Lecture 6

MRRW Bound and Isoperimetric Problems

Feb 18, 2005
Lecturer: Nati Linial
Notes: Ethan Phelps-Goodman and Ashish Sabharwal

6.1 Preliminaries

First we recall the main ideas from the last lecture. Let

$$g = 1_C, \qquad f = \frac{g * g}{|C|}.$$

Then we can bound the code size $A(n, d)$ using Delsarte's linear program:

$$A(n,d) \leq \max_f \sum_{x \in \{0,1\}^n} f(x)$$

subject to

$$f \geq 0 \qquad f(\mathbf{0}) = 1$$
$$\widehat{f} \geq 0 \qquad f_{|1,\ldots,d-1} = 0$$

By averaging over a solution f, we can get an equivalent solution that is symmetric about permutations of the input bits. That is, we can assume w.l.o.g. that f that depends only on the hamming weight of the input. f is then determined by $n+1$ coordinate weights A_j by

$$A_j = \sum_{x \mid |x|=j} f(x)$$

Or equivalently,

$$f = \sum_{j=0}^{n} \frac{A_j}{\binom{n}{j}} 1_{L_j}$$

Central to our proof will be the Krawtchouk polynomials, which are related to our linear program by

$$\widehat{1}_{L_r} = K_r(x) = \sum_{j=0}^{r} (-1)^j \binom{x}{j}\binom{n-x}{r-j}$$

$$\widehat{f} = \sum_{j=0}^{n} \frac{A_j}{\binom{n}{j}} K_j$$

6.2 Primal and Dual Programs

Making the substitutions above we can now write Delsarte's program in terms of Krawtchouk polynomials and symmeterized f.

$$A(n,d) \leq \max_{A_0,\ldots,A_n} \sum_{i=0}^{n} A_i$$

subject to

$$\begin{aligned} & A_0 = 1 \\ & A_1, \ldots, A_{d-1} = 0 \\ \forall k \in \{0,\ldots,n\} \quad & \sum_{i=0}^{n} \frac{A_i}{\binom{n}{i}} K_i(k) \geq 0. \end{aligned}$$

This can be further simplified with the following identity for Krawtchouk polynomials.

Fact 6.1.

$$\frac{K_i(k)}{\binom{n}{i}} = \frac{K_k(i)}{\binom{n}{k}}$$

Proof.

$$\begin{aligned} \frac{1}{\binom{n}{i}} \sum_{j=0}^{i} (-1)^j \binom{k}{j}\binom{n-k}{i-j} &= \sum_{j}^{i} (-1)^j \frac{i!(n-i)!k!(n-k)!}{n!j!(k-j)!(i-j)!(n-k-i+j)!} \\ &= \frac{1}{\binom{n}{k}} \sum_{j}^{i???} (-1)^j \binom{i}{j}\binom{n-i}{k-j} \end{aligned}$$

□

Using this in the last constraint, and removing the $1/\binom{n}{k}$ term, which pulls out of the sum and doesn't affect the sign, we get the constraints

$$\forall k \in \{0,\ldots,n\} \quad \sum_{i=0}^{n} A_i K_k(i) \geq 0.$$

Our approach will be to use LP duality to give a bound on the maximum of this program. Recall that duality tells us that the maximum value of the primal is at most the minimum value of the dual. Strong duality states that the optima are exactly equal, but we will not use this.

Start by multiplying each of the $\sum_{i=0}^{n} A_i K_k(i) \geq 0$ constraints by β_k, and summing all of the constraints. This gives

$$\sum_{k=1}^{n} \beta_k \sum_{i=0}^{n} A_i K_k(i) = \sum_{i=0}^{n} A_i \sum_{k=1}^{n} \beta_k K_k(i) \geq 0$$

Let $\gamma(x) = \sum_{k=1}^{n} \beta_k K_k(i)$. If we add the constraint that $\forall x, \gamma(x) \leq -1$, then using $A_0 = 1, A_1, \ldots, A_{d-1} = 0$, we get

$$\sum_{i=0}^{n} A_i \gamma(i) = \gamma(x) + \sum_{i=d}^{n} A_i \gamma(i) \geq 0$$

$$\gamma(0) \geq -\sum_{i=d}^{n} A_i \gamma(i)$$

$$\geq \sum_{i=d}^{n} A_i$$

$$\gamma(0) + 1 \geq \sum_{i=1}^{n} A_i \geq A(n, d)$$

What we have done here is just an explicit construction of the dual. The reader can check that this dual can be arrived at by any standard method for computing the dual.

