Bounds for Codes
(Handout March 11, 2013)

Let $\mathcal{A}$ be an alphabet consisting of $q$ letters. A word of length $n$ is a string $w \in \mathcal{A}^n$. Given two words $u = u_1u_2\cdots u_n$ and $v = v_1v_2\cdots v_n$ where $u_i, v_i \in \mathcal{A}$, the (Hamming) distance between $u$ and $v$, denoted $\delta(u, v)$, is the number of coordinates $i \in \{1, 2, \ldots, n\}$ such that $u_i \neq v_i$. A code is a subset $\mathcal{C} \subseteq \mathcal{A}^n$. Elements of $\mathcal{C}$ are called codewords. The minimum distance of $\mathcal{C}$ is the minimum distance between two of its codewords:

$$d = \min_{u \neq v \in \mathcal{C}} \delta(u, v).$$

The error-correcting capability of $\mathcal{C}$ is $e = \lfloor \frac{d-1}{2} \rfloor$. The information rate of $\mathcal{C}$ is

$$R = \frac{\log_q |\mathcal{C}|}{n}.$$ 

(In the case of a linear $[n, k]$-code, $\mathcal{C} \subseteq \mathbb{F}_q^n$ is a $k$-dimensional subspace so $|\mathcal{C}| = q^k$ which has information rate $k/n$.) The information rate indicates what proportion of the transmitted letters carry actual information; the remaining letters are redundant, allowing for error correction.

We seek codes which attain an optimal balance between two competing considerations: high information rate (i.e. large number of codewords) and high error-correcting capability (i.e. large minimum distance). This optimum is subject to certain known bounds. We first review two upper bounds which we have already seen; and we introduce for the first time a lower bound.

**Theorem (Singleton bound).** Every $q$-ary code $\mathcal{C}$ of length $n$ with minimum distance $d$ satisfies $|\mathcal{C}| \leq q^{n-d+1}$. Its information rate is bounded above by $R \leq 1 - \frac{d-1}{n}$.

*Proof.* We puncture the code $\mathcal{C}$ by deleting the last $d-1$ coordinates of all codewords. Because the original codewords had minimum distance $\geq d$, the resulting codewords of length $n - (d-1)$ are distinct, so $|\mathcal{C}| \leq q^{n-(d-1)}$. \hfill $\square$

For each word $w \in \mathcal{A}^n$ and $r \geq 0$, denote the closed ball of radius $r$ centered at $w$ by

$$B_r(w) = \{v \in \mathcal{C} : \delta(v, w) \leq r\}.$$
Observe that
\[ |B_r(w)| = \sum_{0 \leq i \leq r} \binom{n}{i} (q - 1)^i, \]
independent of the choice of \( w \in \mathcal{A}^n \); it is convenient to denote this value by \( |B_r(w)| = |B_r(0)| \) where \( 0 \in \mathcal{A}^n \) is a fixed reference word such as \( 000 \cdots 0 \) (if \( 0 \in \mathcal{A} \)).

**Theorem (Hamming bound; sphere-packing bound).** If a \( q \)-ary code \( C \) of length \( n \) corrects \( e \) errors, then
\[ |C| \leq \frac{q^n}{|B_e(0)|}. \]

**Proof.** The closed balls \( B_e(w) \) of radius \( e \) centered at the codewords \( w \in C \) are disjoint. Their union \( \bigcup_{w \in C} B_e(w) \subseteq \mathcal{A}^n \) therefore has size \( |C||B_e(w)| \leq |\mathcal{A}^n| = q^n. \)

If we switch viewpoint from the packing problem (trying to pack as many non-overlapping balls of fixed radius as possible in the space \( \mathcal{A}^n \)) to the dual problem, a covering problem (trying to cover \( \mathcal{A}^n \) with as few balls as possible, if the balls have fixed radius but are allowed to overlap) we obtain the following lower bound for the existence of good codes.

