

Focused Stochastic Local Search and the Lovász Local Lemma

Dimitris Achlioptas *
Department of Computer Science
University of California Santa Cruz

Fotis Iliopoulos†
Department of Electrical Engineering and Computer Science
University of California Berkeley

Abstract

We develop tools for analyzing focused stochastic local search algorithms. These are algorithms which search a state space probabilistically by repeatedly selecting a constraint that is violated in the current state and moving to a random nearby state which, hopefully, addresses the violation without introducing many new ones. A large class of such algorithms arise from the algorithmization of the Lovász Local Lemma, a non-constructive tool for proving the existence of satisfying states. Here we give tools that provide a unified analysis of such algorithms and of many more, expressing them as instances of a general framework.

*Research supported by NSF grant CCF-1514128.

†Research supported by NSF grant CCF-1514434.

1 Introduction

Let Ω be a large finite set of objects and let $F = \{f_1, f_2, \dots, f_m\}$ be a collection of subsets of Ω . We will refer to each $f_i \in F$ as a *flaw* to express that its elements have negative feature $i \in [m]$. For example, given a CNF formula on n variables with clauses c_1, c_2, \dots, c_m , we can define for each clause c_i the flaw (subcube) $f_i \subseteq \{0, 1\}^n$ whose elements violate c_i . Following linguistic rather than mathematical convention we say that f is present in σ if $f \ni \sigma$ and that $\sigma \in \Omega$ is *flawless* (perfect) if no flaw is present in σ .

Our goal is to develop tools for analyzing stochastic local search algorithms for finding perfect objects. Importantly, our analysis *will not* assume that Ω contains perfect objects, but rather will establish their existence by proving that an algorithm converges (quickly) to one. The general idea in stochastic local search is that Ω is equipped with a neighborhood structure and that the search starts at some element (state) of Ω and moves from state to state along the neighborhood structure. *Focused* local search corresponds to the case where each state change can be attributed to an effort to rid the state of some specific present flaw.

Concretely, for each $\sigma \in \Omega$, let $U(\sigma) = \{f \in F : \sigma \in f\}$, i.e., $U(\sigma)$ is the set of flaws present in σ . For every flaw $f_i \in U(\sigma)$, let $A(i, \sigma) \neq \{\sigma\}$ be a non-empty subset of Ω . We call the elements of $A(i, \sigma)$ *actions* and we consider the multi-digraph D on Ω which has an arc $\sigma \xrightarrow{i} \tau$ for every $\tau \in A(i, \sigma)$. We will consider walks on D which start at a state σ_1 , selected according to some probability distribution θ , and which at each non-sink vertex σ first select a flaw $f_i \ni \sigma$, as a function of the trajectory so far, and then select a next state $\tau \in A(i, \sigma)$ with probability $\rho_i(\sigma, \tau)$. Whenever a flaw $f_i \ni \sigma$ is selected we will say flaw f_i was addressed (which will not necessarily mean that f_i was eliminated, i.e., potentially $A(i, \sigma) \cap f_i \neq \emptyset$).

A large class of algorithms for dealing with the setting at this level of generality arise by algorithmizations of the Lovász Local Lemma (LLL). This is a non-constructive tool for proving the *existence* of flawless objects by introducing a probability measure μ on Ω , along the lines of the Probabilistic Method (throughout we assume that products devoid of factors evaluate to 1, i.e., $\prod_{x \in \emptyset} g(x) = 1$ for any g).

General LLL. *Let $\mathcal{A} = \{A_1, A_2, \dots, A_m\}$ be a set of m events. For each $i \in [m]$, let $D(i) \subseteq [m] \setminus \{i\}$ be such that $\mu(A_i | \cap_{j \in S} \overline{A_j}) = \mu(A_i)$ for every $S \subseteq [m] \setminus (D(i) \cup \{i\})$. If there exist positive real numbers $\{\psi_i\}_{i=1}^m$ such that for all $i \in [m]$,*

$$\frac{\mu(A_i)}{\psi_i} \sum_{S \subseteq \{i\} \cup D(i)} \prod_{j \in S} \psi_j \leq 1, \quad (1)$$

then the probability that none of the events in \mathcal{A} occurs is at least $\prod_{i=1}^m 1/(1 + \psi_i) > 0$.

Remark 1. *Condition (1) above is equivalent to the more well-known form $\mu(A_i) \leq x_i \prod_{j \in D(i)} (1 - x_j)$, where $x_i = \psi_i / (1 + \psi_i)$. As we will see, formulation (1) facilitates refinements.*

Erdős and Spencer [6] noted that independence in the LLL can be replaced by *negative correlation*, yielding the stronger Lopsided LLL. The difference is that each set $D(i)$ is replaced by a set $L(i) \subseteq [m] \setminus \{i\}$ such that $\mu(A_i | \cap_{j \in S} \overline{A_j}) \leq \mu(A_i)$ for every $S \subseteq [m] \setminus (L(i) \cup \{i\})$, i.e., “=” is replaced by “≤”.

In a landmark work [17], Moser and Tardos made the general LLL constructive for *product* measures over explicitly presented variables. Specifically, in the *variable setting* of [17], each event A_i is determined by a set of variables $\text{vbl}(A_i)$ so that $j \in D(i)$ iff $\text{vbl}(A_i) \cap \text{vbl}(A_j) \neq \emptyset$. Moser and Tardos proved that if (1) holds, then repeatedly selecting *any* occurring event A_i (flaw present) and resampling every variable in $\text{vbl}(A_i)$ independently of all others, leads to a flawless object after a linear expected number of resamplings. Pegden [20] extended the result of [17] to the cluster expansion criterion of Bissacott et al. [3],

and Kolipaka and Szegedy [13] extended it to Shearer’s criterion (the most general LLL criterion). Beyond the variable setting, Harris and Srinivasan in [11] algorithmized the general LLL for the uniform measure on permutations while, very recently, Harvey and Vondrák [12] algorithmized the Lopsided LLL up to Shearer’s criterion assuming efficiently implementable *resampling oracles*. Resampling oracles, introduced in [12], elegantly capture a common constraint in all prior algorithmizations of LLL, namely that state transitions must be “compatible” with the measure μ . Below we give the part of the definition of resampling oracles that exactly expresses this notion of compatibility, which we dub (measure) *regeneration*.

Regeneration. ([12])[*Resampling Oracles*] Say that (D, ρ) regenerate μ at flaw f_i if for every $\tau \in \Omega$,

$$\frac{1}{\mu(f_i)} \sum_{\sigma \in f_i} \mu(\sigma) \rho_i(\sigma, \tau) = \mu(\tau) . \quad (2)$$

Observe that the l.h.s. of (2) is the probability of reaching τ after first sampling a state $\sigma \in f_i$ according to μ and then addressing f_i at σ according to (D, ρ) . The requirement that this probability equals $\mu(\tau)$ for every $\tau \in \Omega$ means that in every state $\sigma \in f_i$ the distribution of actions for addressing f_i must be such that it removes the conditional $f_i \ni \sigma$. A trivial way to satisfy this (very stringent) requirement is to sample a new state σ' according to μ in each step (assuming μ is efficiently sampleable). Doing this, though, removes any sense of progress, as the set of flaws present in σ' are completely unrelated to those in σ . Instead, we would like to achieve (2) while limiting the set of flaws that may be present in σ' that were not present in σ . For example, note that in the variable setting resampling every variable in $\text{vbl}(f_i)$ independently satisfies (2), while only having the potential to introduce flaws that share at least one variable with f_i . It is, thus, natural to consider the following “projection” of the action digraph introduced in the flaws/actions framework of [1].

Potential Causality. For an arc $\sigma \xrightarrow{i} \tau$ in D and a flaw f_j present in τ we say that f_i causes f_j if $f_i = f_j$ or $f_j \not\ni \sigma$. If D contains any arc in which f_i causes f_j we say that f_i potentially causes f_j .

Potential Causality Digraph. The digraph $C = C(\Omega, F, D)$ on $[m]$ where $i \rightarrow j$ iff f_i potentially causes f_j is called the potential causality digraph. The neighborhood of a flaw f_i is $\Gamma(i) = \{j : i \rightarrow j \text{ exists in } C\}$.

In the interest of brevity we will call C the causality digraph, instead of the potential causality digraph. It is important to note that C contains an arc $i \rightarrow j$ if there exists *even one* state transition aimed at addressing f_i that causes f_j to appear in the new state.

As mentioned, very recently, Harvey and Vondrák [12] made the Lopsided LLL algorithmic, given resampling oracles for μ . Their result actually makes no reference to the lopsidedness condition, which they prove is implied by the existence of resampling oracles, and can be stated as follows.

Theorem 1 (Harvey-Vondrák [12]). *Let Ω, F, μ, D, ρ be such that (D, ρ) regenerate μ at flaw f_i for every $i \in [m]$. If $\theta = \mu$ and there exist positive real numbers $\{\psi_i\}_{i=1}^m$ such that for every $i \in [m]$,*

$$\frac{\mu(f_i)}{\psi_i} \sum_{S \subseteq \Gamma(i)} \prod_{j \in S} \psi_j < 1 , \quad (3)$$

then a perfect object can be found after polynomially many steps on D .

In fact, in [12] it was shown that the conclusion of Theorem 1 holds also if (3) is replaced by Shearer’s condition [21]. Thus, the work of Harvey and Vondrák [12] marks the end of the road for the derivation and analysis of focused stochastic local search algorithms by algorithmizations of the Lovász Local Lemma.

In this work we extend the flaws/actions algorithmic framework of [1] to arbitrary measures and action graphs and connect it to the Lovász Local Lemma. The result is a theorem that subsumes both Theorem 1 and all results of [1], establishing a method for designing and analyzing focused stochastic local search algorithms that goes far beyond algorithmizing the LLL.

Concretely, in [1] we introduced and analyzed focused local search algorithms where:

- D is *atomic*, i.e., for every $\tau \in \Omega$ and every $i \in [m]$ there exists at most one arc $\sigma \xrightarrow{i} \tau$.
- μ is the *uniform* measure on Ω , i.e., $\mu(\cdot) = |\Omega|^{-1}$.
- ρ assigns *equal* probability to every action in $A(i, \sigma)$, for every $f_i \in U(\sigma)$, at every flawed $\sigma \in \Omega$.

Here we generalize to **arbitrary** D, ρ, μ , allowing one to trade the sophistication of the measure μ against the sparsity of the causality graph (while removing the need to sample from μ , or to regenerate μ). Moreover, for the special case of uniform μ , we improve the condition of [1] for convergence. We state our results formally in the next section.

