N93-29595

TDA Progress Report 42-113 May 15, 1993

The General Theory of Convolutional Codes

R. J. McEliece[1]
Communications Systems Research Section

R. P. Stanley[2]

This article presents a self-contained introduction to the algebraic theory of convolutional codes, which is partly tutorial, but at the same time contains a number of new results which will prove useful for designers of advanced telecommunication systems. Among the new concepts introduced here are the Hilbert series *for a convolutional code and the class of* compact *codes.*

I. Introduction

Convolutional codes have played a central part in NASA's deep-space telecommunications systems for many years. In all such applications to date, the convolutional codes have been codes of *dimension 1*, which are commonly, but not strictly correctly, referred to as "rate $1/n$" codes. However, as systems become more sophisticated, the coding subsystems must keep pace, and this article is an outline of algebraic theory for the most general class of convolutional codes known, the so-called "(n, k, m)" codes, of which the usual "rate $1/n$" codes form the special case $k = 1$. Much of this theory was originally developed by Forney [1–4], but this article adds to what is already known, as well as placing many of the older results into a modern, "system-theoretic" context. In particular, introduced here for the first time is the "Hilbert series" for a convolutional code, which is a generating function from which the dimensions of certain polynomial subcodes can be easily computed in terms of the "Forney indices" of the code. The Forney indices provide a derivation of an upper bound on the free distance of a convolutional code which in some cases improves the bounds previously known, and whose derivation makes no use of the structure of any particular encoder structure. Finally, the notion of "compact" and "noncompact" convolutional codes is introduced, and it is argued that only compact codes are likely to be interesting for applications.

II. Convolutional Codes: Polynomial Generator Matrices

Let F be a field, usually $GF(2)$, and let $F(D)$ be the field of rational functions over F. An (n, k) *convolutional code over* F is a k-dimensional subspace of $F(D)^n$. The elements of the code are called its *codewords*. A codeword is thus an n-tuple of rational functions over $F(D)$. The *weight* of a codeword is defined to be the sum of the weights of its components, where the weight of a component (i.e., rational function) is the number of nonzero coefficients in its expansion as a Laurent series in increasing powers of D. The *free distance* of a convolutional code is defined to be the minimum nonzero weight of any codeword.

[1] Consultant, California Institute of Technology, Engineering Department.

[2] Consultant, Massachusetts Institute of Technology, Department of Mathematics.

If C is an (n, k) convolutional code over F, a *generator matrix* $G(D)$ for C is a $k \times n$ matrix over $F(D)$ whose rows form a basis for C. If the entries of $G(D)$ are polynomials, then $G(D)$ is called a *polynomial generator matrix* (PGM) for C. Any convolutional code has a polynomial generator matrix, since if G is an arbitrary generator matrix for C, the matrix obtained from G by multiplying each row by the least common multiple of the denominators of the entries in that row is a PGM for C.

Let $G(D) = (g_{ij}(D))$ be a $k \times n$ PGM for C. The ith row of G, i.e., the n-vector $(g_{i1}, \ldots, g_{in})$, is denoted by g_i, and the *degree* of g_i is defined as the maximum degree of its components. In a similar way, the degree of any n-tuple of polynomials is defined as the maximum degree of any component. The *internal degree* and *external degree* of $G(D)$ are defined as follows:

$$\text{int.deg.}\, G(D) = \text{maximum degree of } G(D)\text{'s } k \times k \text{ minors}$$

$$\text{ext.deg.}\, G(D) = \text{sum of the row degrees of } G(D)$$

The following two definitions will be essential in the discussion of convolutional codes.

A. Definition 1

A $k \times n$ polynomial matrix $G(D)$ is called *basic* if, among all polynomial matrices of the form $T(D)G(D)$, where $T(D)$ is a nonsingular $k \times k$ matrix over $F(D)$, it has the minimum possible internal degree.

B. Definition 2

A $k \times n$ polynomial matrix $G(D)$ is called *reduced* if, among all matrices of the form $T(D)G(D)$, where $T(D)$ is unimodular,[3] $G(D)$ has the minimum possible external degree. Since any unimodular matrix is a product of elementary matrices, an equivalent definition is that a matrix is reduced if its external degree cannot be reduced by a sequence of elementary row operations.

Before continuing along the main line, it is helpful to present a simple theorem that provides several useful facts about the internal and external degrees of a polynomial matrix.

Theorem 1. Let $G(D)$ be a $k \times n$ polynomial matrix.

(1) If $T(D)$ is any nonsingular $k \times k$ polynomial matrix, then $\text{int.deg.}\, T(D)G(D) = \text{int.deg.}\, G(D) + \text{deg.det.}T(D)$. In particular $\text{int.deg.}\, T(D)G(D) \geq \text{int.deg.}\, G(D)$, with equality if and only if $T(D)$ is unimodular.