**Theorem (Gilbert-Varshamov bound).** For a fixed alphabet size \( q \geq 2 \) and positive integers \( d \leq n \), there exists a \( q \)-ary code \( C \) of length \( n \) and minimum distance at least \( d \) of size
\[ |C| \geq \frac{q^n}{|B_{d-1}(0)|}. \]

**Proof.** Let \( \mathcal{A} \) be an alphabet of size \( q \), and let \( C \subseteq \mathcal{A}^n \) be maximal subject to the condition that codewords have minimum distance at least \( d \). Such codes always exist, and may be constructed by the following greedy algorithm:

- **Index the words as** \( \mathcal{A} = \{w_1, w_2, \ldots, w_{q^n}\} \). Initialize \( C := \{\} \). Then
- **For each** \( i = 1, 2, \ldots, q^n \) **do:**
  - **If** \( w_i \) **has distance** \( \geq d \) **from each of the previously chosen words in** \( C \), **set** \( C := C \cup \{w_i\} \).

By construction, every word \( w \in \mathcal{A}^n \) has distance \( \leq d - 1 \) from some codeword. Equivalently, every word lies in at least one of the balls \( B_{d-1}(w) \) for some \( w \in C \). Thus \( \bigcup_{w \in C} B_{d-1}(w) = \mathcal{A}^n \), in which some of the balls may overlap; so
\[ |C||B_{d-1}(w)| \geq |\mathcal{A}^n| = q^n. \]
Unfortunately the latter existence proof does not yield efficient codes in practice. While they have good lower bounds on information rate and minimum distance, they are not canonical: the code $\mathcal{C}$ obtained by the greedy construction above, and even the number of resulting codewords $|\mathcal{C}|$, depend on the initial ordering of words chosen. This problem is somewhat ameliorated if a natural ordering of words is chosen, such as lexicographical order; but still, the resulting code $\mathcal{C}$ in general is obtained as merely a (long) list of codewords, so that decoding would require looking through the entire list of codewords and performing a large number of letter comparisons. So in general, these codes are not practical. It is a challenge to construct practical codes that do nearly as well.

Because we are interested in the reliable and efficient transformation of large amounts of information, we want to study the bounds above for large $n$. We consider a channel which carries a stream of letters over the $q$-ary alphabet $\mathcal{A}$, for which each letter is altered with probability $p$, and errors in different locations occur independently. (This is not the only reasonable assumption, and not necessarily the most realistic, since—in practice—errors often occur in bursts. We adopt it as a first approximation, to facilitate computation.) When we transmit $n$ letters using our best available code of length $n$, the expected number of errors is $pn$. So our code should have minimum distance $d \geq 2pn + 1$. If our interest is in transmitting long messages over this channel, we therefore require codes with minimum distance at least $d = \lfloor \delta n \rfloor$ where $0 < \delta < 1$ and the constant $\delta$ depends on the channel (for the channel above, we should take $\delta > 2p$). I am writing ‘$\delta$’ to conform with the notation in van Lint’s expository article *Algebraic Geometric Codes*, hoping that this does not cause confusion with the notation we have used for the Hamming distance.

To compare the bounds above, we would like to have good estimates for $|B_r(0)|$. Such estimates rely on Stirling’s formula

$$n! \sim n^{n+\frac{1}{2}}e^{-n\sqrt{2\pi}}$$

as $n \to \infty$. Recall that this means $\frac{n!}{n^{n+\frac{1}{2}}e^{-n\sqrt{2\pi}}} \to 1$ as $n \to \infty$. For reasons explained above, we want a family of codes with minimum distance at least $d \geq \lfloor \delta n \rfloor$ growing in direct proportion to the length $n$. Such a family of codes will have $\frac{d}{n} \to \delta$ as $n \to \infty$.

