A faster combinatorial algorithm for triangle detection

Sanjana Das

April 23, 2025

§1 Introduction

In the problem of *triangle detection*, we're given a tripartite graph, and we want to determine whether it contains a triangle. To fix some notation, we'll define it as follows (where A, B, and C are the adjacency matrices between the three parts).

Problem 1.1 (Triangle detection)

- **Input:** Matrices $A \in \{0,1\}^{X\times Y}$, $B \in \{0,1\}^{Y\times Z}$, and $C \in \{0,1\}^{X\times Z}$.
- **Output:** Whether there exists $(x,y,z) \in X \times Y \times Z$ with $A(x,y) = B(y,z) = C(x,z) = 1$.

In this report, we'll explain the proof of the following result of [1].

Theorem 1.2 (Abboud–Fischer–Kelley–Lovett–Meka 2024)

There is an $n^3 \cdot 2^{-\Omega((\log n)^{1/6})}$ time combinatorial algorithm for triangle detection.

(The bound in [1] has $(\log n)^{1/7}$ instead of $(\log n)^{1/6}$, but one can show their algorithm actually achieves $(\log n)^{1/6}$ by very slightly tweaking their analysis, so I'll do that here.)

The historical context for this result is that it's known that any subcubic algorithm for triangle detection would imply one for *Boolean matrix multiplication* (BMM) [4]. Although BMM can be solved in $O(n^\omega)$ time using integer matrix multiplication, there's interest in specifically finding *combinatorial* BMM algorithms. Brute force takes time n^3; in 1970, [2] found a $n^3 \cdot (\log n)^{-2}$ time algorithm (now called the 'Four-Russians' algorithm); and since then, the best we've been able to do is $n^3 \cdot (\log n)^{-4}$ [5]. Theorem 1.2, along with the reduction of [4], provides an $n^3 \cdot 2^{-\Omega((\log n)^{1/6})}$ time algorithm for BMM, which is a substantial improvement.

§2 Regularity and triangle detection

The approach of [1] is based on *graph regularity* — the paper defines certain notions of regularity and shows that every graph has a decomposition into a controlled number of regular pieces, on which solving triangle detection is much easier. Here, we'll define the notions of regularity that the paper uses and discuss why these notions are relevant to triangle detection; then we'll state [1]'s decomposition result and explain why it gives an $2^{-\Omega((\log n)^{1/6})}$ speedup for triangle detection.

§2.1 Definitions

In this section, we'll define the notions of regularity that [1] uses; these will be conditions on a *single* matrix $A \in \{0,1\}^{X\times Y}$ (which it'll often be useful to think of as a bipartite graph). To fix some notation:

- We write $\mathbb{E}[A] := \mathbb{E}_{x\in X,\, y\in Y}[A(x,y)]$ (intuitively, this is the density of A).
- For each $x \in X$, we write $d_A(x) := \mathbb{E}_{y\in Y}[A(x,y)]$ (intuitively, this is a normalized version of the degree of x). Similarly, for each $y \in Y$, we write $d_A(y) = \mathbb{E}_{x\in X}[A(x,y)]$.

We'll have two regularity conditions, defined as follows.

Definition 2.1. For $\varepsilon \in (0,1)$, we say A is ε-min-degree if for all $x \in X$, we have $d_A(x) \geq (1-\varepsilon)\mathbb{E}[A]$.

Intuitively, this states that no vertices in X have degree substantially lower than average.

Definition 2.2. For $k, \ell \in \mathbb{N}$, we define the (k,ℓ)-grid norm of A as

$$\|A\|_{U(k,\ell)} = \left(\mathbb{E}_{x_1,\ldots,x_k\in X,\, y_1,\ldots,y_\ell\in Y}\left[\prod_{i=1}^{k}\prod_{j=1}^{\ell} A(x_i,y_j)\right]\right)^{1/k\ell}.$$

We say A is (ε,k,ℓ)-grid regular if $\|A\|_{U(k,\ell)} \leq (1+\varepsilon)\mathbb{E}[A]$.

Intuitively, $\|A\|_{U(k,\ell)}^{k\ell}$ is the probability that randomly chosen $x_1,\ldots,x_k \in X$ and $y_1,\ldots,y_\ell \in Y$ form a $K_{k,\ell}$ (allowing repeated vertices). If A were a random graph with density $\mathbb{E}[A]$, we'd expect this to be roughly $\mathbb{E}[A]^{k\ell}$, so we'd have $\|A\|_{U(k,\ell)} \approx \mathbb{E}[A]$. So grid regularity essentially states that A doesn't have too many more copies of $K_{k,\ell}$ than a random graph of the same density.

§2.2 Relevance to triangle detection

The reason these notions of regularity are useful for triangle detection is the following theorem of [3], which states that a *product* of two ε-min-degree $(\varepsilon,2,d)$-grid regular matrices is close to uniform.

Theorem 2.3 (Kelley–Lovett–Meka 2024)
Let $\varepsilon \in (0, \frac{1}{80})$ and $d \geq 2/\varepsilon$. Suppose that $A \in \{0,1\}^{X\times Y}$ and $B \in \{0,1\}^{Y\times Z}$ are such that A and $B^\intercal$ are ε-min-degree and $(\varepsilon,2,d)$-grid regular. Then for all but a $2^{-\varepsilon d/2}$-fraction of $(x,z) \in X \times Z$, we have

$$(1-80\varepsilon)\mathbb{E}[A]\mathbb{E}[B] \leq \mathbb{E}_{y\in Y}[A(x,y)B(y,z)] \leq (1+80\varepsilon)\mathbb{E}[A]\mathbb{E}[B].$$

There isn't enough space to prove this theorem, but we'll prove a 'baby' version — a one-sided version whose proof is much simpler — in order to hopefully provide some intuition for why it's plausible that these notions of regularity could lead to such a statement.

