# Coding Theory: main definitions and theorems

Robin Chapman

14 April 2008

An *alphabet* is a finite set $\Omega$. A *word* over $\Omega$ is a finite string $\mathbf{a} = a_1 a_2 \cdots a_n$ of letters $a_i \in \Omega$. Its *length* is $n$. The set of all words of length $n$ over $\Omega$ is denoted by $\Omega^n$. If $\mathbf{a}$, $\mathbf{b} \in \Omega^n$ the *Hamming distance* $d(\mathbf{a}, \mathbf{b})$ is the number of subscripts $j$ with $a_j \neq b_j$.

**Theorem 1** *The Hamming distance satisfies*

1. $d(\mathbf{a}, \mathbf{a}) = 0$ *for all* $\mathbf{a}$,
2. $d(\mathbf{a}, \mathbf{b}) > 0$ *for all* $\mathbf{a}$, $\mathbf{b}$ *with* $\mathbf{a} \neq \mathbf{b}$,
3. $d(\mathbf{a}, \mathbf{b}) = d(\mathbf{b}, \mathbf{a})$ *for all* $\mathbf{a}$, $\mathbf{b}$,
4. $d(\mathbf{a}, \mathbf{c}) \leq d(\mathbf{a}, \mathbf{b}) + d(\mathbf{b}, \mathbf{c})$ *for all* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$.

A *code* of length $n$ over an alphabet $\Omega$ is a subset of $\Omega^n$. Its elements are called *codewords*. Its *minimum distance* is the least value of $d(\mathbf{a}, \mathbf{b})$ where $\mathbf{a}$ and $\mathbf{b}$ range over distinct codewords. the minimum distance of a code $C$ is denoted by $d(C)$. An $(n, k, d)$-*code* over $\Omega$ is a subset of $\Omega^n$ consisting of $k$ words with minimum distance $d$.

The notion of *equivalence* of codes is better conveyed by example than by a brief definition. There are various maps $\phi : \Omega^n \to \Omega^n$ which preserve Hamming distance: $d(\phi(\mathbf{a}), \phi(\mathbf{b})) = d(\mathbf{a}, \mathbf{b})$. One class of $\phi$ is obtained by taking a permutation $\sigma$ of $\{1, \ldots, n\}$ and setting $\phi(a_1 \cdots a_i \cdots a_n) = a_{\sigma(1)} \cdots a_{\sigma(i)} \cdots a_{\sigma(n)}$. Another is got by taking a permutation $\tau$ of $\Omega$ and a fixed $k$ with $1 \leq k \leq n$ and setting setting $\phi(a_1 \cdots a_{k-1} a_k a_{k+1} \cdots a_n) = a_1 \cdots a_{k-1} \tau(a_k) a_{k+1} \cdots a_n)$. If $C \subseteq \Omega^n$ then the image $C'$ of $C$ under a sequence of operations of these two types is said to be a code *equivalent* to $C$. Then $C'$ has the same number of words and the same minimum distance as $C$. See the slides for examples; the above description makes the concept sound more difficult than it really is.

*Minimum distance decoding* of a code $C \subseteq \Omega^n$ decodes a received word $\mathbf{b} \in \Omega^n$ with an $\mathbf{a} \in C$ minimizing $d(\mathbf{a}, \mathbf{b})$. A code $C$ is an *e-error-correcting* code if minimum distance decoding works correctly whenever at most $e$ errors are made, that is if for all $\mathbf{a} \in C$ and $\mathbf{b} \in \Omega^n$ with $d(\mathbf{a}, \mathbf{b}) \leq e$ then the only $\mathbf{c} \in C$ with $d(\mathbf{c}, \mathbf{b})$ is $\mathbf{c} = \mathbf{a}$.

**Theorem 2** *A code $C$ is e-error-correcting if and only if $d(C) \geq 2e + 1$.*

A code $C$ is *e-error-detecting* if and only if for all $\mathbf{a} \in C$ and $\mathbf{b} \in C$ with $0 < d(\mathbf{a}, \mathbf{b}) \leq e$ then $\mathbf{b} \notin C$.

**Theorem 3** *A code $C$ is e-error-detecting if and only if $d(C) \geq e + 1$.*

**Theorem 4 (Sphere packing bound)** *If the code $C \subseteq \Omega^n$ is an e-error-correcting code then it has at most*

$$\frac{q^n}{\sum_{j=0}^{e} \binom{n}{j} (q-1)^j} \qquad (*)$$

*words where $q = |\Omega|$.*

A *perfect* $e$-error-correcting code is one having precisely the number of words in $(*)$.