(2) $\text{int.deg.}\, G(D) \leq \text{ext.deg.}\, G(D)$.

Proof:

(1) The $k \times k$ submatrices of $T(D)G(D)$ are just the $k \times k$ submatrices of $G(D)$, each multiplied by $T(D)$. Thus the $k \times k$ minors of $T(D)$ are just the $k \times k$ minors of $G(D)$, each multiplied by $\det T(D)$. The result now follows.

(2) Denote the degree of the ith row of $G(D)$ by e_i. In the expansion of any $k \times k$ minor of $G(D)$, each term is the product of k entries of $G(D)$, one from each row (and column). Since each entry from the ith row has degree $\leq e_i$, it follows that the degree of any $k \times k$ minor is at most $e_1 + \cdots + e_k = \text{ext.deg.}\, G(D)$. ❐

Basic and reduced polynomial matrices enjoy many useful and surprising properties. The Appendix gives two theorems, Theorem A-1 and Theorem A-2, delineating these properties. These Theorems will be referenced constantly in the rest of this article.

Example 1. Here are eight generator matrices for a $(4, 2)$ convolutional code over $GF(2)$. Of these eight, six, G_2 through G_7, are PGMs.

$$G_1 = \begin{pmatrix} \frac{1}{1+D+D^2} & 1 & \frac{1+D^2}{1+D+D^2} & \frac{1+D}{1+D+D^2} \\ 1 & \frac{1+D+D^2}{D} & D & \frac{1}{D} \end{pmatrix}$$

$$G_2 = \begin{pmatrix} 1 & 1+D+D^2 & 1+D^2 & 1+D \\ D & 1+D+D^2 & D^2 & 1 \end{pmatrix}$$

[3] A unimodular matrix is a square polynomial matrix whose determinant is a nonzero scalar.

$$G_3 = \begin{pmatrix} 1 & 1+D+D^2 & 1+D^2 & 1+D \\ 0 & 1+D & D & 1 \end{pmatrix}$$

$$G_4 = \begin{pmatrix} 1 & D & 1+D & 0 \\ 0 & 1+D & D & 1 \end{pmatrix}$$

$$G_5 = \begin{pmatrix} 1+D & 0 & 1 & D \\ D & 1+D+D^2 & D^2 & 1 \end{pmatrix}$$

$$G_6 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1+D & D & 1 \end{pmatrix}$$

$$G_7 = \begin{pmatrix} 1+D & 0 & 1 & D \\ 1 & D & 1+D & 0 \end{pmatrix}$$

$$G_8 = \begin{pmatrix} 1 & 0 & \frac{1}{1+D} & \frac{D}{1+D} \\ 0 & 1 & \frac{D}{1+D} & \frac{1}{1+D} \end{pmatrix}$$

Table 1 lists the properties of the generator matrices $G_1, \ldots, G_8$. (Note that only the polynomial generator matrices, viz., G_2–G_7, have external or internal degrees, or can be basic or reduced.) These properties are easily verified by referring to Theorems A-1 and A-2. For example, G_3 is *basic* because the gcd of its 2×2 minors is 1 (Theorem A-1, condition (2)), and it is *not reduced* because its internal and external degrees are unequal (Theorem A-2, condition (2)). ❐

It follows from Definition 1 that among all PGMs for a given convolutional code, those for which the *internal* degree is as small as possible are exactly the *basic* PGMs. It turns out, however, that the set of PGMs for which the *external* degree is as small as possible form a much more interesting class, the class of *minimal* PGMs.

C. Definition 3

Among all PGMs for a given convolutional code C, those for which the external degree is as small as possible are called *minimal* PGMs. This minimal external degree is called the *degree* of the code C, and is denoted $\deg C$.

It will be shown that minimal generator matrices have many remarkable properties. The key to these properties is the following theorem:

Theorem 2. A PGM $G(D)$ for the convolutional code C is minimal if and only if it is both basic and reduced.

Proof: First it will be shown that a minimal PGM must be both basic and reduced. Then it will be shown that a PGM that is both basic and reduced must be minimal.