Claim 2.4 — Suppose that $A \in \{0,1\}^{X\times Y}$ and $B \in \{0,1\}^{Y\times Z}$ are such that A and $B^\intercal$ are $(\varepsilon,2,d)$-grid regular. Then for all but a $2^{-\varepsilon d}$-fraction of $(x,z) \in X\times Z$, we have

$$\mathbb{E}_{y\in Y}[A(x,y)B(y,z)] \leq (1+10\varepsilon)\mathbb{E}[A]\mathbb{E}[B].$$

Proof. Assume for contradiction that this is not the case. Then we have

$$\mathbb{E}_{x\in X,\, z\in Z}\left[(\mathbb{E}_{y\in Y}[A(x,y)B(y,z)])^d\right] \geq 2^{-\varepsilon d}\cdot(1+10\varepsilon)^d\mathbb{E}[A]^d\mathbb{E}[B]^d \geq (1+8\varepsilon)^d\mathbb{E}[A]^d\mathbb{E}[B]^d.$$

On the other hand, we can write the left-hand side as $\mathbb{E}_{y_1,\ldots,y_d\in Y}[\mathbb{E}_{x\in X}[\prod_{i=1}^d A(x,y_i)][\mathbb{E}_{z\in Z}\prod_{i=1}^d B(y_i,z)]]$ (taking a dth power corresponds to taking d independent copies of y, and once we fix y_1, ..., y_d, then the parts corresponding to x and z become independent), and Cauchy–Schwarz gives that this is at most

$$\left(\mathbb{E}_{x_1,x_2\in X,\, y_1,\dots,y_d\in Y}\left[\prod_{i=1}^{d} A(x_1,y_i)A(x_2,y_i)\right]\cdot \mathbb{E}_{z_1,z_2\in Z,\, y_1,\dots,y_d\in Y}\left[\prod_{i=1}^{d} B(y_i,z_1)B(y_i,z_2)\right]\right)^{1/2}$$

(the squares that arise from Cauchy–Schwarz correspond to taking two independent copies of x and z). But these expressions exactly correspond to the $(2,d)$-grid norms of A and $B^\intercal$! So we've shown that

$$(1+8\varepsilon)^d\mathbb{E}[A]^d\mathbb{E}[B]^d \le \mathbb{E}_{x\in X,\, z\in Z}\left[(\mathbb{E}_{y\in Y}[A(x,y)B(y,z)])^d\right] \le \|A\|_{U(2,d)}^d \cdot \|B^\intercal\|_{U(2,d)}^d\,,$$

and since we assumed $\|A\|_{U(2,d)} \le (1+\varepsilon)\mathbb{E}[A]$ and $\|B^\intercal\|_{U(2,d)} \le (1+\varepsilon)\mathbb{E}[B]$, this is a contradiction. □

For the actual statement of Theorem 2.3, we don't just care about $\mathbb{E}_{y\in Y}[A(x,y)B(y,z)]$ being too *big*; we also care about it being too *small*. This is intuitively why we need the ε-min-degree condition — if A had a bunch of low-degree vertices $x \in X$ and B were random, then $\mathbb{E}_{y\in Y}[A(x,y)B(y,z)]$ would be small for these $x \in X$. It turns out that these two conditions are enough to get the two-sided bound in Theorem 2.3; but the proof requires significantly more complicated analytic arguments, so we won't present it here.

Why is Theorem 2.3 useful for triangle detection? Imagine that we're given input (A,B,C), and we know that A and $B^\intercal$ are ε-min-degree and $(\varepsilon,2,d)$-grid regular (think of ε as a small absolute constant and d as slowly growing with n). Then Theorem 2.3 guarantees that all but a $2^{-\varepsilon d/2}$-fraction of the entries of AB are positive. To solve triangle detection, we just want to know whether some positive entry of AB coincides with a 1 in C. So if more than a $2^{-\varepsilon d/2}$-fraction of the entries of C are 1's, then we automatically know the answer is yes, without doing any computations. Meanwhile, if at most a $2^{-\varepsilon d/2}$-fraction of the entries of C are 1's, then we can simply brute force over the edges $(x,z)\in X\times Z$ given by C (of which there are at most $2^{-\varepsilon d/2}|X||Z|$) and all $y \in Y$; this costs $2^{-\varepsilon d/2}|X||Y||Z|$, giving a $2^{-\varepsilon d/2}$ speedup.

§2.3 A regularity decomposition and triangle detection algorithm

In general, our matrices won't satisfy these regularity conditions. However, [1] shows that we can always decompose AB into not too many smaller matrix products A_kB_k which *do* satisfy these conditions.

Theorem 2.5

Let $\varepsilon \in (0,1)$ and $d \in \mathbb{N}$. There is an $n^2 \cdot 2^{O_\varepsilon(d^6)}$ time algorithm $\mathsf{DecomposeProduct}(A,B)$ which takes matrices $A \in \{0,1\}^{X\times Y}$ and $B \in \{0,1\}^{Y\times Z}$ and outputs $\{(X_k,Y_k,Z_k,A_k,B_k)\}_{k\in\mathcal{K}}$, where $X_k \subseteq X$, $Y_k \subseteq Y$, $Z_k \subseteq Z$, $A_k \in \{0,1\}^{X_k\times Y_k}$, and $B_k \in \{0,1\}^{Y_k\times Z_k}$ are such that:

(i) We have $AB = \sum_{k\in\mathcal{K}} A_kB_k$ (padding each A_kB_k with 0's to get a matrix in $\{0,1\}^{X\times Z}$).

(ii) For each $k \in \mathcal{K}$, either $\mathbb{E}[A_k] \le 2^{-d}$, $\mathbb{E}[B_k] \le 2^{-d}$, or A_k and $B_k^\intercal$ are ε-min-degree and $(\varepsilon,2,d)$-grid regular.

(iii) We have $\sum_{k\in\mathcal{K}} |X_k||Y_k||Z_k| \le 2(d+2)^2|X||Y||Z|$.

(iv) We have $|\mathcal{K}| \le 2^{O_\varepsilon(d^6)}$.

We'll prove this in Section 3; here, we'll see how to use it to get a $2^{-\Omega((\log n)^{1/6})}$ speedup for triangle detection.

