Codes, matroids and trellises

Peter J Cameron
(with many contributions from C. Papadopoulos,
R. A. Bailey and C. G. Rutherford)
School of Mathematical Sciences
Queen Mary and Westfield College
London E1 4NS
`p.j.cameron@qmw.ac.uk`

Combinatorics 2000, Gaeta

Who discovered the Hamming codes?

Was it

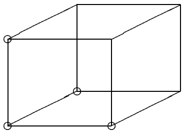
- R. W. Hamming?
- M. J. E. Golay?
- R. A. Fisher?
- J. J. Sylvester?

See “Hamming and Golay, Fisher and Bose” on this Web page for more about this.

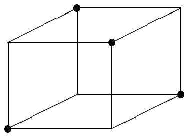
Factorial design

You are investigating a process whose yield is affected by a number of factors, each of which can occur at several levels. Do you

(a) change one factor at a time?



(b) Use a design?



Connections

Linear codes and *factorial designs* are almost the same concept, even though their theories have developed quite separately.

Similarly, *representations of matroids* and *point sets in projective spaces* are almost the same concept.

The theme of these lecture is that in fact *the two concepts just descried are almost the same*.

Codes

A *linear code* C of length n and dimension k over a field F is a k-dimensional subspace of F^n. The *weight* $\mathrm{wt}(v)$ of a word $v \in F^n$ is the number of non-zero coordinates, and the *minimum weight* of C is the smallest weight of a non-zero vector in C.

Codes C and C' are *monomial equivalent* if C' is obtained from C by permuting the coordinates and multiplying them by non-zero scalars.

Theorem 1 *A code with minimum weight d can correct up to $\lfloor (d-1)/2 \rfloor$ errors.*

Matroids

A *matroid* M on a set E is a family I of subsets of E called *independent sets*, closed under taking subsets and satisfying the exchange property.

The *rank* $\rho(A)$ of a subset A of E is the size of the largest independent subset of A. An independent subset of E of size $\rho(E)$ is called a *basis* of M.

Example. The *uniform matroid* $U(k,n)$: the independent sets are all subsets having cardinality at most k.

A *representation* of E over a field F is a map of E into an F-vector space which preserves independence. Two representations are *equivalent* if they are related by an invertible linear transformation between the vector spaces.

Exchange axiom

The *exchange axiom* states: If A and B are independent sets such that $|B| > |A|$, then there exists $b \in B \setminus A$ such that $A \cup \{b\}$ is independent.

This guarantees that all bases have the same cardinality, and so makes the definition of rank sensible.

The code-matroid connection

Let A be a $k \times n$ matrix over a field F having rank k. From A we construct

- a code $C(A)$ generated by the rows of A;
- a matroid $M(A)$ represented in F^k by the columns of A.

The equivalence relation on such matrices given by arbitary row operations and monomial column operations mirrors the natural notions of equivalence for linear codes and representations of matroids.

Note to geometers: Linear *MDS codes* in projective space over $\mathrm{GF}(q)$ correspond to representations of the uniform matroid $U(n,k)$ over $\mathrm{GF}(q)$.

First and last base

Let the ground set E of the matroid M be totally ordered, and let $\rho(E) = k$. Let $\mathcal{B}$ be the set of bases of M. When we write a base as $\{b_1, \dots, b_k\}$, we assume that $b_1 < \cdots < b_k$.

The (lexicographically) *first base* $F = \{f_1, \dots, f_k\}$ satisfies $f_i \leq b_i$ for any base $B = \{b_1, \dots, b_k\}$.

Dually the *last base* $L = \{l_1, \dots, l_k\}$ satisfies $b_i \leq l_i$ for any base $B = \{b_1, \dots, b_k\}$.

These properties express the relationship of matroids to the *greedy algorithm*.

Weight enumerator and Tutte polynomial

The *weight enumerator* of a code C of length n is given by

$$W_C(x,y) = \sum_{c \in C} x^{n-\mathrm{wt}(c)} y^{\mathrm{wt}(c)}.$$

The *Tutte polynomial* of a matroid M on E with rank function ρ is given by

$$T(M;x,y) = \sum_{A \subseteq E} (x-1)^{\rho E - \rho A} (y-1)^{|A| - \rho A}.$$

Internal and external activity

There is an equivalent definition as follows. Suppose that M is a matroid on the set E, which is totally ordered. Let B be a base of M. An element $b \in B$ is *internally active with respect to* B if, for all $c \in B$, we have $B \cup \{c\} \setminus \{b\} \in \mathcal{B} \Rightarrow c < b$. The *internal activity* of a base is the number of internally active elements associated with it.

