ESTIMATING THE SHANNON CAPACITY OF A GRAPH

JOHN THICKSTUN

1. CHANNELS AND GRAPHS

Consider a stationary, memoryless channel that maps elements of discrete alphabets $\mathcal{X}$ to $\mathcal{Y}$ according to a distribution $p_{y|x}$. Recall the channel capacity C of $p_{y|x}$ is defined by

$$C \equiv \liminf_{n\to\infty} \frac{1}{n} \sup_{p_{x^n}} I(p_{y^n|x^n} p_{x^n}).$$

By the memoryless and stationary conditions respectively, this reduces to just

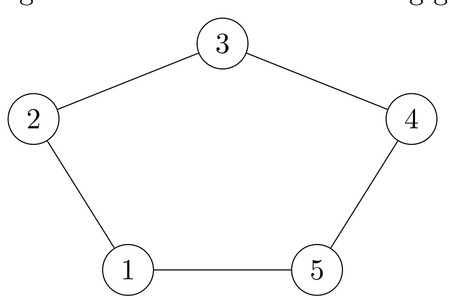
$$C = \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sup_{p_{x^n}} I(p_{y_i|x_i} p_{x_i}) = \liminf_{n\to\infty} \sup_{p_x} I(p_{y|x} p_x) = \sup_{p_x} I(p_{y|x} p_x).$$

Shannon's noisy channel theorem[1] asserts that this capacity is equivalent to the Shannon capacity: the supremum of achievable transmission rates on $p_{y|x}$.

To each discrete channel we will associate a graph $G = \{\mathcal{X}, E\}$. Vertices represent the input alphabet $\mathcal{X}$ and $x_1x_2 \in E$ iff for some y, $p_{y|x}(y|x_1) > 0$ and $p_{y|x}(y|x_2) > 0$. Intuitively, this captures the idea that x_1 and x_2 can be confused in the channel; G has an edge between x_1 and x_2 iff x_1 and x_2 are confusable. For example, consider a noisy channel defined by the following confusion matrix (i.e. transition probabilities):

$$p_{y|x} = \begin{pmatrix} .9 & 0 & 0 & 0 & .1 \\ .1 & .9 & 0 & 0 & 0 \\ 0 & .1 & .9 & 0 & 0 \\ 0 & 0 & .1 & .9 & 0 \\ 0 & 0 & 0 & .1 & .9 \end{pmatrix}.$$

We can associate this "pentagon" channel with the following graph:



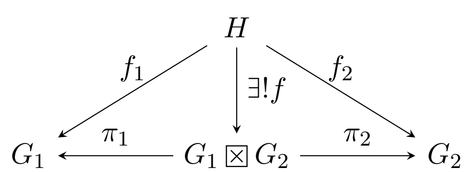
Suppose we attempt to transmit information over this channel, one symbol at a time. If we hope to infer x given y, we must restrict our alphabet $\mathcal{X}$ to a subset of symbols that cannot be confused with one another. That is, we must restrict our transmissions to a stable set in G. The largest such codebook is given by the stability number $\alpha(G)$, the size of a maximal stable set in G. Computing $\alpha(G)$ for a general graph is NP-hard, but in our example we easily see that $\alpha(G) = 2$, so one optimal codebook would be $\{1, 3\}$, which has bit rate 1.

Now suppose we transmit information over the channel in blocks of size two. Clearly we can transmit 4 symbols per block using the codewords $\{11, 13, 31, 33\}$. But we can do better. Consider the code $\{11, 23, 35, 42, 54\}$. It is routine to check that none of these codes can be confused with another. Our bit rate using this codebook will now be $.5 \log 5$; we are effectively transmitting an alphabet of size $\sqrt{5}$. Can we do better than $\sqrt{5}$ using larger blocks? This problem was first considered by Shannon[2]. We will show in section 4 that the answer is no.

This paper is organized as follows. In section two we will characterize the Shannon capacity of a channel in terms of the associated graph. Unfortunately, this quantity is difficult to analyze and its computational complexity is currently unknown. In section three we will introduce a graph invariant that upper bounds the Shannon capacity based on Lovasz's work[3]. We will apply this theory in section four to the pentagon channel described above. In section five we will turn our attention to the problem of computing the Lovasz invariant. In the final section we will consider an interpretation of our upper bound as an SDP relaxation of the stability number.

