6 Lovász local lemma

The Lovász local lemma (LLL), introduced in the paper of Erdős and Lovász (1975) is a powerful tool in the probabilistic method. It is some form of interpolation between the following two extreme (easy) scenerios

- Complete independence: if we have an arbitrary number of independent bad events, each occurring with probability < 1, then it is possible to avoid all of them (although with tiny probability)
- Union bound: if we have a collection of bad events whose total probability is < 1 (but usually much smaller), then it is possible to avoid all of them (often with high probability)

The local lemma deals with the case when each bad event is independent with most other bad events, but possibly dependent with a small number of other events.

We saw an application of the Lovász local lemma back in Section 1.1, where we used it to lower bound Ramsey numbers. This chapter we will explore the local lemma and its applications in depth.

6.1 Statement and proof

Here is the **setup** for the local lemma:

- We have "bad events" $A_1, A_2, \ldots, A_n$
- For each i there is some subset $N(i) \subseteq [n]$ such that A_i is independent from $\{A_j : j \notin N(i) \cup \{i\}\}$.

Here we say that event A_0 is **independent** from $\{A_1, \ldots, A_m\}$ if A_0 is independent of every event of the form $B_1 \wedge \cdots \wedge B_m$ where each B_i is either A_i or $\overline{A}_i$, i.e.,

$$\mathbb{P}(A_0 B_1 \cdots B_m) = \mathbb{P}(A_0)\mathbb{P}(B_1 \cdots B_m),$$

or, equivalently, using Bayes's rule: $\mathbb{P}(A_0 | B_1 \cdots B_m) = \mathbb{P}(A_0)$. (Here $\wedge$ = 'and' and $\vee$ = 'or', and we may omit $\wedge$ symbols, similar to multiplication)

We can represent the above relations by a **dependency (di)graph** whose vertices are indexed by the events (or equivalently $V = [n]$), and the (out-)neighbors of i are $N(i)$. (Mostly we'll just work with undirected dependency graphs for simplicity, but in general it may be helpful to think of them as directed—hence digraphs.)

Remark 6.1.1 (Important!). **Independence ≠ pairwise independence**
The dependency graph is *not* made by joining $i \sim j$ whenever A_i and A_j are not independent (i.e., $\mathbb{P}(A_iA_j) \neq \mathbb{P}(A_i)\mathbb{P}(A_j)$).

Example: suppose one picks $x_1, x_2, x_3 \in \mathbb{Z}/2\mathbb{Z}$ uniformly and independently at random and set, for each $i = 1, 2, 3$ (indices taken mod 3), A_i the event that $x_{i+1} + x_{i+2} = 0$. Then these events are pairwise independent but not independent. So the empty graph on three vertices is not a valid dependency graph (on the other hand, having at least two edges makes it a valid dependency graph).

A related note: there could be more than one choices for dependency graphs. So we speak of "a dependency graph" instead of "the dependency graph."

Remark 6.1.2 (**Random variable model** / **hypergraph coloring**). Many common applications of the local lemma can be phrased in the following form:

- A collection of independent random variables $x_1, \ldots, x_N$
- Each event A_i only depends on $\{x_j : j \in S_i\}$ for some subset $S_i \subseteq [N]$

In this case, valid dependency graph can be formed by placing an edge $i \sim j$ whenever $S_i \cap S_j \neq \emptyset$.

We can also view the above as coloring a hypergraph with vertices labeled by $[N]$, using independent random colors $x_1, \ldots, x_N$ for each vertex, so that various constraints on edges $S_1, S_2, \cdots \subseteq [N]$ are satisfied.

An example of such a problem is the **satisfiability problem (SAT)**: given a **CNF formula** (conjunctive normal norm = *and-of-or*'s), e.g.,

$$(x_1 \vee x_2 \vee x_3) \wedge (\overline{x_1} \vee x_2 \vee x_4) \wedge (\overline{x_2} \vee x_4 \vee x_5) \wedge \cdots$$

the problem is to find a satisfying assignment with boolean variables $x_1, x_2, \ldots$. Many problems in computer science can be modeled using this way.

The following formulation of the local lemma is easiest to apply and is the most commonly used.

Theorem 6.1.3 (Lovász local lemma; symmetric form)**.** Let $A_1, \ldots, A_n$ be events, with $\mathbb{P}[A_i] \leq p$ for all i. Suppose that each A_i is independent from a set of all other A_j except for at most d of them. If

$$ep(d+1) \leq 1,$$

then with some positive probability, none of the events A_i occur.

Remark 6.1.4. The constant e is best possible (Shearer 1985).

Theorem 6.1.5 (Lovász local lemma; general form)**.** Let $A_1, \dots, A_n$ be events. For each $i \in [n]$, let $N(i)$ be such that A_i is independent from $\{A_j : j \notin \{i\} \cup N(i)\}$. If $x_1, \dots, x_n \in [0, 1)$ satisfy

$$\mathbb{P}(A_i) \le x_i \prod_{j \in N(i)} (1 - x_j) \quad \forall i \in [n],$$

then with probability $\ge \prod_{i=1}^n (1 - x_i)$, none of the events A_i occur.