Proof of Theorem 1.2. Set $\varepsilon = \frac{1}{160}$ and $d = c(\log n)^{1/6}$, where c is a constant small enough that both $2^{O_\varepsilon(d^6)}$ terms in Theorem 2.5 are at most $n^{1/4}$. Then to solve triangle detection on input (A,B,C), we first compute

$$\{(X_k,Y_k,Z_k,A_k,B_k)\}_{k\in\mathcal{K}} \leftarrow \mathsf{DecomposeProduct}(A,B),$$

and set $C_k \leftarrow C[X_k,Z_k]$ for each $k\in\mathcal{K}$. Then (A,B,C) has a triangle if and only if (A_k,B_k,C_k) does for some $k\in\mathcal{K}$. So we loop through all $k\in\mathcal{K}$, doing the following:

- If $\mathbb{E}[A_k] > 2^{-d}$ and $\mathbb{E}[B_k] > 2^{-d}$, then Theorem 2.5(ii) states that A_k and $B_k^\intercal$ are ε-min-degree and $(\varepsilon, 2, d)$-grid regular, so by Theorem 2.3, all but a $2^{-\varepsilon d/2}$-fraction of entries of $A_k B_k$ are positive.
  - If $\mathbb{E}[C_k] > 2^{-\varepsilon d/2}$, then we automatically know (A_k, B_k, C_k) has a triangle.
  - If $\mathbb{E}[C_k] \le 2^{-\varepsilon d/2}$, then we can brute force over edges $(x, z) \in X_k \times Z_k$ given by C_k and then over $y \in Y_k$; this costs $2^{-\varepsilon d/2} \left|X_k\right| \left|Y_k\right| \left|Z_k\right|$.
- If $\mathbb{E}[A_k] \le 2^{-d}$, then we can brute force over edges $(x, y) \in X_k \times Y_k$ given by A_k and then $z \in Z_k$; this costs $2^{-d} \left|X_k\right| \left|Y_k\right| \left|Z_k\right|$. Similarly, if $\mathbb{E}[B_k] \le 2^{-d}$, then we can brute force over edges $(y, z) \in Y_k \times Z_k$ given by B_k and then $y \in Y_k$ (with the same cost).

We spend at most $2^{-\varepsilon d/2} \left|X_k\right| \left|Y_k\right| \left|Z_k\right|$ time on each $k \in \mathcal{K}$, so by Theorem 2.5(iii), we have total runtime

$$2^{-\varepsilon d/2} \sum_{k \in \mathcal{K}} \left|X_k\right| \left|Y_k\right| \left|Z_k\right| \le 2^{-\varepsilon d/2} \cdot \mathsf{poly}(d) \left|X\right| \left|Y\right| \left|Z\right| = 2^{-\Omega((\log n)^{1/6})} \left|X\right| \left|Y\right| \left|Z\right|. \qquad \square$$

§3 Constructing a regularity decomposition

In this section, we'll prove Theorem 2.5. We'll mostly ignore questions of runtime; but for most steps of the argument, it should be clear how to implement them with good runtime. (The one exception is Lemma 3.2, which we'll comment on in Remark 3.4.)

§3.1 Regularity vs. density increments

As a first step towards Theorem 2.5, we'll consider a much simpler problem: Given a matrix $A \in \{0, 1\}^{X \times Y}$, how do we find a *single* regular piece? For this, we'll use a *density increment* strategy — we'll show that we can either enforce regularity or find subsets $X' \subseteq X$ and $Y' \subseteq Y$ on which A is substantially denser. Then we can imagine replacing A with $A[X', Y']$ and iterating. Since our density can't go above 1, this must terminate after a controlled number of steps, giving a regular piece.

Lemma 3.1

There is an algorithm $\mathsf{MinDegree}(A, \varepsilon, \gamma)$ which takes in $A \in \{0, 1\}^{X \times Y}$ and $\varepsilon, \gamma \in (0, 1)$ and outputs $X' \subseteq X$ such that $|X'| \ge (1 - \gamma) |X|$ and $\mathbb{E}[A[X', Y]] \ge \mathbb{E}[A[X, Y]]$, and either:

- (Regular case) $A[X', Y]$ is ε-min-degree; or
- (Increment case) $\mathbb{E}[A[X', Y]] \ge (1 + \varepsilon\gamma)\mathbb{E}[A]$.

Proof. We initialize $X' \leftarrow X$. Then we repeatedly find vertices $x \in X'$ with $d_A(x) < (1 - \varepsilon)\mathbb{E}[A[X', Y]]$ and remove them from X'. We stop when there are no such vertices (which means $A[X', Y]$ is ε-min-degree) or $\mathbb{E}[A[X', Y]] \ge (1 + \varepsilon\gamma)\mathbb{E}[A]$.

We're only removing vertices of below-average degree, so our density increases at every step. To see that we must stop by the time we've removed a γ-fraction of vertices, note that each removed vertex satisfies

$$d_A(x) < (1 - \varepsilon)(1 + \varepsilon\gamma)\mathbb{E}[A].$$

So if we've removed a γ-fraction of vertices, we'll have

$$\mathbb{E}[A[X', Y]] \ge \frac{\mathbb{E}[A] - \gamma(1 - \varepsilon)(1 + \varepsilon\gamma)\mathbb{E}[A]}{1 - \gamma} \ge (1 + \varepsilon\gamma)\mathbb{E}[A],$$

which means we'll terminate (if we haven't already terminated). $\square$

Lemma 3.2

There is an algorithm $\mathsf{GridRegular}(A, \varepsilon, k, \ell)$ taking $A \in \{0,1\}^{X \times Y}$, $\varepsilon \in (0,1)$, and $k, \ell \in \mathbb{N}$, which either:

- (Regular case) Correctly outputs that A is (ε, k, ℓ)-grid regular; or
- (Increment case) Outputs $X' \subseteq X$ and $Y' \subseteq Y$ with $|X'| \geq \frac{\varepsilon}{8} \cdot \mathbb{E}[A]^{k\ell} \cdot |X|$ and $|Y'| \geq \mathbb{E}[A]^{k^2} \cdot |Y|$, such that $\mathbb{E}[A[X', Y']] \geq (1 + \frac{\varepsilon}{2})\mathbb{E}[A]$.

Proof. For $x \in X$, we write $Y_x = \{y \in Y \mid A(x, y) = 1\}$ and $A_x = A[X, Y_x]$.