2. SHANNON CAPACITY OF A GRAPH

Consider the category of undirected, reflexive graphs **rGraph**. Note that the graphs discussed in section 1 belong to this category; by definition, any element of $\mathcal{X}$ can be confused with itself and therefore each vertex of the associated graph has a loop at each vertex. Define the strong product $\boxtimes$ to be the categorical product in **rGraph**. In particular suppose $G_1, G_2, H \in \textbf{rGraph}$, $f_1 \in \text{Hom}_{\textbf{rGraph}}(H, G_1)$, and $f_2 \in \text{Hom}_{\textbf{rGraph}}(H, G_2)$. Then the strong product is characterized by the following commutative diagram:



Therefore $V_{G_1 \boxtimes G_2} = V_{G_1} \times V_{G_2}$ and $e = uv = (i, k)(j, l) \in E_{G_1 \boxtimes G_2}$ iff $ij \in E_{G_1}$ and $kl \in E_{G_2}$.

The strong product of G_1 and G_2 exactly characterizes the confusable codewords of these graphs. Observe that a codeword $x_1 x_2$ passed through the channel $G_1(x_1)G_2(x_2)$ can be confused with $x_1' x_2'$ iff G_1 has edge $x_1 x_1'$ and G_2 has edge $x_2 x_2'$, which occurs iff $(x_1, x_2)(x_1', x_2')$ is an edge in $G_1 \boxtimes G_2$. By induction, two codewords of block size n are confusable in a graphical channel G iff there is an edge between them in $G^{\boxtimes n}$.

Recall that the Shannon capacity Θ of a channel is the supremum over all codes of the transmission rate. We saw in the previous section that the maximum size of a codebook transmitted over a graphical channel G is given by the stability number of G. And we saw above that transmitting a block over a graphical channel is equivalent to transmitting a symbol over the strong power graphical channel. Therefore the Shannon capacity of G is given by

$$\Theta(G) = \sup_{n \geqslant 1} \sqrt[n]{\alpha\left(G^{\boxtimes n}\right)}.$$

3. Bounding the Shannon capacity

The Shannon capacity is bounded below by $\alpha(G)$. Establishing an an upper bound on $\Theta(G)$ is more difficult. We will proceed to define the Lovasz number $\vartheta(G)$. The construction is somewhat unintuitive; the goal of the remainder of this section is to show that $\Theta(G) \leqslant \vartheta(G)$, which will give us a reason to care about ϑ. At the end of the paper, we will revisit the construction and attempt to offer some insight.

We say that an indexed set of unit vectors $U = (u_1, ..., u_n)$ together with a vector c (called the handle of U) is an orthonormal representation of a graph $G = ([n], E)$ iff $u_i^T u_j = 0$ whenever $ij \notin E$ and $\|c\| = 1$; i.e. we associate vectors with vertices such that vectors corresponding to non-adjacent vertices are orthogonal. Define the theta body $\mathrm{TH}(G)$[4] to be

$$\mathrm{TH}(G) \equiv \left\{ x \geqslant 0 \in \mathbb{R}^n : \sum_{i=1}^{n} (c^T u_i)^2 x_i \leqslant 1 \text{ for all orthonormal representations} \right\}$$

The Lovasz number ϑ of G is defined in terms of the theta body as

$$\vartheta(G) \equiv \max_{x \in \mathrm{TH}(G)} \mathbf{1}^T x = \min_{U,c} \max_{i} \frac{1}{(c^T u_i)^2}.$$

For example, suppose G is the complete graph with n vertices. Then we can take the vectors U all equal and the max over indices falls out. Clearly the denominator is maximized over c when $c = u_i$, so $\vartheta(G) = 1$. As another example, suppose G is the edgeless graph with n vertices. Then U must consist of n orthogonal vectors; without loss of generality, let them be the standard basis of $\mathbb{R}^n$. Picking c as equally far from each u_i we see that c has components $1/\sqrt{n}$. Then $c^T u_i = 1/\sqrt{n}$ for each i so $\vartheta(G) = n$.