We also make a conceptual contribution by identifying for each measure μ certain pairs (D, ρ) as special.

Harmonic Walks. (D, ρ, μ) are harmonic if for every $i \in [m]$ and every transition $(\sigma, \tau) \in f_i \times A(i, \sigma)$,

$$\rho_i(\sigma, \tau) = \frac{\mu(\tau)}{\sum_{\sigma' \in A(i, \sigma)} \mu(\sigma')} \propto \mu(\tau) . \quad (4)$$

In words, when (D, ρ, μ) are harmonic ρ_i assigns to each state in $A(i, \sigma)$ probability proportional to its probability under μ . It is easy to see that (D, ρ, μ) are harmonic both in the algorithm of Moser and Tardos [17] for the variable setting and in the algorithm of Harris and Srinivasan [11] for the uniform measure on permutations. There are two reasons why harmonic (D, ρ, μ) combinations are interesting.

1.1 Resampling Oracles via Atomic Actions

If we start with the Probabilistic Method setup, i.e., Ω, F , and μ , then to get a constructive result by LLL algorithmization we must design (D, ρ) that regenerate μ at every flaw $f_i \in F$ (note that Theorem 1 *assumes* such (D, ρ) as input). While, in general, this can be a daunting task, we show that if we restrict our attention to D that are atomic matters are dramatically simplified:

- (D, ρ, μ) *must* be harmonic, yielding a *local* characterization of (D, ρ) at every state. (Theorem 6)
- The probability of every sequence of states is *characterized* by the flaws addressed. (Theorem 5)
- The initial state can be *arbitrary*, i.e., we can have $\theta \neq \mu$. (Theorems 2 – 4)

We note that *all* previous LLL algorithmizations, including [12], require $\theta = \mu$. We remove this at a mere cost of adding $O(\log |\Omega|)$ to the running time. This is beneficial in settings where sampling from μ (a global property) is hard, but nonetheless we can regenerate μ at every flaw (a local property).

Atomicity may initially seem artificial and/or restrictive. In reality, it is a very natural way to promote search space exploration, as it is equivalent to the following: $A(i, \sigma) \cap A(i, \sigma') = \emptyset$ for every $\sigma \neq \sigma' \in f_i$. Moreover, in most settings atomicity can be achieved in a straightforward manner. For example, in the variable setting the following two conditions combined imply atomicity:

1. Each flaw forbids exactly *one* joint value assignment to its underlying variables, i.e., is a subcube.
2. Each state transition modifies *only* the variables of the flaw it addresses.

Condition 1 expresses a primarily syntactic requirement: compound constraints must be broken down to constituent parts akin to satisfiability constraints. In most settings, not only is such a breakdown straightforward, but is also advantageous, as it affords a more refined accounting of conflict between constraints.

Condition 2 on the other hand reflects “focusing”, i.e., that every state transformation should be the result of attempting to eradicate some specific flaw f_i without interfering with variables not in $\text{vbl}(f_i)$.

1.2 Beyond Measure Regeneration

Designing (D, ρ) to regenerate μ at every flaw can be highly restrictive. This is commonly demonstrated by LLL’s inability to establish that a graph with maximum degree Δ can be colored with $q = \Delta + 1$ colors, one of the oldest and most vexing concerns about the LLL (see the survey of Szegedy [22]). This is because the regeneration of the uniform measure implies that to recolor a vertex v we must select uniformly among all colors, rather than colors not appearing in v ’s neighborhood, inducing a requirement of $q > e\Delta$ colors.

In [1] we introduced the flaws/actions framework to initiate the study of local search algorithms whose actions can depend arbitrarily on the state. In the aforementioned example this could mean choosing only among colors not appearing in v ’s neighborhood, so that as soon as $q \geq \Delta + 1$, the causality digraph becomes empty and rapid termination follows trivially. In [1], we required D to be atomic and actions to be chosen uniformly, a setting that in our current framework can be seen as the special case where μ is uniform, D is atomic, and (D, ρ, μ) are harmonic. Here we consider general harmonic (D, ρ, μ) , i.e., D need not be atomic and μ can be arbitrary, while (D, ρ) need not regenerate μ . Rather, μ is only used as a gauge of progress and deviation from regeneration is traded-off against sparsity of the causality graph.

The reason we focus on harmonic (D, ρ, μ) is that, as we will see, (i) they are optimal with respect to the aforementioned trade-off, and (ii) for every D and μ there exists ρ such that (D, ρ, μ) are harmonic.

2 Statement of Results

Definition 1. For $i \in [m]$ and $\tau \in \Omega$, let $b_i^\tau = |\{\sigma \in f_i : \tau \in A(i, \sigma)\}|$. For $i \in [m]$, let $b_i = \max_{\tau \in \Omega} b_i^\tau$. If $b_i = 1$ for all $i \in [m]$, then we say that D is atomic. (Note that $b_i > 0$ since $A(i, \sigma) \neq \emptyset$ for $\sigma \in f_i$.)

2.1 Setup

We establish general conditions under which focused stochastic local search algorithms find flawless objects quickly. Recall that any such algorithm performs a random walk on a multi-digraph D which (i) starts at a state $\sigma_1 \in \Omega$ selected according to a distribution θ , and which (ii) at each flawed state σ first selects some $f_i \ni \sigma$ to address and then selects $\tau \in A(i, \sigma)$ as the next state, where each $\tau \in A(i, \sigma)$ is selected with probability $\rho_i(\sigma, \tau)$. As one may expect the flaw-choice mechanism does have a bearing on the running time of such algorithms and we discuss this point in Section 2.6. Our results focus on conditions for rapid termination that do not require sophisticated flaw choice (but can be used in conjunction with such choice).

To establish a walk’s capacity to rid the state of flaws we introduce a measure μ on Ω . Without loss of generality, and to avoid certain trivialities, we assume that $\mu(\sigma) > 0$ for all $\sigma \in \Omega$. The choice of μ is entirely ours and can be trivial, i.e., $\mu(\cdot) = |\Omega|^{-1}$. Typically, μ assigns only exponentially small probability to flawless objects, yet allows us to prove that the walk reaches such an object in polynomial expected time. Its role is to define a “charge” $\gamma_i = \gamma_i(D, \theta, \rho, \mu)$ for each flaw $f_i \in F$, ideally as small as possible.

2.2 Definition of Flaw Charges

Regenerative case. If (D, ρ) regenerate μ at every $f_i \in F$ and either $\theta = \mu$ or D is atomic, then

$$\gamma_i = \mu(f_i) .$$

General case. Otherwise, $\gamma_i = b_i \max_{\sigma \in f_i} \lambda_i^\sigma$, where

$$\lambda_i^\sigma = \max_{\tau \in A(i, \sigma)} \left\{ \rho_i(\sigma, \tau) \frac{\mu(\sigma)}{\mu(\tau)} \right\} . \quad (5)$$

We will discuss several aspects of the definition of γ_i in Section 3. The main point is that the notion of charge allows us to state our results *without* having to distinguish between the regenerative and the general case, by simply substituting the appropriate charge. This also serves to highlight that the standard (Probabilistic Method) formulation of the LLL (and its algorithmizations) is, in fact, only a facet of a far more general picture, for which our results provide the first analytical tools. In the regenerative case, since $\gamma_i = \mu(f_i)$, our conditions will parallel those of the LLL (and its algorithmizations). In the general (non-regenerative) case, we will have $\gamma_i \geq \mu(f_i)$ always, but a potentially far sparser causality graph. To gain some first intuition for γ_i as a notion of congestion in the general case observe that if μ is uniform and D is atomic, then γ_i is simply the greatest transition probability $\rho_i(\sigma, \tau)$ on any arc originating in f_i . In general, it is the ergodic flow from σ to τ divided by the capacity, $\mu(\tau)$, of τ (and scaled by b_i).

To state our results we need a last definition regarding the distribution θ of the initial state.

Definition 2. The span of a probability distribution $\theta : \Omega \rightarrow [0, 1]$, denoted by $\mathcal{S}(\theta)$, is the set of flaws that may be present in a state selected according to θ , i.e., $\mathcal{S}(\theta) = \bigcup_{\sigma \in \Omega: \theta(\sigma) > 0} U(\sigma)$.

2.3 A Simple Markov Chain

Our first result concerns the simplest case where, after choosing an arbitrary permutation π of the flaws, the algorithm in each flawed state σ simply addresses the greatest flaw present in σ according to π . Observe that substituting $\gamma_i = \mu(f_i)$ for the regenerative case to Theorem 2 recovers the condition of Theorem 1.

Theorem 2. If there exist positive real numbers $\{\psi_i\}$ such that for every $i \in [m]$,

$$\zeta_i := \frac{\gamma_i}{\psi_i} \sum_{S \subseteq \Gamma(i)} \prod_{j \in S} \psi_j < 1 , \quad (6)$$

then for every π the walk reaches a sink within $(T_0 + s)/\delta$ steps with probability at least $1 - 2^{-s}$, where $\delta = 1 - \max_{i \in [m]} \zeta_i > 0$, and

$$T_0 = \log_2 \left(\max_{\sigma \in \Omega} \frac{\theta(\sigma)}{\mu(\sigma)} \right) + \log_2 \left(\sum_{S \subseteq \mathcal{S}(\theta)} \prod_{j \in S} \psi_j \right) .$$

Theorem 2 has two features worth discussing, both direct consequences of the generality of our framework, i.e., of abandoning the Probabilistic Method viewpoint and measure regeneration.

Arbitrary initial state. Since θ can be arbitrary in the general case, any foothold on Ω suffices to apply the theorem, without needing to sample from Ω according to some measure. This can also be interesting when we can not sample from μ , but can regenerate μ at every flaw on an atomic D , i.e., the second case of the

regenerative setting. Note also that T_0 captures the trade-off between the fact that when $\theta = \mu$ the first term in T_0 vanishes, but the second term grows to reflect the uncertainty of the set of flaws present in σ_1 .

Arbitrary number of flaws. The running time depends only on the span $|\mathcal{S}(\theta)|$, not the total number of flaws $|F|$. This has an implication analogous to the result of Haeupler, Saha, and Srinivasan [10] on core events: even when $|F|$ is very large, e.g., super-polynomial in the problem’s encoding length, we can still get an efficient algorithm if, for example, we can find a state σ_1 such that $|U(\sigma_1)|$ is small, e.g., by proving that in every state only polynomially many flaws may be present, or θ such that $|\mathcal{S}(\theta)|$ is small.