Claim 3.3 — Let $\alpha, \varepsilon \in (0,1)$, and suppose that $\|A\|_{U(k,\ell)} \geq (1+\varepsilon)\alpha$. Then either:

- (Good case) At least an $\varepsilon\alpha^{k\ell}$-fraction of $x \in X$ satisfy $d_A(x) \geq \alpha$; or
- (Drop-k case) $k \geq 2$, and there is $x \in X$ with $d_A(x) \geq \alpha^k$ and $\|A_x\|_{U(k-1,\ell)} \geq (1+\varepsilon)\alpha$.

(When proving Lemma 3.2, we'll take α and ε to be $(1 + \frac{\varepsilon}{2})\mathbb{E}[A]$ and $\frac{\varepsilon}{8}$, respectively.)

Proof. First, we claim that we can write

$$\|A\|_{U(k,\ell)}^{k\ell} = \mathbb{E}_{x \in X}[f(x)] \quad \text{where} \quad f(x) = \begin{cases} d_A(x)^\ell & \text{if } k = 1 \\ d_A(x)^\ell \cdot \|A_x\|_{U(k-1,\ell)}^{(k-1)\ell} & \text{if } k \geq 2. \end{cases} \tag{3.1}$$

This is because $\|A\|_{U(k,\ell)}^{k\ell}$ is the probability that randomly chosen $x_1, \ldots, x_k \in X$ and $y_1, \ldots, y_\ell \in Y$ form a $K_{k,\ell}$. If we imagine first choosing $x_1 = x$, then $y_1, \ldots, y_\ell$ need to all land in Y_x; this happens with probability $d_A(x)^\ell$. And conditional on this, we're choosing $x_2, \ldots, x_k \in X$ and $y_1, \ldots, y_\ell \in Y_x$ and checking whether they form a $K_{k-1,\ell}$; this has probability $\|A_x\|_{U(k-1,\ell)}^{(k-1)\ell}$.

We'd like to use (3.1) and the assumption $\|A\|_{U(k,\ell)}^{k\ell} \geq (1+\varepsilon)^{k\ell}\alpha^{k\ell}$ to say that there are many $x \in X$ for which $f(x)$ is large. For this, for any $t \in [0,1]$, we can write

$$\mathbb{E}_{x \in X}[f(x)] \leq \mathbb{P}_{x \in X}[f(x) \geq t] \cdot 1 + \mathbb{P}_{x \in X}[f(x) < t] \cdot t \leq \mathbb{P}_{x \in X}[f(x) \geq t] + t.$$

So we say $x \in X$ is *interesting* if $f(x) \geq (1+\varepsilon)^{(k-1)\ell}\alpha^{k\ell}$; then plugging in $t = (1+\varepsilon)^{(k-1)\ell}\alpha^{k\ell}$ gives that

$$\mathbb{P}_{x \in X}[x \text{ is interesting}] \geq (1+\varepsilon)^{k\ell}\alpha^{k\ell} - (1+\varepsilon)^{(k-1)\ell}\alpha^{k\ell} \geq \varepsilon\alpha^{k\ell}.$$

Now, if every interesting $x \in X$ satisfies $d_A(x) \geq \alpha$, then we're in the good case. Otherwise, suppose there's some interesting $x \in X$ with $d_A(x) < \alpha$. Then we must have $k \geq 2$, and

$$\|A_x\|_{U(k-1,\ell)} = \left(\frac{f(x)}{d_A(x)^\ell}\right)^{1/(k-1)\ell} > \left(\frac{(1+\varepsilon)^{(k-1)\ell}\alpha^{k\ell}}{\alpha^\ell}\right)^{1/(k-1)\ell} = (1+\varepsilon)\alpha.$$

Also, we have $\alpha^{k\ell} \leq f(x) \leq d_A(x)^\ell$, which ensures $d_A(x) \geq \alpha^k$. So we're in the drop-k case. □

Now to prove Lemma 3.2, note that if A is not (ε, k, ℓ)-grid regular, then the hypothesis of Claim 3.3 holds with $\alpha = (1 + \frac{\varepsilon}{2})\mathbb{E}[A]$ and ε replaced by $\frac{\varepsilon}{8}$. So we initialize $Y' \leftarrow Y$, $A' \leftarrow A$, and $k' \leftarrow k$, and repeat:

- If at least an $\frac{\varepsilon}{8}\alpha^{k'\ell}$-fraction of $x \in X$ satisfy $d_{A'}(x) \geq \alpha$, let X' be the set of such $x \in X$, and return (X', Y'). This guarantees $\mathbb{E}[A[X', Y']] \geq \alpha$, and $|X'| \geq \frac{\varepsilon}{8}\alpha^{k'\ell} \cdot |X| \geq \frac{\varepsilon}{8} \cdot \mathbb{E}[A]^{k\ell} \cdot |X|$.

- Otherwise, Claim 3.3 guarantees there is some $x \in X$ with $d_{A'}(x) \geq \alpha^{k'}$ (which means $|Y'_x| \geq \alpha^{k'} \cdot |Y'| \geq \mathbb{E}[A]^k \cdot |Y'|$) and $\|A'_x\|_{U(k'-1,\ell)} \geq (1 + \frac{\varepsilon}{8})\alpha$. Set $Y' \leftarrow Y_x$, $A' \leftarrow A_x$, and $k' \leftarrow k' - 1$; the first statement means that this shrinks Y' by at worst a factor of $\mathbb{E}[A]^k$, and the second statement means that the hypothesis of Claim 3.3 continues to hold.

We land in the second case at most k times (because each time, we drop k' by 1) and each shrinks Y' by a factor of $\mathbb{E}[A]^k$, so in the end we'll have $|Y'| \geq \mathbb{E}[A]^{k^2} \cdot |Y|$, as desired. □

Remark 3.4. Implementing this algorithmically seems to require us to compute grid norms (since we need to find $x \in X$ for which $\|A'_x\|_{U(k'-1,\ell)}$ is large), and it's not clear how to do so efficiently. In fact, it's good enough to be able to *estimate* these grid norms. There's a simple randomized algorithm to do so — $\|A\|_{U(k,\ell)}^{k\ell}$ is the probability that randomly chosen vertices form a $K_{k,\ell}$, and we can estimate this probability by randomly sampling. In fact, [1] shows that this can be derandomized using *oblivious samplers*, but we won't discuss that here.

