”) of $G$ is then reduced by 1.

We study the relation between the properties of $R$ and those of the connected fragment $R^{(\ell)}$ created by a cut of an edge $S_j$. We show that if $R$ is state-trim and (internally) controllable, then $R^{(\ell)}$ is externally controllable. For a linear realization $R$, we give a local merging (resp. trimming) procedure similar to that in [7] that makes $R$ (internally) controllable (resp. observable) at $S_j$; however, in general this procedure fails in the group case.
7.1 Cutting cycles

We consider a cyclic normal realization $\mathcal{R}$ as in Figure 4 with internal behavior $\mathcal{B}$ and external behavior $\mathcal{C}$. As usual, we assume that $\mathcal{R}$ is trim and proper. We may also assume without loss of generality (see Section 4.8) that $\mathcal{R}$ is leafless; i.e., the normal graph $\mathcal{G}$ of $\mathcal{R}$ is a generalized cycle.

We assume that the fragment $\mathcal{R}^{(j)}$ that results from cutting one edge $S_j$ is connected—i.e., the edge $S_j$ is not a cut set. Then $\mathcal{R}^{(j)}$ is a trellis fragment with two external state variables with values $s_j \in S_j$ and $s_j' \in S_j$, symbol configurations $a \in A$, and external behavior $\mathcal{C}^{(j)} \subseteq A \times S_j \times S_j$.

The trellis fragment $\mathcal{R}^{(j)}$ is illustrated in Figure 25(a). As in Figure 22(a), we have introduced a reduced symbol configuration space $\tilde{A} = (\mathcal{C}^{(j)})_{|A}/(\mathcal{C}^{(j)})_{|A}$, and a reduced external behavior $\tilde{C}^{(j)} \subseteq \tilde{A} \times S_j \times S_j$. The dual realization of the dual fragment $(\mathcal{R}^{(j)})^\circ$ is shown in Figure 25(b).

![Figure 25: Realizations of (a) fragment $\mathcal{R}^{(j)}$; (b) dual fragment $(\mathcal{R}^{(j)})^\circ$.](image)

The transition space $T^{(j)} \subseteq S_j \times S_j$ of $\mathcal{R}^{(j)}$ is the projection $(\mathcal{C}^{(j)})_{|S_j \times S_j}$, and its unobservable transition space $\mathcal{U}^{(j)} \subseteq S_j \times S_j$ is the cross-section $(\mathcal{C}^{(j)})_{|S_j \times S_j}$. $\mathcal{R}^{(j)}$ is externally controllable if $T^{(j)} = S_j \times S_j$, and externally observable if $\mathcal{U}^{(j)}$ is trivial.

The original realization $\mathcal{R}$ may be recovered by imposing an equality constraint $C_{=S_j} = \{(s_j, s_j') \in S_j \times S_j : s_j = s_j'\}$ on $s_j$ and $s_j'$. Thus $\mathcal{R}$ may be viewed as a realization with one constraint code $\mathcal{C}^{(j)}$ and one internal state space $S_j$.

7.2 State-trimness

We first consider the relation between the state-trimness of $\mathcal{R}$ at $S_j$ and the properties of $\mathcal{R}^{(j)}$.

A realization $\mathcal{R}$ is state-trim at $S_j$ if every state $s_j \in S_j$ lies on a valid configuration $(a, s, s)$ in the extended behavior $\mathcal{B}$ of $\mathcal{R}$. In terms of $\mathcal{R}^{(j)}$, $\mathcal{R}$ is thus state-trim at $S_j$ if every pair $(s_j, s_j) \in C_{=S_j}$ lies on a valid configuration in $\mathcal{R}^{(j)}$; i.e., if $C_{=S_j} \subseteq T^{(j)}$.

We can then generalize [10] Theorem 5.2 to show that, at least when $\mathcal{R}$ is internally controllable, state-trimness of $\mathcal{R}$ at $S_j$ is equivalent to external controllability of $\mathcal{R}^{(j)}$, as follows:

**Theorem (state-trimness).** Let $\mathcal{R}^{(j)}$ be a connected fragment of a normal linear or group realization $\mathcal{R}$ that results from cutting an edge $S_j$.