2.4 A Non-Markovian Algorithm

Our next results concerns the common setting where the subgraph induced by the neighborhood of each flaw in the causality graph contains several arcs. We improve Theorem 2 in such settings by employing a *recursive* algorithm. The flaw addressed in each step thus depends on the *entire* trajectory up that point not just the current state, i.e., the walk is non-Markovian. It is for this reason that we required a non-empty set of actions for every flaw present in a state, and why the definition of the causality digraph does not involve flaw choice. The improvement is that rather than summing over all subsets of $\Gamma(i)$ as in (6), we now only sum over *independent* such subsets, where f_i, f_j are dependent if $f_i \rightarrow f_j$ and $f_j \rightarrow f_i$. This improvement is similar to the cluster expansion improvement of Bissacot et al. [3] of the general LLL. As a matter of fact, Theorem 3 implies the algorithmic aspects of [3] (see [20] and [12]).

The use of a recursive algorithm affords an additional advantage, as it enables “responsibility shifting” between flaws. Specifically, for a fixed action digraph D with causality digraph C , the recursive algorithm (and Theorem 3), take as input *any* digraph $R \supseteq C$, i.e., allow for arcs to be added to the causality digraph. The reason for this as follows. While adding, say, arcs $f_i \rightarrow f_j$ and $f_j \rightarrow f_i$ may make the sums corresponding to f_i and f_j greater, if f_k is such that $\{f_i, f_j\} \subseteq \Gamma(k)$, then its sum may become smaller, as f_i, f_j are now dependent. As a result, such arc addition may enable a sufficient condition for rapid convergence to a perfect object, e.g., in our application on Acyclic Edge Coloring in Section 6. An analogous counter-intuitive phenomenon is also true in the improvement of Bissacot et al. [3] where denser dependency graphs may result to a better analysis.

Below, for $S \subseteq F$, we let $I_\pi(S)$ denote the greatest element of S according to π . For any fixed ordering π of F the recursive walk is the non-Markovian random walk on Ω that occurs by invoking procedure ELIMINATE. Observe that if in line 8 we do not intersect $U(\sigma)$ with $\Gamma_R(f_i)$ the recursion is trivialized and we recover the simple walk of Theorem 2.

Recursive Walk

- 1: **procedure** ELIMINATE
 - 2: $\sigma \leftarrow \theta(\cdot)$ ▷ Sample σ from θ
 - 3: **while** $U(\sigma) \neq \emptyset$ **do**
 - 4: ADDRESS ($I_\pi(U(\sigma)), \sigma$)
 - 5: **return** σ
 - 6: **procedure** ADDRESS(i, σ)
 - 7: $\sigma \leftarrow \tau \in A(i, \sigma)$ with probability $\rho_i(\sigma, \tau)$
 - 8: **while** $B = U(\sigma) \cap \Gamma_R(f_i) \neq \emptyset$ **do** ▷ Note $\cap \Gamma_R(f_i)$
 - 9: ADDRESS($I_\pi(B), \sigma$)
-

Definition 3. Given a digraph R on F let $G = G(R) = (F, E)$ be the undirected graph where $\{f, g\} \in E$ iff both $f \rightarrow g$ and $g \rightarrow f$ exist in R . For $S \subseteq F$, let $\text{Ind}(S) = \{S' \subseteq S : S' \text{ is an independent set in } G\}$.

Theorem 3. Let $R \supseteq C$ be arbitrary. If there exist positive real numbers $\{\psi_i\}$ such that for every $i \in [m]$,

$$\zeta_i := \frac{\gamma_i}{\psi_i} \sum_{S \in \text{Ind}(\Gamma_R(i))} \prod_{j \in S} \psi_j < 1, \quad (7)$$

then for every π the recursive walk reaches a sink within $(T_0 + s)/\delta$ steps with probability at least $1 - 2^{-s}$, where $\delta = 1 - \max_{i \in [m]} \zeta_i > 0$, and

$$T_0 = \log_2 \left(\max_{\sigma \in \Omega} \frac{\theta(\sigma)}{\mu(\sigma)} \right) + \log_2 \left(\sum_{S \subseteq \text{Ind}(\mathcal{S}(\theta))} \prod_{j \in S} \psi_j \right).$$

Remark 2. Theorem 3 strictly improves Theorem 2 since by taking $R = C$ (i) the summation in (7) is only over the subsets of $\Gamma_R(f)$ that are independent in G , instead of all subsets of $\Gamma_R(f)$ as in (6), and (ii) similarly for T_0 , the summation is only over the independent subsets of $\mathcal{S}(\theta)$, rather than all subsets of $\mathcal{S}(\theta)$.

Remark 3. Theorem 3 can be strengthened by introducing for each flaw $f \in F$ a permutation π_f of $\Gamma_R(f)$ and replacing π with π_f in line 9 of Recursive Walk. With this change in (7) it suffices to sum only over $S \subseteq \Gamma_R(f)$ satisfying the following: if the subgraph of R induced by S contains an arc $g \rightarrow h$, then $\pi_f(g) \geq \pi_f(h)$. As such a subgraph can not contain both $g \rightarrow h$ and $h \rightarrow g$ we see that $S \in \text{Ind}(\Gamma_R(f))$.

2.5 A General Theorem

Theorems 2 and 3 are instantiations of a general theorem we develop for establishing the success of focused local search algorithms by local considerations. Before presenting the theorem itself, we first briefly discuss its derivation, as that helps motivate and digest the theorem's form.

To bound the probability of not reaching a sink within t steps we partition the set of all t -trajectories into equivalence classes, bound the total probability of each class, and sum the bounds for the different classes. The partition is according to the t -sequence of flaws addressed, which acts as a *statistic* of the state distribution. Formally, for a trajectory $\Sigma = \sigma_1 \xrightarrow{w_1} \sigma_2 \xrightarrow{w_2} \dots$ we let $W(\Sigma) = w_1, w_2, \dots$ denote its *witness* sequence, i.e., the sequence of flaws addressed along Σ . We let $W_t(\Sigma) = \perp$ if Σ has fewer than t steps, otherwise we let $W_t(\Sigma)$ be the t -prefix of $W(\Sigma)$. Slightly abusing notation we let $W_t = W_t(\Sigma)$ be the random variable when Σ is the trajectory of the walk, i.e., selected according to (D, ρ, θ) and the flaw-choice mechanism. If $\mathcal{W}_t = \mathcal{W}_t(\mathcal{A})$ denotes the range of W_t for an algorithm \mathcal{A} , then the probability that \mathcal{A} takes t or more steps, trivially, is $\sum_{W \in \mathcal{W}_t} \Pr[W_t = W]$.

Key to our analysis is deriving upper bounds for $\Pr[W_t = W]$ that factorize over the elements of W . Specifically, for an arbitrary sequence of flaws $A = a_1, \dots, a_t$, let us denote by $[i]$ the index $j \in [m]$ such that $a_i = f_j$. Lemma 1 holds for both the regenerative and the general case, with the corresponding γ_i . Moreover, we will see that it can be tight, up to the prefactor ξ .

Lemma 1. Let $\xi = \xi(\theta, \mu) = \max_{\sigma \in \Omega} \{\theta(\sigma)/\mu(\sigma)\}$. For every sequence of flaws $A = a_1, \dots, a_t$,

$$\Pr[W_t = A] \leq \xi \prod_{i=1}^t \gamma_{[i]}.$$

The product form of the bound in Lemma 1 allows us to combine it with different collections of forests, each collection expressing an upper bound for (superset of) the set of all possible witness sequences \mathcal{W}_t . The formulation of the supersets as forests enables the combinatorial enumeration of their elements (flaw sequences) which, combined with Lemma 1, yields Theorem 4 below. While \mathcal{W}_t depends on flaw-choice, the main and common feature of all bounds (forests) is the enforcement of the following idea: while the very first occurrence of a flaw f_j in a witness sequence W may be attributed to $f_j \ni \sigma_1$, every subsequent occurrence of f_j must be preceded by a distinct earlier occurrence of a flaw f_i that can “assume responsibility” for f_j , e.g., a flaw f_i that potentially causes f_j . In this way, the set \mathcal{W}_t is bounded *syntactically* by differently sophisticated considerations of flaw-choice and responsibility. Specifically, Definition 4 below (i) imposes a modicum of control over flaw-choice, while (ii) generalizing the subsets of flaws for which a flaw f may be responsible from subsets of $\Gamma(f)$ to arbitrary subsets of flaws, thus enabling responsibility shifting.

Definition 4. *Given (D, ρ, θ) , a flaw-choice mechanism is traceable if there exist sets $\text{Roots}(\theta) \subseteq 2^F$ and $\text{List}(f_1) \subseteq 2^F, \dots, \text{List}(f_m) \subseteq 2^F$ such that for every $t \geq 1$, the set of all possible witness sequences \mathcal{W}_t can be injected into unordered rooted forests with t vertices that have the following properties:*

1. *Each vertex of the forest is labeled by a flaw $f_i \in F$.*
2. *The flaws labeling the roots of the forest are distinct and form an element of $\text{Roots}(\theta)$.*
3. *The flaws labeling the children of each vertex are distinct.*
4. *If a vertex is labelled by flaw f_i , then the labels of its children form an element of $\text{List}(f_i)$.*

To recover the witness sequence from a forest, thus demonstrating the injection of \mathcal{W}_t , we make use of the specificity of the mechanism for selecting which flaw to address at each step. For example, the forests that correspond to the algorithm of Theorem 3 are “recursion forests”, having one node for each recursive call of ADDRESS, labelled by the flaw that is the call’s argument. To recover the sequence of addressed flaws, we order the trees in the forest and the progeny of each vertex using knowledge of π and then traverse each tree in the recursive forest in postorder. We explain why the algorithms of Theorems 2 and 3 are traceable in Appendix A, where we describe the set of witness forests that correspond to each theorem.

Theorem 4 below implies both Theorem 2 and Theorem 3. While those two theorems do not care about the flaw ordering π , Theorem 4 also captures the “LeftHanded Random Walk” result of [1] (motivated by the LeftHanded version of the LLL introduced by Pedgen [19]), under which the flaw order π can be chosen in a *provably* beneficial way, yielding a “responsibility” digraph. That is, both in the regenerative case and in the general case, one can use our γ_i as the charge for each event in the responsibility digraph of [1].