**Theorem 5 (Singleton bound)** *If $C$ is an $(n, k, d)$ code over $\Omega$ then*

$$k \leq q^{n-d+1}$$

*where $q = |\Omega|$.*

For a positive integer $n$, $Z_n$ denotes the set $\{0, 1, \ldots, n-1\}$ equipped with the operations of addition, subtraction and multiplication modulo $n$. Always $Z_n$ is a commutative ring, but when $n = p$ is prime $Z_p$ is also a *field*: every nonzero element $a \in Z_p$ has a *reciprocal* $b$ with $ab = 1$.

**From now on we always let $p$ denote a prime number.**

We can regard a word $\mathbf{a} = a_1 \cdots a_n \in (Z_p)^n$ as a row vector $(a_1, \ldots, a_n)$. In this way $(Z_p)^n$ is a vector space over the field $Z_p$. The *weight* $w(\mathbf{a})$ of $\mathbf{a} \in (Z_p)^n$ is the number of nonzero symbols in $\mathbf{a}$.

**Theorem 6**

$$w(\mathbf{a}) = d(\mathbf{a}, 00 \cdots 0) \qquad \text{and} \qquad d(\mathbf{a}, \mathbf{b}) = w(\mathbf{a} - \mathbf{b})$$

*for all $\mathbf{a}$, $\mathbf{b} \in (Z_p)^n$.*

A *linear code* of length $n$ over $Z_p$ is a vector subspace of $(Z_p)^n$, that is $C \subseteq (Z_p)^n$ if and only if

1. $C$ is nonempty,
2. if $\mathbf{a}$, $\mathbf{b} \in C$ then $\mathbf{a} + \mathbf{b} \in C$ and
3. if $\lambda \in Z_p$ and $\mathbf{a} \in C$ then $\lambda\mathbf{a} \in C$.

However the last of these conditions is redundant. The *minimum distance* of a linear code is the least weight of its nonzero elements.

**Theorem 7** *The minimum weight of a linear code equals its minimum distance.*

Each linear code $C$, being a vector space has a *basis* $\mathbf{a}_1, \ldots, \mathbf{a}_k$, that is the elements of $C$ are the sums $\sum_{i=1}^{k} \lambda_i \mathbf{a}_i$ and that each element of $C$ has precisely one representation in this form. Then $C$ has *dimension* $k$ as a vector space and has $p^k$ codewords. An $[n, k, d]$-linear code is a linear code of length $n$, dimension $k$ and minimum distance $d$. A *generator matrix* for a linear code $C$ is a matrix $A$ whose rows form a basis for $C$. Then the code $C$ is the set of all vectors $\mathbf{x}A$ where $\mathbf{x}$ runs through $(Z_p)^k$. One can use $A$ to transform a word $\mathbf{x} \in (Z_p)^k$ into a codeword $\mathbf{x}A$ in $C$ by multiplication by $A$.

**Theorem 8** *If $A$ is a generator matrix for a linear code $C$ then any matrix $A'$ obtained from $A$ by elementary row operations is also a generator matrix for $C$.*

We say that a generator matrix $A$ is in *standard form* if $A = (I \mid B)$ where $I$ is an identity matrix.

**Theorem 9** *If $A$ is a generator matrix for a linear code $C$ then any matrix $A'$ obtained from $A$ by permuting its columns or multiplying its columns by nonzero scalars is a generator matrix for a code $C'$ equivalent to $C$.*

By reducing a generator matrix to reduced echelon form then permuting its columns one can obtain a generator matrix $A'$ for an equivalent code $C'$ with $A'$ in standard form. If $A = (I \mid B)$ is in standard form, then $\mathbf{x}A = (\mathbf{x} \mid \mathbf{x}B)$ consists of the message $\mathbf{x}$ with some extra digits $\mathbf{x}B$ appended; these are called *check digits*.

If $C \subseteq (Z_p)^n$ is a linear code then a *coset* of $C$ is a set $\mathbf{a}+C = \{\mathbf{a}+\mathbf{c} : \mathbf{c} \in C\}$ where $\mathbf{a} \in (Z_p)^n$. Each element of $(Z_p)^n$ is in exactly one coset. Choose a word of least weight in each coset and call it a *coset leader*. *Coset decoding* decodes a received message $\mathbf{b}$ as $\mathbf{b} - \mathbf{e}$ where $\mathbf{e}$ is the coset leader of $\mathbf{b} + C$. Coset decoding performs correctly if $\mathbf{b} - \mathbf{a}$ is a coset leader where $\mathbf{a}$ and $\mathbf{b}$ are the sent and received messages.