Proof that the general form implies the symmetric form. Set $x_i = 1/(d+1) < 1$ for all i. Then

$$x_i \prod_{j \in N(i)} (1 - x_j) \ge \frac{1}{d+1} \left(1 - \frac{1}{d+1}\right)^d > \frac{1}{(d+1)e} \ge p$$

so the hypothesis of general local lemma holds. □

Here is another corollary of the general form. It says that the local lemma works if the total probability of any neighborhood in a dependency graph is small.

Corollary 6.1.6. In the setup of Theorem 6.1.5, if $\mathbb{P}(A_i) < 1/2$ and $\sum_{j \in N(i)} \mathbb{P}(A_j) \le 1/4$ for all i, then with positive probability none of the events A_i occur.

Proof. In Theorem 6.1.5, set $x_i = 2\mathbb{P}(A_i)$ for each i. Then

$$x_i \prod_{j \in N(i)} (1 - x_j) \ge x_i \left(1 - \sum_{j \in N(i)} x_j\right) = 2\mathbb{P}(A_i) \left(1 - \sum_{j \in N(i)} 2\mathbb{P}(A_i)\right) \ge \mathbb{P}(A_i).$$

(The first inequality is by "union bound.") □

Proof of Lovász local lemma (general case). We will prove that

$$\mathbb{P}\left(A_i \,\middle|\, \bigwedge_{j \in S} \overline{A}_j\right) \le x_i \quad \text{whenever } i \notin S \subseteq [n] \tag{6.1}$$

Once (6.1) has been established, we then deduce that

$$\begin{aligned} \mathbb{P}(\overline{A}_1 \cdots \overline{A}_n) &= \mathbb{P}(\overline{A}_1)\mathbb{P}\left(\overline{A}_2 \mid \overline{A}_1\right) \mathbb{P}\left(\overline{A}_3 \mid \overline{A}_1\overline{A}_2\right) \cdots \mathbb{P}\left(\overline{A}_n \mid \overline{A}_1 \cdots \overline{A}_{n-1}\right) \\ &\ge (1 - x_1)(1 - x_2) \cdots (1 - x_n), \end{aligned}$$

which is the conclusion of the local lemma.

Now we prove (6.1) by induction on $|S|$. The base case $|S| = 0$ is trivial.

Let $i \notin S$. Let $S_1 = S \cap N(i)$ and $S_2 = S \setminus S_1$. We have

$$\mathbb{P}\left(A_i \,\middle|\, \bigwedge_{j\in S} \overline{A}_j\right) = \frac{\mathbb{P}\left(A_i \bigwedge_{j\in S_1} \overline{A}_j \,\middle|\, \bigwedge_{j\in S_2} \overline{A}_j\right)}{\mathbb{P}\left(\bigwedge_{j\in S_1} \overline{A}_j \,\middle|\, \bigwedge_{j\in S_2} \overline{A}_j\right)} \tag{6.2}$$

For the RHS of (6.2),

$$\text{numerator} \le \mathbb{P}\left(A_i \,\middle|\, \bigwedge_{j\in S_2} \overline{A}_j\right) = \mathbb{P}(A_i) \le x_i \prod_{j\in N(i)} (1 - x_i) \tag{6.3}$$

and, writting $S_1 = \{j_1, \dots, j_r\}$,

$$\begin{aligned}
\text{denominator} &= \mathbb{P}\left(\overline{A}_{j_1} \,\middle|\, \bigwedge_{j\in S_2} \overline{A}_j\right) \mathbb{P}\left(\overline{A}_{j_2} \,\middle|\, \overline{A}_{j_1} \bigwedge_{j\in S_2} \overline{A}_j\right) \cdots \mathbb{P}\left(\overline{A}_{j_r} \,\middle|\, \overline{A}_{j_1} \cdots \overline{A}_{j_{r-1}} \bigwedge_{j\in S_2} \overline{A}_j\right) \\
&\ge (1 - x_{j_1}) \cdots (1 - x_{j_r}) \qquad \text{[by induction hypothesis]} \\
&\ge \prod_{j\in N(i)} (1 - x_i)
\end{aligned}$$

Thus (6.2) $\le x_i$, thereby finishing the induction proof of (6.1). □

6.2 Algorithmic local lemma

The local lemma tells you that some good configuration exists, but the proof is non-constructive. The probability that a random sample avoids all the bad events is often very small (usually exponentially small, e.g., in the case of a set of independent bad events). It had been an open problem for a long time whether there exists some efficient algorithm to sample a good configuration in applications of the local lemma.