Theorem 4 (Main result). *If \mathcal{A} results by applying a traceable flaw-choice mechanism on (D, ρ, θ) and there exist positive real numbers $\{\psi_i\}$ such that for every flaw $f_i \in F$,*

$$\zeta_i := \frac{\gamma_i}{\psi_i} \sum_{S \in \text{List}(f_i)} \prod_{j \in S} \psi_j < 1, \quad (8)$$

then a sink is reached within $(T_0 + s)/\delta$ steps with probability at least $1 - 2^{-s}$, where $\delta = 1 - \max_{i \in [m]} \zeta_i$ and

$$T_0 = \log_2 \left(\max_{\sigma \in \Omega} \frac{\theta(\sigma)}{\mu(\sigma)} \right) + \log_2 \left(\sum_{S \in \text{Roots}(\theta)} \prod_{j \in S} \psi_j \right).$$

2.5.1 Proofs of Theorems 2 and 3 from Theorem 4

In Appendix A, we describe the *Break Forests* and *Recursive Forests*, into which we inject the witness sequences of the algorithms of Theorems 2 and 3, respectively. Theorem 2 follows from Theorem 4 as Break Forests satisfy the conditions of Theorem 4 with $\text{Roots}(\theta) = 2^{\mathcal{S}(\theta)}$ and $\text{List}(f) = 2^{\Gamma_R(f)}$. Theorem 3 follows as Recursive Forests satisfy the conditions with $\text{Roots}(\theta) = \text{Ind}(\mathcal{S}(\theta))$ and $\text{List}(f) = \text{Ind}(\Gamma_R(f))$.

2.6 A Sharp Analysis and the Role of Flaw Choice

In Section 4.3 we prove that Lemma 1 is tight for a rather large class of algorithms, including the algorithm of Moser-Tardos [17] when each flaw fixes the values of its variables, as in SAT, and the algorithm of Harris and Srinivasan for permutations [11].

Theorem 5. *Let $\beta = \min_{\sigma \in \Omega} \mu(\sigma) > 0$. If (D, ρ) regenerate μ at every flaw and D is atomic, then for every $W = w_1, w_2, \dots, w_t \in \mathcal{W}_t$,*

$$\beta \leq \frac{\Pr[W_t = W]}{\prod_{i=1}^t \mu(w_i)} \leq \beta^{-1} . \quad (9)$$

Equation (9) tells us that an algorithm will converge to a perfect object in polynomial time *if and only if* the sum $\sum_{W \in \mathcal{W}_t} \prod_{i=1}^t \mu(w_i)$ converges to a number less than 1 as t grows. In that sense, the quality of the algorithm’s analysis depends solely on how well we approximate the set of possible witness sequences \mathcal{W}_t .

The set \mathcal{W}_t is clearly a function of how we choose which flaw to address in each step and therefore algorithmic performance clearly depends on the flaw choice mechanism (even more so, in this “tight” case). However, in the Moser-Tardos analysis, as well as in the work of Harris and Srinivasan on permutations [11], the flaw choice mechanism “is swept under the rug” [22] and is allowed to be arbitrary. This can be explained as follows. In those two settings, due to the symmetry of Ω , we can afford to approximate \mathcal{W}_t in a way that completely ignores flaw choice, i.e., considering it adversarial, and still recover the LLL condition. In a very recent paper [14], Kolmogorov gives a more general symmetry condition under which the results of [1] for the flaws/actions framework hold with arbitrary flaw choice. However, such symmetries can not be expected to hold in general settings, something reflected in Theorems 2 and 3 in the specificity of the flaw-choice mechanism, while in Theorem 4 it is reflected in the requirement of traceability.

2.7 Applications: Incorporating Global Conditions

To demonstrate the power of our framework we derive a novel bound for acyclic edge colorings, aimed at graphs of bounded degeneracy, a class including graphs of bounded treewidth. To get the result we heavily use the fact that we do not have to regenerate a measure (and so the result cannot be captured by the LLL). Unlike recent work on the problem [7, 9] that also goes beyond the LLL, our result is established without any problem specific elements, but rather as a direct application of Theorem 3.

3 Charging Flaws

In Section 2.1 we defined how to assign to each flaw f_i a charge γ_i , depending on whether (D, ρ) regenerate μ or not. We also stated that in the non-regenerative case $\gamma_i \geq \mu(f_i)$ always. Thus, ideally, we would like to use a sophisticated measure μ that assigns minimal probability mass to the flawed states and, at the same time, have (D, ρ, μ) that regenerate μ at every flaw. In reality, the more sophisticated μ is the harder regeneration becomes. Therefore, realistically, we can either employ (D, ρ, μ) that regenerate μ at every

flow and get charges as small as possible, but for unsophisticated measures (like product measures), or we can forgo regeneration, pay the price $\gamma_i > \mu(f_i)$ to reflect the distortion of μ by (D, ρ) , and use more sophisticated measures. Crucially, making the latter choice typically also means that we can get a sparser causality graph by exploiting the flexibility afforded by not having to design D so as to regenerate μ , as in the non-regenerative case D can be arbitrary.

Lemma 2. $\gamma_i \geq \mu(f_i)$.

Proof. By the definition of γ_i , in the regenerative case we have equality, while in the non-regenerative case,

$$\frac{\mu(f_i)}{\gamma_i} \leq \sum_{\sigma \in f_i} \frac{\mu(\sigma)}{b_i \lambda_i^\sigma} = \sum_{\sigma \in f_i} \sum_{\tau \in A(i, \sigma)} \frac{\mu(\sigma)}{b_i \lambda_i^\sigma} \rho_i(\sigma, \tau) \leq \sum_{\sigma \in f_i} \sum_{\tau \in A(i, \sigma)} \frac{\mu(\tau)}{b_i} \leq 1, \quad (10)$$

where for the last inequality we used that $\sum_{\sigma \in f_i} \sum_{\tau \in A(i, \sigma)}$ enumerates every $\tau \in \Omega$ at most b_i times. \square

Observe that for any pair (D, μ) taking ρ so that (D, ρ, μ) are harmonic, i.e., taking $\rho_i(\sigma, \tau) \propto \mu(\tau)$, minimizes λ_i^σ for all $\sigma \in f_i$ *simultaneously*. This optimality is the main reason that motivates harmonic algorithms. The other reason is the realization that designing (D, ρ) that regenerate μ at every flow is often achieved by (D, ρ, μ) being harmonic. As matter of fact, as we state in Theorem 6 below, if D is atomic then (D, ρ, μ) being harmonic is *necessary* for regeneration, a fact that also yields a characterization of the local structure of atomic digraphs regenerating a measure.

As a final remark, we note that all results that correspond to “algorithmizations of the LLL” correspond to the (very) special case where (D, ρ) regenerate μ at every $f_i \in F$.

3.1 The atomic case

Atomic digraphs capture algorithms that appear in several settings, e.g., the Moser-Tardos algorithm [17] when constraints are in CNF form, the algorithm of Harris and Srinivasan for permutations [11], and others (see [1]). Theorem 6 below asserts that (D, ρ, μ) being harmonic is a necessary condition for regeneration when D is atomic. Observe that (D, ρ, μ) being harmonic means that we need not be concerned with the design of ρ as it is implied by (D, μ) . As for D itself, the theorem implies that to build $A(i, \sigma)$ we must “collect” arcs that satisfy (11) (while, presumably, keeping the causality graph as sparse as possible). These two facts offer guidance in designing D so that (D, ρ) regenerate μ at every flow in atomic digraphs.

Theorem 6. *If D is atomic and (D, ρ) regenerate μ at every flow, then (D, ρ, μ) are harmonic. Moreover, for every $i \in [m]$ and every $\sigma \in f_i$,*

$$\sum_{\tau \in A(i, \sigma)} \mu(\tau) = \frac{\mu(\sigma)}{\mu(f_i)}. \quad (11)$$

Proof. If D is atomic, $\mu > 0$, and (D, ρ) regenerate μ at every flow f_i , it follows that for every $\tau \in \Omega$ there is *exactly* one $\sigma \in f_i$ such that $\rho_i(\sigma, \tau) > 0$. (And also that $\bigcup_{\sigma \in f_i} A(i, \sigma) = \Omega$). Therefore, regeneration at f_i in this setting is equivalent to

$$\text{for every } \sigma \in f_i \text{ and the unique } \tau \in A(i, \sigma): \quad \rho_i(\sigma, \tau) = \mu(\tau) \frac{\mu(f_i)}{\mu(\sigma)}. \quad (12)$$

(Note that for given D, μ there may be no ρ satisfying (12), as we also need that $\sum_{\tau \in A(i, \sigma)} \rho_i(\sigma, \tau) = 1$.)

Since $\rho_i(\sigma, \tau) \propto \mu(\tau)$ in (12) we see that ρ is harmonic. Summing (12) over $\tau \in A(i, \sigma)$ yields (11). \square

3.2 Improved charges for the uniform measure case

As mentioned, the framework of [1] amounts to the case where D is atomic, μ is uniform, and (D, ρ, μ) is harmonic, so that in every step a uniformly random element of $A(i, \sigma)$ is selected. When μ is uniform, we prove in Section 4.2 that γ_i is the inverse of $\min_{\sigma \in f_i} a_i^\sigma$, where $a_i^\sigma = |A(i, \sigma)|$ and, thus, our Theorems 2 and 3 recover perfectly the results of [1].

To deal with the case where D is not naturally atomic, e.g., when a flaw occurs under more than one value assignments to its variables, one can proceed to refine each $f_i \in F$ into $b_i = \max_{\sigma \in f_i} b_i^\sigma$ flaws so that D becomes atomic. In subsection 4.4, using the machinery developed for the general case in Section 4.2, we remove the need to “atomize” D and derive bounds dominating those that come from atomization. To achieve this we modify the walk so that each element $\tau \in A(i, \sigma)$ is selected with probability proportional to the inverse of its in-degree, i.e., $\rho_i(\sigma, \tau) = (b_i^\tau \sum_{\sigma' \in A(i, \sigma)} 1/b_i^{\sigma'})^{-1}$. Doing this, we get that the charge we should assign to each flaw f_i (assuming we are in the general case with uniform measure) is:

$$\phi_i = \max_{(\sigma, \tau) \in D_i} \frac{b_i^\tau}{a_i^\sigma} \leq \frac{b_i}{a_i} ,$$

where the right hand side above is the charge on flaw f_i that would be assigned by atomization, potentially much greater than our bound $\max_{(\sigma, \tau) \in D_i} b_i^\tau / a_i^\sigma$.