§3.2 Decomposing a single matrix

In this section, as a first step towards Theorem 2.5, we'll construct a regularity decomposition of a *single* matrix. First, we'll show how to find *one* regular piece (which we refer to as a 'good rectangle'), using the density increment strategy discussed at the beginning of the previous section.

Lemma 3.5

Let $\varepsilon \in (0, 1)$ and $d \in \mathbb{N}$. There is an algorithm $\mathsf{GoodRectangle}(A)$ which takes a matrix $A \in \{0,1\}^{X \times Y}$ with $\mathbb{E}[A] \geq 2^{-d}$ and outputs (X_*, Y_*), where $X_* \subseteq X$ and $Y_* \subseteq Y$ are such that:

(i) $A[X_*, Y_*]$ is ε-min-degree and $(\varepsilon, 2, d)$-grid regular.

(ii) $\mathbb{E}[A[X_*, Y_*]] \geq \mathbb{E}[A]$.

(iii) $|X_*| \geq 2^{-O_\varepsilon(d^3)}\,|X|$ and $|Y_*| \geq 2^{-O_\varepsilon(d^2)}\,|Y|$.

Proof. We initialize $X_* \leftarrow X$ and $Y_* \leftarrow Y$, and repeat the following loop:

(1) Update $X_* \leftarrow \mathsf{MinDegree}(A[X_*, Y_*], \varepsilon, \frac{1}{2})$. If we're in the regular case (meaning that $A[X_*, Y_*]$ is ε-min-degree), we proceed to the next step. Otherwise we say this step has *failed* and go back to the start of this loop.

(2) Run $\mathsf{GridRegular}(A[X_*, Y_*], \varepsilon, 2, d)$. If we're in the regular case, we output (X_*, Y_*) and terminate. Otherwise, we say this step has *failed*; we update $X_* \leftarrow X'_*$ and $Y_* \leftarrow Y'_*$ (where X'_* and Y'_* are as given by Lemma 3.2) and go back to the start of the loop.

It's clear that in the end $A[X_*, Y_*]$ is ε-min-degree (as ensured by Step (1)) and $(\varepsilon, 2, d)$-grid regular (as ensured by Step (2)), and each of these steps can only increase $\mathbb{E}[A[X_*, Y_*]]$.

To prove (iii), note that each time Step (1) fails, we've shrunk X_* by a factor of $\frac{1}{2}$ and multiplied the density $\mathbb{E}[A[X_*, Y_*]]$ by $(1 + \frac{\varepsilon}{2})$. Since we start out with density at least 2^{-d}, and our density can't go above 1, this occurs at most $O_\varepsilon(d)$ times; so it causes X_* to shrink by a factor of $2^{-O_\varepsilon(d)}$.

Meanwhile, each time Step (2) fails, the entire round has shrunk X_* by a factor of $\frac{1}{2} \cdot \frac{\varepsilon}{8} \cdot \mathbb{E}[A]^{2d} \geq \frac{\varepsilon}{16} \cdot 2^{-2d^2}$ (the factor of $\frac{1}{2}$ accounts for Step (1), and the rest comes from the bound in Lemma 3.2, noting that $\mathbb{E}[A[X_*, Y_*]] \geq \mathbb{E}[A] \geq 2^{-d}$) and Y_* by a factor of $\mathbb{E}[A]^4 \geq 2^{-4d}$. And again, we get a multiplicative density increment of $(1 + \frac{\varepsilon}{2})$ each time this happens, so it happens at most $O_\varepsilon(d)$ times. So in total, this shrinks X_* by $2^{-O_\varepsilon(d^3)}$ and Y_* by $2^{-O_\varepsilon(d^2)}$.

So overall, X_* shrinks by $2^{-O_\varepsilon(d)} \cdot 2^{-O_\varepsilon(d^3)} = 2^{-O_\varepsilon(d^3)}$ and Y_* by $2^{-O_\varepsilon(d^2)}$, proving (iii). □

Now we'll obtain a full decomposition of A by iteratively removing good rectangles.

Lemma 3.6

Let $\varepsilon \in (0,1)$ and $d \in \mathbb{N}$. There is an algorithm $\mathsf{Decompose}(A)$ which takes $A \in \{0,1\}^{X\times Y}$ and outputs $\{(X_\ell, Y_\ell, A_\ell)\}_{\ell\in\mathcal{L}}$, where $X_\ell \subseteq X$, $Y_\ell \subseteq Y$, and $A_\ell \in \{0,1\}^{X_\ell \times Y_\ell}$ are such that:

(i) We have $A = \sum_{\ell\in\mathcal{L}} A_\ell$ (padding each A_ℓ with 0's to get a matrix in $\{0,1\}^{X\times Y}$).

(ii) For each $\ell \in \mathcal{L}$, either $\mathbb{E}[A_\ell] \le 2^{-d}$, or A_ℓ is ε-min-degree and $(\varepsilon, 2, d)$-grid regular.

(iii) We have $\sum_{\ell\in\mathcal{L}} |X_\ell|\,|Y_\ell| \le (d+2)\,|X|\,|Y|$.

(iv) For each $\ell \in \mathcal{L}$, we have $|Y_\ell| \ge 2^{-O_\varepsilon(d^2)}\,|Y|$.

(v) We have $|\mathcal{L}| \le 2^{O_\varepsilon(d^3)}$.

Proof. We initialize $A' \leftarrow A$, and then repeatedly do the following (where ℓ is the current index):

- If $\mathbb{E}[A'] \ge 2^{-d}$, we compute $(X_\ell, Y_\ell) \leftarrow \mathsf{GoodRectangle}(A')$, add $(X_\ell, Y_\ell, A'[X_\ell, Y_\ell])$ to our decomposition, and update $A' \leftarrow A' - A'[X_\ell, Y_\ell]$.
- If $\mathbb{E}[A'] < 2^{-d}$, we add (X, Y, A') to our decomposition and terminate.