Moser (2009), during his PhD, achieved a breakthrough by coming up with the first efficient algorithmic version of the local lemma. Later, in a beautiful paper by Moser and Tardos (2010) extended the algorithm to a general framework for the local lemma.

The Moser–Tardos algorithm considers problems in the random variable model (Re-

mark 6.1.2). The algorithm is surprisingly simple.

Algorithm: Moser–Tardos local lemma algorithm

Initialize all the random variables;
while *there are violated events* **do**
  Pick an arbitrary violated event and resample its variables;

Theorem 6.2.1 (Moser and Tardos 2010)**.** If there are $x_1, \dots, x_n \in [0,1)$ such that

$$\mathbb{P}(A_i) \le x_i \prod_{j \in N(i)} (1 - x_j) \quad \forall i \in [n],$$

then the above randomized algorithm resamples each A_i at most $x_i/(1-x_i)$ times in expectation for each i.

Remark 6.2.2. The above theorem shows that the Moser–Tardos algorithm is an *Las Vegas* algorithm with polynomial expected runtime. A Las Vegas algorithm is a randomized algorithm that always terminates a successful result, but it might take a long time to terminate. Contrast this to a *Monte Carlo* algorithm, which runs in bounded time but may return a bad result with some small probability, and there may not be an efficient way to check whether the output is correct–e.g., randomly 2-coloring the edges of K_n to avoid a monochromatic $2\log_2 n$-clique. A Las Vegas algorithm can be converted into a Monte Carlo algorithm by cutting off the algorithm after some time (significantly larger than the expected running time) and applying Markov's inequality to bound the probability of failure. On the other hand, there is in general no way to convert a Monte Carlo algorithm to a Las Vegas algorithm unless there is an efficient way to certify the correctness of the output of the algorithm.

Remark 6.2.3. The Moser–Tardos algorithm assumes the random variable model. Some assumption on the model is necessary since the problem can be computationally hard in general.

For example, let $q = 2^k$, and $f\colon [q] \to [q]$ be some fixed bijection. Let $y \in [q]$ be given. The goal is find x such that $f(x) = y$.

For each $i \in [k]$, let A_i be the event that $f(x)$ and y disagree on i-th bit. Then $A_1, \dots, A_k$ independent (check!). Also, $f(x) = y$ if and only if no event A_i occurs.

A trivial version of the local lemma (with empty dependency graph) guarantees the existence of some x such that $f(x) = y$.

However, finding x may be computationally hard for certain functions f. In fact, the existence of such one-way functions (easy to compute but hard to invert) is the bedrock of cryptography. A concrete example is $f\colon \mathbb{F}_q \to \mathbb{F}_q$ is given by $f(0) = 0$, and for $x \neq 0$, set

$f(x) = g^x$ for some multiplicative generator. Then inverting f is the **discrete logarithm problem**, which is believed to be computationally difficult.

6.3 Coloring hypergraphs

Previously, in Theorem 1.3.1, we saw that every k-uniform hypergraph with fewer than 2^{k-1} edges is 2-colorable. The next theorem gives a sufficient local condition for 2-colorability.

Theorem 6.3.1. A k-uniform hypergraph is 2-colorable if every edge intersects at most $e^{-1}2^{k-1} - 1$ other edges

Proof. For each edge f, let A_f be the event that f is monochromatic. Then $\mathbb{P}(A_f) = p := 2^{-k+1}$. Each A_f is independent from all $A_{f'}$ where f' is disjoint from f. Since at most $d := e^{-1}2^{k-1} - 1$ edges intersect every edge, and $e(d+1)p \le 1$, so the local lemma implies that with positive probability, none of the events A_f occur. □

Corollary 6.3.2. For $k \ge 9$, every k-uniform k-regular hypergraph is 2-colorable. (Here k-regular means that every vertex lies in exactly k edges)

Proof. Every edge intersects $\le d = k(k-1)$ other edges. And $e(k(k-1)+1)2^{-k+1} < 1$ for $k \ge 9$. □

Remark 6.3.3. The statement is false for $k = 2$ (triangle) and $k = 3$ (Fano plane) but actually true for all $k \ge 4$ (Thomassen 1992).

Here is an example where the asymmetric form of the local lemma is insufficient (why is it insufficient? No bound on the the number of dependent events).

Theorem 6.3.4. Let H be a (non-uniform) hypergraph where every edge has size 3. Suppose

$$\sum_{f \in E(H) \setminus \{e\} : e \cap f \neq} 2^{-|f|} \le \frac{1}{8}, \quad \text{for each edge } e,$$

then H is 2-colorable.

Proof. Consider a uniform random 2-coloring of the vertices. Let A_e be the event that

edge e is monochromatic. Then $\mathbb{P}(A_e) = 2^{-|e|+1} \le 1/4$ since $|e| \ge 3$. Also, also

$$\sum_{f \in E(H)\setminus\{e\}: e\cap f\neq} \mathbb{P}(A_f) = \sum_{f \in E(H)\setminus\{e\}: e\cap f\neq} 2^{-|f|+1} \le 1/4.$$

Thus by Corollary 6.1.6 one can avoid all events A_e, and hence H is 2-colorable. □

Remark 6.3.5. A sign for when you should look beyond the symmetric local lemma is when there are bad events of very different nature (in particular, they have very different probabilities).