Let $\beta(x) = 1 + \sum_{k=1}^{n} \beta_k K_k(x)$. Then our final program is given by

$$A(n, d) \leq \min_{\beta_k} \beta(0)$$

subject to:

$$\forall k = 1, \ldots, n, \quad \beta_k \geq 0$$
$$\forall j = d, \ldots, n, \quad \beta(j) \leq 0$$

6.3 The Upper Bound

To show an upper bound on $A(n, d)$ we need to demonstrate a feasible solution β and bound $\beta(0)$. First we need two additional facts about Krawtchouk polynomials.

Fact 6.2 (Christoffel-Darboux). Let $P_1, P_2, \ldots$ be a family of orthonormal polynomials, and let a_i be the leading coefficient of P_i. Then

$$\frac{P_k(x)P_{k+1}(y) - P_{k+1}(x)P_k(y)}{y - x} = \frac{a_{k+1}}{a_k} \sum_{i=0}^{k} P_i(x)P_i(y)$$

For the case of Krawtchouk polynomials, the leading term of $K_r(x)$ is $\frac{-2^r}{r!}$. Also, to normalize we need to divide K_r by $\sqrt{\binom{n}{r}}$. Putting these together, we get

$$\frac{K_{r+1}(x)K_r(y) - K_r(x)K_{r+1}(y)}{y - x} = \frac{2}{r+1}\binom{n}{r} \sum_{i=0}^{r} \frac{K_i(x)K_i(y)}{\binom{n}{i}}.$$

The second fact we need is that the product of two Krawtchouk polynomials can be expressed as a non-negative combination of Krawtchouk polynomials.

Fact 6.3. For any p, q, there exist $\alpha_0, \ldots, \alpha_{p+q} \geq 0$ such that

$$K_p \cdot K_q = \sum_{j=0}^{p+q} \alpha_j K_j$$

This can be seen easily from the harmonic analysis perspective since $K_p \cdot K_q = \widehat{1}_{L_p} \cdot \widehat{1}_{L_q} = \widehat{1_{L_p} * 1_{L_q}}$, and the convolution is a positive combination.

We can now present the feasible solution for the dual. Let

$$\alpha(x) = \frac{\big(K_t(a)K_{t+1}(x) - K_{t+1}(a)K_t(x)\big)^2}{a - x}.$$

Then set $\beta(x) = \frac{\alpha(x)}{\alpha_0}$, where α_0 is chosen to make the constant term equal 1. Now we need to set values for a and t. Denote by $x_r^{(l)}$ the leftmost root of K_r. We know from last lecture that the roots of the Krawtchouk polynomials are real, lie in $[0, n]$, and interleave with one another. Therefore we can pick a t such that $0 < x_{t+1}^{(l)} < x_t^{(l)} < d$. In the region $(x_{t+1}^{(l)}, x_t^{(l)})$, K_{t+1} is negative and K_t is positive, so we can pick an a such that $K_t(a) = -K_{t+1}(a)$.

Now we need to show that $\alpha(x)$ satisfies the two constraints from the dual. First, note that at all $x > d$, $\alpha(x) < 0$. Then we just need to show that $\alpha(x)$ is non-negative combination of Krawtchouk polynomials. Using the above settings, and Christoffel-Darboux, we can factor $\alpha(x)$ as

$$\begin{aligned}
\alpha(x) &= \big(K_t(a)K_{t+1}(x) - K_{t+1}(a)K_t(x)\big)\left[\frac{K_t(a)K_{t+1}(x) - K_{t+1}(a)K_t(x)}{a - x}\right] \\
&= K_t(a)(K_{t+1}(x) + K_t(x))\left[\frac{K_t(a)K_{t+1}(x) - K_{t+1}(a)K_t(x)}{a - x}\right] \\
&= K_t(a)(K_{t+1}(x) + K_t(x))\left[\frac{2}{r+1}\binom{n}{r}\sum_{i=0}^{r}\frac{K_i(x)K_i(y)}{\binom{n}{i}}\right]
\end{aligned}$$

Since all terms are positive, this can be expanded as a positive combination of Krawtchouk polynomials.