It's clear that this satisfies (i) (we're repeatedly finding pieces of A, inserting them to our decomposition, and removing them from A) and (ii) (all but the last piece will be regular, and the last will be sparse).

To see (iii), the key insight is that every piece $A'[X_\ell, Y_\ell]$ that we remove is at least as dense as the current A' (by Lemma 3.5(ii)), so removing it drops the density of A' by a factor of at least $1 - \frac{|X_\ell||Y_\ell|}{|X||Y|}$. This density starts out at most 1, and it can't go below 2^{-d} until the end. This means

$$\prod_\ell \left(1 - \frac{|X_\ell|\,|Y_\ell|}{|X|\,|Y|}\right) \ge 2^{-d},$$

where the product ranges over all but the last two indices ℓ (the second-last ℓ drops our density below 2^{-d}, and the last is the one where we take everything that's left). Using the fact that $1 - t \le e^{-t} \le 2^{-t}$ (for $t \ge 0$) gives that $\sum_\ell |X_\ell|\,|Y_\ell| \le d\,|X|\,|Y|$, and replacing d with $d+2$ accounts for the last two indices.

Finally, Lemma 3.5(iii) states that $|X_\ell| \ge 2^{-O_\varepsilon(d^3)}\,|X|$ and $|Y_\ell| \ge 2^{-O_\varepsilon(d^2)}\,|Y|$ for all $\ell \in \mathcal{L}$, which implies (iv). It also means that $|X_\ell|\,|Y_\ell| \ge 2^{-O_\varepsilon(d^3)}\,|X|\,|Y|$ for all $\ell \in \mathcal{L}$, and together with (iii), this implies (v). □

§3.3 Decomposing a product

In this section, we'll prove Theorem 2.5. First, we'll prove the following analog of Lemma 3.5.

Lemma 3.7

Let $\varepsilon \in (0,1)$, $\gamma \in (0, \frac{1}{2})$, and $d \in \mathbb{N}$. There is an algorithm $\mathsf{GoodCube}(A, B)$ which takes $A \in \{0,1\}^{X\times Y}$ and $B \in \{0,1\}^{Y\times Z}$ with $\mathbb{E}[B] \ge 2^{-d}$ and outputs $(Y_*, Z_*, \{(X_{*\ell}, Y_{*\ell}, Z_{*\ell}, A_{*\ell})\}_{\ell\in\mathcal{L}})$ where:

(i) $Y_* \subseteq Y$, $Z_* \subseteq Z$, $\{(X_{*\ell}, Y_{*\ell}, A_{*\ell})\}_{\ell\in\mathcal{L}}$ is a decomposition of $A[X, Y_*]$ with the properties given by Lemma 3.6, and for every $\ell \in \mathcal{L}$ we have $Z_{*\ell} \subseteq Z_*$ and $|Z_{*\ell}| \ge (1-\gamma)\,|Z_*|$.

(ii) For each $\ell \in \mathcal{L}$, $B[Y_{*\ell}, Z_{*\ell}]^\intercal$ is ε-min-degree and $(\varepsilon, 2, d)$-grid regular.

(iii) $\mathbb{E}[B[Y_*, Z_*]] \ge \mathbb{E}[B]$.

(iv) We have $|Y_*| \ge 2^{-O_\varepsilon(d^3/\gamma)}\,|Y|$ and $|Z_*| \ge 2^{-O_\varepsilon(d^3)}\,|Z|$.

Here's some intuition for how to think about this statement: When we decomposed a single matrix A, the idea was to repeatedly take big bites out of A (using Lemma 3.5). To decompose a product AB, we'd like to take big bites out of B. If we want to take out a bite (Y_*, Z_*), then we need to account for $A[X, Y_*]B[Y_*, Z_*]$. We can imagine doing so by taking a regularity decomposition $\{(X_{*\ell}, Y_{*\ell}, A_{*\ell})\}_{\ell\in\mathcal{L}}$ of $A[X, Y_*]$, and adding $\{(X_{*\ell}, Y_{*\ell}, Z_*, A_{*\ell}, B[Y_{*\ell}, Z_*])\}_{\ell\in\mathcal{L}}$ to our decomposition of AB. Lemma 3.7 gives us a way to do not exactly this, but something very 'close' — we find subsets $Z_{*\ell} \subseteq Z_*$ which cover all but a tiny fraction of Z_* (as quantified by γ) and add $\{(X_{*\ell}, Y_{*\ell}, Z_{*\ell}, A_{*\ell}, B[Y_{*\ell}, Z_{*\ell}])\}_{\ell\in\mathcal{L}}$ to the decomposition instead. This misses the contributions of $A[X_{*\ell}, Y_{*\ell}]B[Y_{*\ell}, Z_* \setminus Z_{*\ell}]$; but these are small, so we'll be able to handle them recursively.

Proof. We initialize $Y_* \leftarrow Y$ and $Z_* \leftarrow Z$ and repeat the following loop:

(1) Update $Y_* \leftarrow \mathsf{MinDegree}(B[Y_*, Z_*], \frac{\varepsilon\gamma}{4}, \frac{1}{2})$. If we're in the regular case (so $B[Y_*, Z_*]$ is $\frac{\varepsilon\gamma}{4}$-min-degree), we proceed to the next step; otherwise we say this step has *failed* and go back to the start of the loop.

(2) Compute $\{(X_{*\ell}, Y_{*\ell}, A_{*\ell})\}_{\ell\in\mathcal{L}} \leftarrow \mathsf{Decompose}(A[X, Y_*])$.

(3) Loop through all $\ell \in \mathcal{L}$ and do the following:

  (3a) Compute $Z_{*\ell} \leftarrow \mathsf{MinDegree}(B[Y_{*\ell}, Z_*]^\intercal, \varepsilon, \gamma)$. If we're in the regular case, we move to the next step. Otherwise we say this step has *failed*, and we update $Y_* \leftarrow Y_{*\ell}$ and $Z_* \leftarrow Z_{*\ell}$, break out of this inner loop, and return to the start of the outer loop (forgetting about all the other pieces of our decomposition of $A[X, Y_*]$).