To prove the first assertion, denote by m_0 the common internal degree of all the basic PGMs for C, and among all the basic PGMs choose one, say $G_0(D)$, for which the *external* degree is as small as possible. Then G_0 must be reduced, since if $T(D)$ is unimodular, $\text{int.deg.}\, T(D)G_0(D) = \text{int.deg.}\, G_0(D) = m_0$ by Theorem 1, and so by the definition of G_0, $\text{ext.deg.}\, T(D)G_0(D) \geq \text{ext.deg.}\, G_0(D)$. Now let $G(D)$ be any minimal PGM for C. Then

$$\text{int.deg.}\, G_0 \leq \text{int.deg.}\, G \leq \text{ext.deg.}\, G \leq \text{ext.deg.}\, G_0 \quad (1)$$

(The first inequality in Definition 1 follows from the fact that G_0 is chosen to have minimum possible internal degree. The second inequality follows from Theorem 1(2). The third inequality is because G, as a minimal PGM, has minimum possible exernal degree.) But since $G_0(D)$ is reduced, by Theorem A-2, condition (2), $\text{int.deg.}\, G_0 = \text{ext.deg.}\, G_0$, so that equality holds throughout Definition 1. Thus $\text{int.deg.}\, G = \text{int.deg.}\, G_0 = m_0$, so G is basic; and $\text{int.deg.}\, G = \text{ext.deg.}\, G$, so that G is reduced, by Theorem A-2, condition (2).

Conversely, suppose that $G(D)$ is basic and reduced, and $G_0(D)$ is any other PGM for C. Then by Theorem 1(2), $\text{ext.deg.}\, G_0(D) \geq \text{int.deg.}\, G_0(D)$. Since $G(D)$ is basic, $\text{int.deg.}\, G_0(D) \geq \text{int.deg.}\, G(D)$; since $G(D)$ is reduced, by Theorem 1 $\text{int.deg.}\, G(D) = \text{ext.deg.}\, G(D)$. Combining these inequalities, $\text{ext.deg.}\, G_0(D) \geq \text{ext.deg.}\, G(D)$, which proves that $G(D)$ is minimal. ❐

In the proof of Theorem 2, m_0, the common internal degree for all PGMs for C, is equal to $\deg C$, i.e., the minimum possible external degree. Thus there are two corollaries to Theorem 2.

Corollary 1. The minimal internal degree of any PGM for a given convolutional code C is equal to the degree of C.

Corollary 2. If G is any basic generator matrix for C, then $\text{int.deg.}\, G = \deg C$.

The following theorem shows that minimal generator matrices are "minimal" in a very strong sense.

Theorem 3. If $e_1 \leq e_2 \leq \cdots \leq e_k$ are the row degrees of a minimal generator matrix for a convolutional code C, and if $f_1 \leq f_2 \leq \cdots \leq f_k$ are the row degrees of any other polynomial generator matrix, say G', for C, then $e_i \leq f_i$, for $i = 1, \ldots, k$.

Proof: If the statement is false, there exists an index j such that $e_1 \leq f_1, \ldots, e_j \leq f_j$, but $e_{j+1} > f_{j+1}$. It then follows from the minimality of G (use the properties in Theorem A-1, condition (5), and Theorem A-2, condition (3)) that the first $j+1$ rows of G' must be polynomial linear combinations of the first j rows of G, which contradicts the fact that the rows of G' are linearly independent. □

Theorem 4. The set of row degrees is the same for all minimal PGMs for a given code.

Proof: This result follows immediately from Theorem 3. □

The row degrees referred to in Theorems 3 and 4, say $(e_1, e_2, \ldots, e_k)$, are called the *Forney indices* of the code. The sum $e_1 + \cdots + e_k$ of the Forney indices, i.e., the minimum possible external degree of any PGM for C, is the degree of the code. The maximum of the Forney indices is called the *memory* of the code. From now on, reserve the letter m to denote the degree of a given convolutional code, and refer to an (n, k) code with degree m as an (n, k, m) code. An (n, k, m) code is called *optimal* if it has the maximum possible free distance among all codes with the same value of n, k, and m.

Example 2: Continuing the study of the $(4, 2)$ code from Example 1, of the eight given generator matrices, only G_6 is minimal (it satisfies condition (2) of Theorem A-1 and condition (2) of Theorem A-2, so it is both basic and reduced), so that the *degree* of the code C is 1, and the *Forney indices* are $(0, 1)$. The code is thus a $(4, 2, 1)$ code. From Example 3, below, it is in fact an *optimal* $(4, 2, 1)$ code. □

III. The Hilbert Series and Free Distance Bounds

If C is a fixed (n, k) convolutional code, a *polynomial codeword* of C is a codeword all of whose components are polynomials. Recalling that the degree of a polynomial vector is defined to be the maximum degree of any component, for any integer $L \geq 0$, C_L is defined as the set of *polynomial codewords of degree* $\leq L$. C_L is a vector space over F. Indeed, it is a subspace of the set of all possible n-dimensional polynomial vectors of degree $\leq L$ over F. The F-dimension of C_L is denoted by δ_L. The following theorem shows that the δ_L's can be computed from the Forney indices.