Proposition. *The Lovasz number of a strong product is bounded by the product of Lovasz numbers:*

$$\vartheta(G \boxtimes H) \leqslant \vartheta(G)\vartheta(H).$$

Proof. Suppose $U = (u_i) \subset \mathbb{R}^n$ and $V = (v_k) \subset \mathbb{R}^m$ are optimal orthonormal representations of G, H. Let $(\mathbb{R}^n \otimes \mathbb{R}^m, \mathrm{Tr})$ denote a tensor product equipped with the trace inner product. If $(i,k) \in V_{G \boxtimes H}$ then we associate (i,k) with $u_i \otimes v_k$. By definition of the trace,

$$(u_i \otimes v_k) \cdot (u_j \otimes v_l) = \mathrm{Tr}((u_i \otimes v_k)^T (u_j \otimes v_l)) = (u_i \cdot u_j)(v_k \cdot v_l).$$

Therefore $u_i \otimes v_k$ and $u_j \otimes v_l$ are orthogonal iff $ij \notin E_G$ or $kl \notin E_H$ iff $(i,j)(k,l) \notin E_{G \boxtimes H}$. Clearly $\|u_i \otimes v_k\| = 1$ so we see that $(u_i \otimes v_k)$ is an orthonormal representation of $G \boxtimes H$.

Suppose that c and d are the handles of U and V respectively. Then

$$\|c \otimes d\|^2 = \mathrm{Tr}((c \otimes d)^T (c \otimes d)) = \|c\|^2 \|d\|^2 = 1.$$

It follows that

$$\vartheta(G \boxtimes H) \leqslant \max_{i,j} \frac{1}{((c \otimes d)^T (u_i \otimes v_j))^2} = \max_{i,j} \frac{1}{(c^T u_i)^2} \frac{1}{(d^T v_j)^2} = \vartheta(G)\vartheta(H).$$

□

Let U be an orthonormal representation of G and let $S \subset V_G$ be a stable set. Then $u_i \cdot u_j = 0$ for all $i, j \in S$, $i \neq j$ and therefore

$$\sum_{i \in S} (c^T u_i)^2 \leqslant \|c\|^2 = 1 \text{ (with equality iff } \{u_i : i \in S\} \text{ is a basis).}$$

If S is a maximal stable set then $|S| = \alpha(G)$ and therefore

$$\sum_{i \in S} (c^T u_i)^2 \leqslant \sum_{i \in S} \frac{1}{\vartheta(G)} = \frac{\alpha(G)}{\vartheta(G)}.$$

So we see that $\alpha(G) \leqslant \vartheta(G)$. By this observation and our earlier proposition,

$$\alpha(G^{\boxtimes n}) \leqslant \vartheta(G^{\boxtimes n}) \leqslant (\vartheta(G))^n.$$

We conclude that $\Theta(G) \leqslant \vartheta(G)$.

4. THE CAPACITY OF THE PENTAGON CHANNEL

Recall the pentagon channel from section 1, which we will denote going forward by C_5. Using the language of Shannon capacity introduced in section 2, we showed in section 1 that $\sqrt{5} \leqslant \Theta(C_5)$. With the results of section 3, we can now show that this is actually an equality. Let $U = (u_i)$ be an orthonormal representation of C_5 defined by

$$u_i = \begin{pmatrix} \cos\theta \\ \sin\theta \cos\phi_k \\ \sin\theta \sin\phi_k \end{pmatrix}, \qquad \cos(\theta) = \frac{1}{\sqrt[4]{5}}, \qquad \phi_k = \frac{2\pi k}{5}.$$

Taking $c = (1, 0, 0)$ we find that

$$\vartheta(C_5) \leqslant \max_i \frac{1}{(c^T u_i)^2} = \sqrt{5}.$$

Therefore we have

$$\sqrt{5} = \sqrt{\alpha(G^{\boxtimes 2})} \leqslant \Theta(G) \leqslant \vartheta(G) \leqslant \sqrt{5}.$$

And we conclude that $\Theta(G) = \sqrt{5}$ exactly.

