Derandomization under Different Resource Constraints

Samuel Epstein*

November 17, 2022

Abstract

We provide another proof to the EL Theorem. We show the tradeoff between compressibility of codebooks and their communication capacity. A resource bounded version of the EL Theorem is proven. This is used to prove three instances of resource bounded derandomization.

1 Introduction

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. This can be formulized by a computable function $V : \Sigma^* \times \Sigma^* \to \Sigma$, where instances $x \in \Sigma^*$ are the first argument and solutions $y \in \Sigma^*$ are the second argument, i.e $\{y : V(x, y) = 1\}$. An example is SAT, where instances are formulas and the solutions are assignment of the variables which satisfies it. Through the recently introduced method of *derandomization*, [Eps22b, Eps22a], 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 fluid, for example in MAXSAT, in [Eps22a], a solution could entail any assignment that satisfies 6/7 the optimal number of possible satisfiable clauses.

The procedure for derandomization is as follows. The first step is to prove solutions to certain instances of problems have solutions that occur with probability at least p, with respect to a simple probability measure P over the solution candidate space. This is done by often employing the Lovasz Local Lemma (Lemma 5). Then, by applying conservation of information (Lemma 1) and the EL Theorem (Corollary 1), bounds on the Kolmogorov complexity of the simpliest solution can be proven. More specifically there exists some solution encoded into $x \in \Sigma^*$, 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 $x \in \Sigma^*$ and the halting sequence $\mathcal{H} \in \Sigma^\infty$. 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. The main step of derandomization is the application of EL Theorem, also known as the Sets Have Simple Members Theorem:

Theorem. (EL) *For finite* $D \subset \Sigma^*, \min_{x \in D} \mathbf{K}(x) <^{\log} \mathbf{m}(D) + \mathbf{I}(D; \mathcal{H})$.

*JP Theory Group. samepst@jptheorygroup.org

The term $\mathbf{m}(D)$ is equal to $\sum_{x \in D} \mathbf{m}(x)$, where $\mathbf{m}$ is the algorithmic probability. There are several proofs in the literature for the EL Theorem [She12, Eps19, Lev16]. In Section 3 of this paper we provide a new proof, which follows analogously to the proof in [Lev16], except left-total machines are not used.

In this paper, we will be applying derandomization to classical channels. Using derandomization, we will show a tradeoff between the compression size of a codebook vs. the communication capacity that the codebook allows. If the codebook is allowed more bits to be compressed to, the more capacity the codebook has in communicating information. Derandomization of codebooks is possible because they can be proven to exist in classical information theory by using the probabilistic method.

1.1 Resource Bounded Derandomization

In this paper, we show a resource bounded version of derandomization. By assuming the verifier V, defined in the previous section, runs in polynomial time, derandomizations can be reformulated using time bounded Kolmogorov complexity. To accomplish this, we introduce a resource bounded EL theorem, Corollary 3. This theorem follows almost directly from Theorem 4.1 in [AF09]. The theorem is as follows. Let $\mathbf{FP}' = \{f : f \in \mathbf{FP},\ \text{if } \|x\| = \|y\| \text{ then } \|f(x)\| = \|f(y)\|\}$.

Theorem. (Resource Bounded EL) *Assume* ***Crypto****. Let $L \in \mathbf{P}$, $A \in \mathbf{FP}'$, and assume $\delta_n = |\Sigma^n \cap A^{-1}(L_n)|/2^n$. Then for some polynomial p, $\min_{x \in L_n} \mathbf{K}^p(x) < -\log \delta_n + O(\log n)$.*

The function $\mathbf{K}^t$ is the t-time bounded Kolmogorov complexity and its formal definition can be found in Section 2. **Crypto** is a cryptographic assumption which ensures the existence of a certain type of pseudorandom generator. It can be found in Assumption 1. This theorem enables resource bounded derandomization, in which certain problems constructed in uniform polynomial time have simple solutions, with respect to $\mathbf{K}^t$. The following theorem is, to our knowledge, the first of its type.

Theorem. (Resource Bounded Derandomization) *Assume* ***Crypto****. then*

1. *Let $\{G_n\}$ be a uniformly computable in polynomial time sequence of k-regular graphs, with $k \geq 5$. There is a polynomial p where for each G_n, there is a partition x of $\lfloor \frac{k}{3 \ln k} \rfloor$ components each containing a cycle with*
$$\mathbf{K}^p(x) < 2n/k^2 + O(\log n).$$

2. *For vector v, $\|v\|_\infty = \max_i |v_i|$. A binary matrix M has entries of 0s or 1s. Let $\{M_n\}$ be a uniformly polynomial time computable sequence of $n \times n$ binary matrices. There is a polynomial p where for each M_n there is a vector $b \in \{-1, 1\}^n$ such that $\|M_n b\|_\infty \leq 4\sqrt{n \ln n}$ and*
$$\mathbf{K}^p(b) = O(\log n).$$

3. *Let Φ_n be a $k(n)$-SAT formula, using n variables, $m(n)$ clauses, uniformly polynomial time computable in n. Furthermore, each variable occurs in at most $2^{k(n)}/k(n)e - 1$ clauses. There is a polynomial p and a satisfying assignment x of Φ_n where*
$$\mathbf{K}^p(x) < 2m(n)e2^{-k(n)} + O(\log n).$$

1.2 Future Work

Future work entails exploring EL Theorems and Derandomizations under different resource constaints and access to random bits. For example if the universal Turing machine has access to some amount of random bits then is a modified EL Theorem such that with high probability, a simple program will produce a member of a set? This paper shows an EL Theorem for polynomial time constraints. A natural area of study would be over $\mathbf{K}^t$ with exponential t. Is there one in which the universal Turing machine has space constraints?. In the recent literature, there are notions of time bound Kolmogorov complexity enhanced by random bits [GKLO, Oli19]. Can they be used to create randomized, time bounded, EL Theorems? Given a new EL Theorem with constraints and/or random bits, are there accompanying derandomization theorems that can be proven?