Theorem 5. If C is an (n, k) convolutional code with Forney indices $(e_1, \ldots, e_k)$ and polynomial subcode dimensions δ_L, then

$$\sum_{L \geq 0} \delta_L t^L = \frac{\sum_{i=1}^{k} t^{e_i}}{(1-t)^2}$$

Note: The power series appearing in Theorem 5 is called the Hilbert series *for the code.*

Proof: Let $G(D)$ be a minimal PGM for C, whose row degrees are the ordered Forney indices, say $e_1 \leq \cdots \leq e_k$, and let $g_1, \ldots, g_k$ be the rows of G. Let $y(D)$ be any polynomial codeword of degree $\leq L$. Then it follows from Theorem A-1, condition (5), that $y(D) = x(D)G(D)$, where $x(D) = (x_1(D), \ldots, x_k(D))$ is a k-vector of polynomials, and it follows from the predictable degree property (Theorem A-2, condition (3)) that $\deg x_i + e_i \leq L$. Thus a basis for the F-space C_L is the set $\{D^j g_i(D) : j + e_i \leq L\}$. Hence

$$\begin{aligned}
\sum_{L \geq 0} \delta_L t^L &= \sum_{i=1}^{k} \sum_{j \geq 0} \left(t^{e_i + j} + t^{e_i + j + 1} + \cdots\right) \\
&= \sum_{i=1}^{k} \sum_{j \geq 0} \frac{t^{e_i + j}}{1 - t} \\
&= \sum_{i=1}^{k} \frac{t^{e_i}}{(1-t)^2}
\end{aligned}$$

□

Corollary 3. The following explicit formula for δ_L holds:

$$\delta_L = \sum_{i=1}^{k} \max(L + 1 - e_i, 0) \tag{3}$$

Proof: By elementary calculus, $(1-t)^{-2} = \sum_{j \geq 0} (j+1) t^j$. Applying this fact to Eq. (2),

$$\begin{aligned}
\sum_{L \geq 0} \delta_L t^L &= \sum_{i=1}^{k} \frac{t^{e_i}}{(1-t)^2} \\
&= \sum_{i=1}^{k} \sum_{j \geq 0} (j+1) t^{e_i + j} \\
&= \sum_{i=1}^{k} \sum_{j \geq e_i} (j + 1 - e_i) t^j
\end{aligned}$$

Thus the coefficient of t^L in the Hilbert series $\sum_{L\geq 0} \delta_L t^L$ is $\sum_{i=1}^{k} \max(L+1-e_i, 0)$, which is the desired proof. ❐

Corollary 4. Let C be an (n,k,m) convolutional code. Then for all $L \geq 0$

$$\delta_L \geq \max((L+1)k - m, 0) \tag{4}$$

Furthermore, there is equality for all $L \geq 0$ in Eq. (4) if and only if the Forney indices assume only the two values $\lfloor m/k \rfloor$ and $\lceil m/k \rceil$. A code for which this is true will be called a *compact* code.

Proof: Since $\max(x,0) \geq x$, Eq. (3) implies that $\delta_L \geq \sum_{i=1}^{k}(L+1-e_i) = (L+1)k - \sum_{i=1}^{k} e_i = (L+1)k - m$. Since $\delta_L \geq 0$, too, it follows that $\delta_L \geq \max((L+1)k-m, 0)$ for all $L \geq 0$. Assuming that the Forney indices are ordered so that $e_1 \leq \cdots \leq e_k$, it follows from Eq. (3) that $\delta_L = (L+1)k - m = \max((L+1)k-m, 0)$ if $L+1 \geq e_k$, and $\delta_L = 0 = \max((L+1)k-m, 0)$ if $L+1 \leq e_1$. If $e_k - e_1 \leq 1$, one of these alternatives must hold for all L. On the other hand, if $e_k - e_1 \geq 2$, then there is at least one value of L for which $e_1 < L+1 < e_k$, in which case $0 < \delta_l < (L+1)k - m$, so that $\delta_L \neq \max((L+1)k-m, 0)$. ❐

Incidentally, it is easy to show that in a compact code, there are exactly (m mod k) Forney indices equal to $\lceil m/k \rceil$, and $k - (m \bmod k)$ Forney indices equal to $\lfloor m/k \rfloor$. Thus for example if $k = 4$ and $m = 13$, a compact code will have Forney indices $(3,3,3,4)$.

Since the subcode C_L forms an $(n(L+1), \delta_L)$ linear block code over F, the *free distance* of C cannot exceed the *minimum distance* of C_L, for $L = 0, 1, \ldots$, which leads to Theorem 6.

