Delsarte's linear programming bound

Jason Li

15-859 Coding Theory, Fall '14

December 5, 2014

- For all n, q, and d, Delsarte's linear program establishes a series of linear constraints that every code in $\mathbb{F}_q^n$ with distance d must satisfy.
- We want to maximize the size of the code, subject to these linear constraints.
- Together, the constraints and the objective function form a linear program.
- Solving this linear program gives an upper bound on the size of a code in $\mathbb{F}_q^n$ with distance d.

Contents

A symmetric association scheme $A = \{X, \mathcal{R}\}$ is a finite set X and a set of relations $\mathcal{R} = \{R_0, R_1, \ldots, R_d\}$ on X such that the R_i satisfy:

- $R_0 = \{(x, x) : x \in X\}$
- If $(x, y) \in R_i$, then $(y, x) \in R_i$. (This condition is weaker in asymmetric association schemes.)
- $\mathcal{R}$ partitions $X \times X$.
- Fix values $h, i, j \in [0, d]$, and consider the relations R_h, R_i, and R_j. For each $(x, y) \in R_h$, the number of elements $z \in X$ such that $(x, z) \in R_i$ and $(z, y) \in R_j$ is always the same, regardless of (x, y).

- We can think of X as the vertices of a graph, and the values of (x, y) are the (undirected) edges of the graph. (Note that (x, x) is allowed, so the graph has self-loops.)
- We can think of the relations $R_0, \ldots, R_d$ as $d + 1$ distinct colors. If an edge (x, y) is in R_i, then we color the edge (x, y) by the color of R_i.
- Since $\{R_0, \ldots, R_d\}$ partitions $X \times X$, we know that each edge is colored exactly one color.

- Recall the following condition:
  - Fix values $i, j, k \in [0, d]$, and consider the relations R_i, R_j, and R_k. For each $(x, y) \in R_i$, the number of elements $z \in X$ such that $(x, z) \in R_j$ and $(z, y) \in R_k$ is always the same, regardless of (x, y).

  Think of the edges (x, y), (x, z), and (z, y) as a triangle in the graph. Then, the condition becomes the following:
  - If we consider all triangles (x, y, z) with $(x, y) \in R_h$, $(x, z) \in R_i$, and $(z, y) \in R_j$, then every edge $(x, y) \in R_h$ takes part in the same number of triangles.

The association scheme that we are interested in is the Hamming scheme. Consider the vector space $\mathbb{F}_q^n$. Our set of elements X will be all coordinates in $\mathbb{F}_q^n$. Then, the Hamming scheme is defined as follows:

- There are $n+1$ relations $R_0, \ldots, R_n$, which correspond to Hamming distances between pairs of points.
- For two coordinates $x, y \in \mathbb{F}_q^n$, (x, y) belongs to the relation indexed by the Hamming distance of x and y. That is, $(x, y) \in R_{\Delta(x,y)}$.

Hamming scheme
... is an association scheme

Let us check that the Hamming scheme satisfies the conditions for a symmetric association scheme.

- $R_0 = \{(x, x) : x \in X\}$
  - Satisfied because $\Delta(x, y) = 0 \Leftrightarrow x = y$.
- If $(x, y) \in R_i$, then $(y, x) \in R_i$.
  - Satisfied because Hamming distance is symmetric.
- $\mathcal{R}$ partitions $X \times X$.
  - Satisfied by definition.
- Fix values $h, i, j \in [0, d]$, and consider the relations R_h, R_i, and R_j. For each $(x, y) \in R_h$, the number of elements $z \in X$ such that $(x, z) \in R_i$ and $(z, y) \in R_j$ is always the same, regardless of (x, y).
  - Intuitively, this is true because the Hamming distance is unaffected by coordinate shifts.

In an association scheme with set X and relations $R_0, \ldots, R_d$, we can define one associate matrix A_i for each R_i as follows:

- Each A_i has rows and columns indexed by elements in X. (So each A_i is an $|X|$-by-$|X|$ matrix.)
- Entry (x, y) of A_i is 1 if $(x, y) \in R_i$, and 0 otherwise.