5. COMPUTING THE LOVASZ INVARIANT

Our working definition of ϑ given in section 3 gives us few clues about how to compute its value, or whether it can be efficiently computed at all. In this section we will reformulate ϑ as a semi-definite program: that is, we will express ϑ as the optimum of linear objective over a spectrahedron. Algorithms for optimizing an SDP are beyond the scope of this paper, but the key result in this area is that any SDP can be optimized in polynomial time. In particular, by expressing ϑ as an SDP it will follow that ϑ can be computed in polynomial time.

Recall from section 3 that ϑ is defined as an optimization over orthonormal representations of $G = ([n], E)$. Feasible handles c are unit vectors, so the feasible region of handles can be expressed by the linear constraint $c^T c = 1$. Feasible orthonormal representations (u_i) can be expressed by the linear constraints $u_i^T u_i = 1$ and $u_i^T u_j = 0$ if $i \neq j$ and $ij \notin E$. Minimizing ϑ is the same as maximizing $\vartheta^{-1/2}$, which we can compute with the following optimization:

$$\begin{array}{ll} \underset{c,(u_i)}{\text{maximize}} & c^T u_i \\ \text{subject to} & c^T c = 1; \\ & u_i^T u_i = 1, \quad i = 1, \ldots, n; \\ & u_i^T u_j = 0, \quad ij \notin E, i \neq j. \end{array}$$

Consider the block matrix

$$Y = \begin{bmatrix} u_1 & \ldots & u_n & c \end{bmatrix}^T \begin{bmatrix} u_1 & \ldots & u_n & c \end{bmatrix} \in \mathbb{S}^{n+1}.$$

Then Y is (by definition) positive semi-definite. We can express our optimization constraints as constraints on Y. The unit constraints $c^T c = 1$ and $u_i^T u_i = 1$ become $\text{Diag}(Y) = \mathbf{1}$. The edge constraints on $u_i^T u_j$ become constraints on $Y_{i,j}$. If $t \leqslant Y_{n+1,j}$ then Y is a feasible solution to the following optimization:

$$\begin{array}{lll} \underset{X}{\text{maximize}} & t & \\ \text{subject to} & \text{Diag}(X) = \mathbf{1}; & \\ & X_{ij} = 0, & ij \notin E, i \neq j; \\ & X_{n+1,j} \geqslant t, & i = 1, \ldots, n; \\ & X \succeq 0. & \end{array}$$

Optimizing this program gives us a maximal $t = c^T u_i$, so Y is not only feasible; Y is optimal. The first three constraints here are affine, and the last constraint confines us to the psd cone $\mathbb{S}_+^{n+1}$. The feasible region is therefore a spectrahedron, the objective function is clearly linear, so this optimization is a semi-definite program that computes $\vartheta^{-1/2}$.

6. SDP relaxation of STAB

One way to interpret $\vartheta(G)$ is as an SDP relaxation of the stability number $\alpha(G)$. We can model the stability number by considering incidence vectors χ^S of stable sets $S \subset G$:

$$\chi_i^S = \begin{cases} 1 & \text{if } i \in S, \\ 0 & \text{otherswise.} \end{cases}$$

In this model, $\alpha(G)$ is an optimal solution to $\mathbf{1}^T x$ over the stable set polytope

$$\mathrm{STAB}(G) = \mathrm{conv.\,hull}\{\chi^S : S \text{ is stable in } G\}.$$

We can alternatively characterize this polytope using linear inequalities. Observe that any $x \in \mathrm{STAB}(G)$ satisfies $x \geqslant 0$ and $x_i + x_j \leqslant 1$. Clearly if $x_i \in \mathbb{Z}$ then x is the incidence vector of a stable set and therefore

$$\mathrm{STAB}(G) = \{x \in \mathbb{R}^n : x \geqslant 0, x_i + x_j \leqslant 1 \text{ for all } ij \in E, \text{ and } x \in \mathbb{Z}^n\}.$$

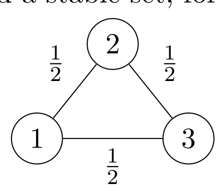
We can therefore formulate $\alpha(G)$ as the solution to an integer program

$$\alpha(G) = \max_{x \in \mathrm{STAB(G)}} \mathbf{1}^T x.$$

Unfortunately optimizing an integer program, and in particular computing $\alpha(G)$, is NP hard. The obvious relaxation is to drop the integral constraint. This turns our optimization into a (polynomial-time) linear program over

$$\mathrm{FRAC}(G) = \{x \in \mathbb{R}^n : x \geqslant 0, x_i + x_j \leqslant 1 \text{ for all } ij \in E\}.$$