**Theorem 10** *A linear code $C$ is $e$-error-correcting if and only if every word of weight at most $e$ is a coset leader.*

For $\mathbf{a} = a_1 \cdots a_n$ and $\mathbf{b} = b_1 \cdots b_n$ define their dot product as $\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + \cdots + a_n b_n = \mathbf{a}\mathbf{b}^t$. If $C \subseteq (Z_p)^n$ is a linear code, its *dual* is

$$C^{\perp} = \{\mathbf{x} \in (Z_p)^n : \mathbf{x} \cdot \mathbf{c} = 0 \text{ for all } \mathbf{c} \in C\}.$$

**Theorem 11** *Let $C$ be a linear code of length $n$ and dimension $k$. Then $C^{\perp}$ is a linear code of length $n$ and dimension $n - k$. Also $(C^{\perp})^{\perp} = C$.*

A generator matrix for $C^{\perp}$ is called a *parity-check matrix* for $C$. Then a generator matrix for $C$ is a parity-check matrix for $C^{\perp}$. If $H$ is a parity-check matrix for $C$ then $C = \{\mathbf{x} \in (Z_p)^n : H\mathbf{x}^t = 0\}$.

**Theorem 12** *If $A = (I \mid B)$ is a generator matrix for $C$ in standard form, then $H = (-B^t \mid I)$ is a parity-check matrix for $C$.*

Let $H$ be a parity check matrix for a linear code $C$. For $\mathbf{x} \in (Z_p)^n$ its *syndrome* is $H\mathbf{x}^t$. Two words have the same syndrome if and only if they lie in the same coset of $C$. *Syndrome decoding* works by first precomputing and tabulating the syndrome of each coset leader, then decoding a received word $\mathbf{b}$ by computing its syndrome $H\mathbf{b}^t$, then identifying the coset leader $\mathbf{e}$ with $H\mathbf{e}^t = H\mathbf{b}^t$ and then decoding $\mathbf{b}$ as $\mathbf{b} - \mathbf{e}$. Syndrome decoding is theoretically equivalent to coset decoding, but is more efficient in practice.

For a prime $p$ and positive integer $r$ the *Hamming code* $\mathrm{Ham}(p, r)$ is defined to be the code with parity check matrix $H$ with $r$ rows and where each nonzero column vector whose top nonzero entry is 1 occurs exactly once as a column of $H$. Then $\mathrm{Ham}(p, r)$ has length $(p^r - 1)/(p - 1)$ dimension $(p^r - 1)/(p - 1) - r$ and is a perfect 1-error-correcting code.

**Theorem 13** *Let $C$ be a linear code with parity check matrix $H$. Then $C$ has minimum weight $k$ if and only if the smallest set of linearly dependent columns of $H$ has size $k$.*

A *cyclic code* of length $n$ over $Z_p$ is a linear code $C$ over $Z_p$ with the additional property:

- if $a_0a_1a_2\cdots a_{n-1} \in C$ then its cyclic shift $a_{n-1}a_0a_1\cdots a_{n-2} \in C$.

To study cyclic codes we introduce the ring $Z_p[x]_n$ (this is **not** a standard notation). It consists of all polynomials

$$a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}.$$

The addition is just like that of ordinary polynomials, noting that since the coefficients lie in $Z_p$, addition is done on them modulo $p$. Multiplication is done similarly to the usual multiplication of polynomials with the extra stipulation that $x^n = 1$ (so that $x^{n+1} = x$, $x^{n+2} = x^2$ etc). As in example, in $Z_5[x]_3$ we have

$$\begin{aligned}(1 + 3x + x^2)(1 + 2x^2) &= 1 + 3x + 3x^2 + 6x^3 + 2x^4 \\ &= 1 + 3x + 3x^2 + 6 + 2x \\ &= 7 + 5x + 3x^2 = 2 + 3x^2.\end{aligned}$$