The following says that for a wide range of $\delta$-values, the value of $|B_{d-1}(0)|$ is approximated rather well by the last term $\binom{n}{d-1}(q-1)^{d-1}$ in the sum:

**Proposition.** If $0 < \delta < \frac{q-1}{q}$ and $d = \lfloor \delta n \rfloor$, then

$$\binom{n}{d-1}(q-1)^{d-1} \leq |B_{d-1}(0)| \leq d\binom{n}{d-1}(q-1)^{d-1}.$$
Proof. The first inequality says that the sum which expresses \(|B_{d-1}(0)|\) bounded below by the last term in the sum

\[ |B_{d-1}(0)| = \sum_{i=0}^{d-1} \binom{n}{i} (q - 1)^i. \]

The ratio of the term for arbitrary \(i\), to the last term, is

\[
\frac{n_i (q - 1)^i}{\binom{n}{d-1} (q - 1)^{d-1}} = \frac{(i + 1)(i + 2)(i + 3) \cdots (d - 1)}{(n - d + 2)(n - d + 3)(n - d + 4) \cdots (n - i)} \times (q - 1)^{i - d + 1}
\]

\[ \leq \left( \frac{d - 1}{n - d + 2} \right)^{d - 1 - i} (q - 1)^{d - 1 - i}
\]

\[ \leq \left( \frac{\delta}{1 - \delta} \right)^{d - 1 - i} (q - 1)^{i - d + 1}
\]

\[ < (q - 1)^{d - 1 - i}(q - 1)^{i - d + 1}
\]

\[ = 1,
\]

i.e. the last term in the sum is the largest term. Since there are \(d\) terms in the sum, the result follows.

This leads to an asymptotic expression for the Gilbert-Varshamov bound as \(n \to \infty\), expressed in terms of the information rate. Denote by \(A_q(n, d)\) the maximum possible information rate for \(q\)-ary codes of length \(n\) with minimum distance \(d\). Then

\[ A_q(n, \lfloor \delta n \rfloor) \geq \frac{1}{n} \log_q \frac{q^n}{|B_{d-1}(0)|}
\]

\[ \geq \frac{1}{n} \left[ \frac{1}{n} - \log_q d - \log_q \left( \frac{n}{d - 1} \right) - (d - 1) \log_q (q - 1) \right]
\]

\[ \sim \frac{1}{n} \left[ \frac{1}{n} - \log_q \delta + \log_q (d - 1) - \frac{1}{2} \log_q n + \frac{1}{2} \log_q (d - 1) \right]
\]

\[ + \frac{1}{n} \left[ \frac{1}{n} - \log_q \delta + \log_q (d - 1) + \sqrt{2\pi} - (d - 1) \log_q (q - 1) \right]
\]

\[ \sim 1 - \log_q n + \delta \log_q (d - 1) + (1 - \delta) \log_q (n - d + 1) - \delta \log_q (q - 1)
\]

\[ = 1 + \delta \log_q \left( \frac{d - 1}{n} \right) + (1 - \delta) \log_q \left( \frac{n - d + 1}{n} \right) - \delta \log_q (q - 1)
\]

\[ = 1 + \delta \log_q \delta + (1 - \delta) \log_q (1 - \delta) - \delta \log_q (q - 1).
\]

We denote the \(q\)-ary entropy function by

\[ H_q(x) = -x \log_q x - (1 - x) \log_q (1 - x) + x \log_q (q - 1); \]
note that $H_q\left(\frac{q-1}{q}\right) = 1$. The asymptotic version of the Gilbert-Varshamov bound is as follows:

$$\limsup_{n \to \infty} \frac{\log_q A_q(n, \lfloor \delta n \rfloor)}{n} \geq 1 - H_q(\delta).$$

whenever $0 < \delta < \frac{q-1}{q}$. This is a lower bound on the asymptotically best information rate for $q$-ary codes of length $n$ with minimum distance $\geq \lfloor \delta n \rfloor$. To compare, the Singleton bound gives the upper bound

$$\limsup_{n \to \infty} \frac{\log_q A_q(n, \lfloor \delta n \rfloor)}{n} \leq 1 - \delta.$$

We illustrate both the lower and upper bound here for $q = 7$: