Packing, coding, and ground states II

Henry Cohn

August 25, 2020

Let's look at a discrete world.

$X = \{0,1\}^n$

<u>Hamming</u> <u>distance</u>

$d_H(x,y) = \#\{i : x_i \neq y_i\}$

<u>weight</u> $w(x) = d_H(0,x)$

This is the setting for binary error-correcting codes.

Keep codewords far apart, so they can't easily be confused.

Isometry group

$$G = S_n \ltimes (\mathbb{Z}/2\mathbb{Z})^n$$

permute coords

translation

2-point homogeneous:

for each i, G acts transitively on

$$\{(x,y) \in (\{0,1\}^n)^2 : d_H(x,y) = i\}.$$

Distance is the only invariant for a pair of points.

Distance distribution

For $\mathcal{C} \subseteq \{0,1\}^n$, let

$$A_i = \frac{1}{|\mathcal{C}|} \#\{ (x,y) \in \mathcal{C}^2 : d_H(x,y) = i \}.$$

($\cong$ pair correlation function in physics)

Key question: How can we characterize $(A_0, \ldots, A_n)$ coming from actual codes?

Why does distance distribution matter?

- It determines the minimal distance d via

$$A_1 = A_2 = \cdots = A_{d-1} = 0.$$

- It determines energy under a pair potential:

$$\frac{1}{|\mathcal{C}|} \sum_{\substack{x,y \in \mathcal{C} \\ x \neq y}} f(d_H(x,y)) = \sum_{i=1}^{n} A_i f(i).$$

- It determines combinatorial design properties (not discussed today).

Energy:

$$E_f(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{\substack{x,y \in \mathcal{C} \\ x \neq y}} f(d_H(x,y)) = \sum_{i=1}^{n} A_i f(i).$$

- Interacting particles in a discrete model of physics. What's their ground state (min. energy given $|\mathcal{C}|$)? Do they crystallize?

- Information theory:

  <u>binary symmetric channel</u>

  each bit flipped independently with probability p

If we send a random elt of $\mathcal{C}$, what's the chance of <u>undetected error</u>? (I.e., receive a different element of $\mathcal{C}$.)

send x
$d_H(x,y) = i$ $\Longrightarrow$ probability of receiving y is $p^i (1-p)^{n-i}$

Let $f(i) = \left(\frac{p}{1-p}\right)^i \cdot (1-p)^n$. Then

$$\boxed{E_f(\mathcal{C}) = \text{prob. of undetected error.}}$$

Goal: choose $\mathcal{C}$ to minimize $E_f(\mathcal{C})$.

$$f(i) = \left(\frac{p}{1-p}\right)^i (1-p)^n$$

<u>Goal</u>: minimize $E_f(\mathcal{C})$
(given $|\mathcal{C}|$, f)

Does the answer depend on p?
Sometimes yes: the optimal code depends on the channel.

But some famous codes (e.g., Hamming or Golay) minimize $E_f(\mathcal{C})$ for all such f, and even more potential functions.

Big Question

How do we distinguish between order and chaos?

When are optimal codes highly structured and symmetric (like crystals), and when are they disordered or pseudorandom?

Probably not answerable in full generality.

Coding theory:

Asymptotically, the best codes known are obtained by probabilistic method.

Many good small codes are constructed algebraically.

Materials science:

What's the difference between crystals and dirt?.

Delsarte inequalities

linear constraints on
$A_0, A_1, \dots, A_n$.

Obvious: $A_0 = 1$
$A_i \geq 0$ for $1 \leq i \leq n$

$$A_0 + \cdots + A_n = |C|$$

Non obvious: Delsarte 1972.

Can be done in very elementary way, but we'll put in context of representations of G.

Let's study functions on $X = \{0,1\}^n$.

$$L^2(X) = \{ f : X \to \mathbb{C} \}$$

(square integrability is automatic since X is finite)

How does $G = S_n \ltimes (\mathbb{Z}/2\mathbb{Z})^n$ act on $L^2(X)$?

Start w/ $(\mathbb{Z}/2\mathbb{Z})^n$:

$$L^2(X) = \bigoplus_{y \in \{0,1\}^n} \mathbb{C}\chi_y$$

where $\chi_y : \{0,1\}^n \to \mathbb{C}$ is defined by $\chi_y(x) = (-1)^{\langle x, y \rangle}$ (characters of $(\mathbb{Z}/2\mathbb{Z})^n$).

$$L^2(X) = \bigoplus_{y \in \{0,1\}^n} \mathbb{C}\chi_y$$

representations of $(\mathbb{Z}/2\mathbb{Z})^n$

S_n just permutes the coord's of y.

representations of G

$$L^2(X) = \bigoplus_{j=0}^{n} V_j, \quad \text{where}$$

$$V_j = \bigoplus_{\substack{y \in \{0,1\}^n \\ w(y) = j}} \mathbb{C}\chi_y.$$

$$V_j = \bigoplus_{\substack{y \in \{0,1\}^n \\ w(y) = j}} \mathbb{C} \chi_y$$

Normalize G-inv't inner product $\langle \cdot , \cdot \rangle$ on V_i so that χ_y orthonormal.