2 Conventions

As noted in the introduction, $\mathbf{K}(x|y)$ is the conditional prefix free Kolmogorov complexity. $\mathbf{m}(x)$ is the algorithmic probability. $\mathbf{I}(x;\mathcal{H}) = \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})$ is the amount of information that the halting sequence $\mathcal{H} \in \Sigma^\infty$ has about x. For some function $t : \mathbb{N} \to \mathbb{N}$, the t-time bounded Kolmogorov complexity is $\mathbf{K}^t(x) = \min\{\|p\| : U(p) = x \text{ in time } t(\|x\|)\}$. A probability is *elementary*, if it has finite support and rational values. The deficiency of randomness of x relative to a elementary probability measure Q is $\mathbf{d}(x|Q) = -\log Q(x) - \mathbf{K}(x|Q)$. We recall for a set $D \subseteq \Sigma^*$, $\mathbf{m}(D) = \sum_{x\in D} \mathbf{m}(x)$. For the nonnegative real function f, we use $<^+ f$, $>^+ f$, and $=^+ f$ to denote $< f+O(1)$, $> f-O(1)$, and $= f \pm O(1)$. We also use $<^{\log} f$ and $>^{\log} f$ to denote $< f+O(\log(f+1))$ and $> f-O(\log(f+1))$, respectively. Derandomization of Section 4 uses the following lemma, which is conservation of mutual information information with the halting sequence over deterministic processing.

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

3 A New Proof to the EL Theorem

This section shows a new proof to the Sets Have Simple Members Theorem [Lev16, Eps19]. We also provide a proof that non-stochastic elements have high mutual information with the halting sequence, a well known result in the literature. This proof also does not rely on left-total machines, which the original proof did.

Definition 1 (Stochasticisty) *A string x is (α,β)-stochastic if there exists an elementary probability measure Q such that*

$$\mathbf{K}(Q) \leq \alpha \text{ and } \mathbf{d}(x|Q) \leq \beta.$$

Theorem 1 (Epstein,Levin) *Let P be a lower-semicomputable semimeasure and c be a large constant. Every (α,β)-stochastic set D with $s = \lceil -\log P(D) \rceil$ contains an element x with*

$$\mathbf{K}(x) < s + \alpha + 2\log\beta + \mathbf{K}(s) + 2\log\mathbf{K}(s) + c.$$

The theorem is directly implied by the following lemma.

Lemma 2 *Let P be a lower-semicomputable semimeasure and c be a large constant. If a set D is (α,β)-stochastic relative to an integer $s=\lceil -\log P(D)\rceil$, then D contains an element x with*

$$\mathbf{K}(x) < s+\alpha+\log\beta+\mathbf{K}(\log\beta)+\mathbf{K}(s)+c.$$

Note that if y is (α,β)-stochastic relative to s, then it is $(\alpha,\beta+\mathbf{K}(s))$-stochastic. Hence the lemma implies the theorem.

Lemma 3 *Let P be a discrete mesure and Q be a measure on sets. There exists a set S of size $\lceil\beta/\gamma\rceil$ such that*

$$Q(\{D : P(D)\geq\gamma \text{ and } D \text{ is disjoint from } S\}) \leq \exp(-\beta).$$

Proof. We use the probabilistic method, and show that if we draw $\lceil\beta/\gamma\rceil$ elements according to the distribution P, then the obtained set S satisfies the inequality with positive probability. The probability that a fixed set D with $P(D)\geq\gamma$ is disjoint from S is

$$\leq (1-\gamma)^{\beta/\gamma}\leq\exp(-\beta).$$

Hence the expected Q-measure of such a D is at most $\exp(-\beta)$ and the required set S exists. □

Proof of Lemma 2 for computable P. Let Q be an elementary probability measure with $\mathbf{K}(Q)\leq\alpha$ and $\mathbf{d}(D|Q,s)\leq\beta$. Without loss of generality, we assume that β is large positive power of 2. Fix a search procedure that on input Q, β, and $\gamma=2^{-s}$ finds a set satisfying the conditions of Lemma 3.

For large β, the set D must intersect the obtained set S. Indeed, consider the Q-test $g(X|Q,s)$ that is equal to $\exp(\beta)$ if X is disjoint from S, and is zero otherwise. This is indeed a test, because the above lemma implies that its expected value for $X\sim Q$ is bounded by 1. Since the test is also computable, it is a lower bound to the optimal test $\mathbf{t}(X|Q,s)$, up to a constant factor. By stochasticity of the set D, $g(D|Q,s)<O(1)\mathbf{t}(D|Q,s)<O(2^{\beta})$, because $2^{\mathbf{d}(X|Q,s)}$ is an optimal Q test relative to s. Thus for large enough β, D intersects Q.

It remains to construct a description of each element in S of the size given in the proposition. We construct a special decompressor that assigns short description to each element in S. On input of a string, the decompressor interprets the string as a concatenation of 4 parts:

1. A prefix-free description of Q of size at most α.
2. A prefix-free description of $\log\beta$ of size $\mathbf{K}(\log\beta)$.
3. A prefix-free description of s of size $\mathbf{K}(s)$.
4. An integer of bitsize $\log(\beta/\gamma)=s+\log\beta$.