Now that we have a feasible solution to the dual, we just need to find the value of $\beta(0)$. We can use the fact that for $t \approx \tau n$, the leftmost root is at $x_t^{(l)} = (1 + o(1))(\frac{1}{2} - \sqrt{\tau(1-\tau)})n$. Given this we can conclude that $R(\delta) \leq H(\frac{1}{2} - \sqrt{\delta(1-\delta)})$. Both the lecture and van Lint [1] seem to imply that this step is obvious, but your scribe has been unable to see any connection.

6.4 More on Isoperimetric Problems on the Cube

We now turn our attention to isoperimetric problems. In a previous lecture, we studied isoperimetric questions on the n-dimensional cube, namely the vertex isoperimetric problem and the edge isoperimetric problem. Why is the study of such problems important? The reason is that Computer Science deals with Boolean functions which are simply partitions of the n-dimensional cube into two parts. Understanding the geometry of the cube is therefore critical to understand Boolean functions. Here is one more isoperimetric problem that is open.

Open Problem 6.1 (Chung-Füredi-Grahan-Seymour, 1988 J.C.T.A.). What is the largest $d = d(n)$ such that for all $S \subseteq \{0,1\}^n, |S| > 2^{n-1}$, there exists $x \in S$ with $d_S(x) \geq d$?

Here $d_S(x)$ denotes the number of neighbors of x in S. Note that for $|S| \leq 2^{n-1}$, S can be an independent set, i.e., $\forall x \in S. d_S(x) = 0$. Further, for $|S| > 2^{n-1}$, S may not be independent. In general, all we know is that $d(n)$ is both $O(\sqrt{n})$ and $\Omega(\log n)$. This leaves a huge gap open.

Consider any Boolean function $f : \{0,1\}^n \to \{0,1\}$ represented as a $0,1$-labeling of the n-dimensional cube seen as a layered lattice. This lattice has four types of edges as depicted in Figure 6.1. Let $S = f^{-1}(0)$. The two edges from 0 to 1 and from 1 to 0 belong to the cut $E(S, S^c)$ and thus contribute to the cut size $e(S, S^c)$.

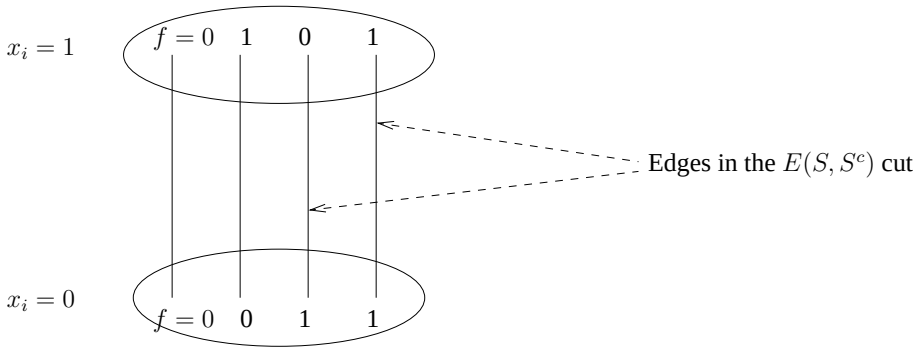

Figure 6.1: The cut defined in terms of the four types of edges in the lattice

If $|S| = 2^{n-1}$, then $e(S, S^c) \geq 2^{n-1}$. This is sharp for $S = \{x \mid x_1 = 0\}$. In the edge isoperimetry problem, given $|S|$, we want to minimize the cut size $e(S, S^c)$. What about trying to *maximize* the cut size instead? The maximum cut size can really be anything. Indeed, when f is the parity function, $e(S, S^c) = n2^{n-1}$.

6.4.1 Maximizing Edge Cut Size for Monotone Functions

Consider the setting of the previous section. How can we maximize the edge cut when f is *monotone*, i.e., $x \succ y \Rightarrow f(x) \geq f(y)$, where $x \succ y$ means $\forall i. x_i \geq y_i$? In the following, we use Parseval's identity to answer this question.

Theorem 6.1. *Let $S \subseteq \{0,1\}^n$ correspond to a monotone Boolean function $f : \{0,1\}^n \to \{0,1\}$. $f =$* majority *maximizes the edge cut size $e(S, S^c)$.*

Proof. It is clear from the lattice corresponding to $f =$ majority (see Figure 6.2) that the size of the cut corresponding to it is $\binom{n}{\lfloor n/2 \rfloor} = \Theta(\sqrt{n}\, 2^n)$. We will use Parseval's identity to prove that this is the optimal.