Consider the Hamming scheme on $\mathbb{F}_2^3$, indexed by $[000, 001, 010, 011, 100, 101, 110, 111]$. We can easily check that

$$A_0 = \begin{bmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&1 \end{bmatrix}, A_1 = \begin{bmatrix} 0&1&1&0&1&0&0&0\\ 1&0&0&1&0&1&0&0\\ 1&0&0&1&0&0&1&0\\ 0&1&1&0&0&0&0&1\\ 1&0&0&0&0&1&1&0\\ 0&1&0&0&1&0&0&1\\ 0&0&1&0&1&0&0&1\\ 0&0&0&1&0&1&1&0 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} 0&0&0&1&0&1&1&0\\ 0&0&1&0&1&0&0&1\\ 0&1&0&0&1&0&0&1\\ 1&0&0&0&0&1&1&0\\ 0&1&1&0&0&0&0&1\\ 1&0&0&1&0&0&1&0\\ 1&0&0&1&0&1&0&0\\ 0&1&1&0&1&0&0&0 \end{bmatrix}, A_3 = \begin{bmatrix} 0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&1&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 1&0&0&0&0&0&0&0 \end{bmatrix}.$$

The associate matrices have several nice properties:

- $A_0 = I$, since R_0 only has elements of the form (x, x)
- $\sum_{i=0}^{d} A_i$ is the all-ones matrix, since R_i partition $X \times X$
- If we multiply two matrices A_i and A_j, then we get a linear combination of A_h for $h \in [0, d]$.
  - In particular, $A_j A_i = \sum_{h=0}^{d} p_{i,j}^h A_h$, where $p_{i,j}^h$ is the number of triangles with one edge in $(x, y) \in R_h$ and other two edges in R_i, R_j. This is easily verified.
- From above, since $p_{i,j}^h = p_{j,i}^h$, we get that $A_j A_i = A_i A_j$, so the matrices are commutative.
- If we think of the matrices as a vector space, then the A_i are linearly independent.
  - because each of the $|X|^2$ entries is 1 in exactly one A_i.

Recall the following property, which is perhaps the most important:

- If we multiply two matrices A_i and A_j, then we get a linear combination of A_h for $h \in [0, d]$.

An algebra is a vector space equipped with a bilinear product.

- The matrices A_i are a basis for a vector space of matrices.
- Moreover, multiplying any two A_i and A_j results in a linear combination of the A_h, which is again an element of the vector space.

Therefore, the vector space spanned by A_i forms an algebra over the matrices, called the Bose-Mesner algebra.

It turns out that the Bose-Mesner algebra always has another basis of pairwise "orthogonal" matrices. Specifically, the vector space spanned by $A_0, \ldots, A_d$ has another basis $E_0, \ldots, E_d$ such that

- $E_i E_j$ is the zero matrix if $i \neq j$
- $E_i^2 = E_i$ (such matrices are called idempotent.)

This is analogous to the spectral theorem of linear algebra.

Define the $(d+1)$-by-$(d+1)$ matrices P and Q as follows:

- The entries of P satisfy $A_i = \sum_{j=0}^{d} P_{ji} E_j$.
- The entries of Q satisfy $E_i = \frac{1}{|X|} \sum_{j=0}^{d} Q_{ji} A_j$.

P is called the first eigenmatrix, and Q is the second eigenmatrix. They are essentially change-of-basis matrices from the basis $A_0, \dots, A_d$ to the basis $E_0, \dots, E_d$ and back.