It interprets the last integer as the index of an element in the set S of size $\lceil\beta/\gamma\rceil$ that is computed by the search procedure on input Q, β, and γ. The element is the output of the decompressor. The proposition is proven for computable P. □

Remark 1 *If P is computable, a set S satisfying the conditions of the lemma can be easily searched. But if P is not computable, then the collection of sets D with $P(D) \geq \gamma$ grows over time. Thus after constructing a good S, it can happen that a large Q-measure of sets D appears that does not contain an element from S, and that new elements to S need to be added. This type of interactive construction leads to an equivalent characterization of the problem in terms of a game which is shown in [She12]. Below, another proof is presented.*

Proof of Lemma 2 for lower-semicomputable P. We still assume that β is a large power of 2. Let $\gamma = 2^{-s}/2$. We can rewrite $P = \frac{\gamma}{\beta}(P_1 + \cdots + P_f + P_*)$, with $f \leq \beta/\gamma$, such that $P_1, \ldots P_f$ are probability measures with finite support obtained by a lower semi-computable approximation of P, and P_* is a lower-semicomputable semimeasure.

Construction of a lower-semicomputable test g over sets. We first construct tests $g_1, \ldots g_f$ together with a list of strings $z_1, \ldots, z_f$. Let $g_0(X) = 1$. Assume we already constructed $z_1, \ldots, z_{i-1}$ and g_{i-1} for some $i = 1, \ldots, f$. Choose z_i such that the test

$$g_i(X) = \begin{cases} g_{i-1}(X) & \text{if } g_{i-1}(X) \geq \exp(\beta) \\ \exp(P_i(X))g_{i-1}(X) & \text{if } g_{i-1}(X) < \exp(\beta) \text{ and } X \text{ is disjoint from } \{z_1, \ldots, z_i\} \\ 0 & \text{otherwise.} \end{cases}$$

satisfies $\mathbf{E}g_i(X) \leq \mathbf{E}g_{i-1}(X)$ where the expectations are taken for $X \sim Q$. Let $g(X)$ be equal to $\exp(\beta)$ if there exists an i such that $g_i(X) \geq \exp\beta$, otherwise let $g(X) = 0$. *End of construction*

We first show that each required string z_i in the construction exists. Suppose $z_1, \ldots, z_{i-1}$ and g_{i-1} have already been constructed. We show the existence of z_i using the probabilistic method. If we draw z_i according to P_i, then for each set X for which the second condition of g_i is satisfied, we have

$$\mathbf{E}_{z_i \sim P_i} g_i(X) \leq (1 - P_i(X))g_{i-1}(X) \exp P_i(X) \leq g_{i-1}(X),$$

because of the inequality $1 + r \leq \exp(r)$ for all reals r. If X satisfies the first or third condition, then $\mathbf{E}g_i(X) \leq \mathbf{E}g_{i-1}(X)$ is trivially true. So

$$\mathbf{E}_{X \sim Q}\mathbf{E}_{z_i \sim P_i} g_i(X) \leq \mathbf{E}_{X \sim Q} g_{i-1}(X),$$
$$\mathbf{E}_{z_i \sim P_i}\mathbf{E}_{X \sim Q} g_i(X) \leq \mathbf{E}_{X \sim Q} g_{i-1}(X),$$

and the required z_i exists.

We have $G(x) \leq O(\mathbf{t}(X|Q, (\gamma, \beta)))$, where $\mathbf{t}$ is the optimal test because the construction implies $\mathbf{E}g \leq 1$ and is effective, thus g is lower semicomputable. Every set X with $P(X) \geq 2^{-s} = 2\gamma$ satisfies $P_1(X) + \cdots + P_f(X) \geq \frac{\beta}{\gamma}P(D) - 1 \geq 2\beta - 1 \geq \beta$ by choice of P_i. Any such X that is disjoint from the set $\{z_1, \ldots, z_f\}$ satisfies

$$g_f(X) = \exp(P_1(X)) \exp(P_2(X)) \ldots \exp(P_f(X)) \geq \exp(\beta).$$

This implies $\mathbf{d}(X|Q, s) > \beta$ for large β, because up to $O(1)$ constants, we have

$$1.44\beta \leq \log g(X) \leq \mathbf{d}(X|Q, (\beta, \gamma)) \leq \mathbf{d}(X|Q, s) + 2\log\beta.$$

By the assumption on (α, β)-stochasticity of D, we have $\mathbf{d}(D|Q, s) \leq \beta$ and hence D must contain some z_j. The theorem follows by constructing a description for each string z_i of bitsize $s + \alpha + \log\beta + \mathbf{K}(\log\beta) + \mathbf{K}(s)$ in a similar way as above. □

3.1 Non-Stochastic Objects

The stochasticity of an object can be measured by

$$\mathbf{Ks}(x) = \min\{\mathbf{K}(P) + O(\log \max\{\mathbf{d}(x|P), 1\}) : P \text{ is an elementary probability measure}\}.$$

This term combines the complexity of the model P with how well it fits x, i.e. the randomness deficiency $\mathbf{d}$. It is well known in the literature that non-stochastic objects have high mutual information with the halting sequence [VS17]. In the following lemma, we reprove this fact, without using left-total machines, which was used in the original proof.

Lemma 4 $\mathbf{Ks}(x) <^{\log} \mathbf{I}(x; \mathcal{H})$.