Theorem 6. If C is an (n,k) convolutional code with Forney indices $(e_1, \ldots, e_k)$, then

$$d_{\text{free}}(C) \leq \min_{L\geq 0} \Delta_F(n(L+1), \delta_L)$$

where $\Delta_F(n,k)$ denotes the maximum possible minimum distance of an (n,k) linear block code over F.

Note from Corollary 4 that δ_L is minimized for all L, and so $\Delta_F(n(L+1), \delta_L)$ is maximized for all L, for a compact code. This suggests, but does not prove, that among all (n,k,m) codes, the compact codes will have the largest free distances. In any case, for applications a bound on d_{free} is needed that applies to all (n,k,m) codes, regardless of their Forney indices. Thus is offered the following corollary to Theorem 6, which gives an upper bound on d_{free} for all (n,k,m) codes, regardless of the Forney indices. It is good to bear in mind, however, that it may be possible to improve the bound if the code is noncompact.

Corollary 5. If C is an (n,k,m) code, then

$$d_{\text{free}}(C) \leq \min_{L\geq 0} \Delta_F(n(L+1), k(L+1) - m)$$

Proof: Combine Theorem 6 with Corollary 4. ❐

The upper bound on d_{free} of Theorem 6 is attained for many, but not all, values of n, k, and m. The following examples will illustrate this. (All examples in this article are over the field $GF(2)$.)

Example 3. Continuing the study of the $(4,2,1)$ code in Examples 1 and 2, since the Forney indices are $(0,1)$, by Theorem 5 the "Hilbert Series" for the code is

$$\frac{1+t}{(1-t)^2} = 1 + 3t + 5t^2 + \cdots + (2L+1)t^L + \cdots$$

Thus the dimension of the Lth subcode C_L is $2L+1$. It then follows from Theorem 6 that the free distance of the code satisfies

$$d_{\text{free}}(C) \leq \min_{L\geq 0} \Delta_{GF(2)}(4(L+1), 2L+1)$$

In particular, for $L = 0$ the above bound gives $d_{\text{free}}(C) \leq \Delta_2(4,1) = 4$. But in fact $d_{\text{free}} = 4$ for this code (use the generator matrix G_4 to check this fact), so this particular code has the largest possible free distance for a $(4,2,1)$ convolutional code, i.e., it is an *optimal* $(4,2,1)$ code. ❐

Example 4. Consider $(n,k,m) = (2,1,2)$ codes. Since $k = 1$, there is only one Forney index, and so any $(2,1,2)$ code is compact. By Corollary 3, $\delta_0 = 0$, and, for $L \geq 1$, $d_L = L - 1$, so that by Theorem 6 (here and hereafter Verhoeff's tables [7] of the values of $\Delta_F(n,k)$ are used when $F = GF(2)$),

$$\begin{aligned} d_{\text{free}} &\leq \min(\Delta(2,0), \Delta(4,0), \Delta(6,1), \Delta(8,2), \Delta(10,3)\ldots) \\ &= \min(\infty, \infty, 6, 5, 5, \ldots) \\ &= 5 \end{aligned}$$

In fact, there is a well-known $(2,1,2)$ code with $d_{\text{free}} = 5$, whose (unique) minimal generator matrix is

$$G(D) = (\,1 + D^2 \quad 1 + D + D^2\,) \tag{5}$$

(See [6], Chapter 9.) It follows then that the $(2,1,2)$ code defined by Theorem 6 is optimal. ❐

Example 5: Consider binary $(4,3,2)$ codes. The Forney indices of such codes must be either $(0,1,1)$ (compact) or $(0,0,2)$ (noncompact). In the first case, by Corollary 3, $\delta_0 = 1$, $\delta_1 = 4$, $\delta_2 = 7$, etc., and so by Theorem 6,

$$\begin{aligned} d_{\text{free}} &\le \min(\Delta(4,1), \Delta(8,4), \Delta(12,7), \Delta(16,10), \ldots) \\ &= \min(4,4,4,4,\ldots) \\ &= 4 \end{aligned}$$

However, it turns out that there is no $(4,3,2)$ code with $d_{\text{free}} = 4$ [9]. The largest possible d_{free} turns out to be $d_{\text{free}} = 3$, which is achieved by the second-order Wyner-Ash code [8]. A minimal PGM for such a code is

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & D \\ D & 0 & 1 & 1 \end{pmatrix}$$

On the other hand, if the Forney indices are $(0,0,2)$, then Corollary 3 tells one that $\delta_0 = 2$, $\delta_1 = 4$, $\delta_2 = 7$, etc., and so by Theorem 6,

$$\begin{aligned} d_{\text{free}} &\le \min(\Delta(4,2), \Delta(8,4), \Delta(12,7), \Delta(16,10), \ldots) \\ &= \min(2,4,4,4,\ldots) \\ &= 2 \end{aligned}$$

And indeed there is a $(4,3,2)$ code with Forney indices $(0,0,2)$ and $d_{\text{free}} = 2$. A minimal PGM for one such code is

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 + D^2 \end{pmatrix}$$

Therefore, among $(4,3,2)$ codes, only the compact ones can be optimal. ❐

References

[1] G. D. Forney, Jr., "Convolutional Codes I: Algebraic Structure," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 720–738, November 1970.