If we instead think of the A_i and E_i as individual entries of a matrix (i.e. pretend that they are numbers), then the definitions can be more concisely written as

- $$\begin{bmatrix} A_0 & A_1 & \cdots & A_d \end{bmatrix} = \begin{bmatrix} E_0 & E_1 & \cdots & E_d \end{bmatrix} \cdot \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ P_{*,0} & P_{*,1} & \cdots & P_{*,d} \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix}$$
- $$\begin{bmatrix} E_0 & E_1 & \cdots & E_d \end{bmatrix} = \frac{1}{|X|} \cdot \begin{bmatrix} A_0 & A_1 & \cdots & A_d \end{bmatrix} \cdot \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ Q_{*,0} & Q_{*,1} & \cdots & Q_{*,d} \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix}$$

If we combine the two equations, then we can see that

- $$\begin{bmatrix} E_0 & E_1 & \cdots & E_d \end{bmatrix} = \frac{1}{|X|} \cdot \left(\begin{bmatrix} E_0 & E_1 & \cdots & E_d \end{bmatrix} \cdot P \right) \cdot Q$$

Since the E_i are linearly independent, we must have

- $P \cdot Q = |X| \cdot I$ (where I is the $(d+1)$-by-$(d+1)$ identity matrix)

- Even for the Hamming scheme, computing the eigenmatrices P and Q is highly non-trivial. Delsarte [Del '73] showed that the eigenmatrix Q for the Hamming scheme on $\mathbb{F}_q^n$ can be represented in terms of the Krawtchouk polynomials, which are defined as:
  - $K_k(x) = \sum_{i=1}^{k} \binom{x}{i} \binom{n-x}{k-i} (-1)^i (q-1)^{k-i}$
- In particular, $Q_{i,k} = K_k(i)$. (This is a highly non-trivial result.)
- As an example, for the Hamming scheme in $\mathbb{F}_2^3$, $Q = \begin{bmatrix} 1 & 3 & 3 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -3 & 3 & -1 \end{bmatrix}.$

- Let $(X, \mathcal{R})$ be an association scheme and let Y be a subset of X.
- The distribution vector of Y is a vector $\mathbf{a}$ of length $d+1$ such that $a_i = \dfrac{|(Y \times Y) \cap R_i|}{|Y|}$.
- In graph notation, we can think of the subgraph induced by the vertices in Y. Then, a_i is simply the average degree of a vertex in Y, where only edges in R_i are considered.
- We can easily see that $\sum_{i=0}^{d} a_i = |Y|$.

- Let us consider the Hamming scheme again, where $X = \mathbb{F}_q^n$.
- Consider a code in $\mathbb{F}_q^n$. We can let Y be the set of codewords.
- Now consider the distribution vector **a**. Recall that a_i is the average degree of a vertex in the subgraph induced by Y, where only edges in R_i are considered. What can we deduce about **a**?
- We know that $a_i \geq 0$ and that $\sum_{i=0}^{d} a_i = |Y|$.
- We can also easily see that $a_0 = 1$.
- Suppose, in addition, the code has distance r. Then, we also know that $a_1 = a_2 = \cdots = a_{r-1} = 0$.
- There is one more key property, which we prove next.

Here is the key theorem on distribution vectors that forms the basis for the linear programming bound.

- Theorem: If $\mathbf{a}$ is a distribution vector of a subset Y of an association scheme with second eigenmatrix Q, then $\mathbf{a}Q \geq \mathbf{0}$. (That is, the vector $\mathbf{a}Q$ has only non-negative entries.)
- Proof:
  - Let $\mathbf{y}$ be the characteristic vector of Y. That is, $y_x = 1$ if $x \in Y$, and 0 otherwise. Then,
    - $a_i = \dfrac{\mathbf{y}A_i\mathbf{y}^T}{|Y|}$.
  - It follows that
    - $0 \leq \|\mathbf{y}E_i\|^2 = (\mathbf{y}E_i)(\mathbf{y}E_i)^T = \mathbf{y}E_iE_i^T\mathbf{y}^T = \mathbf{y}E_i\mathbf{y}^T$, where the last step is true because E_i is idempotent and symmetric.