6.3.1 Compactness argument

Now we highlight an important **compactness argument** that allows us to deduce the existence of an infinite object, even though the local lemma itself is only applicable to finite systems.

Theorem 6.3.6. Let H be a (non-uniform) hypergraph on a possibly infinite vertex set, such that each edges is finite, has at least k vertices, and intersect at most d other edges. If $e2^{-k+1}(d+1) \le 1$, then H has a proper 2-coloring.

Proof. From a vanilla application of the symmetric local lemma, we deduce that for any finite subset X of vertices, there exists an 2-coloring X so that no edge contained in X is monochromatic (color each vertex iid uniformly, and consider the bad event A_e that the edge $e \subset X$ is monochromatic).

Next we extend the coloring to the entire vertex set V by a compactness argument. The set of all colorings is $[2]^V$. By Tikhonov's theorem (which says a product of a possibly infinite collection of compact topological spaces is compact), $[2]^V$ is compact under the product topology (so that open subsets are those defined by restriction to a finite set of coordinates).

For each finite subset X, let $C_X \subset [2]^V$ be the subset of colorings where no edge contained in X is monochromatic. Earlier from the local lemma we saw that $C_X \neq$. Furthermore,

$$C_{X_1} \cap \cdots \cap C_{X_\ell} \supseteq C_{X_1 \cup \cdots \cup X_\ell},$$

so $\{C_X : |X| < \infty\}$ is a collection of closed subsets of $[2]^V$ with the finite intersection property. Hence by compactness of $[2]^V$, we have $\bigcap_{X\subset V:|X|<\infty} \neq \emptyset$, and observe that any element of this intersection is a valid coloring of the hypergraph. □

Note that we may have $\mathbb{P}[\bigwedge_i A_i] = 0$ while $\bigwedge_i A_i \neq \emptyset$.

The same compactness argument tell us that: in the **random variable model** (Remark 6.1.2), **if it is possible to avoid every finite subset of bad events, then it is possible to avoid all bad events simultaneously**. (Again, one needs to be working in the random variable model for the compactness argument to work.

The next application appears in the paper of Erdős and Lovász (1975) where the local lemma originally appears.

Consider k-coloring the real numbers, i.e., a function $c\colon \mathbb{R} \to [k]$. We say that $T \subset \mathbb{R}$ is **multicolored** with respect to c if all k colors appear in T

Question 6.3.7. For each k is there an m so that for every $S \subset \mathbb{R}$ with $|S| = m$, one can k-color $\mathbb{R}$ so that every translate of S is multicolored?

The following theorem shows that this can be done whenever $m > (3+\epsilon)k \log k$ (and $k > k_0(\epsilon)$ sufficiently large).

Theorem 6.3.8. The answer to the above equation is yes if

$$e(m(m-1)+1)k\left(1-\frac{1}{k}\right)^m \le 1.$$

Proof. By the compactness argument, it suffices to check the result for every finite $X \subset \mathbb{R}$.

Each translate of S is not multicolored with probability $p \le k(1-1/k)^m$, and each translate of S intersects at most $m(m-1)$ other translates. Consider a bad event for each translate of S contained in X, and conclude by the symmetric version of the local lemma. □

6.4 Decomposing coverings

We say that a collection of disks in $\mathbb{R}^d$ is a **covering** is their union is $\mathbb{R}^d$. We say that it is a **k-fold covering** if every point of $\mathbb{R}^d$ is contained in at least k disks (so 1-fold covering is the same as a covering).

We say that a k-fold covering is **decomposable** if it can be partitioned into two coverings.

In $\mathbb{R}^d$, is a every k-fold covering by unit balls decomposable if k is sufficiently large?

A fun exercise: in $\mathbb{R}^1$, every k-fold covering by intervals can be partitioned into k coverings.

Mani-Levitska and Pach (1986) showed that every 33-fold covering of $\mathbb{R}^2$ is decomposable.

What about higher dimensions?

Surprising, they also showed that for every k, there exists a k-fold indecomposable covering of $\mathbb{R}^3$ (and similarly for $\mathbb{R}^d$ for $d \geq 3$).

However, it turns out that indecomposable coverings must cover the space quite unevenly:

Theorem 6.4.1 (Mani-Levitska and Pach 1986)**.** Every k-fold nondecomposable covering of $\mathbb{R}^3$ by open unit balls must cover some point $\gtrsim 2^{k/3}$ times.

Remark 6.4.2. In $\mathbb{R}^d$, the same proof gives $\geq c_d 2^{k/d}$.

