COVERING SYSTEMS WITH RESTRICTED DIVISIBILITY

ROBERT D. HOUGH and PACE P. NIELSEN

Abstract
We prove that every distinct covering system has a modulus divisible by either 2 *or* 3.

1. Introduction

A covering system of congruences is a collection

$$a_i \bmod m_i, \quad i = 1, 2, \ldots, k$$

such that every integer satisfies at least one of them. A covering system is distinct if the moduli m_i are distinct and greater than 1. Erdős [3] introduced the idea of a distinct covering system of congruences in constructing an arithmetic progression of odd numbers, none of whose members is the sum of a prime and a power of two, answering a question of Romanoff [11]. Using similar ideas, Sierpinski [15] proved that there are infinitely many odd integers k such that $k2^n + 1$ is composite for all natural numbers n; it is still open whether 78557 is the smallest such number. Some other uses of covering systems include proofs of irreducibility of special classes of polynomials and other positive density results.

In [3], Erdős asked whether the least modulus of a distinct covering system of congruences can be arbitrarily large. The largest known minimum modulus is 42, given by Tyler Owens [10], improving the record of 40 obtained previously by the second author [9]. The first author [8] recently answered Erdős's question in the negative, proving that the least modulus of a distinct covering system of congruences is at most 10^{16}. The proof relies on the Lovász local lemma (LLL); the paper [5] seems to be the first to explicitly mention this probabilistic tool in the context of covering systems. In the present paper we improve the techniques related to the LLL, which may be of independent interest.

A second old problem of Erdős and Selfridge asks whether there exists a distinct covering system of congruences with all moduli odd. According to [4], Erdős

DUKE MATHEMATICAL JOURNAL
Advance publication—final volume, issue, and page numbers to be assigned.
© 2019 DOI 10.1215/00127094-2019-0058
Received 13 March 2017. Revision received 10 April 2019.
2010 *Mathematics Subject Classification*. Primary 11A07; Secondary 11B25, 11K99.

has offered \$25 for the proof that no odd distinct covering system of congruences exists, while Selfridge has offered \$2000 for a construction of an odd distinct covering system. Schinzel proved that a negative answer to the odd modulus problem has applications to the irreducibility of families of polynomials. While the odd modulus problem remains open, Simpson and Zeilberger [16] proved that a distinct covering system consisting of odd square-free numbers involves at least 18 primes, which was improved to 22 primes by Guo and Sun [6]. This paper makes further negative progress toward the odd modulus problem.

THEOREM 1
Every distinct covering system of congruences has a modulus divisible by either 2 *or* 3.

This answers a problem raised in [7]. We have stated Theorem 1 in a succinct form, although further minor improvement is possible.

2. Setup

Suppose that we are given a finite set of moduli $\mathcal{M}$, and, for each $m \in \mathcal{M}$, we are given a set of residues $\mathbf{a}_m$ modulo m. Let

$$Q = \mathrm{LCM}(m : m \in \mathcal{M})$$

and

$$R = \mathbb{Z} \Big\backslash \bigcup_{m \in \mathcal{M}} (\mathbf{a}_m \bmod m),$$

which is a set defined modulo Q. One way to show that the congruences

$$(\mathbf{a}_m \bmod m), \quad m \in \mathcal{M}$$

do not cover the integers is to give a positive lower bound for the density of R. The proof of Theorem 1 gives such a lower bound, although it estimates some related quantities. (For results on uncovered sets in other contexts the reader is directed to [5], which contains some strong general purpose bounds.)

If we let $\mathbb{Z}/Q\mathbb{Z}$ have the uniform probability measure, then the density of R is equal to its probability. For $m \in \mathcal{M}$, let A_m be the event $(\mathbf{a}_m \bmod m)$, which has probability $\frac{|\mathbf{a}_m|}{m}$, and extend this to $m \mid Q$ with $m \notin \mathcal{M}$ by setting $A_m = \emptyset$ for these m. Then

$$\mathbf{P}(R) = \mathbf{P}\Big(\bigcap_{m \mid Q} A_m^c\Big). \tag{1}$$

To apply the Lovász local lemma, we think of these events as vertices in a graph. The graph is *valid* (or, in other words, it meets the requirements of the lemma) if the edges are chosen so that each vertex v is independent of the σ-algebra generated by the collection of vertices *not* in the immediate neighborhood of v.

The Chinese remainder theorem implies that A_m is independent of any set of congruences to moduli coprime to m. Thus, a valid dependency graph for the events $\{A_m : m \mid Q\}$ has edge (m, m') if and only if $\mathrm{GCD}(m, m') > 1$.

A family of results connected to the Lovász local lemma gives worst-case lower bounds for the probability of an intersection as in (1), taking as input only the events' probabilities and their dependency graph. In principle, we could hope to prove Theorem 1 by directly applying one of these results to claim that the uncovered set always has a nonzero density, but, as we will see, such a lower bound cannot be given, and further input is needed. Two methods of Lovász type do figure into our argument, however, as we will describe.