There is a map $\Phi : (Z_p)^n \to Z_p[x]_n$ defined by

$$\Phi(a_0a_1a_2\cdots a_{n-1}) = a_0 + a_1x + a_2x^2 \cdots + a_{n-1}x^{n-1}.$$

Then $\Phi$ is a bijection, $\Phi(\mathbf{a} + \mathbf{b}) = \Phi(\mathbf{a}) + \Phi(\mathbf{b})$ and $\Phi(c\mathbf{a}) = c\Phi(\mathbf{a})$ for $\mathbf{a}$, $\mathbf{b} \in (Z_p)^n$ and $c \in Z_p$. Most importantly, for $\mathbf{a} = a_0a_1a_2\ldots a_{n-1}$ and its cyclic shift $\mathbf{a}' = a_{n-1}a_0a_1\ldots a_{n-2}$ we have

$$\Phi(\mathbf{a}') = x\Phi(\mathbf{a}). \tag{†}$$

An *ideal* of $Z_p[x]_n$ is a nonempty subset $I$ of $Z_p[x]_n$ satisfying

- if $f$, $g \in I$ then $f + g \in I$,
- if $f \in I$ and $h \in Z_p[x]_n$ then $hf \in I$.

**Theorem 14** *Let $C$ be a subset of $(Z_p)^n$. Then $C$ is a cyclic code if and only if $\Phi(C)$ is an ideal of $Z_p[X]_n$.*

The proof of this theorem uses (†) crucially. The importance of this result lies in the fact that there is a complete theory of ideals of rings like $Z_p[x]_n$. Indeed every ideal in this ring is principal. A *principal ideal* in $Z_p[x]_n$ is an ideal of the form

$$\langle f\rangle = \{hf : h \in Z_p[x]_n\}.$$

Recall that a *monic polynomial* is a polynomial whose leading coefficient is 1.

**Theorem 15** *Let $I$ be an ideal of $Z_p[x]_n$. Then $I = \langle f \rangle$ is a principal ideal where $f$ is a monic polynomial which is a factor of the polynomial $x^n - 1$ over $Z_p$. This polynomial $f$ is uniquely determined by the ideal $I$.*

If $C$ is a cyclic code, the we call the polynomial $f$ prescribed by the above theorem the *generator polynomial* of $C$. If $f(X) = \sum_{j=0}^{d} a_j x^j$ has degree $d$ then $a_d = 1$ and $C$ has generator matrix

$$\begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_{d-1} & 1 & 0 & 0 & \cdots & 0 \\ 0 & a_0 & a_1 & \cdots & a_{d-2} & a_{d-1} & 1 & 0 & \cdots & 0 \\ 0 & 0 & a_0 & \cdots & a_{d-3} & a_{d-2} & a_{d-1} & 1 & \cdots & 0 \\ & & & \ddots & & & & & \ddots & \\ 0 & 0 & 0 & \cdots & a_0 & a_1 & a_2 & a_3 & \cdots & 1 \end{pmatrix}$$

with $n - d$ rows. Hence $C$ has dimension $n - d$. The *parity-check* polynomial of $C$ is $g = (x^n - 1)/f$. Then $g = \sum_{j=0}^{n-d} b_j x^j$ where $b_{n-d} = 1$ and $C$ has parity-check matrix

$$\begin{pmatrix} 1 & b_{n-d-1} & b_{n-d-2} & \cdots & b_1 & b_0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & b_{n-d-1} & \cdots & b_2 & b_1 & b_0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & b_3 & b_2 & b_1 & b_0 & \cdots & 0 \\ & & & \ddots & & & & & \ddots & \\ 0 & 0 & 0 & \cdots & 1 & b_{n-d-1} & b_{n-d-2} & b_{n-d-3} & \cdots & b_0 \end{pmatrix}.$$

The *weight enumerator* of a code $C$ of length $n$ over $Z_p$ is the polynomial

$$W_C(z) = \sum_{k=0}^{n} A_k z^k$$

where $A_k$ is the number of words in $C$ having weight $k$. We can rewrite this definition as

$$W_C(z) = \sum_{\mathbf{a} \in C} z^{w(\mathbf{a})}.$$

**Theorem 16 (MacWilliams identity)** *Let $C$ be a linear code of length $n$ over $Z_2$. Then*

$$W_{C^\perp}(z) = \frac{(1+z)^n}{|C|} W_C\left(\frac{1-z}{1+z}\right).$$

*Alternatively*

$$W_C(z) = \frac{|C|(1+z)^n}{2^n} W_{C^\perp}\left(\frac{1-z}{1+z}\right).$$