We will need the following combinatorial geometric fact:

Lemma 6.4.3. A set of $n \geq 2$ spheres in $\mathbb{R}^3$ cut $\mathbb{R}^3$ into at most n^3 connected components.

Proof. Let us first consider the problem in one dimension lower. Let $f(m)$ be the maximum number of connected regions that m circles on a sphere in $\mathbb{R}^3$ can cut the sphere into.

We have $f(m+1) \leq f(m) + 2m$ for all $m \geq 1$ since adding a new circle to a set of m circles creates at most $2m$ intersection points, so that the new circle is divided into at most $2m$ arcs, and hence its addition creates at most $2m$ new regions.

Combined with $f(1) = 2$, we deduce $f(m) \leq m(m-1) + 2$ for all $m \geq 1$.

Now let $g(m)$ be the maximum number of connected regions that m spheres in $\mathbb{R}^3$ can cut $\mathbb{R}^3$ into. We have $g(1) = 2$, and $g(m+1) \leq g(m) + f(m) \leq g(m)$ by a similar argument as earlier. So $g(m) \leq f(m-1) + f(m-2) + \cdots + f(1) + g(0) \leq m^3$. □

Proof. Suppose for contradiction that every point in $\mathbb{R}^3$ is covered by at most $t \leq c2^{k/3}$ unit balls from F (for some sufficiently small c that we will pick later).

Construct an infinite hypergraph H with vertex set being the set of balls and edges having the form $E_x = \{\text{balls containing } x\}$ for some $x \in \mathbb{R}^3$. Note that $|E_x| \geq k$ since we have a k-fold covering.

Claim: every edge of intersects at most $d = O(t^3)$ other edges

Proof of claim: Let $x \in \mathbb{R}^3$. If $E_x \cap E_y \neq \emptyset$, then $|x - y| \leq 2$, so all the balls in E_y lie in the radius-4 ball centered at x. The volume of the radius-4 ball is 4^3 times the unit ball. Since every point lies in at most t balls, there are at most $4^3 t$ balls appearing among those E_y intersecting x, and these balls cut the radius-2 centered at x into $O(t^3)$ connected regions by the earlier lemma, and two different y's in the same region produce the same E_y. So E_x intersects $O(t^3)$ other edges. ■

With c sufficiently small, we have $e2^{-k+1}(d+1) \leq 1$. It then follows by Theorem 6.3.6

(local lemma + compactness argument) that this hypergraph is 2-colorable, which corresponds to a decomposition of the covering, a contradiction. □

6.5 Large independent sets

Every graph with maximum degree Δ contains an independent set of size $\geq |V|/(\Delta+1)$ (choose the independent set greedily). The following lemma shows that by decreasing the desired size of the independent set by a constant factor, we can guarantee a certain structure on the independent set.

Theorem 6.5.1. Let $G = (V, E)$ be a graph with maximum degree Δ and let $V = V_1 \cup \cdots \cup V_r$ be a partition, where each $|V_i| \geq 2e\Delta$. Then there is an independent set in G containing one vertex from each V_i.

This example is instructive because it is not immediately obvious what to choose as bad events (even if you are already told to apply the local lemma).

We may assume that $|V_i| = k := \lceil 2e\Delta \rceil$ for each i, or else we can remove some vertices from V_i.

Pick $v_i \in V_i$ uniformly at random, independently for each i.

What do we choose as bad events $A_\bullet$? It turns out that some choices work better than others.

Attempt 1:
$A_{i,j} = \{v_i \sim v_j\}$ for each $1 \leq i < j \leq r$ where there is an edge between V_i and V_j
$\mathbb{P}(A_{i,j}) \leq \Delta/k$
Dependency graph: $A_{i,j} \sim A_{k,\ell}$ if $\{i,j\} \cap \{k,\ell\} \neq \emptyset$. Max degree $\leq 2\Delta k$ (starting from (i,j), look at the neighbors of all vertices in $V_i \cup V_j$). The max degree is too large compared to the bad event probabilities.

Attempt 2:
$A_e = \{\text{both endpoints of } e \text{ are chosen}\}$ for each $e \in E$
$\mathbb{P}(A_e) = 1/k^2$
Dependency graph: $A_e \sim A_f$ if some V_i intersects both e and f. Max degree $\leq 2k\Delta$ (if e is between V_i and V_j, then f must be incident to $V_i \cup V_j$).
We have $e(1/k^2)(2k\Delta + 1) \leq 1$, so the local lemma implies the with probability no bad event occurs, and hence $\{v_1, \ldots, v_r\}$ is an independent set.

6.6 Directed cycles of length divisible by k

Theorem 6.6.1 (Alon and Linial 1989)**.** For every k there exists d so that every d-regular directed graph has a directed cycle of length divisible by k.

(d-regular means in-degree and out-degree of every vertex is d)

Corollary 6.6.2. For every k there exists d so that every $2d$-regular graph has a cycle of length divisible by k.

