Martingales, Tic-Tac-Toe, and Algorithmic Aspects of the Local Lemma

Petr Škoda, FMF, Ljubljana

January 24, 2008

1 Martingales

Let us start by a motivation. A drunkard goes along a wide road. Due to his unsure steps, with each step he moves with probabilty p by ten inches to the left, with probability p by ten inches to the right, and with probability $1-2p$ he goes straight. The question is: With what probability will he after n steps fall over the edge of the road?

We describe the situation as a probabilistic problem. Let X_i be the result of i-th step, namely 1 and -1 both with probability p and 0 with probability $1-2p$. Now, the distance of the drunkard from the middle of the road after n steps corresponds to the sum of the steps $X = \sum_{i=1}^{n} X_i$. By linearity of expectation, we have $\mathbb{E}[X] = \sum_{i=0}^{n} \mathbb{E}[X_i] = 0$ because $\mathbb{E}[X_i] = 0$ for all i. After few steps he will be probably still somewhere near the middle of the road but he will certainly travel away on a long road. How to describe these notions?

We will get a basic insight into the problem using the second moment.

Definition 1. *The* variance $\mathrm{Var}[X]$ *of a real random variable* X *is*

$$\mathrm{Var}[X] = \mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

(The first equality is a definition, and the second one follows by an easy computation.) The standard deviation of X *is* $\sigma = \sqrt{\mathrm{Var}[X]}$.

Theorem 2 (Chebyshev's Inequality)**.** *Let* X *be a random variable with a finite variance. Then for any* $t > 0$

$$P[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2}.$$

Proof.

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] \geq t^2 P[|X - \mathbb{E}[X]| \geq t].$$

□

The Chebyshev's inequality is a very simple tool but it gives us first useful bounds. Return to the motivation problem. Since X_i are independent, the variance of X is

$$\mathrm{Var}[X] = \sum_{i=0}^{n} \mathrm{Var}[X_i] = \sum_{i=0}^{n} \mathbb{E}[X_i^2] = 2pn \leq n.$$

The Chebyshev's inequality gives us

$$P[|X| \geq t] \leq \frac{2pn}{t^2} \leq \frac{n}{t^2}.$$

For $t = \sqrt{20n}$, there is probability at least 95% that the drunkard will stay at the road of the width t.

Denote Y_i, $0 \leq i \leq n$, the distance of the drunkard from the middle of the road after i steps, $Y_i = \sum_{k=0}^{i} X_k$. Observe that no matter what the earlier history of the walk has been, the expected position after t steps equals the actual position after $t-1$ steps. This is the defining property of a martingale.

Definition 3. *A* martingale *is a list of random variables $Y_0, \ldots, Y_n$ such that, for all i, the expectation of Y_i, given the values of $Y_0, \ldots, Y_{i-1}$, equals Y_{i-1}.*

Martingales can make it easy to show that a random variable is highly concetrated around its expected value. We shall see that this follows from its inability to move more than one unit in each step. The statement is called Azuma's Inequality. This inequality states that if successive random variables in a martingale always differ by at most 1, then the probability that $Y_n - Y_0$ exceeds $\lambda\sqrt{n}$ is bounded by $e^{-\lambda^2/2}$. We first prove two lemmas. These statements hold for continuous random variables, but we consider only discrete variables.

Lemma 4. *Let X be a random variable such that $\mathbb{E}[X] = 0$ and $|X| \leq 1$. If f is a convex function on $[-1, 1]$, then $\mathbb{E}[f(X)] \leq \frac{1}{2}[f(-1) + f(1)]$. In particular, $\mathbb{E}[e^{tX}] \leq \frac{1}{2}[e^t + e^{-t}]$ for all $t > 0$.*

Proof. When X takes only the values ± 1, each with probability $\frac{1}{2}$, we have $\mathbb{E}[f(X)] = \frac{1}{2}[f(-1)+f(1)]$. For other distributions, pushing probability "out to the edges" increase $\mathbb{E}[f(X)]$. For discrete variables, we can use induction on the number of values with non-zero probability. Convexity implies that $f(a) \le \frac{1-a}{2}f(-1) + \frac{1+a}{2}f(1)$. If $P[X = a] = \alpha$, then we can decrease the probability at a to 0, increase $P[X = -1]$ by $\alpha\frac{1-a}{2}$ and increase $P[X = 1]$ by $\alpha\frac{1+a}{2}$ to obtain a new variable X' with the same expectation. By the convexity inequality and induction hypothesis,

$$\mathbb{E}[f(X)] \le \mathbb{E}[f(X')] \le \frac{1}{2}[f(-1) + f(1)]$$

. □

Definition 5. *For events A and B, the* conditional probability *of A given B is obtained by treating the event B as the full probability space, which means normalizing by $P[B]$. Thus we define $P[A|B] = \frac{P[AB]}{P[B]}$.*

When Y, X are random variables, we write $Y|X$ for "Y given X". This defines a random variable for each value of X; we treat X as a constant i and normalize the resulting distribution for Y by $P[X = i]$.

For Azuma's Inequality, we use expectation of conditional variables. For each i, we compute the expected value of Y when restricted to the sample points where $X = i$. The expectation $\mathbb{E}[\mathbb{E}[Y|X]]$ is the expectation of $\mathbb{E}[Y|X = i]$ over the choices for i, which occur with probability $P[X = i]$. The result is an expectation over entire sample space. It removes the effect of conditioning, and we obtain $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$.

Lemma 6. $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$.

Proof. Let $p_{i,j} = P(X = i \text{ and } Y = j)$. Since

$$\mathbb{E}[Y|X = i] = \frac{\sum_j jp_{i,j}}{P[X = i]},$$

the following holds:

$$\mathbb{E}[\mathbb{E}[Y|X]] = \sum_i \mathbb{E}[Y|X = i]P[X = i] = \sum_i \sum_j jp_{i,j} = \mathbb{E}[Y].$$

□

We will use the following simple Markov's Inequality in the proof of the Azuma's Inequality.

Theorem 7 (Markov's Inequality)**.** *For a real positive parameter t and a non-negative random variable X, the following holds:*

$$P[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

Proof.

$$\mathbb{E}[X] = \sum_i iP[X = i] \geq \sum_{i \geq t} tP[X = i] = tP[X \geq t]$$

□

Theorem 8 (Azuma's Inequality)**.** *If $X_0, \ldots, X_n$ is a martingale with $|X_i - X_{i-1}| \leq 1$, then $P[X_n - X_0 \geq \lambda\sqrt{n}] \leq e^{-\lambda^2/2}$.*

Proof. By translation, we may assume that $X_0 = 0$. For $t > 0$, we have $X_n \geq \lambda\sqrt{n}$ if and only if $e^{tX_n} \geq e^{t\lambda\sqrt{n}}$, and hence $P[X_n \geq \lambda\sqrt{n}] = P[e^{tX_n} \geq e^{t\lambda\sqrt{n}}]$. Applied to e^{tX_n}, Markov's Inequality yields $P[e^{tX_n} \geq e^{t\lambda\sqrt{n}}] \leq \mathbb{E}[e^{tX_n}]/e^{t\lambda\sqrt{n}}$. This bound holds for each $t > 0$, and later we will choose t to minimize the bound.

First we prove by induction on n that $\mathbb{E}[e^{tX_n}] \leq [\frac{1}{2}(e^t + e^{-t})]^n$. We introduce X_{n-1} to condition on it. Lemma 6 yields

$$\mathbb{E}[e^{tX_n}] = \mathbb{E}[e^{tX_{n-1}}e^{t(X_n - X_{n-1})}] = \mathbb{E}[\mathbb{E}[e^{tX_{n-1}}e^{t(X_n - X_{n-1})}|X_{n-1}]].$$

When we condition on X_{n-1}, the value of X_{n-1} is constant for the inner expectation. Hence we can remove $e^{tX_{n-1}}$ from the inner expectation to obtain $\mathbb{E}[e^{tX_n}] = \mathbb{E}[e^{tX_{n-1}}\mathbb{E}[e^{tY}|X_{n-1}]]$, where $Y = X_n - X_{n-1}$. Because $\{X_n\}$ is a martingale, $\mathbb{E}[Y] = 0$, and by hypothesis $|Y| \leq 1$. Hence Lemma 4 applies, yielding $\mathbb{E}[e^{tY}|X_{n-1}] \leq \frac{1}{2}(e^t + e^{-t})$. This itself is now a constant, yielding $\mathbb{E}[e^{tX_n}] \leq \frac{1}{2}(e^t + e^{-t})\mathbb{E}[e^{tX_{n-1}}]$. The induction hypothesis completes the proof.

We weaken the bound to a more useful form by observing that $\frac{1}{2}(e^t + e^{-t}) \leq e^{t^2/2}$. This holds because the left side is $\sum t^{2k}/(2k)!$ and the right side is $\sum t^{2k}/(2^k k!)$. Hence our original probability is bounded by $e^{nt^2/2 - \lambda t\sqrt{n}}$ for each $t > 0$. We obtain the best bound minimizing over t. The exponent is quadratic; we minimize it by choosing t to solve $tn - \lambda\sqrt{n} = 0$, or $t = \lambda/\sqrt{n}$. The resulting bound is $e^{-\lambda^2/2}$. □

The Azuma's Inequality gives us a better bound on the concentration of drunkard's steps. The bound is only from above but we get the other bound by symmetry. Denote by α the probability bound $\alpha = 2e^{-\lambda^2/2}$. By an easy computation, we get $\lambda = \sqrt{-2\ln(\alpha/2)}$. For $\alpha = 95\%$, $\lambda < \sqrt{8}$. Hence after n steps, the drunkard will not get farther from the middle of the road than $\sqrt{8n}$ with probability at least 95%.

2 Tic-Tac-Toe and the Method of Conditional Expectations

The children's game Tic-Tac-Toe will be familiar to many readers (actually, a variant of this game was played in Egypt 1400 BC, and a related game, called renju, in China 2500 BC). Two players Nought (O) and Cross (X) alternately place their symbols in the squares of a 3x3 grid. A player can only place his symbol in an unoccupied grid square so the game lasts for at most nine moves. The first player to place his symbol on all the squares in a line (row, column, or diagonal) wins. If all the squares are occupied and no player has covered a line, then the game is a draw. We can reformulate this game in terms of 2-coloring hypergraphs. The Tic-Tac-Toe hypergraph has 9 vertices and 8 edges. Two players Blue and Red alternately color an uncolored vertex. The first player to monochromatically color an edge wins. If the players complete a proper 2-coloring of the hypergraph, then the game is a draw.

We can, of course, play this game on any hypergraph H. For some hypergraphs, the first player wins while for others the second player can force a draw. The second player can never win because the game is so symmetric that the first player can steal any winning strategy for the second player. In general, it is PSpace-complete to determine if the second player can force a draw for an input hypergraph H. However, as we shall see, it can be shown that for certain hypergraphs, the second player can force a draw.

We will recall a simple theorem that a hypergraph is 2-colorable if it has only few edges.

Theorem 9. *If H is a hypergraph with fewer than 2^{k-1} hyperedges, each of size at least k, then H is 2-colorable.*

We repeat the proof because the idea will be reused later.

Proof. Color the vertices at random, assigning to each vertex the color red with probability $\frac{1}{2}$ and blue otherwise, and making each such choice independently of the choices of all other vertices. In other words, choose a uniformly random 2-coloring of vertices. For each hyperedge e, define the random variable X_e to be 1 if e is monochromatic and 0 otherwise. (X_e is called an *indicator variable*). Let $X = \sum_{e \in H} X_e$, and note that X is the number of monochromatic edges. Any one hyperedge e is monochromatic with probability at most $2^{-(k-1)}$, and so $\mathbb{E}[X_e] \leq 2^{-(k-1)}$. Therefore, by the Linearity of Expectation, $\mathbb{E}[X] = \sum_{e \in H} \mathbb{E}[X_e] \leq |E(H)| \times 2^{-(k-1)} < 1$. Therefore, the probability that $X = 0$, i.e. that there is no monochromatic hyperedge, is positive. □

Notice that this theorem is based on the fact that, if the expected number of monochromatic hyperedges in a uniformly random 2-coloring is lower than one, then there exists a coloring where is no monochromatic hyperedge. In this class of hypergraphs, we present an algorithm, due to Erdős and Selfridge for finding proper 2-colorings of such hypergraphs.

We then condsider a game similar to Tic-Tac-Toe in which two players Red and Blue alternately color vertices of a hypergraph until all the vertices are colored. If any edge is monochromatically colored then the first player to have monochromatically colored an edge wins. Otherwise, the game is a draw.

We shall see that for the hypergraphs under consideration optimal play by both players ensures that the game ends in a draw, and provide an effeciently computable strategy that players can use to ensure this outcome.

Note that the final position in a drawn game is a proper 2-coloring, so the second result implies the first.

2.1 The Algorithm

Let $X = X(C)$ be the number of monochromatic edges under a coloring C. If C is a random coloring then X is a random variable. For a partial coloring P, we let $\mathbb{E}_P(X)$ be the expected value of X for a uniformly chosen random completition of P. I.e. $\mathbb{E}_P(X) = \mathbb{E}[X(C)]$ where C is the coloring obtained by coloring each vertex uncolored by P, independently, red with probability $\frac{1}{2}$ and blue with probability $\frac{1}{2}$.

For a hypergraph H on vertex set $\{v_1, \ldots, v_n\}$, we will iteratively construct partial colorings $P_0, P_1, \ldots, P_n$ where for each i:

(a) the set of vertices colored under P_i is $\{v_1, \dots, v_i\}$ (thus P_0 colors no vertices),

(b) P_i and P_{i-1} agree on $\{v_1, \dots, v_{i-1}\}$, and

(c) $\mathbb{E}_{P_i}(X) < 1$.

Clearly, $\mathbb{E}_{P_n}(X) = X(P_n)$. Since, by (c), $X(P_n) < 1$, it must be zero. Thus P_n is the desired proper 2-coloring.

Having constructed P_{i-1}, there are two possible choices for P_i, we can extend P_{i-1} by coloring v_i either red or blue. We denote the first possibility by P_i^r and the second by P_i^b. We shall show:

Observation 10. *We can compute* $\mathbb{E}_P(X)$ *in polynomial time for any partial coloring* P.

And, we will show also:

Observation 11. $\min(\mathbb{E}_{P_i^r}(X), \mathbb{E}_{P_i^b}(X)) \leq \mathbb{E}_{P_{i-1}}(X)$.

With these results in hand, it is easy to iteratively construct $P_1, \dots, P_n$. Given P_{i-1} we simply compute $\mathbb{E}_{P_i^r}(X)$ and $\mathbb{E}_{P_i^b}(X)$ and choose for P_i one of these possibilities which minimizes $\mathbb{E}_{P_i}(X)$.

The proof of Observation 10 is straightforward. To compute $\mathbb{E}_P(X)$ we simply compute the conditional probability for each edge e that e will be monochromatic and sum these values. If e contains vertices of both colors this probability is zero. If e contains no colored vertices, this probability is $2^{1-|e|}$. Finally, if e contains colored vertices of only one color and u uncolored vertices then this probability is 2^{-u}.

The proof of Observation 11. Actually a stronger fact is true: $\mathbb{E}_{P_{i-1}}(X)$ lies between $\mathbb{E}_{P_i^r}(X)$ and $\mathbb{E}_{P_i^b}(X)$. We can even compute this directly. Let B_2^n denote set of all colorings of n vertices and introduce a relation $A \prec B$ between two partial colorings which says that B is an extension of A.

$$\begin{aligned}
\mathbb{E}_{P_{i-1}}(X) &= \frac{1}{2^{n-i+1}} \sum_{C \in B_2^n, P_{i-1} \prec C} X(C) \\
&= \frac{1}{2^{n-i} \cdot 2} \left(\sum_{C \in B_2^n, P_i^r \prec C} X(C) + \sum_{C \in B_2^n, P_i^b \prec C} X(C) \right) \\
&= \frac{1}{2} \left(\mathbb{E}_{P_i^r}(X) + \mathbb{E}_{P_i^b}(X) \right).
\end{aligned}$$

Lemma 12. *If the expected number of monochromatic edges in a uniformly random 2-coloring of H is less than 1, then the second player can force a draw in Generalized Tic-Tac-Toe on H.*

Remark 13. *The natural converse to this statement is false. I.e., there is no lower bound on the expected number of monochromatic edges in a random coloring which guarantees that the first player wins. To see this, consider the hypergraph H_n which consists of n disjoint edges each with two vertices. The expected number of monochromatic edges in a random coloring of H_n is $\frac{n}{2}$. However, the second player can draw by always playing in the edge which the first player just played in.*

Remark 14. *The bound in Lemma 12 is tight, as the following example shows. Let F_n be the hypergraph with $2n+1$ vertices $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n, c$ whose edge set contains exactly the 2^n subsets of vertices consisting of c and n other vertices all with different indices. The expected number of monochromatic edges in a uniformly random 2-coloring of H is exactly one. On the other hand, the first player can win Generalized Tic-Tac-Toe on F_n by first picking c and then always picking a vertex with the same index as that just picked by the second player.*

2.2 Proof of Lemma 12

To ease the exposition in this section, we assume that the players in a game of Generalized Tic-Tac-Toe continue to alternate turns until the whole graph is colored, even if one obtains a monochromatic edge.

Lemma 12 asserts that a certain condition ensures that the second player can prevent the first player from obtaining a monochromatic edge first. We find it more convenient to prove that this condition allows the second player to prevent the first player from obtaining a monochromatic edge at all. That is, we prove the following strengthening of Lemma 12.

Lemma 15. *Suppose Blue and Red play Generalized Tic-Tac-Toe (to completion) on a hypergraph H for which the expected number of monochromatic edges in a uniformly random 2-coloring is less than one. Then, if Blue plays second, he has an efficiently computable strategy to prevent Red from coloring all of any edge.*

Now, the expected number of all red edges in a uniformly random 2-coloring of a hypergraph is clearly exactly half the expected number of monochromatic edges. Furthermore, any initial move can at most double the expected number of all red edges. Combining these two facts, with lemma below yields Lemma 15.

Lemma 16. *Suppose Blue and Red play Generalized Tic-Tac-Toe (to completion) on a hypergraph H for which the expected number of all red edges in a uniformly random 2-coloring is less than one. Then, if Blue plays first, he has an efficiently computable strategy to prevent Red from coloring all of any edge.*

Proof. We use $R = R(C)$ to denote the number of all red edges in a coloring C. For any partial coloring P, we use $\mathbb{E}_P(R)$ to denote the expected number of all red edges in uniformly chosen random completion of P. Blue simply colors the vertex v which minimizes the value of $\mathbb{E}_P(R)$ for the resultant partial coloring P.

Clearly, Blue can efficiently compute the value of $\mathbb{E}_P(R)$ for any candidate partial coloring P, using the algorithm similar that for computing $\mathbb{E}_P(X)$. We claim that no matter what vertex w Red chooses to color on his next turn, letting P_1 be the partial coloring before Blue's turn and P_2 the partial coloring after Red's turn, we have: $\mathbb{E}_{P_2}(R) \le \mathbb{E}_{P_1}(R)$. Iteratively applying this claim proves Lemma 16, as the initial condition implies that for the original coloring P_0, $\mathbb{E}_{P_0}(R) < 1$ and hence this will also be true for the final coloring.

It remains only to prove the claim. To do so, we consider the set E_1 of edges which contain v and contain no blue vertex, and the set E_2 of edges which contain w and contain no blue vertex. For every edge e in $E_1 \cup E_2$, we let $u(e)$ be the number of uncolored vertices in e under P_1.

We note that coloring v blue decreases the conditional expected value of R by $\sum_{e \in E_1} 2^{-u(e)}$. On the other hand, if Blue had colored w blue, he would have decreased the conditional expected value of R by $\sum_{e \in E_2} 2^{-u(e)}$. It follows that

$$\sum_{e \in E_1} 2^{-u(e)} \ge \sum_{e \in E_2} 2^{-u(e)}.$$

Now, $\mathbb{E}_{P_2}(R)$ is clearly $\mathbb{E}_{P_1}(R) - \sum_{e \in E_1} 2^{-u(e)} + \sum_{e \in E_2 \setminus E_1} 2^{-u(e)}$. The desired claim follows by the above inequality. □

3 Algorithmic Aspects of the Local Lemma

In this section, we discuss finding proper 2-colorings of hypergraphs where each edge intersects a bounded number of other edges. We present the following theorem of Beck:

Theorem 17 (Beck)**.** *There is a deterministic polytime algorithm which will find a proper 2-coloring of any hypergraph H in which each edge has size at least k and intersects at most $d \leq 2^{k/16-2}$ other edges.*

Recall the following application of the Local Lemma.

Theorem 18. *If H is a hypergraph such that each hyperedge has size at least k and intersects at most 2^{k-3} other hyperedges, then H is 2-colorable.*

Hence, we know that the 2-coloring exists but the approach taken here to actually construct the 2-coloring effeciently requires the smaller bound on d.

We will first present a randomized algorithm, and then show how to use the techniques of the previous chapter to derandomize it.

Since there is no bound on the number of edges in our hypergraph, the expected number of monochromatic edges in a uniformly random 2-coloring can be albitrarily large. So picking such a random coloring will not work, and a direct application of the Erdős-Selfridge approach is also doomed to failure. Instead, we will take a different approach.

As in the previous section, we will color the vertices one-at-a-time, but here our choice of colors is different. We will not specify how to make the color choices until later in our discussion as how we do so depends on whether we are using a randomized algorithm or a deterministic one. However, the reader should be aware that these choices are not made to avoid monochromatic edges. Rather, we avoid monochromatic edges by permitting ourselves to leave some vertices uncolored. Specifically, when we consider a vertex v, if it lies in an edge half of whose vertices have been colored, all with the same color, then we will not color v. This ensures that after this first pahse, no completely colored edge is monochromatic. Indeed, if an edge does not contain vertices of both colors, then at least half of its vertices are uncolored.

In our second phase, we will color the vertices which were passed over in the first phase so that every edge contains a vertex of both colors. Since the only edges with which we need to concern ourselves have at least $k/2$ uncolored vertices, and since $d \times 2^{1-k/2} < \frac{1}{4}$, a simple application of the Local

Lemma, ensures that there exists a completition of the partial 2-coloring to a proper 2-coloring.

We have thus reduced our problem to a similar smaller problem which, at first sight, does not seem any easier to solve. It turns out however, that if we make a judicious choice of the color assignments in our first pass, then the smaller problem will have a very simple structure and so we will be able to quickly find a completion of our coloring in a straightforward manner.

One way to make judicious choices is to simple assign a uniformly random color to each vertex that we color. This ensures that the probability of the first $k/2$ vertices of an edge all being given the same color is at most $2^{-k/2}$. Using this fact, we can show that with sufficiently high probability, the subhypergraph we need to color in the second phase has very small components. So we can carry out the second phase by using, on each component, an exhaustive search through all possible colorings.

This yields a fairly simple randomized algorithm. To derandomize the algorithm, we must find a deterministic way of carrying out the first phase so that the components of the resulting subhypergraph are all small. We can do so by applying the Erdős-Selfridge technique; when coloring a vertex, we choose the color which minimizes the conditional expected number of large components.

In this introductory discussion, we have oversimplified the procedure somewhat. For example, we usually repeat the first phase twice to reduce even further the size of the components which we color using brute force. Also, when derandomizing the algorithm, we do not compute the conditional expected number of large components in the subhypergraph considered in the second phase. Instead, we bound this number by focusing on a larger variable which is simpler to deal with. These and other details will be given more fully in the next section.

3.1 The Algorithm

We start by presenting the randomized form of our algorithm in the case where k is fixed. We are given a hypergraph H on n vertices and m hyperedges satisfying the conditions of Theorem 17. Since we are doing an asymptotic analysis of the running time, we can assume that m is large.

In the first phase, we arbitrarily order the vertices $v_1, \ldots, v_n$. We go through the vertices in this order, assigning a uniformly random color to each vertex. After coloring v_i, if an edge e containing v_i has half its vertices colored, all with the same color, then we say that e is *bad*, and for each $v_j \in e$ with $j > i$, at step j we will pass over v_j without assigning it a color.

We let U denote the set of vertices which are not colored during the first phase, and we let M denote the set of edges which don't yet have vertices of both colors. For each $e \in M$ we define e' to be the set of uncolored vertices in e. An important consequence of our procedure is that every e' has size at least $\frac{1}{2}k$.

In our second phase, we color U. No edge outside M can become monochromatic, no matter how we color U, so we can ignore all such edges in this phase. Thus, we focus our attention on the hypergraph H' with vertex set U, and edge set $\{e' | e \in M\}$. As we observed, each edge of H' has size at least $\frac{1}{2}k$, and no edge intersects more than $2^{k/16-2}$ other edges. Therefore the Local Lemma implies that there exists a proper 2-coloring of H'. Clearly, using such a 2-coloring to complete the partial coloring formed in the first phase will yield a proper 2-coloring of H.

The main part of our analysis is to show that the components of H' will all be small. In particular, letting m be the number of edges in H, we will prove the following.

Lemma 19. *With probability at least $\frac{1}{2}$, every component of H' has at most $5(d+1)\log m$ vertices.*

So we can run Phase I, and if H' has any component with more that $5(d+1)\log m$ vertices, then we start over. The expected number of times we must do this is at most 2. Having obtained such an H', we can find a proper 2-coloring of H' in polynomial time using exhaustive search. For, to find such a 2-coloring of H', we need only find a proper 2-coloring of each of its components. Since $k = O(1)$ we have $d = O(1)$ and so there are only $2^{O(\log m)} = \text{poly}(m)$ candidate 2-colorings for each of these components.

The main step is to prove Lemma 19, which we do now. We will bound the expected number of large components of H' by showing that every such component must have many disjoint bad edges. Because disjoint edges become bad independently, the probability that a specific large collection of disjoint edges all turn bad is very small. This will help show that the probability of H' having a big component is small.

We use $L(H)$ to denote the line graph of H. $L^{(a,b)}(H)$ is the graph with vertex set $V(L(H))(=E(H))$ in which two vertices are adjacent if they are at distance exactly a or b in $L(H)$. We call $T \subseteq E(H)$ a $(1,2)$-*tree* if the subgraph induced by T in $L^{(1,2)}(H)$ is connected. We call $T \subseteq E(H)$ a $(2,3)$-*tree* if the subgraph induced by T in $L^{(2,3)}$ is connected and T is a stable set in $L(H)$ (i.e. no two edges of T intersect in H). We call an (a,b)-tree *bad* if it contains only bad edges from H.

Lemma 20. *Every component C of H' with l vertices contains a bad $(2,3)$-tree with at least $l/(k(d+1))$ edges.*

This lemma follows immediately from two simple facts:

Fact 21. *Every component C of H' with l vertices contains a bad $(1,2)$-tree with at least l/k edges.*

Proof. Note that every vertex of C lies in a bad edge, by definition of H'. Since each edge contains at most k vertices, the number of bad edges in C is at least l/k.

