Computing Non-Repetitive Sequences Using the Lovász Local Lemma

Daniel Mourad

University of Connecticut

Daniel.Mourad@Uconn.edu

North Eastern Recursion and Defineability Seminar
April 24th, 2022

Introduction

Background

- The Lovász local lemma (LLL) is an existence theorem with many uses within the probabilistic method (Erdős and Lovász, 1975).
- There is a probabilistic algorithm for finding witnesses to the LLL (Moser and Tardos, 2010).
- This algorithm can be simulated to compute infinite witnesses (Rumyantsev and Shen, 2014).
- This effective version has been applied in complexity theory (Csima, Dzhafarov, Hirschfeldt, Jockusch, Solomon, and Westrick, 2019; Liu, Monin, and Patey, 2018).
- The LLL can by extended in myriad ways. We effectivise a version of the LLL inspired by the Lefthanded LLL (Pegden, 2011).

Non-Repetitive Sequences

Classical Existence of Non-Repetitive Sequences

The following theorem of classical combinatorics, says "there exists of a sequence such that repetitions of long blocks are far apart."

Theorem (Beck, 1981)

For each $\varepsilon > 0$ there is an N_ε and an infinite $\{0,1\}$-valued sequence such that any two identical blocks $[k, k+n)$ and $[\ell, \ell+n)$ of length $n > N_\varepsilon$ have distance $\ell - k$ greater than $(2-\varepsilon)^n$.

Example

In the string

$$a_0 a_1 a_2 \ldots a_{11} = 011010010001,$$

the only pair of identical blocks of size 4 are $[3, 7)$ and $[6, 10)$.

Question: Can we compute such a sequence?

Classical Existence of Non-Repetitive Sequences

Theorem (Beck, 1981)

For each $\varepsilon > 0$ *there is an* N_ε *and an infinite* $\{0, 1\}$*-valued sequence such that any two identical intervals of length* $n > N_\varepsilon$ *have distance greater than* $(2 - \varepsilon)^n$.

- Question: How to compute such a sequence?
- Existence is given by the infinite LLL.
- Natural choice: use effective version of the LLL given by Rumyantsev and Shen (2014).

Effective Local Lemma Setup

- Let $\mathcal{X} = \{x_1, x_2, \ldots\}$ be a set of computable random variables with finite ranges and uniformly computable probability distributions.
- Let $\mathcal{A} = \{A_1, A_2, \ldots\}$ be a set of events such that
  - Each $A \in \mathcal{A}$ is determined by a finite set of variables $\mathrm{vbl}(A) \subset \mathcal{X}$.
  - The code numbers for $\mathrm{vbl}(A_i)$ are uniformly computable with respect to i.
  - For each $A \in \mathcal{A}$, the set of neighbors $\Gamma(A) = \{B \in \mathcal{A} : \mathrm{vbl}(A) \cap \mathrm{vbl}(B) \neq \emptyset\}$ is finite.
  - For each x_i, $\{A_j : x_i \in \mathrm{vbl}(A_j)\}$ is finite and has code number uniformly computable with respect to i.

Computable Local Lemma

Recall that $\Gamma(A) = \{B \in \mathcal{A} : \text{vbl}(A) \cap \text{vbl}(B) \neq \emptyset\}$

Theorem (Rumyantsev and Shen, 2014)

Suppose there exists a rational constant $\alpha \in (0,1)$ and a computable real-valued function $z : \mathcal{A} \to (0,1)$ such that, for each $A \in \mathcal{A}$,

$$\Pr(A) \leq \alpha z(A) \prod_{B \in \Gamma(A)} (1 - z(B)).$$

Then there exists a computable assignment of the variables in $\mathcal{X}$ that makes all events $A \in \mathcal{A}$ false.

Setup for Building a Non-Repetitive Sequence

Theorem (Beck, 1981)

For each $\varepsilon > 0$ there is an N_ε and an infinite $\{0,1\}$-valued sequence such that any two identical intervals of length $n > N_\varepsilon$ have distance greater than $(2-\varepsilon)^n$.