Dually, an element $e \notin B$ is *externally active with respect to* B if, for all $f \in B$, we have $f \in C(e,B) \Rightarrow f < e$.The *external activity* of a base is the number of externally active elements associated with it.

Then we have

$$T(M;x,y) = \sum_{B \in \mathcal{B}} t_{i,j} x^i y^i$$

where $t_{i,j}$ is the number of bases with i internally active elements and j externally active elements.

First and last; internal and external

A *loop* in a matroid is an element $e \in E$ which is contained in no basis.

A *coloop* is an element $e \in E$ which is contained in every basis.

Note that

(a) The internal activity of the first base is the number of coloops of M, while its external activity is equal to $|E| - \rho(E)$.

(b) The internal activity of the last base is $\rho(E)$, while its external activity is equal to the number of loops of M.

Greene’s Theorem

Curtis Greene showed in 1975 that the weight enumerator of $C = C(A)$ is a specialisation of the Tutte polynomial of $M = M(A)$:

Theorem 2

$$W_C(x,y) = y^{n-\dim(C)}(x-y)^{\dim(C)}T\left(M; \frac{x+(q-1)y}{x-y}, \frac{x}{y}\right).$$

In particular, the Tutte polynomial of $M(A)$ determines the minimum weight of $C(A)$.

Duality

The *dual* of a matroid M on E is the matroid M^* on E whose bases are the complements of the bases of M.

The *dual* of a code C is the code

$$C^\perp = \{v \in F^n : v \cdot c = 0 \text{ for all } c \in C\},$$

where $\cdot$ is the usual dot product.

Under the code–matroid connection, dual codes correspond to dual matroids. Also, it is trivial that

$$T(M^*; x, y) = T(M; y, x),$$

from which we obtain the *MacWilliams relation*

$$W_{C^\perp}(x,y) = \frac{1}{|C|} W_C(x+(q-1)y, x-y).$$

An example

Suppose that we are using the binary dual Hamming code of length 7 to send information. The codewords are:

0000000
0011011
0101101
0110110
1001110
1010101
1100011
1111000

The minimum weight is 4, so we can correct one error and detect two errors.

Analog errors

In practice, the received word is an analog signal, sampled at seven time points, i.e. seven real numbers. Suppose that we receive

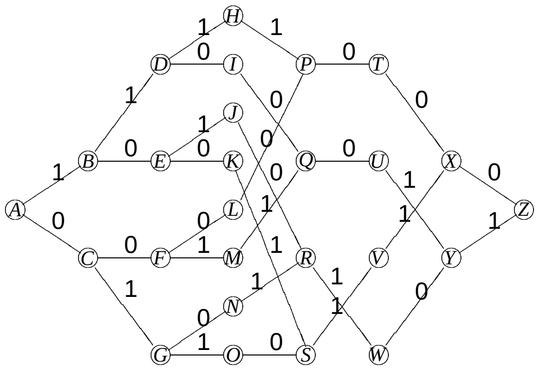
$$w = (-0.1, 0.0, 0.2, 0.9, 1.8, 0.9, 1.4) \in \mathbb{R}^7.$$

If we round each value to the nearest of zero and one, we obtain 0001111, which is at distance 2 from the second, third and fifth codewords in the list, so we have a decoding failure.

If we make the (physically realistic) assumptions that the errors at the sampling points are independent identically distributed Gaussian variables, then it can be shown that the most likely codeword to have been transmitted is the one at smallest Euclidean distance from w in $\mathbb{R}^7$, which turns out to be 0101101.

