Today

Limitations on performance of codes.

- Lower bounds on n.
- Upper bounds in R, δ.

The course so far

- Seen various codes: Hamming, Hadamard, Reed Solomon, Reed-Muller.
- Concatenation, and using it to build asymptotically good codes.
- Aside on Justesen codes: Let $\mathbb{F}$ be a field of size 2^ℓ and assume its elements are written as ℓ bit vectors so as to preserve addition. Then the Justesen code has as its messages $c_0 \ldots . c_{k-1} \in \mathbb{F}$ and maps it to $\langle p(\alpha), \alpha p(\alpha) \rangle_{\alpha \in \mathbb{F}^*}$ where $p(x) = \sum_i c_i x^i$. Maps $k\ell$ bits to $\ell \cdot 2^{\ell+1}$ bits. Very explicit. (Thanks to Johan Hastad for pointing this out.)

Summary (contd.)

- Seen some impossibility results too: Hamming = volume bound. Singleton = projection bound.
- While we can construct decent binary codes, and show existence of even better ones, our constructions are from bounds. Would like to know what is right.
- Eventual hope: Upper bounds = Lower bounds.
- Don't have this yet - so how to get qualitative understanding of results?
- One focus: Pick the best result (upper bound/lower bound) for every δ.

- Our focus: Look at the extreme cases $\delta \rightarrow 0$ or $\delta \rightarrow ?$.
- What is the right limit on δ subject to positive rate?

Plotkin bound

(Stated only for binary case)

Plotkin Bound - 1:

If $d \geq (1+\epsilon) \cdot \frac{n}{2}$ then
$\#$ codewords $\leq 1 + \frac{1}{\epsilon}$.

Plotkin Bound - 2:

If $d \geq \frac{n}{2}$ then
$\#$ codewords $\leq 2n$.

Plotkin Bound - 3:

$k \leq n - 2d + \log_2 n$

Interpretation:

- Parts 1 & 2: Address $\frac{1}{2} \leq \delta \leq 1$.
- Part 3: Gives continuity at $\delta = \frac{1}{2}$.
- Part 3: Reduction to Part 2 by restricting.

Plotkin bound: Main idea

- Map Hamming spaces to Euclidean spaces. Use geometric intuition.
- Simplest reduction: $0 \to 1$, $1 \to -1$. Maps $\mathbb{F}_2 \to \mathbb{R}$, $\mathbb{F}_2^n \to \mathbb{R}^n$.
- If $x \to v_x$ and $y \to v_y$, then $\langle x, y \rangle = n - 2\Delta(x,y)$ Hamming distance related to inner products.
- Code C with m codewords and distance $d > n/2$:
  - Normalize inner product by n ...
  - Codewords map to unit vectors.
  - Inner product $\leq 1 - (2d/n) < 0$.

- Can't have too many vectors in R^n with angle $> 90^o$ - Can we?

Geometric fact

- Fact: If m vectors of length ≤ 1 have pair wise inner product less than $-\alpha$, then $m \leq 1 + \frac{1}{\alpha}$.
- Tedious, but intuitive, inductive proof:
  - Let $v_1, \ldots, v_m$ be the vectors.
  - Note $v_1 \neq 0$.
  - W.l.o.g. $v_1 = \langle 1, 0, \ldots, 0 \rangle$.
  - Therefore $v_i = \langle -\alpha_i, v_i' \rangle$, where $\alpha_i \geq \alpha$.
  - Project remaining vectors to last $n-1$ coordinates and scale up by $1/\sqrt{1-\alpha^2}$.
  - Induction, preceded by tedious careful calculation, implies number of vectors at most $\frac{1}{\alpha}$.

Nicer proof?

- Use linear algebra. Cleaner proof. Less intuitive.
- Let $v = v_1 + \cdots + v_m$.
- Then $0 \leq \langle v, v\rangle \leq m - m(m-1)\alpha$.
- QED
- Moral: Guess statement by intution, Prove by linear algebra.

Elias-Bassalygo-Johnson Bounds

Motivation: Hamming bound better for small δ, Plotkin better for large δ. Any way to get a combined proof?

Elias-Bassalygo Bound: $R \leq 1 - H(\tau)$ where τ comes from Johnson bound below.

Johnson Bound: If $\mathcal{C}$ is an $(n, ?, \delta n)_2$-code, then any Hamming ball of radius τn has at most $O(n)$ codewords, where

$$\tau = \frac{1}{2} \cdot \left(1 - \sqrt{1 - 2\delta}\right).$$

- τ vs. δ?

- $\delta/2 \leq \tau \leq \delta$: So E-B bound always better than Hamming, but never better than GV (which is sane).
- $\delta \to 0$, $\tau \approx \delta/2$: So for small rel. distance, don't improve much on Hamming.
- $\delta \to \frac{1}{2}$, $\tau \approx \delta$: So for large δ, approach GV bound.