  (3b) Run $\mathsf{GridRegular}(B[Y_{*\ell}, Z_{*\ell}]^\intercal, \varepsilon, 2, d)$. If we're in the regular case, we move on (moving to the next value of ℓ). If we're in the increment case (so it returns some $Z'_{*\ell} \subseteq Z_{*\ell}$ and $Y'_{*\ell} \subseteq Y_{*\ell}$), we say this step has *failed*. We update $Y_* \leftarrow Y'_{*\ell}$ and $Z_* \leftarrow Z'_{*\ell}$, break out of this inner loop, and return to the start of the outer loop.

If we get through all $\ell \in \mathcal{L}$ without failing, we return $(Y_*, Z_*, \{(X_{*\ell}, Y_{*\ell}, Z_{*\ell}, A_{*\ell})\}_{\ell\in\mathcal{L}})$ and terminate.

For the analysis, first note that on every iteration of the outer loop, since we ensure $B[Y_*, Z_*]$ is $\frac{\varepsilon\gamma}{4}$-min-degree in Step (1), we'll have $\mathbb{E}[B[Y_{*\ell}, Z_*]] \geq (1 - \frac{\varepsilon\gamma}{4})\mathbb{E}[B[Y_{*\ell}, Z_*]]$ for all $\ell \in \mathcal{L}$. Lemma 3.1 guarantees that every time Step (3a) fails, it'll give us a density increment of $(1 + \varepsilon\gamma)$ over $B[Y_{*\ell}, Z_*]$; this means we get a density increment of at least $(1 + \varepsilon\gamma)(1 - \frac{\varepsilon\gamma}{4}) \geq (1 + \frac{\varepsilon\gamma}{2})$ over the $B[Y_*, Z_*]$ we started the round with. Similarly, every time Step (3b) fails, we get a density increment of at least $(1 + \frac{\varepsilon}{2})(1 - \frac{\varepsilon\gamma}{4}) \geq (1 + \frac{\varepsilon}{8})$.

This in particular means the density of $B[Y_*, Z_*]$ only ever increases, proving (iii). It also means the density of every $B[Y_{*\ell}, Z_{*\ell}]^\intercal$ in Step (3b) is at least $(1 - \frac{\varepsilon\gamma}{4}) \cdot 2^{-d} \geq 2^{-2d}$ (when proving (iv), we'll plug this into the bounds in Lemma 3.2 when analyzing how much Step (3b) shrinks our sets by).

To prove (iv), every time Step (1) fails, we've shrunk Y_* by $\frac{1}{2}$ and gotten a multiplicative density increment of $(1 + \frac{\varepsilon\gamma}{4})$. Since our density starts out at least 2^{-d}, this occurs at most $O_\varepsilon(d/\gamma)$ times, so in total, it causes Y_* to shrink by $2^{-O_\varepsilon(d/\gamma)}$.

Every time Step (3a) fails, we've shrunk Y_* by $\frac{1}{2} \cdot 2^{-O_\varepsilon(d^2)}$ (where $\frac{1}{2}$ comes from Step (1), and $2^{-O_\varepsilon(d^2)}$ comes from the fact that we're passing to a piece $Y_{*\ell}$ of the regularity decomposition of $A[X, Y_*]$, using Lemma 3.6(iv)) and Z_* by $(1 - \gamma)$. And we get a density increment of $(1 + \varepsilon\gamma)$, so this occurs at most $O_\varepsilon(d/\gamma)$ times, and in total it shrinks Y_* by $2^{-O_\varepsilon(d^3)}$ and Z_* by $(1 - \gamma)^{O_\varepsilon(d/\gamma)} = 2^{-O_\varepsilon(d)}$.

Finally, every time Step (3b) fails, we've shrunk Y_* by $\frac{1}{2} \cdot 2^{-O_\varepsilon(d^2)} \cdot 2^{-8d} = 2^{-O_\varepsilon(d^2)}$ (these factors come from Step (1), passing from Y_* to a piece $Y_{*\ell}$, and running $\mathsf{GridRegular}$, respectively — Lemma 3.2 gives that $\mathsf{GridRegular}$ causes it to shrink by $\mathbb{E}[B[Y_{*\ell}, Z_{*\ell}]]^4$, and we have $\mathbb{E}[B[Y_{*\ell}, Z_{*\ell}]] \geq 2^{-2d}$). Similarly, we've shrunk Z_* by a factor of $(1 - \gamma) \cdot \frac{\varepsilon}{8} \cdot 2^{-4d^2}$ (these factors come from running $\mathsf{MinDegree}$ and $\mathsf{GridRegular}$, respectively). And we get a density increment of $(1 + \frac{\varepsilon}{2})$, so this occurs at most $O_\varepsilon(d)$ times; so in total, it causes Y_* to shrink by $2^{-O_\varepsilon(d^3)}$ and Z_* by $2^{-O_\varepsilon(d^3)}$.

So in total, Y_* shrinks by $2^{-O_\varepsilon(d/\gamma)} \cdot 2^{-O_\varepsilon(d^3/\gamma)} \cdot 2^{-O_\varepsilon(d^3)} = 2^{-O_\varepsilon(d^3/\gamma)}$, and Z_* shrinks by $2^{-O_\varepsilon(d)} \cdot 2^{-O_\varepsilon(d^3)} = 2^{-O_\varepsilon(d^3)}$, proving (iv). □

Finally, we'll use this to prove Theorem 2.5 (in the manner discussed after the statement of Lemma 3.7).

Proof of Theorem 2.5. Set $\gamma = \frac{1}{4(d+2)^2}$. We initialize $B' \leftarrow B$, and repeat (where m is the current index):

- If $\mathbb{E}[B'] \geq 2^{-d}$, then we compute $(Y_m, Z_m, \{(X_{m\ell}, Y_{m\ell}, Z_{m\ell}, A_{m\ell})\}_{\ell \in \mathcal{L}}) \leftarrow \mathsf{GoodCube}(A, B')$, and we add $(X_{m\ell}, Y_{m\ell}, Z_{m\ell}, A_{m\ell}, B'[Y_{m\ell}, Z_{m\ell}])$ to our decomposition for all $\ell \in \mathcal{L}$.