Given the problem of estimating from below the probability of the intersection of the complements of some events given only their probabilities and their dependency graph, the best possible estimate has been given by Shearer [14]. The estimate is best possible in the sense that the argument constructs a probability space and events having the prescribed probabilities and dependency graph, and such that the lower bound holds with equality. However, the condition with which Shearer's result holds can be difficult to verify, and so the following result is useful because it is easy to check. Note that this is essentially due to [16] in this context.

THEOREM 2 (Shearer-type theorem)
Suppose that we have a probability space. Let $[n] = \{1, 2, \ldots, n\}$, and assume that for each $1 \le i \le n$ there is a weight π_i assigned, satisfying $1 \ge \pi_1 \ge \pi_2 \ge \cdots \ge \pi_n \ge 0$. Let the sets $\emptyset \ne T \subset [n]$ index events A_T, each having probability

$$0 \le \mathbf{P}(A_T) \le \prod_{t \in T} \pi_t := \pi_T.$$

Assume that A_T is independent of the σ-algebra generated by $\{A_S : S \subset [n], S \cap T = \emptyset\}$, so that a valid dependency graph for the events $\{A_T : \emptyset \ne T \subset [n]\}$ has an edge between $S \ne T$ whenever $S \cap T \ne \emptyset$.

Define $\rho(\emptyset) = 1$, and given $\emptyset \ne T \subset [n]$, set

$$\rho(T) = 1 - \sum_{\emptyset \ne S_1 \subset T} \pi_{S_1} + \sum_{\substack{\emptyset \ne S_1, S_2 \subset T \\ S_1 < S_2 \text{ disjoint}}} \pi_{S_1}\pi_{S_2}$$

$$- \sum_{\substack{\emptyset \ne S_1, S_2, S_3 \subset T \\ S_1 < S_2 < S_3 \text{ disjoint}}} \pi_{S_1}\pi_{S_2}\pi_{S_3} + \cdots.$$

(Here, "$<$" is an arbitrary total ordering on $2^{[n]}$, used to avoid overcounting terms.) Suppose that $\rho([1]) \geq \rho([2]) \geq \cdots \geq \rho([n]) > 0$. Then for any $\emptyset \neq T \subset [n]$,

$$\mathbf{P}\Big(\bigcap_{\emptyset \neq S \subset T} A_S^c\Big) \geq \rho(T) > 0 \tag{2}$$

and, for any $T_1 \subset T_2 \subset [n]$,

$$\frac{\mathbf{P}(\bigcap_{\emptyset \neq S \subset T_2} A_S^c)}{\mathbf{P}(\bigcap_{\emptyset \neq S \subset T_1} A_S^c)} \geq \frac{\rho(T_2)}{\rho(T_1)}. \tag{3}$$

We prove a slightly more general version of this theorem in Appendix C.

To apply the Shearer-type theorem in the context of Theorem 1, order the primes greater than 3 as $p_1 = 5$, $p_2 = 7$, $p_3 = 11, \dots$. Suppose that we are given a distinct congruence system with moduli formed with the primes $p_1, \dots, p_n$. Identify $S \subset [n]$ with the square-free number $m_S = \prod_{i \in S} p_i$, and form the event A_S which is the union of all congruences having square-free part m_S,

$$A_S = \bigcup_{m\colon \mathrm{sqf}(m) = m_S} (a_m \bmod m),$$

where $\mathrm{sqf}(m) = \prod_{p : p | m} p$. Then A_S is an event with probability

$$\mathbf{P}(A_S) < \prod_{i \in S} \frac{1}{p_i - 1}.$$

In particular, we may appeal to Theorem 2 with $\pi_i = \frac{1}{p_i - 1}$. Arguing in this way, we may check that there is no covering composed of only the primes between 5 and 631, but at this point the Shearer function becomes negative, and no further result can be drawn from that estimate.

What allows us to make further progress is that, within the range in which Shearer's theorem holds, estimate (3) of Theorem 2 gives substantial information about the structure of the uncovered set. To see this, suppose that we have a congruence system as above with uncovered set R, and that Theorem 2 applies. We can estimate the proportion of the set R that lies in a given congruence class $(b \bmod m)$ for $m | Q$ by

$$\begin{aligned}\frac{\mathbf{P}((b \bmod m) \cap R)}{\mathbf{P}(R)} &= \frac{\mathbf{P}((b \bmod m) \cap \bigcap_{m' \in \mathcal{M}} (a_{m'} \bmod m')^c)}{\mathbf{P}(\bigcap_{m' \in \mathcal{M}} (a_{m'} \bmod m')^c)} \\ &\leq \mathbf{P}\big((b \bmod m)\big) \frac{\mathbf{P}(\bigcap_{m' \in \mathcal{M}, (m,m')=1} (a_{m'} \bmod m')^c)}{\mathbf{P}(\bigcap_{m' \in \mathcal{M}} (a_{m'} \bmod m')^c)} \\ &= \frac{1}{m} \frac{\mathbf{P}(\bigcap_{m' \in \mathcal{M}, (m,m')=1} (a_{m'} \bmod m')^c)}{\mathbf{P}(\bigcap_{m' \in \mathcal{M}} (a_{m'} \bmod m')^c)}.\end{aligned}$$