Elias-Bassalygo Bound

- Pushes the packing bound.
- Go to larger radius.
- Suppose: Can prove that at most 4 balls of radius $e = 2d/3$ contain any one given point.
- Prveious argument gives:

$$V(n, 2d/3, q)q^k \leq 4q^n.$$

- Lose almost nothing on RHS.
- Improve LHS (significantly).

Motivates the Johnson question.

Johnson Bound

Question: Given $\mathbf{r} \in \Sigma^n$, $(n,k,d)_q$ code $\mathcal{C}$.
How many codewords in $B(\mathbf{r}, e)$?

Motivation: (for binary alphabet)
How to pick a bad configuration?
I.e. many codewords in small ball.
W.l.o.g. set $\mathbf{r} = \mathbf{0}$.
Pick c_i's at random from $B(\mathbf{0}, e)$.

Expected' dist. between codewords = ?
Let $\epsilon = e/n$.
Codewords simultaneously non-zero on ϵ^2 fraction of coordinates;
Thus distance $\approx (2\epsilon - 2\epsilon^2)n$.

Johnson bound shows you can't do better!

Hamming to Euclid

- Map $\Sigma \to \mathbb{R}^q$: ith element $\mapsto 0^{i-1}\ 1\ 0^{q-i}$.
- Induces natural map $\Sigma^n \to \mathbb{R}^{qn}$:
  – Maps vectors into Euclidean space.
  – Hamming distance large implies Euclidean distance large.

Argue: Can't have many large vectors with pairwise small inner products.

Hamming to Euclid (contd).

In our case:

Given: $c_1, \ldots, c_m$ codewords in Σ^n and $\mathbf{r} \in \Sigma^n$, s.t.

- $\Delta(c_i, \mathbf{r}) \leq e$
- $\Delta(c_i, c_j) \geq d$

Want: Upper bound on m.

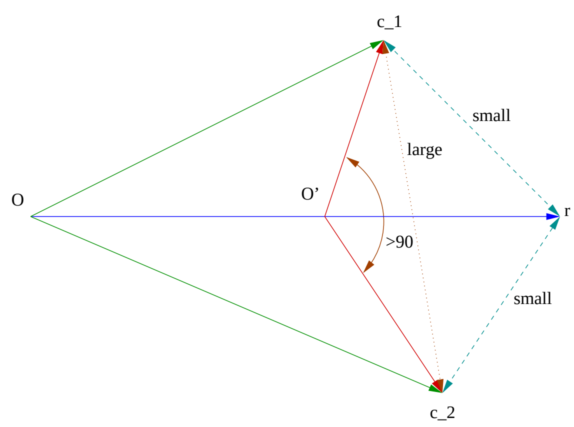
After mapping to $\mathbb{R}^{nq}$
(and abusing notation)

Given: $c_1, \ldots, c_m$ $\mathbb{R}^{nq}$ and $\mathbf{r} \in \mathbb{R}^{nq}$, s.t.

- $\langle \mathbf{r}, \mathbf{r} \rangle = n.$
- $\langle c_i, c_i \rangle = n.$
- $\langle c_i, \mathbf{r} \rangle \geq n - e$
- $\langle () c_i, c_j \rangle \leq n - d$

Want: Upper bound on m.

Hamming to Euclid (contd).



Main idea: Find a new point O' to set as origin, such that the angle subtended by C_i and C_j at O' is at least $90°$.

Conclude: # vectors $\leq$ dimension $= nq$.

Johnson bound (contd).

How to pick the new origin?

Idea 1: Try some point of the form $\alpha\mathbf{r}$.

Then $\langle c_i - \alpha\mathbf{r}, c_j - \alpha\mathbf{r}\rangle$
$$= \langle c_i, c_j\rangle - \alpha\langle c_i \mathbf{r}\rangle - \alpha\langle c_j, \mathbf{r}\rangle + \alpha^2\langle \mathbf{r}, \mathbf{r}\rangle$$
$$\leq (1-\alpha)^2 n + 2\alpha e - d$$

Setting $\alpha = 1$, says: Need $e \leq d/2$.

Setting $\alpha = 1 - e/n$ yields:
Need $e/n \leq 1 - \sqrt{1-\delta}$.

(Not quite what was promised.)

Johnson bound (contd).

A better choice for origin.

Idea 2: Try some point of the form
$\alpha\mathbf{r} + (1-\alpha)\mathbf{Q}$,
where $\mathbf{Q} = (\frac{1}{q})^{qn}$.

Appropriate setting of $\alpha = 1 - e/n$ yields, the desired bound.

Back to Elias Bound

Plugging Johnson bound into earlier argument:

$$k \leq (1 - H_q(\epsilon))n + o(n),$$

where ϵ such that the Johnson bound holds for $e = \epsilon n$.

Importance:

- Proves e.g. No codes of exponential growth with distance $(1-1/q)n$.
- Decently comparable with existential lower bound on rate from random code.