Proof. Every $2d$-regular graph can be made into a d-regular digraph by orientating its edges according to an Eulerian tour. And then we can apply the previous theorem. □

More generally they proved:

Theorem 6.6.3 (Alon and Linial 1989)**.** Every directed graph with min out-degree δ and max in-degree Δ contains a cycle of length divisible by $k \in \mathbb{N}$ as long as

$$k \le \frac{\delta}{1 + \log(1 + \delta\Delta)}.$$

Proof. By deleting edges, can assume that every every vertex has out-degree exactly δ.

Assign every vertex v an element $x_v \in \mathbb{Z}/k\mathbb{Z}$ iid uniformly at random.

We will look for directed cycles where the labels increase by 1 (mod k) at each step. These cycles all have length divisible by k.

For each vertex v, let A_v be the event that there is nowhere to go from v (i.e., if no outneighbor is labeled $x_v + 1 \pmod k$). We have

$$\mathbb{P}(A_v) = (1 - 1/k)^\delta \le e^{-\delta/k}.$$

The following is a valid dependency graph, noting that A_v only depends on $\{x_w : w \in \{v\} \cup N^+(v)\}$, where $N^+(v)$ denotes the out-neighbors of v and $N^-(v)$ the in-neighbors of v:

$A_v \sim A_w$ if $\{v\} \cup N^+(v)$ intersects $\{w\} \cup N^+(w)$.

The maximum degree in the dependency graph is at most $\Delta + \delta\Delta$ (starting from v, there are (1) at most Δ choices stepping backward (2) δ choices stepping forward, and (3) at most $\delta(\Delta - 1)$ choices stepping forward and then backward to land somewhere other than

v). So an application of the local lemma shows that, as long as $e^{1-\delta/k}(1+\Delta+\delta\Delta)$, i.e.,

$$k \le \delta/(1+\log(1+\Delta+\delta\Delta)),$$

then we are done. This is almost, but not quite the result (though, for most application, we would be perfectly happy with such a bound).

The final trick is to notice that we actually have an even smaller dependency digraph:

A_v is independent of all A_w where $N^+(v)$ is disjoint from $N^+(w) \cup \{w\}$.

Indeed, even if we fix the colors of all vertices outside $N^+(v)$, the conditional probability that A_v is still $(1-1/k)^\delta$.

The number of w such that $N^+(v)$ intersects $N^+(w) \cup \{w\}$ is at most $\delta\Delta$ (no longer need to consider (1) in the previous count). And we have

$$ep(\delta\Delta+1) \le e^{1-\delta/k}(\delta\Delta+1) \le 1.$$

So we are done by the local lemma. □

6.7 Lopsided local lemma

In the dependency graph, intuitively, the neighbors of A_i consists of all the events dependent on A_i (again, same warning as earlier: it is insufficient to simply check for pairwise dependence). However, if the there is a positive dependence among the bad events—avoiding some bad events make it easier to avoid others—then perhaps it would actually make it easier to avoid all bad events. For example, in an extreme scenario, if several bad events are identical, so that they are perfectly positively correlated, then it is much easier to avoid them compared to avoiding independent bad events. In the opposite extreme, if several bad events are disjoint, then it would be harder to avoid all of them. Thus, intuitively, it seems reasonable that in the local lemma, we are primarily concerned about negative dependencies and but not positive dependencies among bad events.

We can make this notion precise by re-examining the proof of the local lemma. Where did we actually use the independence assumption in the hypothesis of the local lemma? It was in the following step, Equation (6.3):

$$\text{numerator} \le \mathbb{P}\left(A_i \,\middle|\, \bigwedge_{j\in S_2} \overline{A}_j\right) = \mathbb{P}(A_i) \le x_i \prod_{j\in N(i)} (1-x_i).$$

If we had changed the middle $=$ to $\le$, the whole proof would remain valid. This observation allows us to weaken the independence assumption. Therefore we have the following

theorem.

Theorem 6.7.1 (Lopsided local lemma — Erdős and Spencer 1991)**.** Let $A_1, \dots, A_n$ be events. For each i, let $N(i) \subset [n]$ be such that

$$\mathbb{P}\left(A_i \,\middle|\, \bigwedge_{j \in S} \overline{A}_j\right) \le \mathbb{P}(A_i) \quad \forall i \in [n] \text{ and } S \subseteq [n] \setminus (N(i) \cup \{i\}) \tag{6.4}$$

Suppose there exist $x_1, \dots, x_n \in [0, 1)$ such that

$$\mathbb{P}(A_i) \le x_i \prod_{j \in N(i)} (1 - x_j) \quad \forall i \in [n].$$

Then with probability $\ge \prod_{i=1}^n (1 - x_i)$ none of the event A_i occur.

Like earlier, we also have a symmetric version that is easier to apply.