(a) $\mathcal{R}$ is state-trim at $S_j$ if $\mathcal{R}^{(j)}$ is externally controllable.

(b) If $\mathcal{R}$ is internally controllable, then $\mathcal{R}^{(j)}$ is externally controllable if $\mathcal{R}$ is state-trim at $S_j$.

**Proof:** (a) If $\mathcal{R}^{(j)}$ is externally controllable, then $T^{(j)} = S_j \times S_j$, so $C_{=S_j} \subseteq T^{(j)}$.

(b) If $\mathcal{R}$ is state-trim at $S_j$, then $C_{=S_j} \subseteq T^{(j)}$, so the dual unobservable transition space $(\mathcal{U}^\circ)^{(j)} = (T^{(j)})^\perp$ satisfies $(\mathcal{U}^\circ)^{(j)} \subseteq (C_{=S_j})^\perp = C_{=S_j}$, the sign inversion constraint over $S_j$. Thus if $(\mathcal{U}^\circ)^{(j)}$ is nontrivial, then there is a configuration with $a = 0$ and $s_j' = -s_j \neq 0$ in the dual realization $\mathcal{R}^\circ$, so it is internally unobservable. But if $\mathcal{R}$ is internally controllable, then $\mathcal{R}^\circ$ must be internally observable; therefore $(\mathcal{U}^\circ)^{(j)}$ must be trivial, which implies that $T^{(j)} = S_j \times S_j$; i.e., $\mathcal{R}^{(j)}$ is externally controllable. □
A realization that is not state-trim may be made so by the local reduction of state-trimming. Therefore we may and will assume that all realizations \( R \) and dual realizations \( R^o \) are state-trim.

Given state-trimmness, part (b) of the theorem above shows that if \( R \) is (internally) controllable, then every connected fragment \( R^{(j)} \) resulting from cutting an edge \( S_j \) is externally controllable; i.e., all transitions \( (s_j, s'_j) \in S_j \times S_j \) are possible. Dually, assuming dual state-trimmness, every fragment \( R^{(j)} \) of an (internally) observable realization \( R \) is externally observable.

### 7.3 Internal observability and controllability at \( S_j \)

We next relate the (internal) observability and controllability of \( R \) and \( R^o \) at \( S_j \) to the structure of the dual trellis fragments \( R^{(j)} \) and \( (R^o)^{(j)} \), and in particular to their unobservable state spaces and controllable subspaces.

We view \( R \) as a realization with one constraint code \( C^{(j)} \) and one internal state space \( S_j \). We say that \( R \) is (internally) observable at \( S_j \) if there is no nontrivial configuration \( (0, s_j, s_j) \) in its extended behavior \( B = C^{(j)} \cap (A \times C_{=S_j}) \), where \( A \times C_{=S_j} = \{(a, s_j, s_j) \in A \times S_j \times S_j \} \) is the validity space of \( R \) in this context.

As in Section 3.7 the unobservable extended behavior of \( R \) is \( \bar{B}^u = C^{(j)} \cap (\{0\} \times C_{=S_j}) \), and the unobservable state space is \( (S_j)^u = \{s_j \in S_j : (0, s_j, s_j) \in \bar{B}^u \} \), which is isomorphic to \( \bar{B}^u \). A realization of \( (S_j)^u \) is shown in Figure 26(a).

Evidently \( (S_j)^u \) is the intersection \( \bar{X}^{(j)} \cap \bar{Y}^{(j)} \) of the unobservable projections \( \bar{X}^{(j)} \) and \( \bar{Y}^{(j)} \), the projections of the unobservable transition space \( \bar{U}^{(j)} \subseteq S_j \times S_j \) onto its two external state variables. Thus \( R \) is (internally) observable at \( S_j \) if and only if \( (S_j)^u = \bar{X}^{(j)} \cap \bar{Y}^{(j)} \) is trivial.