There's a unique S_n-inv't vector, up to scaling:

$$\sum_{w(y)=j} \chi_y .$$

Given a representation V of G, an H-fixed vector $v \in V$ gives a G-equivariant map

$$\begin{aligned} G/H &\longrightarrow V \\ g &\longmapsto gv. \end{aligned}$$

In V_j, our S_n-fixed vector gives the map

$$\{0,1\}^n \xrightarrow{\varphi_j} V_j$$

$$x \longmapsto \sum_{w(y)=j} (-1)^{\langle x,y \rangle} \chi_y$$

Now we can use the fundamental inequality

$$\left| \sum_{x \in \mathcal{C}} \varphi_j(x) \right|^2 \geq 0 .$$

i.e.,

$$\sum_{x,y \in \mathcal{C}} \langle \varphi_j(x), \varphi_j(y) \rangle \geq 0 .$$

Note: $\langle \varphi_j(x), \varphi_j(y) \rangle$ depends only on j and $d_H(x,y)$, by G-invariance.

Call it $K_j^n(d_H(x,y))$.

We have seen that

$$\sum_{x,y \in C} K_j^n(d_H(x,y)) \geqslant 0.$$

Equivalently,

$$\sum_{i=0}^{n} K_j^n(i) A_i \geqslant 0.$$

These are the <u>Delsarte inequalities</u>.

What is K_j^n? Krawtchouk polynomial.

$$K_j^n(i) = \sum_{w(y)=j} (-1)^{\langle x,y \rangle}$$

where $w(x) = i$.

$$= \sum_{k=0}^{j} (-1)^k \underbrace{\binom{i}{k}\binom{n-i}{j-k}}_{\text{count } y\text{'s based on overlap w/ } x}$$

polynomial of degree j in i

orthogonality inherited from χ_y:

$$\sum_{i=0}^{n} \binom{n}{i} K_j(i) K_k(i) = 0$$
if $j \neq k$.

Inequalities on $(A_0, A_1, \ldots, A_n)$

$$A_0 = 1$$

$$A_i \geq 0$$

$$A_0 + \cdots + A_n = |\mathcal{C}|$$

$$\sum_{i=0}^{n} K_j(i) A_i \geq 0$$

The <u>linear programming bound</u> is optimization subject to these inequalities.

Energy $E_f(\mathcal{C})$

Minimize $\displaystyle\sum_{i=1}^{n} A_i f(i)$

subject to Delsarte ineq's
(given $|\mathcal{C}|$, f).

Size of code given min. dist.

Maximize
$$|\mathcal{C}| = \sum_{i=0}^{n} A_i$$

given $A_1 = A_2 = \cdots = A_{d-1} = 0$

and Delsarte ineq's.

How good are these bounds?

We don't have a general sol'n.

Can solve for fixed parameters by linear programming.

Suboptimal bounds can be derived in general by ad hoc methods. (Best bounds known!)

Important unsolved problem.

When do LP bounds apply?

When pairwise distances within the code determine everything.

Packing: yes.

Covering: no.

(SDP bounds incorporate higher-order correlations.)

When are LP bounds sharp?

Usually not, but sharp in some noteworthy cases, such as Hamming and Golay codes.

Why these cases and not others? Some explanations known, but still somewhat mysterious.

$$\Delta f(i) = f(i+1) - f(i)$$

<u>Def.</u> $f : \{1, 2, \ldots, n\} \to \mathbb{R}$ is <u>completely monotonic</u> if

$$(-1)^k \Delta^k f(i) \geq 0$$

for $k \geq 0$ and $1 \leq i \leq n-k$.

I.e., decreasing, convex, etc.

Nicest class of potential functions.

"etc." is more natural than it seems.

__Def.__ $\mathcal{C} \subseteq \{0,1\}^n$ is __universally optimal__ if it minimizes E_f for each completely monotonic f.

All known cases of sharp LP bounds (aside from certain degenerate cases) are universally optimal (Cohn and Zhao, Ashikhmin and Barg).

Analogue of Cohn, Kumar, Miller, Radchenko, Viazovska for E_8, Λ_{24}.

Why does this happen?

Def. An N-point quasicode is a vector $(A_0, \dots, A_n)$ with $A_0 + \dots + A_n = N$ satisfying the Delsarte inequalities.

Maybe there is always a universally optimal quasicode for all n, N?
(Then the only question is whether it is attained by some code.)

True for $n \le 11$, all N.

False for $n = 12$, $24 < N < 40$.

???

Questions

- Why does universal optimality occur so often? Is there any (interesting) sharp case in which it fails?

- Hamming, Golay are analogous to E_8, Λ_{24}.

  Does Viazovska's approach have a discrete analogue?

Modular forms $\overset{?}{\longleftrightarrow}$ invariant polynomials for certain finite groups

$\{0,1\}^n$ is closely analogous to $\mathbb{R}^n$, with one exception so far:

(i.e., proved by LP bounds)

<u>Thm (Cohn and Zhao)</u>

If C is LP-universally optimal in $\{0,1\}^n$, then it remains universally optimal if any one point is removed from it.

Does this have a continuous analogue?

Exercises

- Prove that $\mathcal{C} \subseteq \{0,1\}^n$ minimizes E_g if and only if $\{0,1\}^n \setminus \mathcal{C}$ does. (Antiparticles!)

- What does LP duality tell us when applied to the linear programs we're studying?

- Prove $$\sum_{x,y \in \mathcal{C}} K_j^n(d_H(x,y)) \geq \binom{n}{j}$$ whenever $|\mathcal{C}|$ is odd.

- Prove the theorem from the previous slide. (Hint: what if you remove a point at random?)

Further reading

P. Delsarte, An algebraic approach to the association schemes of coding theory, 1973

P. Delsarte and V. I. Levenshtein, Association schemes and coding theory, 1998

(These papers deal with association schemes, a beautiful framework for LP bounds.)

H. Cohn and Y. Zhao, Energy-minimizing error-correcting codes, 2014