  Then we recursively run $\mathsf{DecomposeProduct}(A[X_{m\ell}, Y_{m\ell}], B'[Y_{m\ell}, Z_m \setminus Z_{m\ell}])$ for all $\ell \in \mathcal{L}$, and add all its outputs to our decomposition as well.

  Finally, this accounts for the entire contribution of $B'[Y_m, Z_m]$ to our matrix product; so we remove it, meaning that we update $B' \leftarrow B' \setminus B'[Y_m, Z_m]$.
- If $\mathbb{E}[B'] < 2^{-d}$, we add (X, Y, Z, A, B') to our decomposition and terminate.

We also cut off the recursion at depth d — when we hit recursion depth d, we don't perform the above procedure. Instead, we split B into 2^d matrices on the *same* host sets, each with density at most 2^{-d}. More precisely, we write $B = \sum_{i=1}^{2^d} B_i$ where $B_i \in \{0,1\}^{Y \times Z}$ and $\mathbb{E}[B_i] \leq 2^{-d}$ for all i, and add (X, Y, Z, A, B_i) to the decomposition for all i. (In the analysis, we'll show that our subproblems at depth d are so small that we can afford to do this.)

To prove (iii), let's first consider the top level of the recursion. Lemma 3.6(iii) gives that $\sum_\ell |X_{m\ell}|\,|Y_{m\ell}| \leq (d+2)\,|X|\,|Y_m|$, and the same argument from its proof also shows $\sum_m |Y_m|\,|Z_m| \leq (d+2)\,|Y|\,|Z|$ — every piece $B'[Y_m, Z_m]$ we remove is at least as dense as the current B', which means that removing it shrinks the density of B' by a factor of $1 - \frac{|Y_m||Z_m|}{|Y||Z|}$, and this density must remain above 2^{-d} until the end. So

$$\sum_{m,\ell} |X_{m\ell}|\,|Y_{m\ell}|\,|Z_{m\ell}| \leq \sum_{m,\ell} |X_{m\ell}|\,|Y_{m\ell}|\,|Z_m| \leq (d+2)\sum_m |X|\,|Y_m|\,|Z_m| \leq (d+2)^2\,|X|\,|Y|\,|Z|\,. \tag{3.2}$$

Then when we recurse, we have a subproblem on the sets $(X_{m\ell}, Y_{m\ell}, Z_m \setminus Z_{m\ell})$ for all m and ℓ, and Lemma 3.5 ensures that $|Z_m \setminus Z_{m\ell}| \leq \gamma\,|Z_m|$. So if we think of $|X|\,|Y|\,|Z|$ as the 'volume' of our original problem, then the total volume of all our subproblems (at recursive depth 1) is at most

$$\sum_{m,\ell} |X_{m\ell}|\,|Y_{m\ell}|\,|Z_m \setminus Z_{m\ell}| \leq \gamma \sum_{m,\ell} |X_{m\ell}|\,|Y_{m\ell}|\,|Z_m| \leq \gamma(d+2)^2\,|X|\,|Y|\,|Z| \leq \frac{1}{4}\,|X|\,|Y|\,|Z|\,.$$

So every time we move a level down in the recursion, our total volume multiplies by $\frac{1}{4}$. This means when se sum (3.2) over all levels of the recursion, we'll get $\sum_k |X_k|\,|Y_k|\,|Z_k| \leq (d+2)^2\,|X|\,|Y|\,|Z|\,(1 + \frac{1}{4} + \frac{1}{16} + \cdots + \frac{1}{4^{d-1}} + \frac{1}{4^d} \cdot 2^d) \leq 2(d+2)^2\,|X|\,|Y|\,|Z|$. (The final term is because we're cutting off the recursion at depth d and doing something that has blowup 2^d there.) This proves (iii).

For (iv), again let's first consider the top level of the recursion. Lemma 3.7(iv) guarantees that $|Y_m|\,|Z_m| \geq 2^{-O_\varepsilon(d^3/\gamma)} \cdot 2^{-O_\varepsilon(d^3)}\,|Y|\,|Z| = 2^{-O_\varepsilon(d^5)}\,|Y|\,|Z|$ for all m, and we have $\sum_m |Y_m|\,|Z_m| \leq (d+2)^2\,|Y|\,|Z|$ (as we saw above), so the total number of indices m is at most $2^{O_\varepsilon(d^5)}$. And each m corresponds to at most $2^{O_\varepsilon(d^3)}$ indices ℓ by Lemma 3.6(v). So in total, we add at most $2^{O_\varepsilon(d^5)} \cdot 2^{O_\varepsilon(d^3)} = 2^{O_\varepsilon(d^5)}$ pieces to the decomposition at the top level of recursion. This also means we spawn at most $2^{O_\varepsilon(d^5)}$ subproblems. So each level of the recursion multiplies our number of subproblems by $2^{O_\varepsilon(d^5)}$, and we have d levels; this means we have $2^{O_\varepsilon(d^6)}$ total subproblems, and therefore the decomposition has $2^{O_\varepsilon(d^6)} \cdot 2^{O_\varepsilon(d^5)} = 2^{O_\varepsilon(d^6)}$ parts. ☐

References

[1] Amir Abboud, Nick Fischer, Zander Kelley, Shachar Lovett, and Raghu Meka. New graph decompositions and combinatorial Boolean matrix multiplication algorithms. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, pages 935–943, 2024. arxiv:2311.09095.

[2] V. L. Arlazarov, Y. A. Dinic, M. A. Kronrod, and I. A. Faradzhev. On economical construction of the transitive closure of an oriented graph. *Doklady Academii Nauk SSSR*, 194(3):487–488, 1970.

[3] Zander Kelley, Shachar Lovett, and Raghu Meka. Explicit separations between randomized and deterministic number-on-forehead communication. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, pages 1299–1310, 2024. arxiv:2308.12451.

[4] Virginia Vassilevska Williams and R. Ryan Williams. Subcubic equivalences between path, matrix, and triangle problems. *Journal of the ACM*, 65(5):1–38, 2018.

[5] Huacheng Yu. An improved combinatorial algorithm for Boolean matrix multiplication. *Information and Computation*, 261(2):240–247, 2018.