Proof. We dovetail all programs to the universal Turing machine U. For $p \in \text{Domain}(U)$, $n(p) \in \mathbb{N}$ is the position in which the program $p \in \Sigma^*$ terminates. Let $\Omega^n = \sum_{p:n(p)<n} 2^{-\|p\|}$ and $\Omega = \Omega^\infty$ be Chaitin's Omega. Let Ω_t^n be Ω^n restricted to the first t digits. Let $x^* \in \Sigma^{\mathbf{K}(x)}$, with $U(x^*) = x$ with minimum $n(x^*)$. Let $k(p) = \max\{\ell : \Omega_\ell^{n(p)} = \Omega_\ell\}$ and $k = k(x^*)$. We define the elementary probability measure $Q(x) = \max\{2^{-\|p\|+k} : k(p) = k, U(p) = x\}$, $Q(\emptyset) = 1 - Q(\Sigma^* \setminus \{\emptyset\})$.

$$\begin{aligned}
\mathbf{d}(x|Q) &= -\log Q(x) - \mathbf{K}(x|Q) <^+ (\mathbf{K}(x) - k) - \mathbf{K}(x|\Omega_k) \\
&<^+ (\mathbf{K}(x|\Omega_k) + \mathbf{K}(\Omega_k) - k) - \mathbf{K}(x|\Omega_k) <^+ (k + \mathbf{K}(k)) - k \\
&<^+ \mathbf{K}(k).
\end{aligned}$$

$$\begin{aligned}
\mathbf{K}(x|\mathcal{H}) &<^+ \mathbf{K}(x|Q) + \mathbf{K}(Q|\mathcal{H}) <^+ \mathbf{K}(x|Q) + \mathbf{K}(\Omega_k|\mathcal{H}) \\
&<^+ -\log Q(x) + \mathbf{K}(k) <^+ (\mathbf{K}(x) - k) + \mathbf{K}(k) \\
k &<^{\log} \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})
\end{aligned}$$

$$\mathbf{Ks}(x) <^+ \mathbf{K}(Q) + O(\log \max\{\mathbf{d}(x|P), 1\}) <^+ k + O(\mathbf{K}(k)) <^{\log} \mathbf{I}(x; \mathcal{H}).$$

□

The following corollary comes from Theorem 1 and Lemma 4.

Corollary 1 (Epstein,Levin) *For finite* $D \subset \Sigma^*$, $\min_{x \in D} \mathbf{K}(x) <^{\log} -\log \mathbf{m}(D) + \mathbf{I}(D; \mathcal{H})$.

4 Classical Channels

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 2 (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.

Definition 3 (Codebook) *A (M, n) codebook for channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ contains the following:*

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

The rate of the codebook is $R = \frac{\log M}{n}$. The conditional probability of error is $\lambda_i = \sum_{y_n} p(y^n|x^n = \mathrm{Enc}(i))[\mathrm{Dec}(y^n) \neq i]$, where $[\cdot]$ is the indicator function. The average error rate of the codebook with respect to a fixed channel p is $P_e^{(n)} = \frac{1}{M}\sum_{i=1}^M \lambda_i$. It is the probability that, given the uniform distribution over $\{1, \ldots, M\}$ for the sending symbols, the receiver decodes a symbol different from the encoded one.

This section shows the following high level description of a communication scheme is possible: 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. This setup is formalized with Theorem 3. To prove this theorem, some results are needed from classical information theory.

4.1 Jointly Typical Sequences

We need the following definition and theorem, which can be found in [CT91], in the proof of Theorem 3. $H(X)$ is the entropy of random variable X, and $I(X:Y)$ is the mutual information between random variables X and Y.

Definition 4 *The set $A_\epsilon^{(n)}$ of jointly typical sequences $\{(x^n, y^n)\}$ with respect to the distribution $p(x, y)$ is the set of n-sequences with empirical entropies ϵ-close to the true entropies. $\mathcal{X}$ and $\mathcal{Y}$ are the finite discrete alphabet of random variables X and Y. Let $p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i)$.*

$$\begin{aligned} A_\epsilon^{(n)} = \{&(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \\ &\left| -\frac{1}{n}\log p(x^n) - H(X) \right| < \epsilon, \\ &\left| -\frac{1}{n}\log p(y^n) - H(Y) \right| < \epsilon, \\ &\left| -\frac{1}{n}\log p(x^n, y^n) - H(X,Y) \right| < \epsilon \Big\}. \end{aligned}$$

'] The following theorem details properties about the set $A_\epsilon^{(n)}$. A proof for it can be found in [CT91].

Theorem 2 (Joint AEP) *Let (X^n, Y^n) be sequences of length n drawn i.i.d. according to $p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i)$. Then*

1. $\Pr\left((X^n, Y^n) \in A_\epsilon^{(n)}\right) \to 1 - o(1)$.
2. *If $(\tilde{X}^n, \tilde{Y}^n) \sim p(x^n)p(y^n)$ ($\tilde{X}^n$ and $\tilde{Y}^n$ are independent with the same marginals as $p(x^n, y^n)$), then* $\Pr\left((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}\right) \leq 2^{-nI(X:Y)-3\epsilon}$.