Clearly $\mathrm{STAB}(G) \subset \mathrm{FRAC}(G)$ and it can be shown that $\mathrm{STAB}(G) = \mathrm{FRAC}(G)$ iff G is bipartite. The relaxation fails to find a stable set, for example, in the complete graph K_3:



We could attempt to remedy this situation by adding in more linear constraints to replace the integrality constraint that we dropped. For example, we could optimize over

$$\mathrm{QSTAB}(G) = \left\{ x \in \mathbb{R}^n : x \geqslant 0, \sum_{i \in Q} x_i \leqslant 1 \text{ where } Q \subset G \text{ is a clique} \right\}.$$

This resolves the problem described above for K_3, at the expense of a polytope description using exponentially many inequalities. Clearly $\mathrm{STAB}(G) \subset \mathrm{QSTAB}(G)$ but Lovasz showed that for many graphs this relation is still strict. The condition $\mathrm{STAB}(G) = \mathrm{QSTAB}(G)$ defines the full subcategory of perfect graphs and examples of non-perfect graphs abound, for example the pentagon C_5 discussed in section 4. A precise characterization of the perfect graphs was given in 2002[5]: these are precisely the graphs without odd holes or anti-holes.

A more promising approach to computing $\alpha(G)$ is to approximate it with an SDP. One popular SDP relaxation of $\mathrm{STAB}(G)$ is the theta body $\mathrm{TH}(G)$ defined in section 3.

Proposition. *For any graph G, $STAB(G) \subset TH(G) \subset QSTAB(G)$.*

Proof. Let U, c be an orthonormal representation of G and suppose $S \subset G$ is a stable set:

$$\sum_{i=1}^{n} (c^T u_i)^2 \chi_i^S = \sum_{i \in S} (c^T u_i) \leqslant \|c\|^2 = 1.$$

Therefore $\chi^S \in \mathrm{TH}(G)$. Because $\mathrm{TH}(G)$ is convex, $\mathrm{STAB}(G) \subset \mathrm{TH}(G)$. Now suppose $x \in \mathrm{TH}(G)$ and let $Q \subset G$ be a clique. Define an orthonormal representation $u_i = c$ for each $i \in Q$ and $u_j \in c^\perp$ pairwise orthogonal for $j \notin Q$. Then because $\|c\| = 1$,

$$\sum_{i \in Q} x_i = \sum_{i=1}^{n} (c^T u_i)^2 x_i \leqslant 1.$$

We conclude that $\mathrm{TH}(G) \subset \mathrm{STAB}(G)$. □

This property of the theta body together with the strong perfect graph theorem allows us to compute the stability number of perfect graphs in polynomial time. In particular,

$$\alpha(G) = \max_{x \in \mathrm{STAB}(G)} \mathbf{1}^T x \leqslant \max_{x \in \mathrm{TH}(G)} \mathbf{1}^T x = \vartheta(G).$$

Equality holds precisely when G is perfect. In this sense, we can interpret ϑ as an SDP relaxation of $\alpha(G)$.

References

[1] Shannon, Claude E., and Warren Weaver. "The mathematical theory of communication." Univ. Illinois 19.7 (1949): 1.

[2] Shannon, Claude E. "The zero error capacity of a noisy channel." Information Theory, IRE Transactions on 2.3 (1956): 8-19.

[3] Lovasz, Laszlo. "On the Shannon capacity of a graph." Information Theory, IEEE Transactions on 25.1 (1979): 1-7.

[4] Grotschel, Martin, Laszlo Lovasz, and Alexander Schrijver. "The ellipsoid method and its consequences in combinatorial optimization." Combinatorica 1.2 (1981): 169-197.

[5] Chudnovsky, Maria, et al. "The strong perfect graph theorem." Annals of mathematics (2006): 51-229.

[6] Schrijver, A. "A Course in Combinatorial Optimization (lecture notes)." (2000).

[7] Knuth, Donald Ervin. The sandwich theorem. Stanford University, Department of Computer Science, 1993.