![Figure 26: Dual normal realizations of (a) \((S_j)^u = \bar{X}^{(j)} \cap \bar{Y}^{(j)} \); (b) \((S_j)^c = (\bar{V}^o)^{(j)} + (\bar{W}^o)^{(j)} \).](image)

By normal realization duality, the dual realization of Figure 26(b) realizes \((S_j)^c \), the controllable subspace \( (S_j)^c \) of the dual realization \( R^o \). By observability/controllability duality, \( R^o \) is (internally) controllable at \( S_j \) if and only if \( (S_j)^c = S_j \). Since \( (S_j)^u = \bar{X}^{(j)} \cap \bar{Y}^{(j)} \), we have

\[
(S_j)^c = (\bar{X}^{(j)})^\perp + (\bar{Y}^{(j)})^\perp = (\bar{V}^o)^{(j)} + (\bar{W}^o)^{(j)},
\]

where \( (\bar{V}^o)^{(j)} \) and \( (\bar{W}^o)^{(j)} \) are the reachable subspaces of the dual external behavior \( (C^{(j)})^\perp \), namely the cross-sections of the dual transition space \( (T^o)^{(j)} \) on its two external state variables.

In summary, we have:

**Theorem (observability/controllability at \( S_j \))** For a realization \( R \) with a single constraint code \( C^{(j)} \) and a single internal state space \( S_j \), the unobservable state space is the intersection \( (S_j)^u = \bar{X}^{(j)} \cap \bar{Y}^{(j)} \) of the unobservable projections \( \bar{X}^{(j)} \) and \( \bar{Y}^{(j)} \), and the controllable subspace is the sum \( (S_j)^c = \bar{V}^{(j)} + \bar{W}^{(j)} \) of the reachable subspaces \( \bar{V}^{(j)} \) and \( \bar{W}^{(j)} \). \( R \) is (internally) observable at \( S_j \) if and only if \( (S_j)^u = \{0\} \), and controllable at \( S_j \) if and only if \( (S_j)^c = S_j \). \( \blacksquare \)
7.4 Reduction of unobservable/uncontrollable realizations

We now show how realizations that are (internally) unobservable or uncontrollable at $\mathcal{S}_j$ may be made observable or controllable by a trimming or merging of $\mathcal{S}_j$ into a reduced state space $\mathcal{S}_j'$. The procedure is a straightforward generalization of the reduction procedure of [7] for unobservable or uncontrollable linear tail-biting trellises. The reduction yields a linear or group realization provided that a certain algebraic condition holds; this condition always holds in the linear case.

Suppose then that $\mathcal{R}$ is a realization with one constraint code $C^{(\langle j \rangle)}$ and one internal state space $\mathcal{S}_j$, and that $\mathcal{R}$ is unobservable at $\mathcal{S}_j$; i.e., $(\mathcal{S}_j)^u = \mathcal{X}^{(\langle j \rangle)} \cap \mathcal{Y}^{(\langle j \rangle)} \neq \{0\}$, where $\mathcal{X}^{(\langle j \rangle)}$ and $\mathcal{Y}^{(\langle j \rangle)}$ are the unobservable projections of $C^{(\langle j \rangle)}$.

Suppose further that $(\mathcal{S}_j)^u$ is a direct summand of $\mathcal{S}_j$; i.e., that there is a complementary subspace $\bar{\mathcal{S}}_j$ such that $\mathcal{S}_j = (\mathcal{S}_j)^u \times \bar{\mathcal{S}}_j$, so every element of $\mathcal{S}_j$ can be expressed uniquely as $s_j = s_j^u + \bar{s}_j$ for some $s_j^u \in (\mathcal{S}_j)^u$, $\bar{s}_j \in \bar{\mathcal{S}}_j$. In the linear case, where $\mathcal{S}_j$ is actually a vector space and $(\mathcal{S}_j)^u$ a subspace, there always exists such a complementary subspace $\bar{\mathcal{S}}_j$; however, in the group case this assumption may fail.