[2] G. D. Forney, Jr., "Structural Analysis of Convolutional Codes via Dual Codes," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 512–518, July 1973.

[3] G. D. Forney, Jr., "Minimal Bases of Rational Vector Spaces with Applications to Multivariable Linear Systems," *SIAM J. Control*, vol. 13, pp. 493–502, May 1975.

[4] G. D. Forney, Jr., "Algebraic Structure of Convolutional Codes, and Algebraic System Theory," in *Mathematical System Theory*, edited by A. C. Antoulas, Berlin, Germany: Springer-Verlag, pp. 527–557, 1991.

[5] T. Kailath, *Linear Systems*. Englewood Cliffs, New Jersey: Prentice Hall, 1980.

[6] R. McEliece, *The Theory of Information and Coding*, Reading, Massachusetts: Addison-Wesley, 1977.

[7] T. Verhoeff, "An Updated Table of Minimum-Distance Bounds for Binary Linear Codes," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 665–680, September 1987.

[8] A. D. Wyner and R. B. Ash, "Analysis of Recurrent Codes," *IEEE Trans. Inform. Theory*, vol. IT-9, pp. 143–156, July, 1963.

[9] Ø. Ytrehus, "A Note on High Rate Binary Convolutional Codes," in *Report in Informatics No. 68*, Bergen, Norway: University of Bergen, August 1992.

Appendix

Basic and Reduced Matrices

This appendix—a reference collection of many useful properties of basic and reduced matrices—begins with the basic matrices.

Theorem A-1. A $k \times n$ polynomial matrix $G(D)$ is basic (see Definition 1) if and only if any one of the following six conditions is satisfied:

(1) The invariant factors of $G(D)$ are all 1.

(2) The gcd of the $k \times k$ minors of $G(D)$ is 1.

(3) $G(\alpha)$ has rank k for any α in the algebraic closure of F.

(4) $G(D)$ has a right $F[D]$ inverse, i.e., there exists an $n \times k$ polynomial matrix $H(D)$ such that $G(D)H(D) = I_k$.

(5) If $y(D) = x(D)G(D)$, and if $y(D) \in F[D]^n$, then $x(D) \in F[D]^k$. ("Polynomial output implies polynomial input.")

(6) $G(D)$ is a submatrix of a unimodular matrix, i.e., there exists an $(n-k) \times n$ matrix $L(D)$ such that the $n \times n$ matrix $\begin{pmatrix} G(D) \\ L(D) \end{pmatrix}$ has determinant 1.

Proof: The proof is logically rather involved. The following implications will be proved: (Basic) $\rightarrow$ (1) $\rightarrow$ (2) $\rightarrow$ (4) $\rightarrow$ (5) $\rightarrow$ (Basic); (2) $\leftrightarrow$ (3); (1) $\leftrightarrow$ (6).

• (Basic) $\rightarrow$ (1): Suppose that Γ is the $k \times n$ invariant-factor form for G, i.e., $\Gamma = \mathrm{diag}(\gamma_1, \gamma_2, \ldots, \gamma_k)$ where $\gamma_i = \Delta_i/\Delta_{i-1}$, Δ_i being the gcd of the $i \times i$ minors of G. (Take $\Delta_0 = 1$ by convention.) Then there is a $k \times k$ unimodular matrix X and an $n \times n$ unimodular matrix Y such that

$$XGY = \Gamma \qquad \text{(A-1)}$$

(For a proof of this "invariant factor decomposition," see [5], Theorem 6.3.16.) Thus if Γ_k denotes the $k \times k$ matrix formed by the leftmost k columns of Γ, the matrix $G' = \Gamma_k^{-1} XG$ is a polynomial matrix equivalent to G. Furthermore, since $\det(\Gamma_k^{-1}X) = \det \Gamma_k^{-1} = (\det \Gamma_k)^{-1} = (\gamma_1 \cdots \gamma_k)^{-1}$, unless the γ_i's are all 1, the internal degree of G' is strictly less than that of Γ. Thus if the invariant factors of G are not all 1, then G is not basic, which proves (Basic) $\rightarrow$ (1).