4 Proof of Lemma 1

In Section 4.1 we give the proof for the regenerative case when the measure μ is sampleable (while (D, ρ, μ) are not necessarily harmonic). The proof for that case mimics that of [9] and [12]. In Section 4.2 we show the proof for the general (non-regenerative) case. The proof for the regenerative case when μ is not sampleable but D is atomic, is given as a special case of that proof in Section 4.3. Finally, in Section 4.4 we show how to get the improved bounds for the general case with uniform measure, described in Section 3.2.

4.1 The Regenerative Case with Sampleable μ

We start by proving Lemma 1 for the case where $\theta = \mu$ and (D, ρ) regenerate μ at every flaw.

Lemma 3. *If $\theta = \mu$ and (D, ρ) regenerate μ at every $f_i \in F$, then for every sequence $W = w_1, \dots, w_t$,*

$$\Pr[W_t = W] \leq \prod_{i=1}^t \mu(w_i) .$$

Proof. To bound $\Pr[W_t = W]$ we will drop the requirement that flaw w_i is selected by the flaw-choice mechanism at σ_i and only require that $\sigma_i \in w_i$. To bound this latter probability we introduce a randomized process C which given as input an arbitrary sequence of flaws a_1, a_2, \dots, a_t proceeds as follows:

- Select a state σ_1 according to θ and set $\text{fail}(0) = 0$
- For i from 1 to t do:
 - If $\sigma_i \notin a_i$, then set $\text{fail}(i) = 1$ else set $\text{fail}(i) = \text{fail}(i - 1)$
 - Address a_i at σ_i according to (D, ρ) , i.e., set $\sigma_{i+1} = \tau \in A(i, \sigma)$ with probability $\rho_i(\sigma_i, \tau)$

Observe that the sequence $\text{fail}(i)$ is non-decreasing. By coupling C with the algorithm we readily get

$$\Pr_C [W_t = w_1, \dots, w_t] \leq \Pr_C \left[\bigcap_{i=1}^t \sigma_i \in w_i \right] . \quad (13)$$

For $t \geq 0$, let $P(t)$ be the proposition: for every $A = a_1, \dots, a_t$ and every $\tau \in \Omega$, on input A

$$\Pr_C [\sigma_{t+1} = \tau | \text{fail}(t) = 0] = \mu(\tau) . \quad (14)$$

We will prove that $P(t)$ holds for all $t \geq 0$ by induction on t . Along (13) this readily implies the lemma.

$P(0)$ follows from $\theta = \mu$. Assume that $P(t)$ holds for all $t < s$ and consider any sequence a_1, \dots, a_s . If $\text{fail}(s) = 0$ then $\text{fail}(s-1) = 0$ as well which, by $P(s-1)$, implies that σ_s is distributed according to μ . Moreover, $\sigma_s \in a_s$ and, therefore, σ_{s+1} is selected by addressing a_s at σ_s . Therefore for every state $\tau \in \Omega$,

$$\Pr_C [\sigma_{s+1} = \tau | \text{fail}(s) = 0] = \sum_{\sigma \in a_s} \frac{\mu(\sigma)}{\mu(a_s)} \rho_{[s]}(\sigma, \tau) = \mu(\tau) ,$$

where the second equality holds because (D, ρ) regenerate μ every flaw in F . \square

4.2 The General Case

Since the flaw addressed in each step depends only on the trajectory up to that point and not on any future randomness, the probability of any specific t -trajectory $\Sigma = \sigma_1 \xrightarrow{w_1} \sigma_2 \xrightarrow{w_2} \dots \sigma_t \xrightarrow{w_t} \sigma_{t+1}$ is

$$\theta(\sigma_1) \prod_{i=1}^t \rho_{[i]}(\sigma_i, \sigma_{i+1}) , \quad (15)$$

where recall that $[i] = j$ such that $w_i = f_j$. To bound $\Pr[W_t = W]$ we will sum (15) over all t -trajectories with witness sequence W .

Recall that $b_i^\tau = |\{\sigma \in f_i : \tau \in A(i, \sigma)\}|$. For each pair $\langle W, \sigma_{t+1} \rangle$, where $\sigma_{t+1} \in \Omega$, we construct an edge-weighted tree as follows. The root of the tree is σ_{t+1} . Let $P_t = \{\sigma : \sigma_{t+1} \in A([t], \sigma)\}$, i.e., P_t contains the $b_{[t]}^{\sigma_{t+1}}$ states in D with an arc to σ_{t+1} labelled $f_{[t]}$ which, thus, are the possible states immediately prior to σ_{t+1} in any trajectory with witness sequence W . The progeny of the root consists of a child for each $\sigma_t \in P_t$, each parent-child edge weighted by $\rho_{[t]}(\sigma_t, \sigma_{t+1})$. Each child vertex acquires progeny in the same manner, i.e., it has one child per possible immediately prior state, the corresponding edge annotated by $\rho_{[t-1]}(\sigma_{t-1}, \sigma_t)$. And so on, until W is exhausted. The constructed tree has the following properties:

- Its root-to-leaf paths are in 1–1 correspondence with the trajectories compatible with $\langle W, \sigma_{t+1} \rangle$.
- The product of the numbers along each root-to-leaf path (trajectory) equals $\prod_{i=1}^t \rho_{[i]}(\sigma_i, \sigma_{i+1})$.

Thus, to compute the probability of all trajectories with witness sequence W and final state σ_{t+1} it suffices to sum, over all root-to-leaf paths, the product of the probability of each path's leaf vertex under θ with the product of the weights along the path's edges.

To bound this sum for a measure μ we define for every $i \in [m]$,

$$\xi_i^\sigma = \xi_i^\sigma(\mu) = \mu(\sigma) \max_{\tau \in A(i, \sigma)} \rho_i(\sigma, \tau) \frac{b_i^\tau}{\mu(\tau)} . \quad (16)$$

Observe that, by definition, for every (σ, τ) such that $\tau \in A(i, \sigma)$ we have $b_i^\tau \geq 1$ and, therefore,

$$\rho_i(\sigma, \tau) \leq \frac{\xi_i^\sigma}{b_i^\tau} \frac{\mu(\tau)}{\mu(\sigma)} . \quad (17)$$

Therefore, we can bound the probability of any t -trajectory (root-to-leaf path) $\Sigma = \sigma_1, \dots, \sigma_{t+1}$ by

$$\theta(\sigma_1) \prod_{i=1}^t \rho_{[i]}(\sigma_i, \sigma_{i+1}) \leq \theta(\sigma_1) \prod_{i=1}^t \frac{\xi_{[i]}^\sigma}{b_{[i]}^{\sigma_{i+1}}} \frac{\mu(\sigma_{i+1})}{\mu(\sigma_i)} = \frac{\theta(\sigma_1)}{\mu(\sigma_1)} \mu(\sigma_{t+1}) \prod_{i=1}^t \frac{\xi_{[i]}^\sigma}{b_{[i]}^{\sigma_{i+1}}} . \quad (18)$$

If in the tree we now replace the weight $\rho_{[i]}(\sigma_i, \sigma_{i+1})$ of every edge by $\xi_{[i]}^\sigma / b_{[i]}^{\sigma_{i+1}}$ and the weight of every leaf by $\xi \mu(\sigma_{t+1})$, where $\xi = \max_{\sigma \in \Omega} \{\theta(\sigma) / \mu(\sigma)\}$, we see by (18) that the aforementioned sum over all root-to-leaf paths will give an upper bound on the total probability of trajectories with witness sequence W and final state σ_{t+1} . A key thing to observe is that after this edge-weight replacement, any vertex of the tree, say one corresponding to a state τ , will have some number $b_{[i]}^\tau$ children, while each of its parent-child edges will have weight $\xi_{[i]}^\sigma / b_{[i]}^\tau$, for some state σ . This fact puts us in a position to perform the summation.

For each $i \in [m]$, let $\phi_i = \max_{\sigma \in f_i} \xi_i^\sigma$. Consider any vertex v of the tree whose children are leaves and let τ be v 's state. Replacing the $b_{[1]}^\tau$ children of v with a single leaf child, connected to v by an edge of weight $\phi_{[1]}$, can only increase the contribution to the sum of the subtree rooted at v since $b_{[1]}^\tau \times \xi_{[1]}^\sigma / b_{[1]}^\tau = \xi_{[1]}^\sigma \leq \phi_i$. Proceeding to collapse the progeny of all other vertices in the same level of the tree as v and then moving on to the next level, etc. collapses the entire tree to a single path whose product of edge-weights is $\prod_{i=1}^t \phi_{[i]}$ and whose leaf was weight $\xi \mu(\sigma_{t+1})$, implying

$$\Pr[W_t = W] \leq \sum_{\sigma_{t+1} \in \Omega} \xi \mu(\sigma_{t+1}) \prod_{i=1}^t \phi_{[i]} = \xi \prod_{i=1}^t \phi_{[i]} .$$

To conclude the proof for the general case observe that for every $i \in [m]$,

$$\phi_i = \max_{\sigma \in f_i} \xi_i^\sigma = \max_{\sigma \in f_i} \left\{ \mu(\sigma) \max_{\tau \in A(i, \sigma)} \rho_i(\sigma, \tau) \frac{b_i^\tau}{\mu(\tau)} \right\} \leq \max_{\tau \in \Omega} b_i^\tau \max_{\sigma \in f_i} \left\{ \mu(\sigma) \max_{\tau \in A(i, \sigma)} \frac{\rho_i(\sigma, \tau)}{\mu(\tau)} \right\} = \gamma_i . \quad (19)$$

4.3 The Atomic Case and Proof of Theorem 5

If D is atomic and (D, ρ) regenerate μ at every f_i , Theorem 6 implies that (D, ρ, μ) are harmonic and thus

$$\xi_i^\sigma = \lambda_i^\sigma = \max_{\tau \in A(i, \sigma)} \left\{ \rho_i(\sigma, \tau) \frac{\mu(\sigma)}{\mu(\tau)} \right\} = \max_{\sigma \in f_i} \frac{\mu(\sigma)}{\sum_{\sigma' \in A(i, \sigma)} \mu(\sigma')} = \mu(f_i) , \quad (20)$$

where the last equality follows from (11). This establishes the regenerative case of Lemma 1 for atomic D .