Under this assumption, we have the following reduction procedure:

**Theorem (unobservable realization reduction).** Let $\mathcal{R}$ be a linear or group realization that is unobservable at $\mathcal{S}_j$, so $(\mathcal{S}_j)^u \neq \{0\}$, and suppose that $(\mathcal{S}_j)^u$ is a direct summand of $\mathcal{S}_j$, with complementary subspace $\bar{\mathcal{S}}_j$. Then $\mathcal{S}_j$ may be trimmed to $\bar{\mathcal{S}}_j$ without changing the code $\mathcal{C}$ realized by $\mathcal{R}$. Moreover, the trimmed realization $\bar{\mathcal{R}}$ is an (internally) observable linear or group realization.

**Proof:** $\mathcal{C}$ is the image of the projection of the behavior $\mathcal{B} \subseteq \mathcal{A} \times \mathcal{S}_j = \mathcal{A} \times (\mathcal{S}_j)^u \times \bar{\mathcal{S}}_j$ onto the symbol configuration space $\mathcal{A}$, whose kernel is the unobservable behavior $\mathcal{B}^u = \{(0, s_j^u, 0) \in \mathcal{B}\} = \{0\} \times (\mathcal{S}_j)^u \times \{0\}$. By the fundamental theorem of homomorphisms, $\mathcal{C} \cong \mathcal{B}/\mathcal{B}^u \cong \mathcal{A} \times \bar{\mathcal{S}}_j$. Thus if $\mathcal{S}_j$ is trimmed to $\bar{\mathcal{S}}_j$, then the trimmed behavior $\bar{\mathcal{B}}$ is the projection of $\mathcal{B} \subseteq \mathcal{A} \times (\mathcal{S}_j)^u \times \bar{\mathcal{S}}_j$ onto $\mathcal{A} \times \bar{\mathcal{S}}_j \cong \mathcal{C}$. Since $\bar{\mathcal{B}}$ is a group, the trimmed realization $\bar{\mathcal{R}}$ is a group realization. Moreover, the projection of $\bar{\mathcal{B}}$ onto $\mathcal{A}$ is an isomorphism, so $\bar{\mathcal{R}}$ is (internally) observable.

Figure 27(a) shows this reduction; $\mathcal{S}_j$ is simply restricted to the complementary subspace $\bar{\mathcal{S}}_j$.

![Figure 27](image)

We remark that even without the direct summand assumption, if $\mathcal{S}_j$ is trimmed to a set of coset representatives $\tilde{\mathcal{S}}_j = [\mathcal{S}_j/(\mathcal{S}_j)^u]$ of the quotient group $\mathcal{S}_j/(\mathcal{S}_j)^u$, then the resulting trimmed realization $\bar{\mathcal{R}}$ realizes $\mathcal{C}$ and is one-to-one. The difficulty is that it may not be possible to find a subgroup of $\mathcal{S}_j$ that can serve as the set of coset representatives $[\mathcal{S}_j/(\mathcal{S}_j)^u]$, so the trimmed state space $\bar{\mathcal{S}}_j$ may not be a group.

---

4 As noted in [10], this difference between groups and vector spaces also underlies why the canonical realizations of group codes in [8, 17] may not be homomorphic, even though they are necessarily linear in the linear case.
As we have seen, \( \mathcal{R} \) is unobservable with unobservable state space \((\mathcal{S}_j)^u\) if and only if the dual realization \( \mathcal{R}^\circ \) is controllable with controllable subspace \((\mathcal{S}_j)^c = ((\mathcal{S}_j)^u)^\perp \). Moreover, it is easy to see that \((\mathcal{S}_j)^u\) is a direct summand of \( \mathcal{S}_j \) with complementary subspace \( \hat{\mathcal{S}}_j \) if and only if \((\mathcal{S}_j)^c\) is a direct summand of \( \hat{\mathcal{S}}_j \) with complementary subspace \( (\hat{\mathcal{S}}_j)^\perp \). Finally, the dual realization \( \mathcal{R}^\circ \) to the trimmed realization \( \tilde{\mathcal{R}} \) is the linear or group realization obtained from \( \mathcal{R}^\circ \) by merging \( \hat{\mathcal{S}}_j \) to its quotient \( \hat{\mathcal{S}}_j/(\hat{\mathcal{S}}_j)^\perp = (\hat{\mathcal{S}}_j)^c \). By normal realization duality, \( \mathcal{R}^\circ \) realizes \( \mathcal{C}_\perp \), and by observability/controllability duality, \( \mathcal{R}^c \) is controllable. We thus have as a corollary:

**Theorem (reduction of uncontrollable realizations).** Let \( \mathcal{R} \) be a linear or group realization that is uncontrollable at \( \mathcal{S}_j \), so \((\mathcal{S}_j)^c \neq \mathcal{S}_j\), and suppose that \((\mathcal{S}_j)^c\) is a direct summand of \( \mathcal{S}_j \). Then \( \mathcal{S}_j \) may be merged to \((\mathcal{S}_j)^c\) without changing the code \( \mathcal{C} \) realized by \( \mathcal{R} \). Moreover, the merged realization \( \tilde{\mathcal{R}} \) is an (internally) controllable linear or group realization. \( \Box \)

Figure 27(b) shows this dual reduction; \( \hat{\mathcal{S}}_j \) is merged to \((\hat{\mathcal{S}}_j)^c = \hat{\mathcal{S}}_j/(\hat{\mathcal{S}}_j)^\perp\) via a natural map.

**Example 2** (self-dual trellis fragment, cont.). Let us take the self-dual trellis fragment of Figure 24 as the external behavior \( \mathcal{C}^{(\cup)} \) of a realization \( \mathcal{R}^{(\cup)} \) that has been cut at a state \( \mathcal{S}_j \) with alphabet \((\mathbb{Z}_2)^2\). We recall that this fragment is both externally unobservable and externally uncontrollable. The realization \( \mathcal{R} \) obtained by constraining the two state spaces to be equal realizes the code \( \mathcal{C} = \langle 11 \rangle \), and is both externally unobservable with unobservable state space \((\mathcal{S}_j)^u = \langle 10 \rangle \), and externally uncontrollable with controllable subspace \((\mathcal{S}_j)^c = \langle 01 \rangle \).

\[
\begin{array}{cccccc}
11 & 11 & 11 & 01,10 & 00,11 & 01,10 \\
00 & 00 & 00 & 00,11 & 01,10 & 00,11 \\
\text{(a)} & \text{(b)} & \text{(c)} & \text{(d)}
\end{array}
\]

**Figure 28:** Reductions of self-dual trellis fragment.

To reduce \( \mathcal{R} \) to an observable realization \( \tilde{\mathcal{R}} \), we may choose either of two complementary subspaces \( \hat{\mathcal{S}}_j \), namely \( \langle 11 \rangle \) or \( \langle 01 \rangle \). If we restrict \( \mathcal{S}_j \) to \( \langle 11 \rangle \), then we obtain a linear observable realization that generates \( \mathcal{C} \), but is uncontrollable, shown in Figure 28(a). If the other hand we restrict \( \mathcal{S}_j \) to \( \langle 01 \rangle \), then we obtain a linear observable realization that generates \( \mathcal{C} \) and moreover is controllable, shown in Figure 28(b).

Dually, to reduce \( \mathcal{R} \) to a controllable realization \( \tilde{\mathcal{R}} \), we may choose either of two complementary subspaces to \( \langle 01 \rangle \), namely \( \langle 11 \rangle \) or \( \langle 10 \rangle \). If we merge \( \mathcal{S}_j \) to \( \mathcal{S}_j/\langle 11 \rangle \), then state 11 maps to 00, so we obtain a linear controllable realization that generates \( \mathcal{C} \), but is unobservable, shown in Figure 28(c). If instead we merge \( \mathcal{S}_j \) to \( \mathcal{S}_j/\langle 10 \rangle \), then state 10 maps to 00, so we obtain a linear controllable realization that generates \( \mathcal{C} \) and is observable, shown in Figure 28(d). \( \Box \)
Example 3 (group trellis fragment). Now let us consider the group trellis fragment of Figure 29(a) as the external behavior $C(j)$ of a realization $R(j)$ that has been cut at a state $S_j$ with alphabet $Z_4$. $C(j)$ is isomorphic to $Z_4 \times Z_2$, with generators $(2, 1, 1)$ and $(2, 0, 2)$. The realization $R$ obtained by constraining $s'_j = s_j$ realizes the code $C = 2Z_4$, and is externally unobservable with unobservable state space $(S_j)^u = 2Z_4$.

