AIT Blog

Derandomization, Lovasz Local Lemma, and Classical Channels

Samuel Epstein*

October 5, 2022

Over time, this blog will present unpublished material and then public material, with the eventual goal of writing a manuscript covering everything presented. This blog will also focus on the connection between AIT and Quantum Mechanics.

This blog entry deals with *derandomization*, [Eps22b, Eps22a], first with a general discussion and then with some new examples. This blog entry will showcase a new application of derandomization to compressing codebooks. In classical information theory, parties that communicate over channels share auxiliary information such as codebooks. Derandomization can be used to show the more bits used to describe the codebook results in a greater rate of communication, similar to the Kolmogorov structure function.

A problem is a collection of instances and the goal is to determine whether the instance satisfies a certain property, i.e. has a solution. An example is SAT, where instances are formulas and the goal is to find an assignment of the variables which satisfies it. Through the recently introduced method of *derandomization*, this approach can be aligned with Algorithmic Information Theory in a new way: bounds on the Kolmogorov complexity of the simplest solutions can be proven. The notion of a "solution" is intentionally vague, for example in MAXSAT, a solution could entail any assignment that satisfies 6/7 the optimal number of possible satisfiable clauses. Derandomization represents a good research topic for students because its framework is flexible and it can be applied to many fields of study.

The procedure for derandomization is as follows. The first step is to prove solutions to certain (or all!) instances of problems occur with probability at least p, with respect to a simple probability measure P over the solution candidate space. This is done by employing standard techniques (to be discussed later). Then, by applying Lemma 1 and Theorem 2, bounds on the Kolmogorov complexity of the simpliest solution can be proven. More specifically there exists some solution encoded into $x \in \{0,1\}^*$, with

$$\mathbf{K}(x) <^{\log} \mathbf{K}(P) - \log p + \mathbf{I}(\langle \text{description of the instance} \rangle ; \mathcal{H}).$$

$\mathbf{K}$ is the prefix-free Kolmogorov complexity function. $\mathbf{I}(x;\mathcal{H}) = \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})$ is the asymmetric mutual information term between x and the halting sequence $\mathcal{H}$. The instance itself can be incredibly complex (for example a formula with exponential number of clauses to variables), but for all non-exotic instances, the term $\mathbf{I}(\langle \text{description of the instance} \rangle ; \mathcal{H})$ will be negligible.

*JP Theory Group. samepst@jptheorygroup.org

Lovasz Local Lemma. One of the main tools to lower bounding the probability of a solution to an instance is the Lovasz Local Lemma. The Lovasz Local Lemma is traditionally used to show that some property is true for a random object with positive probability, and thus objects with this property exists. In fact, there is an exact lower bound on the probability of the property occurring and thus when LLL is applied to instances of problems, it produces a lower bound on the probability that a solution exists. Examples of this can be seen in [Eps22a]. Thus, going forward, one can look to where LLL is applied in the literature as well as apply LLL to new cases where LLL provides no traditional benefit, such as when the desired object property is trivially present.

Compressing Codebooks

There are deep connections between classical information theory and algorithmic information theory, with many theorems of the former appearing in an algorithmic form in the latter. In this section we revisit this connection. In particular we prove properties about the compression size of shared codebooks. A standard setup in information theory is two parties Alice and Bob who want to communicate over a noisy channel and share a codebook over a noiseless channel. However one might ask is how many bits did it take to communicate the codebook? By using derandomization, the tradeoff between codebook complexity and communication capacity can be proven.

Definition 1 (Discrete Memoryless Channel) *The input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ are finite. The channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ is represented by a conditional probability distribution $p(y|x)$. To send multiple symbols, we have $p(y^n|x^n) = \prod_{i=1}^n p(y_i|x_i)$. The capacity of channel with respect to a distribution Q over $\mathcal{X}$ is*

$$C_Q = I(X:Y) \text{ where random variables } (X,Y) \text{ are distributed according to } Q(x)p(y|x).$$

The term I is the mutual information between random variables. The capacity of a channel is

$$C = \max_{Q(x)} C_Q.$$

Definition 2 (Codebook) *A (M,n) codebook for channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ contains the following:*

1. *An encoder* $\mathrm{Enc}_n : \{1,\dots,M\} \to \mathcal{X}^n$.
2. *A decoder* $\mathrm{Dec}_n : \mathcal{Y}^n \to \{1,\dots,M\}$.

The rate of the codebook is $R = \frac{\log M}{n}$. The error rate of the codebook with respect to a fixed channel is the probability that, given the uniform distribution over $\{1,\dots,M\}$ for the sending symbols, the receiver decodes a symbol different from the encoded one.

Imagine the following setup: there is a sender Alice and a receiver Bob that communicate through a noisy memoryless discrete channel and Alice can send a codebook to Bob once on a side noiseless channel. Bob has oracle acess to the channel function $p(y|x)$ but Alice does not. Given a computable distribution Q over the input alphabet, and assuming the channel is non-exotic, Alice can hypothetically send $\sim\mathbf{K}(Q)$ bits plus some encoded parameters describing a codebook to Bob on the side channel. Then Alice and Bob can communicate with any rate R less than the capacity C_Q over the noisy channel.