Let f be any monotone Boolean function in n dimensions. Recall that for characters $\chi_T(Z) = (-1)^{|Z \cap T|}$, the function f can be represented as $f = \sum_T \hat{f}(T)\chi_T$ where $\hat{f}(T) = \langle f, \chi_T \rangle$. What is $\hat{f}(\{i\})$?

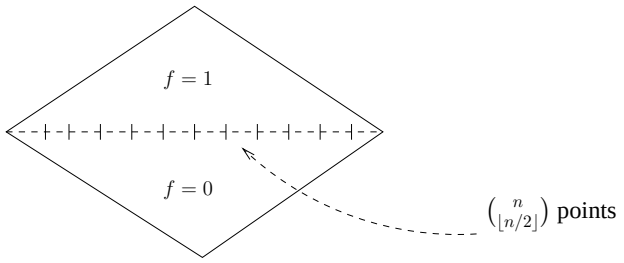

Figure 6.2: The lattice corresponding to the majority function

$\chi_{\{i\}}(Z) = (-1)^{|Z \cap \{i\}|}$ which is $+1$ if $i \notin Z$ and -1 if $i \in Z$. Therefore

$$\begin{aligned}
\hat{f}(\{i\}) &= \langle f, \chi_{\{i\}} \rangle \\
&= \frac{1}{2^n} \sum_Z f(Z) \chi_{\{i\}}(Z) \\
&= \frac{1}{2^n} \left(\sum_{Z \not\ni i} f(Z) - \sum_{Z \ni i} f(Z) \right) \\
&= -\frac{1}{2^n} \cdot \text{(number of cut edges in the } i\text{-direction)}
\end{aligned}$$

For ease of computation, convert everything from the $\{0,1\}$ basis to the $\{-1,+1\}$ basis. This quantity is then $(2/2^n)$ times the number of cut edges in the i-direction. Using Parseval's identity and Cauchy-Schwartz inequality, $1 = \|f\|_2^2 = \sum_S \left(\hat{f}(S)\right)^2 \geq \sum_i \left(\hat{f}(\{i\})\right)^2 \geq (1/n) \left(\sum_i \hat{f}(\{i\})\right)^2$. Hence $\sqrt{n} \geq \sum_i \hat{f}(\{i\}) = (2/2^n)\, e(S, S^c)$, which finishes the proof. □

We give an alternative *combinatorial* proof of the fact that $e(S, S^c) = 2^{n-1} \sum_i \hat{f}(\{i\})$ based on the following claim.

Claim 6.1. Let f be a monotone Boolean function. If the expectation of f is given and fixed, then to maximize $e(f^{-1}(0), f^{-1}(1))$, it is best to take f symmetric.

Proof of claim. Consider $\sum_{x:f(x)=0}(n - 2|x|)$. This is the sum of the first Krawtchouk polynomials and is equal to the cut size $e(f^{-1}(0), f^{-1}(1))$ because $(n - |x|)$ edges in the lattice corresponding to f that go upwards from x contributing $+1$ each while $|x|$ edges go downward from x contributing -1 each (see Figure 6.3. Maximizing this quantity means minimizing $\sum_{x:f(x)=0} |x|$ which happens exactly when f is "pushed down" as much as possible.

Formally, let us change the basis from $\{0,1\}$ to $\{-1,+1\}$ and reinterpret the summation. It is equal to $\sum_{x:f(x)=1}(n - 2|x|) - \sum_{x:f(x)=-1}(n - 2|x|) = 2^n \langle f, K_1 \rangle$. Observe however that $\sum_x (n - 2|x|) = \langle K_1, K_0 \rangle = 0$. Therefore $\sum_{x:f(x)=1}(n - 2|x|) = 2^{n-1} \langle f, K_1 \rangle$, which is the same as $\sum_i \hat{f}(\{i\})$ by the properties of Krawtchouk polynomials. □

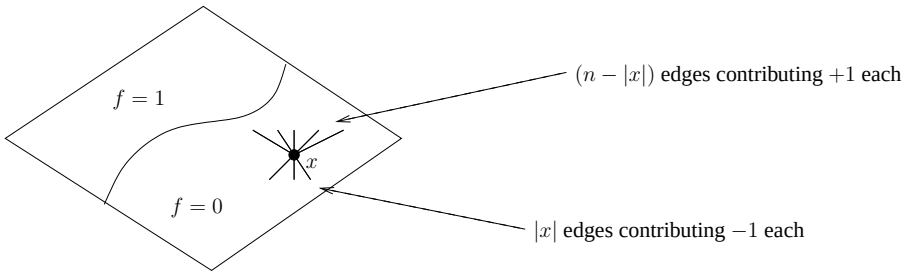

Figure 6.3: Contribution of f to the cut

6.4.2 The Brunn-Minkowski Inequality

Let v be a volume measure on subsets of $\mathbb{R}^n$.