To prove Theorem 5 we note that Lemma 1, valid for any (D, ρ, μ, θ) , readily yields the upper bound. For the lower bound we observe that in order for $W \in \mathcal{W}_t$ there must exist at least one trajectory Σ^* such that $W_t(\Sigma^*) = W$. Since, by (20), we have $\lambda_i^\sigma = \rho_i(\sigma, \tau) \mu(\sigma) / \mu(\tau)$ we can conclude that

$$\Pr[W_t = W] \geq \Pr[\Sigma = \Sigma^*] = \theta(\sigma_1^*) \prod_{i=1}^t \rho_{[i]}(\sigma_i^*, \sigma_{i+1}^*) = \theta(\sigma_1^*) \prod_{i=1}^t \lambda_{[i]}^{\sigma_i^*} \frac{\mu(\sigma_{i+1}^*)}{\mu(\sigma_i^*)} = \mu(\sigma_{t+1}) \prod_{i=1}^t \mu(w_i) .$$

4.4 The Special Case of the Uniform Measure

To get the improved bounds for the uniform measure recall that $a_i = \min_{\sigma \in f_i} a_i^\sigma = \min_{\sigma \in f_i} |A(i, \sigma)|$. If D_i is the subgraph of D comprising the arcs labeled by f_i , then (19) yields

$$\phi_i = \max_{\sigma \in f_i} \xi_i^\sigma = \max_{\sigma \in f_i} \max_{\tau \in A(i, \sigma)} \rho_i(\sigma, \tau) b_i^\tau = \max_{(\sigma, \tau) \in D_i} \frac{b_i^\tau}{a_i^\sigma} \leq \frac{b_i}{a_i}.$$

5 Proof of Theorem 4

Per the hypothesis of Theorem 4, each bad t -trajectory on D is associated with a rooted labeled witness forest with t vertices such that given the forest we can reconstruct the sequence of flaws addressed along the t -trajectory. Recall that neither the trees, nor the nodes inside each tree in the witness forest are ordered. To prove Theorem 4 we will give T_0 such that the probability that a $(T_0 + s)$ -trajectory on D is bad is exponentially small in s . Let $\widetilde{\mathcal{W}}_t \supseteq \mathcal{W}_t$ be the set of witness sequences of size t that correspond to these forests (sometimes, when it is clear from the context, we abuse the notation and use \mathcal{W}_t to denote the set of witness forests themselves). Per our discussion above (see Lemma 1) to prove the theorem it suffices to prove that $\max_{\sigma \in \Omega} \frac{\theta(\sigma)}{\mu(\sigma)} \sum_{W \in \widetilde{\mathcal{W}}_t} \prod_{i=1}^t \gamma_{[i]}$ is exponentially small in s for $t = T_0 + s$.

To facilitate counting we fix an arbitrary ordering π of F and map each witness forest into the unique ordered forest that results by ordering the trees in the forest according to the labels of their roots and similarly ordering the progeny of each vertex according to π (recall that both the flaws labeling the roots and the flaws labeling the children of each vertex are distinct).

Having induced this ordering for the purpose of counting, we will encode each witness forest as a rooted, ordered d -ary forest T with exactly t nodes, where $d = \max_{f \in F} |\text{List}(f)|$. In a rooted, ordered d -ary forest both the roots and the at most d children of each vertex are ordered. We think of the root of T as having reserved for each flaw $f \in \text{Roots}(\theta)$ a slot. If $f \in \text{Roots}(\theta)$ is the i -th largest flaw in F according to ψ then we fill the i -th slot (recall that the flaws labeling the roots of the witness forest are distinct and that, as a set, belong in the set $\text{Roots}(\theta)$).

Each node v of T corresponds to a node of the witness forest and therefore to a flaw f that was addressed at some point in the t -trajectory of the algorithm. Recall now that each node in the witness forest that is labelled by a flaw f has children labelled by distinct flaws in $\text{List}(f)$. We thus think of each node v of T as having precisely $|\text{List}(f)|$ slots reserved for each flaw $g \in \text{List}(f)$ (and, thus, at most d reserved slots in total). For each $g \in \text{List}(f)$ we fill the slot reserved for g and make it a child of v in T . Thus, from T we can reconstruct the sequence of flaws addressed with the algorithm. To proceed, we use ideas from [20]. Specifically, we introduce a branching process that produces only ordered d -ary forests that correspond to witness forests and bound $\sum_{W \in \widetilde{\mathcal{W}}_t} \prod_{i=1}^t \gamma_{[i]}$ by analyzing it.

Given any real numbers $0 < \psi_i < \infty$ we define $x_i = \frac{\psi_i}{\psi_i + 1}$ and write $\text{Roots}(\theta) = \text{Roots}$ to simplify notation. To start the process we produce the roots of the labeled forest by rejection sampling as follows: For each flaw $g \in F$ independently, with probability x_g we add a root with label g . If the resulting set of roots is in Roots we accept the birth. If not, we delete the roots created and try again. In each subsequent round we follow a very similar procedure. Specifically, at each step, each node u with label ℓ “gives birth”, again, by rejection sampling: For each flaw $g \in \text{List}(\ell)$ independently, with probability x_g we add a vertex with label g as a child of u . If the resulting set of children of u is in $\text{List}(\ell)$ we accept the birth. If not, we delete the children created and try again. It is not hard to see that this process creates every possible witness forest with positive probability. Specifically, for a vertex labeled by ℓ , every set $S \notin \text{List}(\ell)$ receives probability

0, while every set $S \in \text{List}(\ell)$ receives probability proportional to

$$w_\ell(S) = \prod_{g \in S} x_g \prod_{h \in (\text{List}(\ell)) \setminus S} (1 - x_h) .$$

To express the exact probability received by each $S \in \text{List}(\ell)$ we define

$$Q(S) := \frac{\prod_{g \in S} x_g}{\prod_{g \in S} (1 - x_g)} = \prod_{g \in S} \psi_g \quad (21)$$

and let $Z_\ell = \prod_{f \in (\text{List}(\ell))} (1 - x_f)$. We claim that $w_\ell(S) = Q(S) Z_\ell$. To see the claim observe that

$$\frac{w_\ell(S)}{Z_\ell} = \frac{\prod_{g \in S} x_g \prod_{h \in (\text{List}(\ell)) \setminus S} (1 - x_h)}{\prod_{f \in \text{List}(\ell)} (1 - x_f)} = \frac{\prod_{g \in S} x_g}{\prod_{g \in S} (1 - x_g)} = Q(S) .$$

Therefore, each $S \in \text{List}(\ell)$ receives probability equal to

$$\frac{w_\ell(S)}{\sum_{B \in \text{List}(\ell)} w_\ell(B)} = \frac{Q(S) Z_\ell}{\sum_{B \in \text{List}(\ell)} Q(B) Z_\ell} = \frac{Q(S)}{\sum_{B \in \text{List}(\ell)} Q(B)} . \quad (22)$$

Similarly, each set $R \in \text{Roots}$ receives probability equal to $Q(R) (\sum_{B \in \text{Roots}} Q(B))^{-1}$.

Lemma 4. *The branching process described above produces every tree $\phi \in \widetilde{\mathcal{W}}_t$ with probability*

$$p_\phi = \left(\sum_{S \in \text{Roots}} \prod_{i \in S} \psi_i \right)^{-1} \prod_{v \in \phi} \frac{\psi_v}{\sum_{S \in \text{List}(v)} Q(S)}$$

Proof. For each node v of ϕ let $N(v)$ denote the set of labels of its children. By (22),

$$\begin{aligned} p_\phi &= \frac{Q(R)}{\sum_{S \in \text{Roots}} Q(S)} \prod_{v \in \phi} \frac{Q(N(v))}{\sum_{S \in \text{List}(v)} Q(S)} \\ &= \frac{Q(R)}{\sum_{S \in \text{Roots}} Q(S)} \cdot \frac{\prod_{v \in \phi \setminus R} \psi_v}{\prod_{v \in \phi} \sum_{S \in \text{List}(v)} Q(S)} \\ &= \left(\sum_{S \in \text{Roots}} Q(S) \right)^{-1} \prod_{v \in \phi} \frac{\psi_v}{\sum_{S \in \text{List}(v)} Q(S)} . \end{aligned}$$

□

Notice now that

$$\begin{aligned} \sum_{W \in \widetilde{\mathcal{W}}_t} \prod_{i=1}^t \gamma^{[i]} &= \sum_{W \in \widetilde{\mathcal{W}}_t} \prod_{i=1}^t \frac{\zeta_{[i]} \psi_{[i]}}{\sum_{S \in \text{List}([i])} Q(S)} \\ &\leq \left(\max_{i \in F} \zeta_i \right)^t \sum_{W \in \widetilde{\mathcal{W}}_t} \prod_{i=1}^t \frac{\psi_{[i]}}{\sum_{S \in \text{List}([i])} Q(S)} \\ &= \left(\max_{i \in F} \zeta_i \right)^t \sum_{W \in \widetilde{\mathcal{W}}_t} \left(p_W \sum_{S \in \text{Roots}} Q(S) \right) \\ &= \left(\max_{i \in F} \zeta_i \right)^t \sum_{S \in \text{Roots}} Q(S) \quad (23) \end{aligned}$$

Using (23) we see that the binary logarithm of the probability that the walk does not encounter a flawless state within t steps is at most $t \log_2(\max_{i \in F} \zeta_i) + T_0$, where

$$T_0 = \log_2 \left(\max_{\sigma \in \Omega} \frac{\theta(\sigma)}{\mu(\sigma)} \right) + \log_2 \left(\sum_{S \in \text{Roots}} \prod_{i \in S} \psi_i \right).$$

Therefore, if $t = (T_0 + s)/\log_2(1/\max_{i \in F} \zeta_i) \leq (T_0 + s)/\delta$, the probability that the random walk on D does not reach a flawless state within t steps is at most 2^{-s} .

6 Application to Acyclic Edge Coloring

6.1 Earlier Works and Statement of Result

An edge-coloring of a graph is *proper* if all edges incident to each vertex have distinct colors. A proper edge coloring is *acyclic* if it has no bichromatic cycles, i.e., no cycle receives exactly two (alternating) colors. Acyclic Edge Coloring (AEC), was originally motivated by the work of Coleman et al. [5, 4] on the efficient computation of Hessians. The smallest number of colors, $\chi'_a(G)$, for which a graph G has an acyclic edge-coloring can also be used to bound other parameters, such as the oriented chromatic number [15] and the star chromatic number [8], both of which have many practical applications. The first general linear upper bound for χ'_a was given by Alon et al. [2] who proved $\chi'_a(G) \leq 64\Delta(G)$, where $\Delta(G)$ denotes the maximum degree of G . This bound was improved to 16Δ by Molloy and Reed [16] and then to $9.62(\Delta - 1)$ by Ndreca et al. [18]. Attention to the problem was recently renewed due to the work of Esperet and Parreau [7] who proved $\chi'_a(G) \leq 4(\Delta - 1)$, via an entropy compression argument, a technique that goes beyond what the LLL can give for the problem. Very recently, Giotis et al. improved the result of [7] to 3.74Δ .