- Let x_i be the value the $i'th$ bit in the sequence.
- Let $A_{k,\ell,n}$ be the event that blocks $[k, k+n)$ and $[\ell, \ell+n)$ are identical (assume $k < \ell$).
- Let $\mathcal{A} = \{A_{k,\ell,n} : \ell - k < (2-\varepsilon)^n\}$.
- $\mathrm{vbl}(A_{k,\ell,n}) = [k, k+n) \cup [\ell, \ell+n)$. $\Pr(A_{k,\ell,n}) = 2^{-n}$.
- $\Gamma(A_{k_0,\ell_0,n_0}) = \{A_{k,\ell,n} \in \mathcal{A} : \mathrm{vbl}(A_{k,l,n}) \cap \mathrm{vbl}(A_{k_0,\ell_0,n_0}) \neq \emptyset\}$

Unsatisfied Conditions

We run into a the following issues with this setup.

- Each x_i appears in $\mathrm{vbl}(A_{k,\ell,n})$ for infinitely many $A_{k,\ell,n} \in \mathcal{A}$.
- Each $A_{k_0,\ell_0,n_0} \in \mathcal{A}$ has infinitely many neighbors $A_{k,\ell,n}$.

There are two sources:

1. Fix k, ℓ. Increase n.
   - Can be fixed by modifying $\mathcal{A}$ to be $\{A_{k,\ell,n} : n \text{ is least such that } \ell - k < (2-\varepsilon)^n\}$
2. Fix k. Increase ℓ and n.

The latter is not as readily fixed. To resolve them, we modify the Moser-Tardos algorithm.

The Resample Algorithm

Moser-Tardos Algorithm

The Moser-Tardos algorithm, also known as the resample algorithm, looks for a valuation of the variables in $\mathcal{X} = \{x_1, x_2, \ldots\}$ that makes each event in $\mathcal{A} = \{A_1, A_2, \ldots\}$ false.

Algorithm

Start with a random sample of the variables in $\mathcal{X}$ and proceed in stages.

- *At each stage, resample each $x \in \mathrm{vbl}(A)$ for some true event A.*
- *If all $A \in \mathcal{A}$ are false at any stage, then the algorithm stops doing anything.*
- *Prioritize events A_i with lower indices.*

Example Stage

Example

Suppose $\{x_0, \ldots, x_{11}\}$ are independent fair coin flips and that $A_{k,\ell,n} \in \mathcal{A}$ for $(k, \ell, n) = (3, 6, 4)$ and $(k, \ell, n) = (0, 5, 4)$. If the current valuation is

$$x_0, x_1, \ldots, x_{11} = 011010010001,$$

then $A_{3,6,4}$ is true. So, the resample algorithm takes new random samples for each $x_i \in \mathrm{vbl}(A_{3,6,4}) = [3, 7) \cup [6, 10) = [3, 10)$. Suppose the resulting valuation is

$$x_0, x_1, \ldots, x_{11} = 011\underline{0101101}01.$$

This valuation makes $A_{0,5,4}$ true.

Resampling $A_{3,6,4}$ caused $A_{0,5,4}$ to go from false (good) to true (bad).

Ingredients of Computable LLL

Theorem (Constructive Lovász Local Lemma (Moser and Tardos, 2010))

Suppose the set of events $\mathcal{A}$ depending on variables $\mathcal{X}$ satisfy the conditions of the local lemma. Let τ_n be the first stage of the resample algorithm at which each of $A_1, A_2, \ldots, A_n$ is false. Then, $\mathbb{E}(\tau_n) < \infty$ for each n.

Lemma (Rumyantsev and Shen, 2014)

Suppose the set of events $\mathcal{A}$ depending on variables $\{X\}$ satisfy the setup and conditions for the computable local lemma. Then, there is a computable function $f : \mathbb{N}^2 \to \mathbb{N}$ such that

$$\Pr(x_i \text{ is resampled after stage } s) \leq f(i, s)$$

and $\lim_{s \to \infty} f(i, s) = 0$ for every i.

A Computable Witness

Lemma (Rumyantsev and Shen, 2014)

Suppose the hypotheses and conclusions of the previous lemma hold. Then,

1. *With probability 1, the resample algorithm converges to a witness to the infinite LLL on $\mathcal{A}$ and $\mathcal{X}$.*
2. *At least one of these witnesses is computable.*

To compute initial segment $x_1, \ldots x_n$ of a witness:

- Simulate the resample algorithm for every possible resampling.
- Do this for enough steps s to approximate the probability distribution on the final values of $x_1, \ldots, x_n$.
- Pick a valuation of $x_1, \ldots, x_n$ that has approximate probability greater than $\sum_{i=1}^{n} f(i, s)$.