We note that the only possible restriction of $S_j$ to a subgroup, namely to $2Z_4$, does not yield a realization of $C = 2Z_4$, but rather of $\{0\}$. Thus our reduction procedure does not work in this case.

![Figure 29: Dual group trellis fragments.](image)

The dual fragment $(C(j))^\perp$, shown in Figure 29(b), is isomorphic to $Z_4 \times Z_2$, with generators $(1, 1, 1)$ and $(2, 0, 0)$. The realization $R$ obtained by constraining $s'_j = -\hat{s}_j$ realizes the code $(C)^\perp = 2Z_4$, and is externally uncontrollable with controllable subspace $(\hat{S}_j)^c = 2Z_4$.

We observe that the only quotient of $Z_4$ is $Z_4 / 2Z_4$, and merging $Z_4 \rightarrow Z_4 / 2Z_4$ yields a realization of $Z_4$, not $(C)^\perp = 2Z_4$; thus our reduction procedure does not work in this case either. □

7.5 Discussion

This is the only place in this paper where we observe a difference between the linear and group cases.

In the linear case, we see that it is always possible to reduce an (internally) unobservable realization $R$ to an observable linear realization (resp. uncontrollable to controllable) by a series of local reductions at state spaces $S_j$ at which $R$ is unobservable. Thus without loss of generality we may assume that every linear realization $R$ has been reduced to be observable and controllable—as well as state-trim, with a state-trim dual. By the state-trim theorem, every connected fragment $R^{(j)}$ is then externally observable and controllable.

However, in the group case, we have no generally valid procedure for reducing an unobservable (resp. uncontrollable) group realization to an observable (resp. controllable) group realization. Finding such a procedure is an open topic for future research.
8 Conclusion

This paper has started to develop a fundamental structure theory for linear and group codes on general graphs, using only elementary group and graph theory, including basic group duality theory.

Remarkably, these tools suffice to develop a structure theory for realizations on cycle-free graphs. For cyclic graphs, on the other hand, while the results of this paper may be a good starting point, there remain many open questions.

Almost none of our results relate to the properties of symbol configurations, which ultimately determine code properties. We have not rederived the well-known “shortest basis theorem” [5] for conventional trellis realizations, much less the related results of Koetter and Vardy [15] and subsequent authors (e.g., [11, 12, 10]) for tail-biting trellis realizations. We believe that the theory of this paper should be extended to cover “granules” as in [8, 9].

Also, although we have a fairly complete structure theory for leaf and trellis fragments, we lack such a theory for cubic and hypercubic fragments. A straightforward extension of the methods used here soon bogs down under the curse of dimensionality.

However, we note that it has been shown in [4, Internal Node Theorem] that linear constraint codes of degree 4 or more may be replaced by two or more connected constraint codes of degree 3 or less, without increasing the maximum constraint code dimension. Thus one may be able to assume that every normal realization or fragment consists only of leaf, trellis and cubic constraint codes, without incurring a complexity penalty in this sense.

Acknowledgments

I thank Heide Gluesing-Luerssen for an ongoing research collaboration under which many of these results were first developed, and for extensive comments on this paper. I also thank Moshe Schwartz for help with elementary graph theory.

References

[1] C. Berge, The Theory of Graphs. New York: Wiley, 1962.
[2] B. Bollobás, “The evolution of sparse graphs,” Graph Theory and Combinatorics (Cambridge, 1983), pp. 35–57. London: Academic Press, 1984.
[3] G. D. Forney, Jr., “Codes on graphs: Normal realizations,” IEEE Trans. Inf. Theory, vol. 47, pp. 520–548, Feb. 2001.
[4] G. D. Forney, Jr., “Codes on graphs: Constraint complexity of cycle-free realizations of linear codes,” IEEE Trans. Inf. Theory, vol. 49, pp. 1597–1610, Jul. 2003.
[5] G. D. Forney, Jr., “Minimal realizations of linear systems: The ‘shortest basis’ approach,” IEEE Trans. Inf. Theory, vol. 57, pp. 726–737, Feb. 2011.
[6] G. D. Forney, Jr., “Codes on graphs: Past, present and future,” Proc. 2012 Allerton Conf. (Monticello, IL), Oct. 2012.
[7] G. D. Forney, Jr. and H. Gluesing-Luerssen, “Codes on graphs: Observability, controllability and local reducibility,” IEEE Trans. Inf. Theory, vol. 59, pp. 223–238, Jan. 2013.
[8] G. D. Forney, Jr. and M. D. Trott, “The dynamics of group codes: State spaces, trellis diagrams and canonical encoders,” *IEEE Trans. Inf. Theory*, vol. 39, pp. 1491–1513, Sept. 1993.