We give a bound of $(2 + o(1))\Delta$ for graphs of bounded degeneracy. This not only covers a significant class of graphs, but demonstrates that our method can incorporate global graph properties. Recall that a graph G is d -degenerate if its vertices can be ordered so that every vertex has at most d neighbors greater than itself. If \mathcal{G}_d denotes the set of all d -degenerate graphs, then all planar graphs are in \mathcal{G}_5 , while all graphs with treewidth or pathwidth at most d are in \mathcal{G}_d . We prove the following.

Theorem 7. *Every d -degenerate graph of maximum degree Δ has an acyclic edge coloring with $\lceil (2 + \epsilon)\Delta \rceil$ colors that can be found in polynomial time, where $\epsilon = 16\sqrt{d/\Delta}$.*

6.2 Background

As will become clear shortly, the main difficulty in AEC comes from the short cycles of G , with 4-cycles being the toughest. This motivates the following definition.

Definition 5. *Given a graph $G = (V, E)$ and a, perhaps partial, edge-coloring of G , say that color c is 4-forbidden for $e \in E$ if assigning c to e would result in either a violation of proper-edge-coloration, or in a bichromatic 4-cycle containing e . Say that c is 4-available if it is not 4-forbidden.*

Similarly to [7, 9] we will use the following observation that the authors of [7] attribute to Jakub Kozik.

Lemma 5 ([7]). *In any proper edge-coloring of G at most $2(\Delta - 1)$ colors are 4-forbidden for any $e \in E$.*

Proof. The 4-forbidden colors for $e = \{u, v\}$ can be enumerated as: (i) the colors on edges adjacent to u , and (ii) for each edge e_v adjacent to v , either the color of e_v (if no edge with that color is adjacent to u), or the color of some edge e' which together with e, e_v and an edge adjacent to u form a cycle of length 4. \square

Armed with Lemma 5, the general idea is to use a palette P of size $2(\Delta - 1) + Q$ colors so that whenever we wish to (re)color an edge e there will be at least Q colors 4-available for e (of course, assigning such a color to e may cause one or more cycles of length at least 6 to become bichromatic). At a high level, similarly to [9], our algorithm will be:

- Start at a proper edge-coloring with no bichromatic 4-cycles.
- While bichromatic cycles of length at least 6 exist, recolor the edges of one with 4-available colors.

Note that to find bichromatic cycles in a properly edge-colored graph we can just consider each of the $\binom{|P|}{2}$ pairs of distinct colors from P and seek cycles in the subgraph of the correspondingly colored edges.

6.3 Applying our Framework

Given $G = (V, E)$ and a palette P of $2(\Delta - 1) + Q$ colors, let Ω be the set of all proper edge-colorings of G with no monochromatic 4-cycle. Fix an arbitrary ordering π of E and an arbitrary ordering χ of P . For every even cycle C of length at least 6 in G fix (arbitrarily) two adjacent edges e_1^C, e_2^C of C .

– Our distribution of initial state θ assigns all its probability mass to the following $\sigma_1 \in \Omega$: color the edges of E in π -order, assigning to each edge $e \in E$ the χ -greatest 4-available color.

– For every even cycle C of length at least 6 we define the flaw $f_C = \{\sigma \in \Omega : C \text{ is bichromatic}\}$. Thus, a flawless $\sigma \in \Omega$ is an acyclic edge coloring of G .

– The set of actions for addressing f_C in state σ , i.e., $A(C, \sigma)$, comprises all $\tau \in \Omega$ that may result from the following procedure: uncolor all edges of C except for e_1^C, e_2^C ; go around C , starting with the uncolored edge that is adjacent to e_2^C , etc., assigning to each uncolored edge $e \in C$ one of the 4-available colors for e at the time e is considered. Thus, by lemma 5, $|A(C, \sigma)| \geq Q^{|C|-2}$.

Lemma 6. *For every flaw f_C and state $\tau \in \Omega$, there is at most 1 arc $\sigma \xrightarrow{C} \tau$, i.e., $b_C^\tau \leq 1$.*

Proof. Given τ and C , to recover the previous state σ it suffices to extend the bicoloring in τ of e_1^C, e_2^C to the rest of C (since C was bichromatic in σ and only edges in $C \setminus \{e_1^C, e_2^C\}$ were recolored). \square

Thus, taking μ to be uniform and ρ such that (D, ρ, μ) is harmonic yields $\gamma_C = Q^{-|C|+2}$.

Let R be the symmetric directed graph with one vertex per flaw where $f_C \rightleftharpoons f_{C'}$ iff $C \cap C' \neq \emptyset$. Since a necessary condition for f_C to potentially cause $f_{C'}$ is that $C \cap C' \neq \emptyset$, we see that R is a supergraph of the causality digraph. Thus, if we run the RECURSIVE WALK algorithm with input R , to apply Theorem 3 we need to evaluate for each flaw f_C a sum over the subsets of $\Gamma_R(C)$ that are independent in R . To carry out this enumeration we observe that independence in R implies edge-disjointness which, in turn, implies that in each (independent) set of cycles to be enumerated, no edge of C appears in multiple cycles. Thus, to perform the enumeration it suffices to enumerate the subsets of edges of C that appear in the cycles and for each appearing edge e to enumerate all even cycles of length at least 6 containing e .

Let $g(k) = \max_{e \in E} |\{k\text{-cycles in } G \text{ that contain } e\}|$. If $\psi_C = \psi(|C|)$, then we can bound (3) as

$$\begin{aligned} \frac{\gamma_C}{\psi_C} \sum_{S \in \text{Ind}(\Gamma_R(C))} \prod_{C' \in S} \psi_{C'} &\leq \frac{1}{\psi_C Q^{|C|-2}} \cdot \sum_{i=0}^{|C|} \left(\binom{|C|}{i} \left(\sum_{j=3}^{\infty} g(2j) \psi(2j) \right)^i \right) \\ &= \frac{1}{\psi_C Q^{|C|-2}} \cdot \left(1 + \left(\sum_{j=3}^{\infty} g(2j) \psi(2j) \right) \right)^{|C|}. \end{aligned} \quad (24)$$

We will prove the following structural lemma relating degeneracy to g .

Lemma 7. *If $G \in \mathcal{G}_d$ has maximum degree Δ , then $g(k) \leq 2(4d\Delta)^{(k-2)/2}$.*

We are thus left to choose ψ such that for every even $|C| \geq 6$, the r.h.s. of (24) is strictly less than 1. Taking $\psi(k) = (8d\Delta)^{-\frac{k-2}{2}}$ and $Q = \lceil 16\sqrt{d\Delta} \rceil$ we see that for all $k \geq 6$,

$$\frac{(8d\Delta)^{\frac{k-2}{2}}}{(16\sqrt{d\Delta})^{k-2}} \cdot \left(1 + 2 \left(\sum_{j=3}^{\infty} (4d\Delta)^{j-1} (8d\Delta)^{-j+1} \right) \right)^k = 2^{-\frac{3k}{2}+5} < 1 .$$

Regarding the running time, notice that $\delta \geq 1 - 2^{-4} = 15/16$ and that it can easily be seen that that T_0 is a polynomial in $|E|$, Δ and the number of colors used.

Proof of Lemma 7. Fix any edge $e = \{u, v\} \in E$. To enumerate the k -cycles containing e we will partition them into equivalence classes as follows. First we orient all edges of G arbitrarily to get a digraph D . Consider now the two possible traversals of the path $C \setminus \{u, v\}$, i.e., the one starting at u and the one starting at v . For each traversal generate a string in $\{0, 1\}^{k-2}$ whose characters correspond to successive vertices of the path, other than the endpoints, and denote whether the corresponding vertex was entered along an edge oriented in agreement (1) or in disagreement (0) with the direction of travel. Observe that each of the $k - 3$ edges of C that have no vertex from $\{u, v\}$ will create a 1 in one string and a 0 in the other. Therefore, at least one of the strings will have at least $\lceil (k - 3)/2 \rceil = (k - 2)/2$ ones. Select that string, breaking ties in favor of the string corresponding to starting at u . Finally, prepend a single bit to the string to designate whether the winning string corresponded to u or to v . The string denotes C 's equivalence class.

To enumerate all k -cycles containing e we can thus enumerate all binary strings of length $k - 1$ and use each string to select the $k - 2$ other vertices of the cycle as follows: after reading the first character to decide whether to start at u or at v , we interpret each successive character to indicate whether we should choose among the out-neighbors or the in-neighbors of the current vertex. By the string's construction, we will chose among out-neighbors $q \geq (k - 2)/2$ times. If Out and In are upper bounds on the out- and in-degree of D , respectively, then the total number of cycles per string (class) is bounded by $\text{Out}^q \text{In}^{k-2-q}$.

To conclude the argument we note that since $G \in \mathcal{G}_d$ we can direct its edges so that every vertex has out-degree at most d by repeatedly removing any vertex v of current degree at most d (it always exists) and, at the time of removal, orienting its current neighbors away from v . \square

Acknowledgements

We are grateful to Dan Kral for providing us with Lemma 7 and to Louis Esperet for pointing out an error in our application of Theorem 3 to yield Theorem 7 in a previous version of the paper. FI is thankful to Alistair Sinclair for many fruitful conversations.

References

- [1] Dimitris Achlioptas and Fotis Iliopoulos. Random walks that find perfect objects and the lovasz local lemma. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 494–503. IEEE Computer Society, 2014.

