Today

Limitations on performance of codes (contd.).

- Elias-Bassalygo/Johnson bound.
- Linear Programming bound.

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.

Motivation for Johnson bound result

- The τ of the Johnson bound comes from the equation: $\delta = 2\tau - 2\tau^2$.
- Why this formula?
  - Pick (exponentially) many points from Hamming ball of radius τn around $\mathbf{0}$.
  - Expected distance between points is $(2\tau - 2\tau^2)n = \delta n$.
  - W.h.p. no pair at distance $(\delta - \epsilon)n$.
- So the Johnson bound is tight.

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.

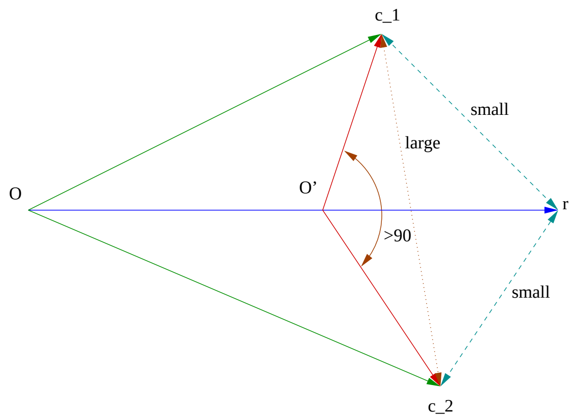
After mapping to $\mathbb{R}^{nq}$
(and abusing notation)

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

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

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

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.

MacWilliams Identities

Defn: Weight distribution of code is $\langle A_0, \ldots, A_n \rangle$, where A_i is # codewords of weight i.

- MacWilliams Identity determines weight distribution of code from weight distribution of its dual.
- Quite magical.
- Many nice consequences.

MacWilliams Identities

Thm:
- Let $A_0, \ldots, A_n$ wt. dist. of $\mathcal{C}$.
- Let $A'_0, \ldots, A'_n$ wt. dist. of $\mathcal{C}^\perp$.
- Let $W(y) = \sum_i A_i y^i$.
- Let $W'(y) = \sum_i A'_i y^i$.
- Then $W'(y) = \frac{(1+(q-1)y)^n}{|\mathcal{C}|} W\left(\frac{1-y}{1+(q-1)y}\right)$.

- Implications: Equating coefficients of y^i, get $n+1$ linear equations in $2(n+1)$ variables.
- Natural use, gives weight distribution of primal given dual or vice-versa.
- Interesting use: Can compute weight distribution of MDS codes!

MacWilliams Identities: Proof

(Will only do the Binary case)

Defn: The verbose generating function

(a) The generating function of a bit:
$W_b(x, y) = (1-b)x + by$

(b) The generating function of a word:
$W_c(x_1, y_1, \ldots, x_n, y_n) = \prod_{i=1}^{b} W_{c_i}(x_i, y_i)$

(c) The generating function of a code:
$$W_{\mathcal{C}}(x_1, y_1, \ldots, x_n, y_n) = \sum_{c \in \mathcal{C}} W_c(x_1, y_1, \ldots, x_n, y_n)$$

E.g. if $\mathcal{C} = \{000, 011, 101, 110\}$, then
$$W_{\mathcal{C}}(x_1, y_1, x_2, y_2, x_3, y_3) = x_1x_2x_3 + x_1y_2y_3 + y_1x_2y_3 + y_1y_2x_3$$

MacWilliams Identities (contd).

Trivial Claim: Given $W_{\mathcal{C}}$, can compute $W_{\mathcal{C}^\perp}$.