Analysis of the Moser-Tardos Algorithm

To understand why the resample algorithm converges, we track the causality behind each resampling.

Tracking "Blame" in the Resample Algorithm

Suppose $\Gamma(A_i) = \{A_{i-1}, A_i, A_{i+1}\}$ and the first few events resampled by the Moser-Tardos algorithm begin with

$$A_1, A_2, A_5, A_4, A_3, A_6, A_5, A_5, A_4, A_6.$$

Reversing this initial segment yields

$$A_6, A_4, A_5, A_5, A_6, A_3, A_4, A_5, A_2, A_1.$$

Then, we can track "blame" for A_6 being true at the last step via the string

$$A_6, \cancel{A_4}, A_5, A_5, A_6, \cancel{A_3}, A_4, A_5, \cancel{A_2}, \cancel{A_1} = A_6, A_5, A_5, A_6, A_4, A_5.$$

We track "blame" for A_5 being true at the third step with the following sequence of length 1

$$A_5, \cancel{A_2}, \cancel{A_1} = A_5.$$

Key Features of Moser-Tardos Algorithm

Let $A_{i_1}, A_{i_2}, A_{i_3}, \ldots$ be the log of events resampled by the a run of the Moser-Tardos algorithm.

- For each A_{i_j} in the log, we can track the "blame" of why A_{i_j} is true via a Moser-Tree
- The probability of large Moser trees appearing approaches zero
- The α in the statement of the computable local lemma gives us a computable bound on the rate.
- If $x_i \in \mathrm{vbl}(A)$ for finitely many A then $\Pr(x_i$ is resampled after stage $s)$ therefore also approaches zero in a uniformly computable way.

What if $x_i \in \mathbf{vbl}(A)$ for infinitely many A?

What if $x_i \in \text{vbl}(A)$ for Infinitely Many A?

If $x_i \in \text{vbl}(A)$ for infinitely many A, then

- Large Moser trees are still just as rare,
- but small Moser trees can still cause x_i to be resampled at late stages s.

Example

Suppose $\mathcal{A}$ and $\mathcal{X}$ satisfy all conditions of the computable local lemma except that

$$[x_1 \in \text{vbl}(A_j)] \Leftrightarrow [j = 2^n].$$

If the singleton Moser tree

$$A_{2^k}$$

can occur for any k, then x_1 can be resampled arbitrarily late despite not being part of a long Moser tree.

Modifying the Resample Algorithm

Restricting the Resample Set

As in the previous slide, suppose that

$$[x_1 \in \mathrm{vbl}(A_j)] \Leftrightarrow [j = 2^n]$$

but also that

$$\Pr(A_j | x_1 = 0) = \Pr(A_j | x_0 = 1) = \Pr(A_j).$$

Idea

We should able to get all of the previous results for a modified resample algorithm in which we only resample $\mathrm{vbl}(A_{2^n}) \setminus \{x_1\}$ *when* A_{2^n} *is the least true event in* $\mathcal{A}$.

Restricting the Resample Set

Idea

In general, for each $A \in \mathcal{A}$, specify a subset $\text{rsp}(A) \subset \text{vbl}(A)$ of variables to resample. Then, we should redefine the neighborhood relation Γ by

$$\Gamma(A) = \{B \in \mathcal{A} : \text{rsp}(A) \cap \text{rsp}(B) \neq \emptyset\}$$

for our analysis of this modified resample algorithm.

- Let $\text{stc}(A) = \text{vbl}(A) \setminus \text{rsp}(A)$.
- We also require that $\max(\text{stc}(A)) < \min(\text{rsp}(A))$.

Priority of Events for Resampling

- Not resampling the full set $\mathsf{vbl}(A)$ of variables that determine A obfuscates the "blame tracking" feature of the Moser-sequences
- To recover this, we resample events in a specific order

Idea

Fix a linear order $\prec$ on $\mathcal{A}$ such that

$$[\max(\mathrm{rsp}(A)) < \max(\mathrm{rsp}(B))] \Rightarrow [A \prec B].$$

Where max *is calculated based on the indices of the variables in $\mathcal{X}$.*

The modified resample algorithm chooses the $\prec$-least true event to resample at each stage.