Theorem 6.2 (Brunn-Minkowski [2]). *For A, B measurable subsets of $\mathbb{R}^n$,*

$$(v(A+B))^{1/n} \geq (v(A))^{1/n} + (v(B))^{1/n}.$$

Moreover, equality holds if and only if A and B are homothetic, i.e. $B = \lambda A + C$ for $\lambda \in \mathbb{R}$.

Here $A+B$ is the Minkowski sum defined as $\{a+b \mid a \in A, b \in B\}$, where $a+b$ is the standard vector sum over $\mathbb{R}^n$. For $\lambda \in \mathbb{R}$, λA is similarly defined as $\{\lambda a \mid a \in A\}$. We will not be using the second part of the theorem.

Let us try to understand what this inequality says. Take a convex body K in $\mathbb{R}^n$ and slide a hyperplane $A_t, t \in \mathbb{R}$, through it (see Figure 6.4). What can we say about the function $f(t) = \mu_{n-1}(A_t \cap K)$ which is the volume of the intersection of the body with the hyperplane? Brunn-Minkowski inequality says that $(f(t))^{1/(n-1)}$ is convex.

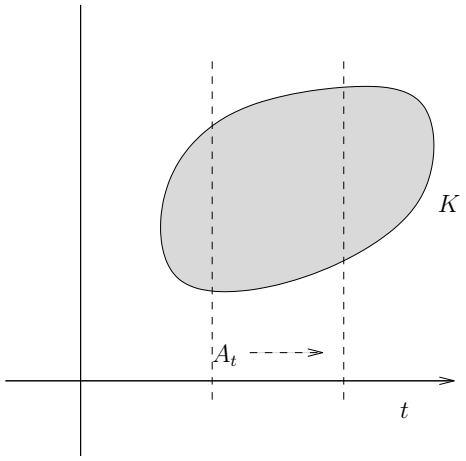

Figure 6.4: Sliding a hyperplane A_k through a convex body K

Theorem 6.3. *Brunn-Minkowski inequality implies the classical n-dimensional isoperimetric inequality.*

Proof. We want to show that if $K \subseteq \mathbb{R}^n$ and B is the unit ball in $\mathbb{R}^n$, then

$$\left(\frac{v(K)}{v(B)}\right)^{\frac{1}{n}} \leq \left(\frac{S(K)}{S(B)}\right)^{\frac{1}{n-1}}$$

where S denotes the surface area. For a 2-dimensional plane, the LHS equals $\sqrt{A/\pi}$ while the RHS equals $L/(2\pi)$. To prove LHS $\geq$ RHS, we need $L^2 \geq 4\pi A$, which we know to be true. Let's try to generalize this to higher dimensions.

The surface area of K is, by definition,

$$S(K) = \lim_{\varepsilon \to 0} \frac{v(K + \varepsilon B) - v(K)}{\varepsilon}.$$

By Brunn-Minkowski inequality,

$$\begin{aligned} S(K) &\geq \lim_{\varepsilon \to 0} \frac{\left((v(K))^{\frac{1}{n}} + \varepsilon \, (v(B))^{\frac{1}{n}} \right)^n - v(K)}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{n\varepsilon \, (v(K))^{\frac{n-1}{n}} \, (v(B))^{\frac{1}{n}} + O(\varepsilon^2)}{\varepsilon} \\ &= n \ (v(K))^{\frac{n-1}{n}} \, (v(B))^{\frac{1}{n}} \\ &= S(B) \ \left(\frac{v(K)}{v(B)}\right)^{\frac{n-1}{n}} \ \frac{n \, v(B)}{S(B)} \end{aligned}$$

The last term $n \, v(B)/S(B)$ is, however, always 1 in any number of dimensions. We have therefore proved the isoperimetric inequality. □

References

[1] J.H. van Lint. *Introduction to Coding Theory.* Springer, 1999.

[2] J. Matousek. *Lectures on Discrete Geometry.* Springer-Verlag, 2002.