4.2 Naive Sender Paradigm

Theorem 3 *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 average error rate $o(1)$ such that there is a program p with $\|p\| <^{\log} \mathbf{K}(n, R, Q) + \mathbf{I}((n, R, Q, \mathfrak{C}); \mathcal{H})$ and*

$$U(p, x) = \mathrm{Enc}_n(x),$$
$$U(p, \mathfrak{C}, x) = \mathrm{Dec}_n(x).$$

Proof. We start by generating a $(2^{nR}, n)$ code randomly according to distribution Q. We generate 2^{nR} codewords $x \in \mathcal{X}$ independently according to the distribution

$$Q(x^n) = \prod_{i=1}^n p(x_i).$$

The codewords can be represented as rows of a matrix

$$\mathcal{C} = \begin{bmatrix} x_1(1) & x_2(1) & \dots & x_n(1) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(2^{nR}) & x_2(2^{nR}) & \dots & x_n(2^{nR}) \end{bmatrix}$$

Each entry is generated i.i.d according to $Q(x)$, with

$$\Pr(\mathcal{C}) = \prod_{w=1}^{2^{nR}} \prod_{i=1}^n p(x_i(w)).$$

Consider the following algorithm for encoding and decoding a message.

1. A random code $\mathcal{C}$ is generated according to $Q(x)$.
2. The code $\mathcal{C}$ is sent to both the sender and the receiver. *Only the receiver is assumed to know the channel transition matrix $p(y|x)$ for the channel.* This differs from the standard literature, which assumes knowledge of p by the sender.
3. A message W is chosen according to the uniform distribution.
$$\Pr(W = w) = 2^{-nR}, \qquad w = 1, 2, \dots, 2^{nR}.$$
4. The wth codeword $X^n(w)$ corresponding to the wth row of $\mathcal{C}$ is sent over the channel.

5. The receiver receives a sequence Y^n according to the distribution

$$P(y^n|x^n(w)) = \prod_{i=1}^{n} p(y_i|x_i(w)).$$

6. The receiver decares that the index $\hat{W}$ was sent if the following conditions are satisfied:

   - $(X^n(\hat{W}), Y^n)$ is jointly typical, i.e. $(X^n(\hat{W}), Y^n) \in A_\epsilon^{(n)}$.
   - There is no other index $W' \neq \hat{W}$ such that $(X^n(W'), Y^n) \in A_\epsilon^{(n)}$.

   If no such $\hat{W}$ exists or if there are more than one, an error is declared, and the decoder outputs 0.

7. There is a decoding error if $\hat{W} \neq W$. Let $\mathcal{E}$ be this event.

We now analyze the probability of the error with respect to the random codebook $\mathcal{C}$.

$$\begin{aligned}
\Pr(\mathcal{E}) &= \sum_{\mathcal{C}} \Pr(\mathcal{C}) P_e^{(n)}(\mathcal{C}) \\
&= \sum_{\mathcal{C}} \Pr(\mathcal{C}) \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \lambda_w(\mathcal{C}) \\
&= \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \sum_{\mathcal{C}} \Pr(\mathcal{C}) \lambda_w(\mathcal{C}) \\
&= \sum_{\mathcal{C}} \Pr(\mathcal{C}) \lambda_1(\mathcal{C}) \qquad (1) \\
&= \Pr(\mathcal{E}|W=1),
\end{aligned}$$

where Equation 1 is due to symmetry of the code construction. We define

$$E_i = \{(X^n(i), Y^n) \in A_\epsilon^{(n)}\}, \qquad i \in \{1, 2, \ldots, 2^{nR}\}.$$

So E_i is the event that the ith code and Y^n are jointly typical, noting that Y^n is the result of sending the first codeword $X^n(1)$ over the channel. So

$$\Pr(\mathcal{E}|W=1) = P(E_1^c \cup E_2 \cup E_3 \ldots E_{2^{nR}}|W=1) \leq P(E_1^c|W=1) + \sum_{i=2}^{2^{nR}} P(E_i|W=1).$$

Due to the code generation procedure, $X^n(1)$ and $X^n(i)$ are independent for $i \neq 1$, and therefore, so are Y^n and $X^n(i)$. Due to Theorem 2 (2), the probability that $X^n(i)$ and Y^n are jointly typical is $\leq 2^{-n(I(X;Y)-3\epsilon)}$, where random variables X and Y are distributed acording to $Q(x)p(y|x)$. So by Theorem 2 (1), for sufficiently large n,

$$\begin{aligned}
\Pr(\mathcal{E}) = \Pr(\mathcal{E}|W=1) &\leq P(E_1^c|W=1) + \sum_{i=2}^{2^{nR}} P(E_i|W=1) \\
&\leq \epsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(X:Y)-3\epsilon)} \\
&= \epsilon + \left(2^{nR-1}\right) 2^{-n(I(X:Y)-3\epsilon)} \\
&\leq \epsilon + 2^{3n\epsilon} 2^{-n(I(X:Y)-R)} \\
&\leq 2\epsilon,
\end{aligned}$$

under the condition $R < I(X:Y) - 3\epsilon = C_Q - 3\epsilon$. Hence if $R < C_Q$ we can choose an ϵ and n so the average probability of error, averaged over codebooks is less than 2ϵ. We now remove the average over codebooks. Since the average error rate $P_e(\mathcal{C})$ is small, there exists at least one codebook $\mathcal{C}^*$ with a small average probability of error, with

$$\Pr(\mathcal{E}|\mathcal{C}*) = \frac{1}{2^{nR}} \sum_{i=1}^{2^{Rn}} \lambda_i(\mathcal{C}^*) \leq 2\epsilon.$$