Explicit version: (non-trivial)
$$W_{\mathcal{C}}(x_1+y_1, x_1-y_1, \ldots, x_n+y_n, x_n-y_n) = |\mathcal{C}| \cdot W_{\mathcal{C}^\perp}(x_1, y_1, \ldots, x_n, y_n)$$

Proof steps:

Bit case:
$$W_{b'}(x+y, x-y) = \sum_{b \in \{0,1\}} (-1)^{\langle b, b' \rangle} W_b(x, y).$$

Vector case:
$$W_c(x_1+y_1, x_1-y_1, \ldots, x_n+y_n, x_n-y_n) = \sum_{b \in \{0,1\}^n} (-1)^{\langle b, c \rangle} W_b(x_1, y_1, \ldots, x_n, y_n).$$

Proof (contd).

Code case:

$$W_{\mathcal{C}}(x_1+y_1, x_1-y_1, \ldots, x_n+y_n, x_n-y_n)$$

$$= \sum_{c\in\mathcal{C}} \sum_{b\in\{0,1\}^n} (-1)^{\langle b,c\rangle} W_b(x_1,y_1,\ldots,x_n,y_n)$$

$$= \sum_{b\in\{0,1\}^n} W_b(x_1,y_1,\ldots,x_n,y_n) \sum_{c\in\mathcal{C}} (-1)^{\langle b,c\rangle}$$

$$= |\mathcal{C}| \cdot W_{\mathcal{C}^\perp}(x_1,y_1,\ldots,x_n,y_n)$$

MacWilliams Identity follows using:

$$(1+y)^n W\left(\frac{1-y}{1+y}\right) = W_{\mathcal{C}}(1+y,1-y,\ldots,1+y,1-y)$$

$$\text{and } W'(y) = W_{\mathcal{C}^\perp}(1,y,\ldots,1,y)$$

MDS Codes

Fact: Dual of MDS code is MDS.

Proof: Along lines of Singleton bound.

Fact: MDS code of dim k has $(q-1)\binom{n}{k}$ codewords of minimum weight.

Proof: By inspection.

Consequence: Have values for $n+1$ variables out of $2(n+1)$ used in M.I. System turns out to have full rank.

Thm: $\#$ poly of degree $< k$ with w non-zero evaluations at n points is:

$$\binom{n}{w} \sum_{j=0}^{w+k-n} (-1)^j \binom{w}{j} (q^{w+k-n-j} - 1)$$

.

LP bound

- One more bound in literature.
- Strongest known bound.
- Analysis hard.
- So hard, one only has upper bounds on the LP bound.
- Current upper bound on LP bound is still far from random code or AG-code (so may not be optimal either).
- Will see LP later.
- However (only) bound proving that if $d = (\frac{1}{2} - \epsilon)n$, then $n = O(k/\epsilon^2)$. (Matches random code for small ϵ.)

LP bound

- Let $A_0, \ldots, A_n$ be dist. of $[n, ?, d]_q$ code.
- $\#$ codewords $= A_0 + \cdots + A_n$.
- Know $A_0 = 1$, $A_1 = \cdots = A_{d-1} = 0$.
- Further $A'_0 = 1, A'_1, \ldots, A'_n \geq 0$.
- How large can $A_0 + \cdots + A_n$ be under above conditions?
- Above is a linear program ... Gives best known bound [MRRW].
- Note: Extends to non-linear codes also. Define $A_i = \mathbf{E}_{c\in\mathcal{C}}[|S(c,i) \cap \mathcal{C}|]$, $S(c,i) =$ sphere of radius i around c.

Alon's proof for ϵ-biased spaces

Thm: Suppose have binary code with K codewords of length n s.t. no two are have distance less than $(\frac{1}{2} - \epsilon)n$ or greater than $(\frac{1}{2} + \epsilon)n$: Then $K \leq 2n$, provided $\epsilon \leq \frac{1}{2\sqrt{n}}$.

Proof:

- Map 0 to 1 and 1 to -1, and normalize so that vectors have unit norm.
- Then inner products lie between -2ϵ and 2ϵ.
- Let M be $K \times K$ matrix of inner products.
- M close to identity matrix and hence has rank close to that of identity matrix. Specifically: rank $\geq \frac{K}{1+4(K-1)\epsilon^2}$.
- On the other hand, rank$(M) \leq n$.