MATH3302 Coding Theory summary

1. Any Binary Code

- If the reliability of a binary symmetric channel is $p$ and if $v$ and $w$ are words of length $n$ that differ in $d$ positions, then $\Phi_p(v, w) = p^{n-d}(1-p)^d$.

- The information rate of a binary code of length $n$ with $|C|$ codewords is $\frac{1}{n} \log_2 |C|$.

- The weight of a word is the number of ones in it. The Hamming distance between two words is the number of positions in which they differ. So $d(v, w) = wt(v + w)$.

- Maximum Likelihood Decoding: Received word is decoded to the closest codeword. For a code $C$ and a codeword $v$, $\Theta_p(C, v) = \sum_{w \in L(v)} \Phi_p(v, w)$ is the probability that if $v$ is transmitted over a BSC with reliability $p$ then IMLD correctly concludes that $v$ was sent.

- Error detection and correction: Let $\delta$ be the minimum distance between any pair of codewords in $C$. $C$ will detect all nonzero error patterns of weight less than or equal to $\delta - 1$. $C$ will correct all error patterns with weight less than or equal to $(\delta - 1)/2$ (for odd $\delta$) or $(\delta - 2)/2$ (for even $\delta$).

- Two codes $C$ and $C'$ are equivalent if the words of $C'$ can be obtained by applying a particular permutation to the bits of each word of $C$.

- The extended code $C^*$ of a code $C$ is obtained by adding a parity check digit to each codeword in $C$ so that the weight of each codeword in $C^*$ is even.

- Given two codes $A$ and $B$ of length $n$, a new code $C$ of length $2n$ can be formed using the $(a | a + b)$ construction.

2. Linear Binary Codes

- For a linear code $C$: if $v, w \in C$, then $v + w \in C$; the distance $\delta$ is the weight of the nonzero codeword of smallest weight; the dimension $k$ is the number of codewords in a basis for $C$; the rate of $C$ is $k/n$; and the number of codewords in $C$ is $|C| = 2^k$.

- If $C$ has dimension $k_1$ and the dual code $C^\perp$ has dimension $k_2$, then $k_1 + k_2 = n$.

- A generating matrix for $C$ is a $k \times n$ matrix whose rows form a basis for $C$. A message word $u$ of length $k$ is encoded as $v = uG$. If $G$ is in standard form then $u$ is the first $k$ bits of the corresponding codeword $v$.

- A parity check matrix for $C$ is an $n \times (n-k)$ matrix whose columns form a basis for $C^\perp$. If $H$ is a parity check matrix for $C$ then $H^T$ is a generating matrix for $C^\perp$. If $H$ is a parity check matrix for $C$ and $v \in C$, then $vH = 0$.

- A linear code $C$ of length $n$ and dimension $k$ has $2^{n-k}$ cosets. The word $u + v \in C$ if and only if $u$ and $v$ are in the same coset.

- IMLD for linear codes is based on the fact that the most likely error pattern and the received word are in the same coset. For a received word $w$, calculate its syndrome $wH$ and the most likely error pattern is the word of least weight in the coset with that syndrome. If $u$ is the error pattern in a received word $w$, then $uH = wH$ is the sum of the rows of $H$ that correspond to the positions where errors occurred in transmission. The reliability of IMLD is the same for all codewords in a linear code.
• If $C$ is a linear code of length $n$ and distance $\delta = 2t + 1$ or $2t + 2$, then

$$|C| \leq \frac{2^n}{\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{\delta - 1}}.$$  

A perfect code is one for which this is an equality, it will correct all error patterns of weight up to and including $t$ and no other error patterns.

• If $\binom{n-1}{0} + \binom{n-1}{1} + \cdots + \binom{n-1}{\delta - 2} < 2^{n-k}$ then there exists a linear code of length $n$, dimension $k$ and distance at least $\delta$. Thus there exists a linear code of length $n$ and distance at least $\delta$ with

$$|C| \geq \frac{2^{n-1}}{\binom{n-1}{0} + \binom{n-1}{1} + \cdots + \binom{n-1}{\delta - 2}}.$$  

• Examples of linear codes:
  - Hamming codes $n = 2^r - 1$, $k = 2^r - 1 - r$, $\delta = 3$.
  - Simplex codes (dual of Hamming) $n = 2^r - 1$, $k = r$, $\delta = 2^{r-1}$.
  - $r^{th}$ order Reed-Muller codes $n = 2^m$, $k = \sum_{i=0}^{m} (m)$, $\delta = 2^{m-r}$.
  - Extended Golay code $n = 24$, $k = 12$, $\delta = 8$.
  - Golay code $n = 23$, $k = 12$, $\delta = 7$.

3. Cyclic Linear Codes

• Words of length $n$ correspond to polynomials of degree at most $n - 1$.

• If $v \in C$, then $\gamma(v) \in C$. If $f(x) \in C$, then $xf(x) \in C$.

• The generator of $C$ is the unique polynomial of least degree, and every polynomial $f(x) \in C$ can be written as a multiple of $g(x)$. $C$ has length $n$ and dimension $k$ iff $g(x)$ has degree $n - k$.

• Generating matrices for $C$ (in non-standard and standard form) are

$$G_1 = \begin{pmatrix} g(x) \\ xg(x) \\ \vdots \\ x^{k-1}g(x) \end{pmatrix} \quad G_2 = \begin{pmatrix} r_{n-k} \\ \vdots \\ r_{n-1} \end{pmatrix} \cdot \text{(where } r_i \leftrightarrow x^i \text{ mod } g(x))$$

• The polynomial $g(x)$ is a generator for a cyclic linear code of length $n$ if and only if $g(x)|(1 + x^n)$.

• Message $a(x)$ encodes to $c(x) = a(x)g(x)$ if you use the generating matrix $G_1$ above.

• IMLD for cyclic linear codes and burst error correction is based on the idea that “closest” means the codeword that differs from the received word in a cyclic burst error pattern of shortest length. Code $C$ is $t$ cyclic burst error correcting if every word of length $n$ that contains a cyclic burst error of length at most $t$ is in a distinct coset of $C$ (so has a distinct syndrome).

• Decoding algorithm involves calculating the syndrome of $w$, $s = wH$ with corresponding polynomial $s_0(x)$, then calculating $s_i(x) = xs_{i-1}(x)$ until an $s_i$ is found which contains a burst error of length at most $t$. Then $e_i = (0, s_i)$ and shifting is done to find the error pattern $e$.

• Interleaving and cross-interleaving are techniques to improve the burst error correction capabilities of codes.