• (1) $\leftrightarrow$ (2): According to the definitions given in the previous paragraph, the product of the invariant factors of G is

$$\gamma_1\gamma_2 \cdots \gamma_k = \frac{\Delta_1}{\Delta_0} \cdot \frac{\Delta_2}{\Delta_1} \cdots \frac{\Delta_k}{\Delta_{k-1}}$$

$$= \frac{\Delta_k}{\Delta_0}$$

$$= \Delta_k = (\text{the gcd of the } k \times k \text{ minors of } G)$$

Hence the g_i's are all 1 if and only if $\Delta_k = 1$.

• (2) $\rightarrow$ (4): Suppose that the gcd of the $k \times k$ minors of $G(D)$ is 1, and denote the individual minors by $\Delta_\nu(D)$, for $\nu = 1, 2, \ldots, \binom{n}{k}$. Then by Cramer's rule for each ν, there will exist a "pseudo-inverse" for $G(D)$, with factor $\Delta_\nu(D)$, i.e., an $n \times k$ matrix $H_\nu(D)$ such that $G(D)H_\nu(D) = \Delta_\nu(D)I_k$. Since the gcd of the $\Delta_\nu(D)$'s is 1, there exists a polynomial linear combination of the $\Delta_\nu(D)$'s equal to 1, say $\sum_\nu \lambda_\nu(D)\Delta_\nu(D) = 1$. It follows that $H(D) = \sum_\nu \lambda_\nu(D)H_\nu(D)$ is an $n \times k$ polynomial inverse for $G(D)$.

• (4) $\rightarrow$ (5): Suppose that $G(D)$ has an $n \times k$ polynomial inverse $H(D)$, and that $x(D) = (x_1(D), \ldots, x_k(D))$ is a k-vector of rational functions such that $y(D) = x(D)G(D)$ is an n-vector of polynomials. Multiplying this equation on the right by $H(D)$, $y(D)H(D) = x(D)$, which implies that $x(D)$ is in fact a polynomial vector.

• (5) $\rightarrow$ (Basic): Suppose that property (5) holds, and let $T(D)$ be an arbitrary nonsingular $k \times k$ matrix of rational functions such that $G' = TG$ is a polynomial matrix. Then by property (5), T must in fact be a polynomial matrix, so that by Theorem 1(1), int.deg. $G' \geq$ int.deg. G, which means that G is basic.

• (2) $\leftrightarrow$ (3): Suppose that the gcd of the $k \times k$ minors of $G(D)$ is 1, let α be an arbitrary element of the algebraic closure of F, and let $p(D)$ be the minimal polynomial of α. Then there must be at least one $k \times k$ subdeterminant of $G(D)$ which is not divisible by $p(D)$, which means that the corresponding $k \times k$ submatrix of $G(\alpha)$ is nonsingular. Thus $G(\alpha)$ must have rank k. Conversely, suppose that the gcd of the $k \times k$ minors of G is not 1, which means

that it is divisible by some irreducible polynomial $p(D)$. If α is a root of $p(D)$ in some extension field of F, it follows that every $k \times k$ minor of $G(\alpha)$ is zero, which in turn means that $G(\alpha)$ has rank less than k.

• (1) $\leftrightarrow$ (6): Suppose the invariant factors of G are all 1. Then the invariant-factor decomposition in Eq. (A-1) can be written as

$$G = A \begin{pmatrix} I_k & 0_{k,n-k} \end{pmatrix} B$$

where $A = X^{-1}$ and $B = Y^{-1}$. Thus if $B = \begin{pmatrix} B_U \\ B_L \end{pmatrix}$ where B_U is $k \times n$ and B_L is $(n-k) \times n$, it follows that $G = AB_U$. But the matrix $\begin{pmatrix} AB_U \\ B_L \end{pmatrix}$ is unimodular, since it is obtained from the unimodular matrix B via a sequence of elementary row operations on the first k rows. Conversely, if $B = \begin{pmatrix} G(D) \\ H(D) \end{pmatrix}$ is unimodular, then the equation $G(D) = I_k \begin{pmatrix} \Gamma_k & 0_{k,n-k} \end{pmatrix} B$ shows that the invariant factors of $G(D)$ are all 1. ❐

Theorem A-2. A $k \times n$ polynomial is reduced (see Definition 2) if and only if one of the following three conditions is satisfied:

(1) If the "matrix of high-order coefficients" $\overline{G}$ is defined by

$$\overline{G}_{ij} = \operatorname*{coeff}_{D^{e_i}} g_{ij}(D)$$

where e_i is the degree of $G(D)$'s ith row, then $\overline{G}$ has rank k.