Connection with Algorithmic Information Theory. We now derive the statements of the theorem. Define P to be the probability over codebooks used in earlier in this proof that uses the distribution Q to generate the codewords. Thus $\mathbf{K}(P) <^+ \mathbf{K}(Q, n, R)$. Let D be the set of encoded codebooks that achieve an error rate less than or equal to 2ϵ. By the arguments above, $P(D) \geq 0.5$. This set D is computable from Q, n, R, and $\mathfrak{C}$, with $\mathbf{K}(D|(Q, n, R, \mathfrak{C})) = O(1)$. Thus by Corollary 1 and Lemma 1, there is a codebook $\mathcal{C}^* \in D$ that has an error rate $\leq 2\epsilon$, with

$$\begin{aligned} \mathbf{K}(\mathcal{C}^*) &<^{\log} \mathbf{K}(P) - \log P(D) + \mathbf{I}(D; \mathcal{H}) \\ &<^{\log} \mathbf{K}(Q, n, R) + \mathbf{I}((Q, r, n, \mathfrak{C}); \mathcal{H}). \end{aligned} \tag{2}$$

Thus the sender can use solely $\mathcal{C}^*$ to send messages to the receiver. The receiver needs to determine if sequences are jointly typical, and thus uses $(\mathcal{C}^*, Q, \mathfrak{C})$ to decode the messages. Note that with careful analysis of the proof of Lemma 2 for computable probabilities, one can construct a short program for $\mathcal{C}^*$ (with size less than that of Equation 2) that can also compute Q. Thus, we can construct a program p with the properties described in the theorem statement. □

5 Resource Bounded EL Theorem

In this section we derive the resource bounded EL theorem. We also derive an interesting corollary to Theorem 4.1 in [AF09] which states to invert a hash function $f^{-1}(x)$, one can find a secret key π of size approximiately equal to x that will efficiently decompress to a pre-image of x with respect to f. The results in this section are not unconditional, they require the existence of the pseudorandom generator, introduced in [Nis94].

Assumption 1 ***Crypto*** *is the assumption that there exists a language in* $\mathbf{DTIME}(2^{O(n)})$ *that does not have size* $2^{o(n)}$ *circuits with* Σ_2^p *gates. This asssumption is need in the proof of Theorem 4 in [AF09] to assume the existence of a pseudorandom generator* $g : \Sigma^{k \log n} \to \Sigma^n$*, computable in time polynomial in* n*.*

Definition 5 $\mathbf{FP}' = \{f : f \in \mathbf{FP}$ *and if* $\|x\| = \|y\|$ *then* $\|f(x)\| = \|f(y)\|\}$*.*

Definition 6 *For* $A \in \mathbf{FP}'$ *we say that* A *samples* $D \subset \Sigma^n$ *with probability* γ*, if* $|\Sigma^n \cap A^{-1}(D)|/2^n > \gamma$*.*

Theorem 4 ([AF09]) *Assume* ***Crypto****. Let* $F \in \mathbf{FP}'$*. Let* $m, n \in \mathbb{N}$ *where* $\Sigma^n \supseteq f(\Sigma^m)$*. Let* $T_y = \{w \in \Sigma^m : F(w) = y\}$ *and* $V_k = \{y : \|y\| = n$ *and* $|T_y| \geq 2^k\}$*. There exists a function*

$$G : \Sigma^{m-k+O(\log m)} \to \Sigma^m$$

computable in polynomial time such that for all $y \in V_k$*,* $\text{range}(G) \cap T_y \neq \emptyset$*.*

Remark 2 *In the previous theorem, the running time of G is a polynomial function of the running time of F. This was noted in [LOZ22]. In addition, in subsequent theorems and corollaries of this section, the polynomial time function p in the resource bounded complexity $\mathbf{K}^p$ is a polynomial function of the running times of the algorithms of the theorem/corollary statements. Furthermore, due to [AF09], G can be encoded in $O(1)$ bits.*

The following corollary implies that to invert x with a hash function f, one can find a secret key π of size approximately equal to x that efficiently expands to an element in $f^{-1}(x)$.

Corollary 2 *Assume **Crypto**. Let $f \in \mathbf{FP}'$, where $f(\Sigma^n) \subseteq \Sigma^{n-k}$. Then for some polynomial p where for $\Sigma^n \supseteq D = f^{-1}(x)$,*

$$\min_{y \in D} \mathbf{K}^p(y) = n - \log |D| + O(\log n).$$

Proof. Follows directly from Theorem 4. □

Corollary 3 (Resource EL) *Assume **Crypto**. Let $L \in \mathbf{P}$, $A \in \mathbf{FP}'$, and assume A samples L_n with probability δ_n. Then for some polynomial p,*

$$\min_{x \in L_n} \mathbf{K}^p(x) < -\log \delta_n + O(\log n).$$

Proof. Let $F \in \mathbf{FP}'$ where $F(\Sigma^n) \subseteq \Sigma^n$ and for $x \in \Sigma^n$, $F(x) = 1^n$ if $A(x) \in L_n$ and $F(x) = 0^n$ otherwise. Let $k \in \mathbb{N}$ be maximal such that $\delta_n \geq 2^{k-n}$. Let $\ell = n - k + O(c \log n)$. By Theorem 4, there exists a function $G : \Sigma^\ell \to \Sigma^n$ running in polynomial time such that there exists $x \in \ell$, with $G(x) = 1^n$. This is because $1^n \in T_k$, using the definition in Theorem 4, because A produces a member of L_n with probability at least δ_n and all of L_n is mapped to 1^n. We define a program P that uses G to map x to a string y, then use A to map y to a string $z \in L_n$. This program P is of size ℓ and runs in polynomial time. □

A verifier $V : \Sigma^* \times \Sigma^* \to \Sigma$ is a function computable in polynomial time with respect to the first argument. For a given x, $\text{Proofs}(x) = \{y : V(x, y) = 1\}$.