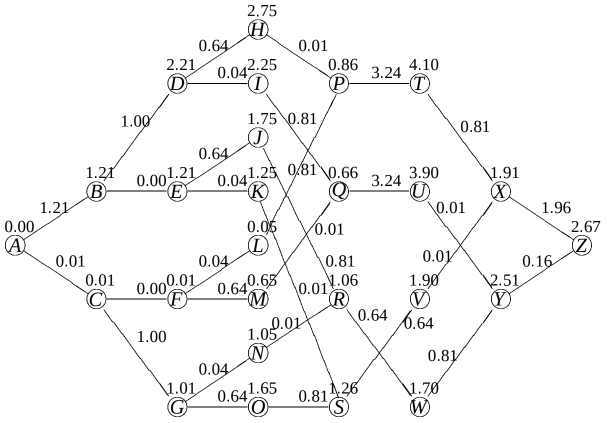
A trellis

A trellis for the dual Hamming code:



The codewords are the sequences of labels on the paths from A to Z.

Trellis decoding



The shortest squared Euclidean distance from received word to codeword is equal to the length of the shortest path from A to Z. The shortest path is ACGNRWYZ, and the decoded word is 0101101.

Muder’s Theorem

The best trellis for a code will have the fewest vertices or edges, or smallest cycle rank.

For a linear code, there is a trellis which is uniformly best:

Theorem 3 *Let C be a linear code of length n. Then there is a trellis T representing C, with layers $V_0,\dots,V_n$, such that, if another proper trellis T' for C has layers $V_0',\dots,V_n'$, then $|V_i'| \ge |V_i|$ for $i = 0,\dots,n$. Moreover, if $|V_i'| = |V_i|$ for $i = 0,\dots,n$, then T' is isomorphic to T. Furthermore, T also minimises the sizes of all the edge layers and the cycle rank.*

Past and future

For $0 \le i \le n$, we define the ith *past subcode* of C to be

$$P_i = \{c \in C : c_j = 0 \text{ for all } j > i\},$$

and the ith *future subcode* to be

$$F_i = \{c \in C : c_j = 0 \text{ for all } j \le i\}.$$

By convention, $P_n = F_0 = C$.

If A and B denote the first and last bases for M, then

$$\begin{aligned}\dim(F_i) &= |A \cap \{i+1,\dots,n\}|,\\ \dim(P_i) &= |B \cap \{1,\dots,i\}|.\end{aligned}$$

Let $V_i = C/(P_i \oplus F_i)$. For each codeword c, put an edge with label c_i from $(P_{i-1} \oplus F_{i-1}) + c \in V_{i-1}$ to the coset $(P_i \oplus F_i) + c \in V_i$. Identify edges with the same label between the same vertices.

This is the Muder trellis for C.

Equivalent codes

Equivalent codes may have Muder trellises of different size. The problem of finding the smallest Muder trellis for a code equivalent to C is NP-complete in general.

However, using the matroid allows us to produce bounds (or exact values) for many important codes. We must order so that the first base is as late as possible, and the last base as early as possible. Of course, these requirements conflict.

For example, a code is MDS if and only the matroid is uniform. In this case, regardless of permutations, the first base is $\{1,\dots,k\}$ and the last base is $\{n-k+1,\dots,n\}$. So the trellis is as large as possible, no matter how we permute coordinates.

Hamming weight hierarchy

The ith *generalised Hamming weight* of a code C is the smallest size d_i of the support of an i-dimensional subcode of C. So, for example, d_1 is the minimum weight of C.

The *Hamming weight hierarchy* is $(d_1,d_2,\dots,d_k)$. Note that it is determined by the Tutte polynomial.

Bounds for d_i can be obtained from the Griesmer bound and other methods; for example,

$$d_{i+1} \geq d_i + \left\lceil \frac{d_i(q-1)}{q(q^i-1)} \right\rceil .$$

Theorem 4 *The first and last bases of the* $[n,k]$ *code* C *satisfy*

$$a_i \leq n - d_{k-i+1}(C) + 1, \qquad b_i \geq d_i(C).$$

If these bounds are attained then the Muder trellis for C *is smaller than that for any equivalent code.*

The binary Golay code

The Hamming weight hierarchy for the extended binary Golay code is

$$\{8,12,14,15,16,18,19,20,21,22,23,24\}.$$

The coordinates can be ordered so that this is the last base, and the first base is its complement. So the bounds of Theorem 4 are attained. Indeed, this will hold as long as we ensure that

$$\{\{1,2,3,4\},\{5,6,7,8\},\dots,\{21,22,23,24\}\}$$

is a *sextet*.

(Since the code is self-dual, the complement of the first base is the last base.)

Hence the smallest trellis for the extended binary Golay code has 2686 vertices.