- Proof (continued):
  - Recall that
$$E_i = \frac{1}{|X|}\sum_{j=0}^{d} Q_{ji}A_j \text{ and } a_i = \frac{\mathbf{y}A_i\mathbf{y}^T}{|Y|}.$$
  - Therefore,
    - $0 \leq \mathbf{y}E_i\mathbf{y}^T = \frac{1}{|X|}\mathbf{y}\left(\sum_{j=0}^{d} Q_{ji}A_j\right)\mathbf{y}^T = \frac{1}{|X|}\left(\sum_{j=0}^{d} Q_{ji}\mathbf{y}A_j\mathbf{y}^T\right) =$
$$\frac{|Y|}{|X|}\sum_{j=0}^{d} a_jQ_{ji} = \frac{|Y|}{|X|}(\mathbf{a}Q)_i.$$
  - So for each i, $(\mathbf{a}Q)_i \geq 0$, as desired. □

- Let us collect all of the conditions that **a** must satisfy:
  - $a_0 = 1$.
  - $a_i = 0$ for $1 \leq i < r$.
  - $a_i \geq 0$ for $r \leq i \leq n$.
  - $\mathbf{a}Q \geq \mathbf{0}$. (This introduces $d + 1$ linear inequalities.)
- At the end, we know that $\sum_{i=0}^{d} a_i = |Y|$. Therefore, to upper bound the set Y of codewords, our objective of the linear program is to maximize $\sum_{i=0}^{d} a_i$.
- That's it for Delsarte's linear program.

A nice property that merits its own slide:

- The linear programming bound works for all codes, not just linear codes. This is because we make no assumption on the set $Y \subseteq X$.

The linear programming bound
Comparison to Hamming bound

For fixed n and q, we can numerically solve the linear program to find the upper bound for codes in $\mathbb{F}_n^q$. Here is a table comparing the Hamming bound and the LP bound for codes in $\mathbb{F}_2^n$ with distance δ.

- Note that the tables suggest that the LP bound is always at most the Hamming bound. This is in fact true: Delsarte [Del '73] showed how to establish the Hamming bound using the LP bound, so the LP bound is always at least as strong.
- Also note the perfect code with $n = 15 = 2^4 - 1$ and $\delta = 3$. As expected, both bounds achieve this perfect code.

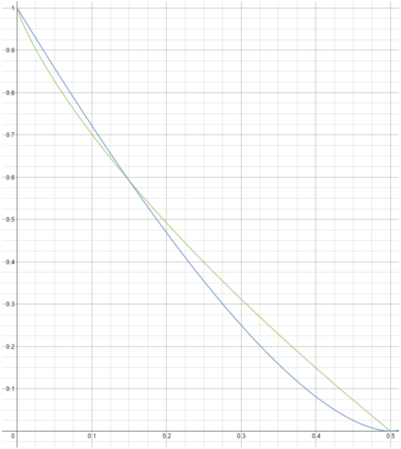
n	δ	Hamming Bound	Linear Programming Bound
11	3	170.7	170.7
11	5	30.6	24
11	7	8.8	4
12	3	315.1	292.6
12	5	51.9	40
12	7	13.7	5.3
13	3	585.1	512
13	5	89.0	64
13	7	21.7	8
14	3	1092.3	1024
14	5	154.6	128
14	7	34.9	16
15	3	2048	2048
15	5	270.8	256
15	7	56.9	32

- What about for higher n? We would like to find asymptotics for the linear programming bound.
- For the rest of this talk, we will focus only on *binary* codes.

- Let $A(n, \lfloor \delta n \rfloor)$ be the maximum size of a binary code with length n and distance δn. We can define the function $R(\delta) = \limsup_{n \to \infty} \frac{\log_2 A(n, \lfloor \delta n \rfloor)}{n}$. Intuitively, this is an asymptotic measure of the best rate possible for a binary code.
- Similarly, let $A_{LP}(n, \lfloor \delta n \rfloor)$ to be the maximum value of $\sum_{i=0}^{d} a_i$ for some $\mathbf{a}$ that satisfies Delsarte's linear program. Since the LP bound is an upper bound, we have $A(n, \lfloor \delta n \rfloor) \leq A_{LP}(n, \lfloor \delta n \rfloor)$.
- We can also define $R_{LP}(\delta) = \limsup_{n \to \infty} \frac{\log_2 A_{LP}(n, \lfloor \delta n \rfloor)}{n}$. We want bounds on $R_{LP}(\delta)$, which is an upper bound for $R(\delta)$.