Corollary 4 *Assume **Crypto**. Let $\{x_n\}$ be uniformly computable in polynomial time. For a verifier $V(x, y)$, let $A \in \mathbf{FP}'$ sample $\text{Proofs}(x_n)$ with probability γ_n. Thus there is a polynomial p and $y \in \text{Proofs}(x_n)$ with*

$$\mathbf{K}^p(y) < -\log \gamma_n + O(\log n).$$

6 Resource Bounded Derandomization

In this section, we use Corollary 4 to produce three examples of resource bounded derandomization. The resource free versions of these theorems can be found in [Eps22a].

Lemma 5 (Lovasz Local Lemma) *Let $E_1, \ldots, E_n$ be a collection of events such that $\forall i : \Pr[E_i] \leq p$. Suppose further that each event is dependent on at most d other events, and that $ep(d+1) \leq 1$. Then, $\Pr\left[\bigcap_i \overline{E}_i\right] > \left(1 - \frac{1}{d+1}\right)^n$.*

Proposition 1 (Mutual Independence Principle) *Suppose that $Z_1, \ldots Z_m$ is an underlying sequence of independent events and suppose that each event A_i is completely determined by some subset $S_i \subset \{Z_1, \ldots, Z_m\}$. If $S_i \cap S_j = \emptyset$ for $j = j_1, \ldots, j_k$ then A_i is mutually independent of $\{A_{j_1}, \ldots, A_{j_k}\}$.*

6.1 Vertex-Disjoint-Cycles

Theorem 5 *Assume **Crypto.** Let $\{G_n\}$ be a uniformly computable in polynomial time sequence of k-regular graphs, with $k \geq 5$. There is a polynomial p where for each G_n, there is a partition x of $\lfloor \frac{k}{3 \ln k} \rfloor$ components each containing a cycle with*

$$\mathbf{K}^p(x) < 2n/k^2 + O(\log n).$$

Proof. We partition the vertices of G into $c = \lfloor k/3 \ln k \rfloor$ components by assigning each vertex to a component chosen independently and uniformly at random. With positive probability, we show that every component contains a cycle. It is sufficient to prove that every vertex has an edge leading to another vertex in the same component. This implies that starting at any vertex there exists a path of arbitrary length that does not leave the component of the vertex, so a sufficiently long path must include a cycle. A bad event $A_v = \{\text{vertex } v \text{ has no neighbor in the same component}\}$. Thus

$$\begin{aligned}\Pr[A_v] &= \prod_{(u,v)\in E} \Pr[u \text{ and } v \text{ are in different components}] \\ &= \left(1 - \frac{1}{c}\right)^k < e^{-k/c} \leq e^{-3 \ln k} = k^{-3}.\end{aligned}$$

A_v is determined by the component choices of itself and of its out neighbors $N^{\text{out}}(v)$ and these choices are independent. Thus by the Mutual Independence Principle, (Proposition 1) the dependency set of A_v consist of those u that share a neighor with v, i.e., those u for which $(\{v\} \cup N(v)) \cap (\{u\} \cup N(u)) \neq 0$. Thus the size of this dependency is at most $d = (k+1)^2$.

Take $d = (k+1)^2$ and $p = k^{-3}$, so $ep(d+1) = e(1 + (k+1)^2)/k^3 \leq 1$, holds for $k \geq 5$. Thus, noting that $k \geq 5$, by Lovasz Local Lemma, (Lemma 5),

$$\Pr\left[\bigcap_{v \in G} \overline{A}_v\right] > \left(1 - \frac{1}{d+1}\right)^n = \left(1 - \frac{1}{(k+1)^2+1}\right)^n > \left(1 - \frac{1}{k^2}\right)^n. \tag{3}$$

Graphs G_n of size n are encoded in strings of size $kn\lceil \log n \rceil$ and partitions are the proofs, encoded in strings of size $n\lceil \log k \rceil$. The verify V returns 1 if each partition contains a cycle. The verifier runs in time $O(n \log n)$. We define a sampling function $A \in \mathbf{FP}'$ over the partition/proofs that is the same as the probability used in the Lovasz Local Lemma, i.e. the uniform distribution. Thus $A(x) = x$. A samples Proofs(G_n) with probability γ_n, where by Equation 3,

$$-\log \gamma_n < -n \log(1 - 1/k^2) < 2n/k^2.$$

Thus by Corolloray 4, there is a polynomial p, where for each graph $G_n \in Q$ of n vertices, there is a partition $x \in \text{Proofs}(G_n)$ with

$$\mathbf{K}^p(x) < 2n/k^2 + O(\log n).$$

□

6.2 Balancing-Vectors

Corollary 5 *Assume **Crypto**. For vector v, $\|v\|_\infty = \max_i |v_i|$. A binary matrix M has entries of 0s or 1s. Let $\{M_n\}$ be a uniformly polynomial time computable sequence of $n \times n$ binary matrices. There is a polynomial p where for each M_n there is a vector $b \in \{-1,1\}^n$ such that $\|M_n b\|_\infty \leq 4\sqrt{n \ln n}$ and*

$$\mathbf{K}^p(b) = O(\log n).$$

Proof. Let $v = (v_1, \ldots, v_n)$ be a row of M. Choose a random $b = (b_1, \ldots, b_n) \in \{-1, +1\}^n$. Let $i_1, \ldots, i_m$ be the indices such that $v_{i_j} = 1$. Thus

$$Y = \langle v, b \rangle = \sum_{i=1}^{n} v_i b_i = \sum_{j=1}^{m} v_{i_j} b_{i_j} = \sum_{j=1}^{m} b_{i_j}.$$

$$\mathbf{E}[Y] = \mathbf{E}[\langle v, b \rangle] = \mathbf{E}\left[\sum_i v_i b_i\right] = \sum_i \mathbf{E}[v_i b_i] = \sum v_i \mathbf{E}[b_i] = 0.$$