Corollary 6.7.2 (Lopsided local lemma; symmetric version)**.** In the previous theorem, if $|N(i)| \le d$ and $\mathbb{P}(A_i) \le p$ for every $i \in [n]$, and $ep(d+1) \le 1$, then with positive probability none of the events A_i occur.

The (di)graph where $N(i)$ is the set of (out-)neighbors of i is called a **negative dependency (di)graph**. Erdős and Spencer called it the **lopsidependency graph**, though I prefer "negative dependency graph" since it is more descriptive.

Remark 6.7.3. Here are several equivalent formulations of (6.4): for every $i \in [n]$ and $S \subseteq [n] \setminus (N(i) \cup \{i\})$,

- $\mathbb{P}\left(\overline{A}_i \,\middle|\, \bigwedge_{j \in S} \overline{A}_j\right) \ge \mathbb{P}(\overline{A}_i)$
- $\mathbb{P}\left(A_i \bigwedge_{j \in S} \overline{A}_j\right) \le \mathbb{P}(A_i)\mathbb{P}\left(\bigwedge_{j \in S} \overline{A}_j\right)$

To put in words, each event is non-negatively dependent on its non-neighbors.

It may be slightly strange to think about at first, but to verify the validity of a *negative* dependency graph, we are actually checking *nonnegative* dependencies (against non-neighbors). Likewise, earlier, to verify a dependency graph, we need to check independence against non-neighbors.

Remark 6.7.4. From the proof of Theorem 6.7.1, we see that we can weaken the negative dependency hypothesis to

$$\mathbb{P}\left(A_i \,\middle|\, \bigwedge_{j \in S} \overline{A}_j\right) \le x_i \prod_{j \in N(i)} (1 - x_j) \quad \forall i \text{ and } S \subseteq [n] \setminus N(i).$$

Though negative dependency is often easier to argue.

6.7.1 Random permutations and positive dependencies

Just like how most applications of the local lemma can be cast in terms of the the random variable model, which makes it easy to produce a valid dependency graph (by looking at shared random variables), a natural setting for applications of the lopsided local lemma is that of random permutations (and, by extending the domain, also random injections).

Here is a model problem: what is the probability that a random permutation π of $[n]$ has no fixed points? (Such permutations are called "derangements")

This problem can be solved exactly: using inclusion-exclusion, one can deduce that probability to be $\sum_{i=0}^{n}(-1)^i/i! = e^{-1} + o(1)$. Suppose that we did not know this answer.

Let A_i be the event that $\pi(i) = i$. It is easy to see that $\mathbb{P}(A_i) = 1/n$. If the events $A_1, \dots, A_n$ were independent, then we would deduce that with probability $(1 - 1/n)^n = 1/e + o(1)$ none of the A_i occur. But these events are not independent.

Intuitively, these events are positively dependent: having some fixed points makes it likes to see other fixed points. The next theorem makes this rigorous, so that we can deduce $\mathbb{P}(\overline{A}_1 \dots \overline{A}_n) \geq \mathbb{P}(\overline{A}_1) \cdots \mathbb{P}(\overline{A}_n) = (1 - 1/n)^n = 1/e - o(1)$, a lower bound that matches the truth.

It may be easier to visualize permutations as perfect matchings in the complete bipartite graph $K_{n,n}$. We will use these two interpretations interchangeably.

Theorem 6.7.5 (Positive dependence for random perfect matchings)**.** Let M be a perfect matching of $K_{n,n}$ chosen uniformly at random. For each matching F, let A_F denote the event that $F \subseteq M$.
Let $F_0, F_1, \dots, F_k$ be matchings such that no edge of F_0 shares a vertex with any edge from $F_1 \cup \cdots \cup F_k$. Then

$$\mathbb{P}\left(A_{F_0} \mid \overline{A}_{F_1} \cdots \overline{A}_{F_k}\right) \leq \mathbb{P}(A_{F_0}).$$

In other words, if $\mathcal{F}$ is a set of matchings in $K_{n,n}$, then the following if a valid negative dependency graph on the events $\{A_F : F \in \mathcal{F}\}$: $A_{F_1} \sim A_{F_2}$ if F_1 and F_2 touch (i.e., some two edges coincide or share an endpoint).

Proof. By relabeling, we may assume that the edges of F_0 are $(1,1), (2,2), \dots, (t,t)$.

For each injection $\tau\colon [t] \to [n]$ (also viewed as a matching with edges $(1, \tau(1)), \dots, (t, \tau(t))$), let $\mathcal{M}_\tau$ denote the set of perfect matchings in $K_{n,n}$ containing τ but not containing any of $F_1, \dots, F_k$.