Asymptotics: upper bound
for the linear programming bound

- We are most interested in an upper bound for $R_{LP}(\delta)$, since this will also be an upper bound for $R(\delta)$.
- McEliece, Rodemich, Rumsey, and Welch [MRRW '77] showed that $R_{LP}(\delta) \le H(\frac{1}{2} - \sqrt{\delta(1-\delta)})$. **This is the best bound known for $\delta \ge 0.273$.**
- Here is a plot of this upper bound with the Elias-Bassalygo bound $R(\delta) \le$ $$1 - H\left(\frac{1-\sqrt{1-2\delta}}{2}\right),$$ which we saw in class:



Asymptotics: upper bound
for the linear programming bound

- Navon and Samorodnitsky [NS '05] established a simpler proof of the same bound, $R_{LP}(\delta) \leq H(\frac{1}{2} - \sqrt{\delta(1-\delta)})$.
- Their method was to construct feasible solutions to the *dual* of Delsarte's linear program, which can be formulated as:
  - minimize $(Q\mathbf{b})_0$, given the constraints
    - $\mathbf{b} \geq \mathbf{0}$.
    - $b_0 = 1$.
    - $(Q\mathbf{b})_i \leq 0$ for $d \leq i \leq n$.
- By linear programming duality, the minimum of the dual equals the maximum of the primal, so any feasible solution to the dual is an upper bound of the optimum $A_{LP}(n, d)$.
- Their construction uses Fourier analysis on $\mathbb{Z}_2^n$.
  - Unfortunately, Fourier analysis is not as nice on $\mathbb{Z}_q^n$ for $q > 2$, so their construction does not generalize to arbitrary q.

- We might also be interested in a lower bound for $R_{LP}(\delta)$.
- A lower bound for $R_{LP}(\delta)$ gives a better measure of how powerful the LP bound actually is. It is essentially a cap on the strength of the bound.
- Navon and Samorodnitsky [NS '05] showed the lower bound $R_{LP}(\delta) \geq \frac{1}{2}H(1-2\sqrt{\delta(1-\delta)})$, which is currently the best known.
- Here is a plot of the lower bound with the upper bound.

Open problems

- Improved lower bounds for binary codes
  - Delsarte's linear program provides asymptotic upper bounds that are the best for $\delta \geq 0.273$. Therefore, any improvement to the upper bound with δ in this range improves upon the best known upper bound.
  - While the lower and upper bounds of [NS '05] converge as $\delta \to \frac{1}{2}$, there is a large gap for smaller δ. This allows for improvement of at least one of the bounds.
- Asymptotic bounds for $q > 2$:
  - The linear programming bound has provided some of the best asymptotic bounds for $q = 2$. Unfortunately, these techniques do not generalize for arbitrary q.
  - However, Delsarte's linear program generalizes to all q.
  - Therefore, an open question remains to show good asymptotic bounds for $R_{LP}(\delta)$ for $q > 2$.

The end

- References:

  - Delsarte, P.: An algebraic approach to the association schemes of coding theory. Philips Res. Rep., Suppl. 10 (1973)
  - McEliece, R.J., Rodemich, E.R., Rumsey, H. Jr., Welch, L.R.: New upper bounds on the rate of a code via the DelsarteMacWilliams inequalities. IEEE Trans. Inf. Theory IT-23, 157166 (1977) of FOCS 46
  - McKinley, S.: The Hamming Codes and Delsarte's Linear Programming Bound, available at `http://www.mth.pdx.edu/~caughman/thesis.pdf`.
  - Navon, M., Samorodnitsky, A.: On Delsarte's linear programming bounds for binary codes. In: Proceedings

- Thank you for your attention. :)