Theorem 1 *For channel $\mathfrak{C} = (\mathcal{X}, p(y|x), \mathcal{Y})$ and every computable distribution Q over $\mathcal{X}$, for every rate $R < C_Q$, there is a $(2^{nR}, n)$ codebook $(\mathrm{Enc}_n, \mathrm{Dec}_n)$ with rate R and error rate $o(1)$ and*

$$\mathbf{K}(\mathrm{Enc}_n) <^{\log} \mathbf{K}(n, R, Q) + \mathbf{I}((n, R, Q, \mathfrak{C}); \mathcal{H}),$$
$$\mathbf{K}(\mathrm{Dec}_n|\mathfrak{C}) <^{+} \mathbf{K}(n, R, Q).$$

The proof comes from copying down pages of the original proof of the capacity of memoryless discrete channels, and noting which construct is computing from which. Future work entails looking at auxiliary information used in other types of channels, such as ones with feedback. Future work also involves proving properties of the following function, which is defined with respect to a channel, with for $0 \leq n \leq \mathbf{K}(\mathfrak{C})$,

$$\Gamma(n) = \max\{C_Q : \mathbf{K}(Q) \leq n\}.$$

Connected Subgraphs

This was an example of derandomization that I was interested in, and thought it might be of some independent interest. Conservation of information has been proven in all standard definitions of information. For example, in the symmetric definition, $\mathbf{I}(x : y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x, y)$, it has been proven that $\mathbf{I}(f(x) : y) <^{+} \mathbf{I}(x : y)$, [Lev84]. The following lemma shows that conservation holds in the asymmetric form of information, as long as the halting sequence $\mathcal{H}$ is used.

Lemma 1 ([Eps22b]) *For partial computable $f : \{0,1\}^* \to \{0,1\}^*$, for all $a \in \{0,1\}^*$,* $\mathbf{I}(f(a); \mathcal{H}) <^{+} \mathbf{I}(a; \mathcal{H}) + \mathbf{K}(f)$.

The following theorem has been called the Sets Have Simple Members theorem. For a set $D \subseteq \{0,1\}^*$, $\mathbf{m}(D) = \sum_{x \in D} \mathbf{m}(x)$, where $\mathbf{m}$ is the algorithmic probability. For the purposes of derandomization, D encodes all solutions to the instance. If $P(D)$ is large for some simple probability measure P, then $-\log \mathbf{m}(D)$ will be small. The error term is $\mathbf{I}(D; \mathcal{H})$, but using Lemma 1, this can be bounded by $\mathbf{I}(\langle \text{description of the instance} \rangle; \mathcal{H})$. This is because given a description of an instance (such as a MaxSat formula) one can define a simple function f that outputs all encoded solutions D (such as assignments that are 6/7 optimal).

Theorem 2 ([Lev16, Eps19])
For finite $D \subset \{0,1\}^*$, $-\log \max_{x \in D} \mathbf{m}(x) <^{\log} -\log \mathbf{m}(D) + \mathbf{I}(D; \mathcal{H})$.

Theorem 3 *Let $G = (V, E)$ be a connected graph with n vertices and m edges. Then there is a set $S \subseteq E$ with $|S| \leq (m+n)/2$ such that when E restricted to S, a connected subgraph of G is created, and $\mathbf{K}(S) <^{\log} \mathbf{K}(m) + 2n + \mathbf{I}(G; \mathcal{H})$.*

Proof. We order the edges E from 1 to m. Let $T \subseteq E$ be a spanning tree of G of size $n-1$. T can be encoded as a set of natural numbers, each number representing an edge of E. We use strings $x \in \{0,1\}^m$ to represent subgraphs, where $x[i] = 1$ if edge i is included in the graph. We define the probability P over strings $x \in \{0,1\}^m$ of length m, with $P(x) = \prod_{i=1}^{m} (3/4)^{1-x[i]} (1/4)^{x[i]}$. The complexity of P is $\mathbf{K}(P) <^{+} \mathbf{K}(m)$. Let $D \subset \{0,1\}^m$, such that if $x \in D$, then $x[i] = 1$ for all $i \in T$ and $\sum_{i \in \{1,\ldots,m\}-T} x[i] \leq (m-(n-1))/2$. Thus by the Markov inequality, $P(D) \geq (1/4)^{n-1} \times (1/2)$. The set D can be produced from a description of the graph, with $\mathbf{K}(D|G) = O(1)$. This is because T is simple relative to G. Thus, using Lemma 1 and Theorem 2, there is a set of $\leq (m+n)/2$ edges $x \in D$, with

$$\mathbf{K}(x) <^{\log} \mathbf{K}(P) - \log P(D) + \mathbf{I}(D; \mathcal{H}) <^{\log} \mathbf{K}(m) + 2n + \mathbf{I}(G; \mathcal{H}).$$

□

References

[Eps19] S. Epstein. On the algorithmic probability of sets. *CoRR*, abs/1907.04776, 2019.

[Eps22a] S. Epstein. 22 examples of solution compression via derandomization. *CoRR*, abs/2208.11562, 2022.

[Eps22b] S. Epstein. The outlier theorem revisited. *CoRR*, abs/2203.08733, 2022.

[Lev84] L. A. Levin. Randomness conservation inequalities; information and independence in mathematical theories. *Information and Control*, 61(1):15–37, 1984.

[Lev16] L. A. Levin. Occam bound on lowest complexity of elements. *Annals of Pure and Applied Logic*, 167(10):897–900, 2016.