(2) ext.deg. $G(D)$ = int.deg. $G(D)$.

(3) The "predictable degree property": For any k-dimensional polynomial vector, i.e., any $x(D) \in F[D]^k$

$$\deg(x(D)G(D)) = \max_{1 \le i \le k} (\deg x_i(D) + \deg g_i(D))$$

Proof: The logical organization of this proof is as follows: It shall be proved that (Reduced) $\rightarrow$ (1) $\rightarrow$ (2) $\rightarrow$ (Reduced), and (1) $\leftrightarrow$ (3).

• (Reduced) $\rightarrow$ (1): Suppose property (1) is false. Then there is a nonzero k-dimensional vector from F, say $\alpha = (\alpha_1, \ldots, \alpha_k)$, such that $\alpha\overline{G} = 0$. Now suppose that the rows of G are $(g_1, \ldots, g_k)$ with $\deg g_i = e_i$, and $e_1 \le e_2 \le \cdots \le e_k$. Then from $\alpha\overline{G} = 0$, it follows that the coefficient of D^{e_k} in the linear combination

$$g'_k = \alpha_1 D^{e_k - e_1} g_1 + \alpha_2 D^{e_k - e_2} g_2 + \cdots + \alpha_k D^{e_k - e_k} g_k$$

is zero, so that the unimodular transformation of $G(D)$ that replaces g_k with g'_k—and leaves the remaining rows of G unchanged—reduces the external degree of G. In other words, if property (1) is false, G is not reduced, which is (the contrapositive of) what the authors have set out to prove.

• (1) $\rightarrow$ (2): Suppose that $\overline{G}$ has rank k, and denote the $k \times k$ submatrices of $\overline{G}$ by $\overline{G}_\nu$, for $\nu = 1, 2, \ldots, \binom{n}{k}$. Then since rank $\overline{G} = k$, there is at least one index ν_0 such that $\det \overline{G}_{\nu_0} \neq 0$. If now the row degrees of G are $e_1, \ldots, e_k$, then (cf. the proof of Theorem 1(2)) the coefficient of $D^{e_1 + \cdots + e_k}$ in $\det G_{\nu_0}$ is $\det \overline{G}_{\nu_0} \neq 0$. Thus int.deg. $G \ge$ ext.deg. G. The opposite inequality is true for any matrix, as was shown in Theorem 1(2).

• (2) $\rightarrow$ (Reduced): Suppose int.deg. $G(D)$ = ext.deg. $G(D)$, and that $T(D)$ is an arbitrary $k \times k$ unimodular matrix. Then ext.deg. $TG \ge$ int.deg. TG by Theorem 1(2); int.deg. TG = int.deg. G, by Theorem 1(2); and int.deg. G = ext.deg. T, by assumption. Combining this string of inequalities and equalities, ext.deg. $TG \ge$ ext.deg. G, which proves that G is reduced.

• (1) $\leftrightarrow$ (3): Let $x(D) = (x_1(D), \ldots, x_k(D))$ be a k-vector of polynomials, and let $y(D) = (y_1(D), \ldots, y_n(D))$ be defined by the equation $y(D) = x(D)G(D)$. If the k rows of $G(D)$ are denoted by $g_1(D), \ldots, g_k(D)$, then

$$\begin{aligned} y(D) &= x(D)G(D) \\ &= x_1(D)g_1(D) + \cdots + x_k(D)g_k(D) \end{aligned} \tag{A-2}$$

If the degree of $x_i(D)$ is d_i for $i = 1, \ldots, k$, and the degree of $g_i(D)$ is e_i for $i = 1, \ldots, k$, it follows from Eq. (A-2) that the degree of $y(D)$ is at most $d = \max_i(d_i + e_i)$. Call d the "prediction" of the degree of $y(D)$. To test the prediction, note that the vector of coefficients of D^d in $y(D)$ is $\alpha = (\alpha_1, \ldots, \alpha_k)\overline{G}$, where α_i is the coefficient of D^{d-e_i} in $x_i(D)$. (At least one of the α_i is nonzero, since $d_i + e_i = d$ must hold for at least one index i.) But $\alpha\overline{G} \neq 0$ for all nonzero α's if and only if $\overline{G}$ has rank k, and so the prediction is true for all $x(D)$'s if and only if rank $\overline{G} = k$. ❐

Table 1. Generator-matrix properties.

Property	Basic?	Reduced?	Int.deg.	Ext.deg.
G_1	–	–	–	–
G_2	No	No	3	4
G_3	Yes	No	1	3
G_4	Yes	No	1	2
G_5	No	Yes	3	3
G_6	Yes	Yes	1	1
G_7	No	Yes	2	2
G_8	–	–	–	–