- [2] Noga Alon. A parallel algorithmic version of the local lemma. *Random Structures & Algorithms*, 2(4):367–378, 1991.
- [3] Rodrigo Bissacot, Roberto Fernández, Aldo Procacci, and Benedetto Scoppola. An improvement of the Lovász local lemma via cluster expansion. *Combinatorics, Probability & Computing*, 20(5):709–719, 2011.
- [4] Thomas F Coleman and Jin Yi Cai. The cyclic coloring problem and estimation of sparse hessian matrices. *SIAM J. Algebraic Discrete Methods*, 7(2):221–235, April 1986.
- [5] Thomas F. Coleman and Moré Jorge J. Estimation of sparse hessian matrices and graph coloring problems. *Mathematical Programming*, 28(3):243–270, 1984.
- [6] Paul Erdős and Joel Spencer. Lopsided Lovász local lemma and latin transversals. *Discrete Applied Mathematics*, 30(2-3):151–154, 1991.
- [7] Louis Esperet and Aline Parreau. Acyclic edge-coloring using entropy compression. *European Journal of Combinatorics*, 34(6):1019–1027, 2013.
- [8] Guillaume Fertin, André Raspaud, and Bruce Reed. Star coloring of graphs. *Journal of Graph Theory*, 47(3):163–182, 2004.
- [9] Ioannis Giotis, Lefteris M. Kirousis, Kostas I. Psaromiligkos, and Dimitrios M. Thilikos. On the algorithmic lovász local lemma and acyclic edge coloring. In Robert Sedgewick and Mark Daniel Ward, editors, *Proceedings of the Twelfth Workshop on Analytic Algorithmics and Combinatorics, ANALCO 2015, San Diego, CA, USA, January 4, 2015*, pages 16–25. SIAM, 2015.
- [10] Bernhard Haeupler, Barna Saha, and Aravind Srinivasan. New constructive aspects of the Lovász local lemma. In *FOCS*, pages 397–406, 2010.
- [11] David G. Harris and Aravind Srinivasan. A constructive algorithm for the Lovász local lemma on permutations. In *SODA*, pages 907–925. SIAM, 2014.
- [12] Nicholas Harvey and Jan Vondrák. An algorithmic proof of the lopsided lovasz local lemma. *CoRR*, abs/1504.02044, 2015. To appear in FOCS’15.
- [13] Kashyap Babu Rao Kolipaka and Mario Szegedy. Moser and Tardos meet Lovász. In *STOC*, pages 235–244. ACM, 2011.
- [14] Vladimir Kolmogorov. Commutativity in the random walk formulation of the lovasz local lemma. *CoRR*, abs/1506.08547, 2015.
- [15] Alexandr. V. Kostochka, Eric Sopena, and Xuding Zhu. Acyclic and oriented chromatic numbers of graphs. *Journal of Graph Theory*, 24(4):331–340, 1997.
- [16] Michael Molloy and Bruce Reed. Further algorithmic aspects of the local lemma. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing, STOC ’98*, pages 524–529, New York, NY, USA, 1998. ACM.
- [17] Robin A. Moser and Gábor Tardos. A constructive proof of the general Lovász local lemma. *J. ACM*, 57(2):Art. 11, 15, 2010.

- [18] Sokol Ndreca, Aldo Procacci, and Benedetto Scoppola. Improved bounds on coloring of graphs. *Eur. J. Comb.*, 33(4):592–609, May 2012.
- [19] Wesley Pegden. Highly nonrepetitive sequences: Winning strategies from the local lemma. *Random Struct. Algorithms*, 38(1-2):140–161, 2011.
- [20] Wesley Pegden. An improvement of the Moser-Tardos algorithmic local lemma. *CoRR*, abs/1102.2853, 2011.
- [21] J.B. Shearer. On a problem of Spencer. *Combinatorica*, 5(3):241–245, 1985.
- [22] Mario Szegedy. The Lovász local lemma - a survey. In Andrei A. Bulatov and Arseny M. Shur, editors, *CSR*, volume 7913 of *Lecture Notes in Computer Science*, pages 1–11. Springer, 2013.

A Mapping Bad Trajectories to Forests

In this section, we show how to represent each sequence of t steps that does not reach a sink as a forest with t vertices, where the forests have different characteristics for each of the walks of Theorems 2, 3.

A.1 Forests of the Permutation Walk (Theorem 2)

For $S \subseteq F$, we denote by $I_\pi(S) = I(S)$ the greatest element of S according to π . We will sometimes write $I(\sigma)$ to denote $I(U(\sigma))$. We first show how to represent the witness sequences of the Permutation Walk as sequences of sets.

Let B_i be the set of flaws “introduced” by the i -th step of the walk, where a flaw f_j is said to “introduce itself” if it remains present after an action from $A(j, \cdot)$ is taken. Formally,

Definition 6. Let $B_0 = U(\sigma_1)$. For $1 \leq i \leq t - 1$, let $B_i = U(\sigma_{i+1}) \setminus (U(\sigma_i) \setminus I(\sigma_i))$.

Let $B_i^* \subseteq B_i$ comprise those flaws addressed in the course of the trajectory. Thus, $B_i^* = B_i \setminus \{O_i \cup N_i\}$, where O_i comprises any flaws in B_i that were eradicated “collaterally” by an action taken to address some other flaw, and N_i comprises any flaws in B_i that remained present in every subsequent state after their introduction without being addressed. Formally,

Definition 7. The Break Sequence of a t -trajectory is $B_0^*, B_1^*, \dots, B_{t-1}^*$, where for $0 \leq i \leq t - 1$,

$$\begin{aligned} O_i &= \{f \in B_i \mid \exists j \in [i + 1, t] : f \notin U(\sigma_{j+1}) \wedge \forall \ell \in [i + 1, j] : f \neq w_\ell\} \\ N_i &= \{f \in B_i \mid \forall j \in [i + 1, t] : f \in U(\sigma_{j+1}) \wedge \forall \ell \in [i + 1, t] : f \neq w_\ell\} \\ B_i^* &= B_i \setminus \{O_i \cup N_i\} . \end{aligned}$$

Given $B_0^*, B_1^*, \dots, B_{i-1}^*$ we can determine the sequence of flaws addressed w_1, w_2, \dots, w_i inductively, as follows. Define $E_1 = B_0^*$, while for $i \geq 1$,

$$E_{i+1} = (E_i - w_i) \cup B_i^* . \quad (25)$$

By construction, the set $E_i \subseteq U(\sigma_i)$ is guaranteed to contain $w_i = I(\sigma_i) = I(U(\sigma_i))$. Since $I = I_\pi$ returns the greatest flaw in its input according to π , it must be that $I_\pi(E_i) = w_i$. We note that this is the only place we ever make use of the fact that the function I is derived by an ordering of the flaws, thus guaranteeing that for every $f \in F$ and $S \subseteq F$, if $I(S) \neq f$ then $I(S \setminus f) = I(S)$.

We now give a 1-to-1 map, from Break Sequences to vertex-labeled unordered rooted forests. Specifically, the *Break Forest* of a bad t -trajectory Σ has $|B_0^*|$ trees and t vertices, each vertex labelled by an element of $W(\Sigma)$. To construct it we first lay down $|B_0^*|$ vertices as roots and then process the sets B_1^*, B_2^*, \dots in order, each set becoming the progeny of an already existing vertex (empty sets, thus, giving rise to leaves).

Break Forest Construction

- 1: Lay down $|B_0^*|$ vertices, each labelled by a different element of B_0^* , and let V consist of these vertices
 - 2: **for** $i = 1$ to $t - 1$ **do**
 - 3: Let v_i be the vertex in V_i with greatest label according to π
 - 4: Add $|B_i^*|$ children to v_i , each labelled by a different element of B_i^*
 - 5: Remove v_i from V ; add to V the children of v_i .
-

Observe that even though neither the trees, nor the nodes inside each tree of the Break Forest are ordered, we can still reconstruct $W(\Sigma)$ since the set of labels of the vertices in V_i equals E_i for all $0 \leq i \leq t - 1$.

A.2 Forests of the Recursive Walk (Theorem 3)

We will represent each witness sequence $W = W(\Sigma)$ of the Recursive Walk as a vertex-labeled unordered rooted forest, having one tree per invocation of procedure ADDRESS by procedure ELIMINATE. Specifically, to construct the *Recursive Forest* $\phi = \phi(\Sigma)$ we add a root vertex per invocation of ADDRESS by ELIMINATE and one child to every vertex for each (recursive) invocation of ADDRESS that it makes. As each vertex corresponds to an invocation of ADDRESS (step of the walk) it is labeled by the invocation's flaw-argument. Observe now that (the invocations of ADDRESS corresponding to) both the roots of the trees and the children of each vertex appear in W in their order according to π . Thus, given the unordered rooted forest $\phi(\Sigma)$ we can order its trees and the progeny of each vertex according to π and recover W as the sequence of vertex labels in the preorder traversal of the resulting ordered rooted forest.

Recall the definition of graph G on F from Definition 3. We will prove that the flaws labeling the roots of a Recursive Forest are independent in G and that the same is true for the flaws labelling the progeny of every vertex of the forest. To do this we first prove the following.

Proposition 1. *If ADDRESS(i, σ) returns at state τ , then $U(\tau) \subseteq U(\sigma) \setminus (\Gamma_R(f_i) \cup \{f_i\})$.*

Proof. Let σ' be any state subsequent to the ADDRESS(i, σ) invocation. If any flaw in $U(\sigma) \cap \Gamma_R(f_i)$ is present at σ' , the “while” condition in line 8 of the Recursive Walk prevents ADDRESS(i, σ) from returning. On the other hand, if $f_h \in \Gamma_R(f_i) \setminus U(\sigma)$ is present in σ' , then there must have existed an invocation ADDRESS(j, σ''), subsequent to invocation ADDRESS(i, σ), wherein addressing f_j caused f_h . Consider the last such invocation. If σ''' is the state when this invocation returns, then $f_h \notin U(\sigma''')$, for otherwise the invocation could not have returned, and by the choice of invocation, f_h is not present in any subsequent state between σ''' and τ . \square

Let $([i], \sigma^i)$ denote the argument of the i -th invocation of ADDRESS by ELIMINATE. By Proposition 1, $\{U(\sigma^i)\}_{i \geq 1}$ is a decreasing sequence of sets. Thus, the claim regarding the root labels follows trivially: for each $i \geq 1$, the flaws in $\Gamma_R(f_i) \cup f_i$ are not present in σ^{i+1} and, therefore, are not present in $U(\sigma^j)$, for any $j \geq i + 1$. The proof for the children of each node is essentially identical. If a node corresponding to an invocation of ADDRESS has q children, corresponding to q (recursive) invocations with arguments $\{(a_i, \tau^i)\}_{i=1}^q$, then the sequence of sets $\{U(\tau^i)\}_{i=1}^q$ is decreasing. Thus, the flaws in $\Gamma_R(a_i) \cup \{a_i\}$ are not present in τ^{i+1} and, therefore, not present in $U(\tau^j)$, for any $j \geq i + 1$.