Constraints on $\mathrm{rsp}(A)$

- Let $A \ll B$ if and only if $A \prec B$ and $A \notin \Gamma(B)$ (i.e. $\mathrm{rsp}(A)$ and $\mathrm{rsp}(B)$ are disjoint).
- To ensure that $\ll$ is transitive, we also impose that $\mathrm{rsp}(A)$ be an interval $[\min(\mathrm{rsp}(A)), \max(\mathrm{rsp}(A))]$.
- We also require analogous conditions on $\mathrm{rsp}(A)$ as we did for $\mathrm{vbl}(A)$ in the effectivisation of the original LLL.

Computable "Lefthanded" LLL

Under all of the conditions previously described,

Theorem (M.)

Suppose there is $P^*(\mathcal{A}) \to [0,1]$ *such that for each* $A \in \mathcal{A}$ *and each valuation* μ *of the variables in* $\mathrm{stc}(A)$,

$$P^*(A) \geq \Pr(A | x = \mu(x) \textit{ for all } x \in \mathrm{stc}(A)).$$

Furthermore, suppose there is computable $z : \mathcal{A} \to (0,1)$ *and* $\alpha \in (0,1)$ *such that, for each* $A \in \mathcal{A}$,

$$P^*(A) \leq \alpha z(A) \prod_{B \in \Gamma(A)} (1 - z(B)).$$

Then, there is a computable valuation of $\mathcal{X}$ *under which each* $A \in \mathcal{A}$ *is false.*

Computable Non-Repetitive Sequences

Thus, we can compute a sequence whose long identical intervals are far apart:

Corollary

For each $\varepsilon > 0$ there is an N_ε and a computable $\{0, 1\}$-valued sequence such that any two identical intervals of length $n > N_\varepsilon$ have distance greater than $(2 - \varepsilon)^n$.

We can also compute a sequence whose adjacent intervals are very different.

Corollary

For each $\varepsilon > 0$, there is an N_ε and a computable $\{0, 1\}$-valued sequence such that any two adjacent intervals of length $n > N_\varepsilon$ have share at most $(\frac{1}{2} - \varepsilon)n$ many entries.

Game versions of these applications appear in (Pegden, 2011).

Outlook

Further Questions

- Can we compute winning strategies to the games studied by Pegden (2011)?
  - Almost: we can compute a winning sequence of moves from an opposing strategy
- Can we make conditions for the computable version of the "Lefthanded" LLL closer to the conditions of the original?
- Is the computable "Lefthanded" LLL useful in complexity theory?

Thank you!

References I

Beck, J. (1981). "An application of Lovász Local Lemma: there exists an infinite 01-sequence containing no near identical intervals". In: *Finite and Infinite Sets* 37.

Csima, Barbara F, Damir D Dzhafarov, Denis R Hirschfeldt, Carl G Jockusch, Reed Solomon, and Linda B Westrick (2019). "The reverse mathematics of Hindman's Theorem for sums of exactly two elements". In: *Computability* 8.3-4, pp. 253–263. ISSN: 22113576. DOI: 10.3233/COM-180094.

Erdős, P. and L. Lovász (1975). "Problems and results on 3-chromatic hypergraphs and some related questions". In: *Infinite and finite sets* 2.2, pp. 609–627.

Liu, Lu, Benoit Monin, and Ludovic Patey (2018). "A computable analysis of variable words theorems". In: *Proceedings of the American Mathematical Society* 147.2, pp. 823–834. ISSN: 0002-9939. DOI: 10.1090/proc/14269.

References II

Moser, Robin A. and Gábor Tardos (2010). "A constructive proof of the general lovász local lemma". In: *Journal of the ACM* 57.2, pp. 1–12. ISSN: 00045411. DOI: 10.1145/1667053.1667060. arXiv: 0903.0544.

Pegden, Wesley (2011). "Highly nonrepetitive sequences: Winning strategies from the local lemma". In: *Random Structures & Algorithms* 38.1-2, pp. 140–161. DOI: https://doi.org/10.1002/rsa.20354. arXiv: 1010.5772v1. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20354.

Rumyantsev, Andrei and Alexander Shen (May 2014). "Probabilistic constructions of computable objects and a computable version of Lovász local lemma". In: *Fundamenta Informaticae* 132.1, pp. 1–14. ISSN: 01692968. DOI: 10.3233/FI-2014-1029. arXiv: 1305.1535. URL: http://arxiv.org/abs/1305.1535.