By the Chernoff inequality and the symmetry Y, for $\tau = 4\sqrt{n \ln n}$,

$$\Pr[|Y| \geq \tau] = 2\Pr[v \cdot b \geq \tau] = 2\Pr\left[\sum_{j=1}^{m} b_{i_j} \geq \tau\right] \leq 2\exp\left(-\frac{\tau^2}{2m}\right) = 2\exp\left(-8\frac{n \ln n}{m}\right) \leq 2n^{-8}.$$

Thus, the probability that any entry in Mb exceeds $4\sqrt{n \ln n}$ is smaller than $2n^{-8}$. Thus, with probability $1 - 2n^{-7}$, all the entries of Mb have value smaller than $4\sqrt{n \ln n}$.

Let $A(x) = x$ be the uniform sampling function. The verifier V takes in a matrix M and a vector b and returns 1 iff $\|Mb\|_\infty \leq 4\sqrt{n \ln n}$. Let $D \subset \Sigma^n$ consist of all strings that encode vectors $b_x \in \{-1, +1\}^n$ in the natural way such that $\|Mb_x\|_\infty \leq 4\sqrt{n \ln n}$. By the above reasoning, A samples D with probability $\geq 1 - 2n^{-7} > 0.5$. So by Corollary 4, there is a polynomial p, where for each $n \times n$ matrix M_n there is a binary vector $b \in \{-1, 1\}^n$ with $\|Mb\|_\infty \leq 4\sqrt{n \ln n}$ and

$$\mathbf{K}^p(b) = O(\log n).$$

□

6.3 k-Sat

Corollary 6 *Assume **Crypto**. Let Φ_n be a $k(n)$-SAT formula, using n variables, $m(n)$ clauses, uniformly polynomial time computable in n. Furthermore, each variable occurs in at most $2^{k(n)}/k(n)e - 1$ clauses. There is a polynomial p and a satisfying assignment x of Φ_n where*

$$\mathbf{K}^p(x) < 2m(n)e2^{-k(n)} + O(\log n).$$

Proof. The sample space is the set of all 2^n assigments. We choose a random assignment, where each variable is independently equally likely to have a true or false assignment. For each clause C_J, E_j is the bad event "C_j is not satisfied". Let $p = 2^{-k(n)}$ and $d = (2^{k(n)}/e) - 1$. Thus $\forall j$, $\Pr[E_j] \leq p$ as each clause has size $k(n)$ and each E_j is dependent on at most d other events since each variable

appears in at most $2^{k(n)}/k(n)e - 1$ other clauses, and each clause has $k(n)$ variables. Thus since $ep(d+1) \leq 1$, by the Lovasz Local Lemma 5, we have that,

$$\Pr\left[\bigcap_j \overline{E_j}\right] > \left(1 - \frac{1}{d+1}\right)^{m(n)} = \left(1 - \frac{e}{2^{k(n)}}\right)^{m(n)}. \tag{4}$$

Let $D_n \subset \Sigma^n$ be the set of all assignments that satisfy ϕ_n. We use a uniform sampler, with $A(x) = x$. By the above reasoning, A samples D_n with probability $\gamma_n > \left(1 - \frac{e}{2^{k(n)}}\right)^{m(n)}$. Thus

$$-\log \gamma_n < -m(n) \log\left(1 - e/2^{k(n)}\right) < 2em(n)2^{-k(n)}.$$

By Corollary 4, there is a polynomial p, where for all n, there is a satisfying assignment $x \in D_n$ of $\Phi(n)$ with

$$\mathbf{K}^p(x) < 2m(n)e2^{-k(n)} + O(\log n).$$

□

References

[AF09] L. Antunes and L. Fortnow. Worst-Case Running Times for Average-Case Algorithms. In *2009 24th Annual IEEE Conference on Computational Complexity*, pages 298–303, 2009.

[CT91] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, NY, USA, 1991.

[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.

[GKLO] H. Goldberg, V. Kabanets, Z. Lu, and I. Oliveira. Probabilistic Kolmogorov complexity with applications to average-case complexity. In *Computational Complexity Conference (CCC)*, volume 234. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, July.

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

[LOZ22] Z. Lu, I. Oliveira, and M. Zimand. Optimal coding theorems in time-bounded kolmogorov complexity. *CoRR*, abs/2204.08312, 2022.

[Nis94] Hardness vs randomness. *Journal of Computer and System Sciences*, 49(2):149–167, 1994.

[Oli19] I. Oliveira. Randomness and intractability in Kolmogorov complexity. In *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, volume 132 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 32:1–32:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019.

[She12] A. Shen. Game Arguments in Computability Theory and Algorithmic Information Theory. In *Proceedings of 8th Conference on Computability in Europe* , volume 7318 of *LNCS*, pages 655–666, 2012.

[VS17] Nikolay K. Vereshchagin and Alexander Shen. Algorithmic statistics: Forty years later. In *Computability and Complexity*, pages 669–737, 2017.