Let T be a maximal subset of the bad hyperedges of C which forms a $(1,2)$-tree. We show that T contains all the bad hyperedges of C. Suppose the contrary and consider some bad hyperedge $e \notin T$. If e intersects T, then e can be added to T to form a larger $(1,2)$-tree, thereby contradicting the maximality of T. Otherwise, since C is connected, there must be a path from T to e; let P be the shortest such path. Let e_0, e_1, e_2 be the first three hyperedges of P, where e_0 is in T. Consider any v in $e_1 \cup e_2$ and any bad edge f containing v. Now, f does not belong to T, by the minimality of P, but f is of distance at most 2 from T. Thus f can be added to T to form a larger bad $(1,2)$-tree, thereby contradicting the maximality of T. $\square$

Fact 22. *For any $(1,2)$-tree with t hyperedges, there is a subset of at least $t/(d+1)$ of these edges which forms a $(2,3)$-tree.*

Proof. Consider any $(1,2)$-tree T and a maximal subset T' of the hyperedges of T which forms a $(2,3)$-tree. Consider any hyperedge $e \in T - T'$; we will show that e must intersect some edge in T'. This implies our fact since every hyperedge intersects at most d hyperedges in $T - T'$.

So suppose that e does not intersect any hyperedges in T'. Since T is a $(1,2)$-tree, there must be a path in $L^{(1,2)}(H)$ from T' to e; let P be a shortest such path, and let $e_0 \in T'$ be the first hyperedge in P. If this path has no internal hyperedge, then e is at distance 2 in $L(H)$ from e_0 and so e can be added to T' to form a bigger $(2,3)$-tree, thereby contradicting the maximality of T'.

Otherwise P has at least 1 internal edge, so let e_0, e_1, e_2 be the first 3 edges on P, where again $e_0 \in T'$. If e_1 does not intersect any edge in T', then we can add e_1 to T' to form a larger $(2,3)$-tree. Otherwise, since e_2 is at distance at most 2 from e_1, it is at distance at most 3 from an edge in T'. By minimality of the path, e_2 does not intersect any edge in T', so we can add e_2 to T' to form a larger $(2,3)$-tree, thereby contradicting the maximality of T'. □

These two facts imply Lemma 20. Along with Claim 23, below, and Markov's Inequality, this yields Lemma 19.

Claim 23. *The expected number of bad* $(2,3)$*-trees with at least* $\frac{5}{k}\log m$ *edges is less than* $\frac{1}{2}$*.*

Proof. We first show that for each $r \geq 1$, H contains fewer than $m \times (4d^3)^r$ different $(2,3)$-trees with r hyperedges. To choose such a $(2,3)$-tree, we will first choose an unlabeled tree T on r vertices. It is well-known that there are at most 4^r choices for T. We then choose an edge of H to map onto each vertex of T, starting with an arbitrary vertex v_1 of T, and then proceeding through the rest of the vertices of T in a breadth-first order. There are m choices of an edge in H to map onto v_1. For every subsequent vertex v_i of T, we have already specified an edge e' which maps onto a neighbor of v_i. Thus, the edge mapping onto v_i must be to one of the at most d^3 neighbors in $L^{(2,3)}(H)$ of e'. Therefore, there are a total of $4^r \times m \times (d^3)^{r-1} < m \times (4d^3)^r$ such $(2,3)$-trees, as required.

Now consider any such a tree. It is easily seen that the probability of a particular edge becoming bad is at most $2^{1-\frac{1}{2}k}$. Furthermore, since no two edges of the tree intersect, the probability that all of them become bad is at most $(2^{1-\frac{1}{2}k})^r$. Therefore, the expected number of bad $(2,3)$-trees of size r is at most $m(4d^3 \times 2^{1-\frac{1}{2}k})^r$. Since $d \leq 2^{k/16-2}$, this expected number is at most $m(2^{3k/16-4} \times 2^{1-\frac{1}{2}k})^r < m \times 2^{-5kr/16}$, which is less than $\frac{1}{2}$ for $r \geq \frac{5}{k}\log m$. Of course, if there is a bad $(2,3)$-tree of size at least $\frac{5}{k}\log m$, then there is one of size exactly $\lceil \frac{5}{k}\log m \rceil$, and so this completes our proof. □

References

[1] N. Alon, J. H. Spencer, *The probabilistic method*, Courant Institute, 2000.

[2] J. Matoušek, J. Vondrák, *The probabilistic method*, Institute for Theoretical Computer Science, 2002.

[3] D. B. West, *Introduction to Graph Theory*, Prentice Hall, Upper Saddle River, 1996.