Let $\tau_0 \colon [t] \to [n]$ be the map sending i to i (i.e., the matching F_0). Then the LHS and RHS of the desired inequality $\mathbb{P}\left(A_{F_0} \mid \overline{A}_{F_1} \cdots \overline{A}_{F_k}\right) \le \mathbb{P}(A_{F_0})$ can be rewritten as

$$\frac{|\mathcal{M}_{\tau_0}|}{\sum_\tau |\mathcal{M}_\tau|} \le \frac{1}{n(n-1)\cdots(n-t+1)},$$

where the sum is taken over all $n(n-1)\cdots(n-t+1)$ injections $\tau \colon [t] \to [n]$. Thus it suffices to prove that

$$|\mathcal{M}_{\tau_0}| \le |\mathcal{M}_\tau| \quad \text{for every injection } \tau \colon [t] \to [n].$$

To show this inequality, we construct an injection $\mathcal{M}_{\tau_0} \to \mathcal{M}_\tau$. Intuitively, this injection is obtained by permuting some of the vertices on the right-half of $K_{n,n}$ so that the matching τ_0 to taken to τ. Let us illustrate this idea in a simple case when $\tau(i) = t + i$ for each $i \in [t]$: we construct $\mathcal{M}_{\tau_0} \to \mathcal{M}_\tau$ by swapping, in $K_{n,n}$, the i-th vertex on the right-half with the $(t+i)$-th vertex on the right-half, for each $i \in [n]$.

More generally, extend $\tau \colon [t] \to [n]$ to a permutation σ on $[n]$ sending $\tau([t]) \setminus [t]$ to $[t] \setminus \tau([t])$ and otherwise leaving $[n] \setminus ([t] \cup \tau([t]))$ fixed as identity.

Then σ acts on the set of matchings in $K_{n,n}$ by permuting the right-endpoints. In particular, σ sends τ_0 to τ. Also σ permutes the set of perfect matchings of $K_{n,n}$.

It remains to show that if $M \in \mathcal{M}_{\tau_0}$, then its image σM lies in $\mathcal{M}_\tau$. By construction $\tau \subset \sigma M$. Suppose $F_i \subset \sigma M$ for some $i \in [k]$. Since F_i does not share any vertex with F_0, all the left-endpoints in F_i lie in $[n] \setminus [t]$. Since $(i, \tau(i))$ is an edge of σM, all the right-endpoints in F_i lie in $[n] \setminus ([t] \cup \tau([t]))$. It follows that $\tau F_i = F_i$, so that $F_i \subset M$, which contradicts $M \in \mathcal{M}_{\tau_0}$.

Thus σ induces an injection from $\mathcal{M}_{\tau_0}$ to $\mathcal{M}_\tau$. □

6.7.2 Latin square transversals

A **Latin square** of order n is an $n \times n$ array filled with n symbols so that every symbol appears exactly once in every row and column. Example:

1	2	3
2	3	1
3	1	2

(Name origin: The name Latin square was inspired by mathematical papers by Leonhard Euler (1707–1783), who used Latin characters as symbols)

Given an $n \times n$ array, a **transversal** is a set of n entries with one in every row and column.

A **Latin transversal** is a transversal with distinct entries. Example:

$$\begin{matrix} \mathbf{1} & 2 & 3 \\ 2 & \mathbf{3} & 1 \\ 3 & 1 & \mathbf{2} \end{matrix}$$

Here is a famous open conjecture about Latin transversals. (Can you see why "odd" is necessary?)

Conjecture 6.7.6. Every odd order Latin square has a transversal.

The next result is the original application of the lopsided local lemma.

Theorem 6.7.7 (Erdős and Spencer 1991)**.** Every $n \times n$ array where every entry appears at most $n/(4e)$ times has a Latin transversal.

Proof. Let (m_{ij}) be the array. Pick a transversal uniformly at random. For each pair of equal entries $m_{ij} = m_{kl}$ in the array in distinct rows and distinct columns, consider the event $A_{ijkl} = A_{klij}$ that the transversal contains both locations (i, j) and (k, l). Then $\mathbb{P}(A_{ijkl}) = 1/(n(n-1))$. (By reinterpreting in the earlier language of matchings, A_{ijkl} is the event that the random perfect matchings contains the two edges (i, j) and (k, l), which are assigned identical edge-labels.)

By the earlier theorem, the following is a negative dependency graph: two pairs of entries are adjacent if they share some row or column, i.e., $A_{ijkl} \sim A_{i'j'k'l'}$ unless $|\{i, k, i', k'\}| = |\{j, l, j', l'\}| = 4$.

Let us count neighbors in this negative dependency graph. Given A_{ijkl}, there are at most $4n-4$ additional locations (x, y) that share a column or row with either of the two chosen entries (i, j) and (k, l). Once we have chosen (x, y), there are at most $n/(4e)-1$ choices for another (z, w) with $m_{xy} = m_{zw}$. Thus the maximum degree in this negative dependence graph is at most $(4n-4)\left(\frac{n}{4e} - 1\right) \le \frac{n(n-1)}{e} - 1$. We can now apply the symmetric lopsided local lemma to conclude that with positive probability, none of the events A_{ijkl} occur. □

18.226 Probabilistic Method in Combinatorics
Fall 2020