[9] G. D. Forney, Jr. and M. D. Trott, “The dynamics of group codes: Dual abelian group codes and systems,” *IEEE Trans. Inf. Theory*, vol. 50, pp. 2935–2965, Dec. 2004.

[10] H. Gluesing-Luerssen and G. D. Forney, Jr., “Local irreducibility of tail-biting trellises,” *IEEE Trans. Inf. Theory*, to appear, 2013.

[11] H. Gluesing-Luerssen and E. A. Weaver, “Linear tail-biting trellises: Characteristic generators and the BCJR construction,” *IEEE Trans. Inf. Theory*, vol. 57, pp. 738–751, Feb. 2011.

[12] H. Gluesing-Luerssen and E. A. Weaver, “Characteristic generators and dualization for tail-biting trellises,” *IEEE Trans. Inf. Theory*, vol. 57, pp. 7418–7430, Nov. 2011.

[13] M. Hall, Jr., *The Theory of Groups*. New York: MacMillan, 1959.

[14] R. Koetter, “On the representation of codes in Forney graphs,” in *Codes, Graphs, and Systems: A Celebration of the Life and Career of G. David Forney, Jr.* (R. E. Blahut, R. Koetter, eds.), pp. 425–450. Kluwer, 2002. On-line: www.comm.csl.uiuc.edu/~koetter/publications/final_FG.ps.

[15] R. Koetter and A. Vardy, “The structure of tail-biting trellises: Minimality and basic principles,” *IEEE Trans. Inf. Theory*, vol. 49, pp. 2081–2105, Sept. 2003.

[16] H.-A. Loeliger, G. D. Forney, Jr., T. Mittelholzer and M. D. Trott, “Minimality and observability of group systems,” *Linear Algebra Appl.*, vol. 205–206, pp. 937-963, July 1994.

[17] H.-A. Loeliger and T. Mittelholzer, “Convolutional codes over groups,” *IEEE Trans. Inf. Theory*, vol. 42, pp. 1660–1686, Nov. 1996.

[18] R. M. Tanner, “A recursive approach to low-complexity codes,” *IEEE Trans. Inf. Theory*, vol. IT–27, pp. 533–547, Sept. 1981.

[19] H. Trentelman, “A truly behavioral approach to the \( H_\infty \) control problem,” in *The Mathematics of Systems and Control: From Intelligent Control to Behavioral Systems* (J. W. Polderman and H. Trentelman, eds.). Groningen, The Netherlands: U. Groningen, 1999, pp. 177–190.

[20] P. O. Vontobel and H-A. Loeliger, “On factor graphs and electrical networks,” in *Mathematical Systems Theory in Biology, Communication, Computation, and Finance* (J. Rosenthal and D. S. Gilliam, eds.), pp. 469–492. New York: Springer, 2003.

[21] N. Wiberg, “Codes and decoding on general graphs,” Ph.D. dissertation, Univ. Linköping, Linköping, Sweden, 1996.

[22] N. Wiberg, H.-A. Loeliger, and R. Kötter, “Codes and iterative decoding on general graphs,” *Euro. Trans. Telecomm.*, vol. 6, pp. 513–525, Sept./Oct. 1995.

[23] J. C. Willems, “Models for dynamics,” in *Dynamics Reported* (U. Kirchgraber and H. O. Walther, eds.), vol. 2, pp. 171–269. New York: Wiley, 1989.