The ratio of probabilities on the right is bounded by the relative conclusion (3) of Theorem 2, which gives a ratio of $\frac{\rho([n]\setminus S_m)}{\rho([n])}$, where $[n]$ again represents the full set of primes dividing Q, and the S_m's are those primes from $[n]$ which divide m. Thus,

$$\frac{\mathbf{P}((b \bmod m)\cap R)}{\mathbf{P}(R)} \leq \frac{1}{m}\frac{\rho([n]\setminus S_m)}{\rho([n])}.$$

If the S_m's are such that $\rho([n]\setminus S_m) \approx \rho([n])$, then we deduce that R is almost uniformly distributed across residues modulo m.

We summarize the above discussion in the following theorem.

THEOREM 3
Let $p_1 < p_2 < \cdots < p_n$ be a sequence of primes, and let the weights $\pi_1, \ldots, \pi_n$ be given by $\pi_i = \frac{1}{p_i - 1}$. For a subset $S \subset [n]$ identify S with $q_S = \prod_{p\in S} p$, and write $\rho(q_S) = \rho(S)$ for the Shearer function associated to S with weights π_i, as in Theorem 2.

Suppose that $\rho(p_1) \geq \rho(p_1 p_2) \geq \cdots \geq \rho(p_1 p_2 \cdots p_n) > 0$. Then any distinct congruence system with moduli composed only of $p_1, \ldots, p_n$ does not cover the integers. Moreover, if R is the uncovered set and if m is a modulus composed of primes corresponding to a set $S \subset [n]$, then

$$\max_{b \bmod m} \frac{|R\cap (b \bmod m)|}{|R|} \leq \frac{1}{m}\frac{\rho(q_{[n]\setminus S})}{\rho(q_{[n]})}. \tag{4}$$

Although the sieving problem described in Theorem 1 concerns systems of congruences in which each congruence set $\mathbf{a}_m$ has size 0 or 1, in the course of our argument we consider congruences with sets $\mathbf{a}_m$ of variable size. In this situation the condition of Theorem 2 becomes unwieldy and we appeal instead to the following theorem, which follows from an improved form of the Lovász local lemma due to [1] (see also [13]).

THEOREM 4
Let $\mathcal{N} \subset \mathbb{N}_{>1}$ be a finite collection of moduli whose prime factors are drawn from a set of primes $\mathcal{P}$. Let $\mathrm{LCM}(n : n \in \mathcal{N}) = Q$. Suppose that for each $n \in \mathcal{N}$ a collection of residues $\mathbf{a}_n \bmod n$ is given. Write

$$R = \mathbb{Z} \setminus \bigcup_{n\in\mathcal{N}} (\mathbf{a}_n \bmod n).$$

Suppose that there exist weights $\{x_p\}_{p\in\mathcal{P}}$ with $x_p \geq 0$, which satisfy the constraints

$$\forall p \in \mathcal{P}, \quad x_p \geq \sum_{n\in\mathcal{N}: p|n} \frac{|\mathbf{a}_n \bmod n| \prod_{p'|n}(1+x_{p'})}{n}.$$

Then the density of R is at least

$$\frac{|R \bmod Q|}{Q} \geq \exp\Big(-\sum_{n\in\mathcal{N}} \frac{|\mathbf{a}_n \bmod n| \prod_{p|n}(1+x_p)}{n}\Big) > 0. \tag{5}$$

Also, for any $n \in \mathcal{N}$,

$$\max_{b \bmod n} \frac{|R \cap (b \bmod n) \bmod Q|}{|R \bmod Q|} \leq \frac{\exp(\sum_{p|n} x_p)}{n}. \tag{6}$$

Remark
Conclusion (5) corresponds to (2) of Theorem 2, and (6) corresponds to (4).

If we write $\underline{x}$ for $\{x_p\}_{p\in\mathcal{P}}$ and $\underline{G}(\underline{x})$ for

$$G_p(\underline{x}) = \sum_{n\in\mathcal{N}:p|n} \frac{|\mathbf{a}_n \bmod n| \prod_{p'|n}(1+x_{p'})}{n},$$

then the condition of Theorem 4 equivalently asks for a nonnegative ($\underline{x} \geq \underline{0}$) fixed point $\underline{G}(\underline{x}) = \underline{x}$, which is relatively easy to determine. Thus, although Theorem 4 is strictly weaker than Theorem 2, it is useful since it is more easily applied.

A proof and further discussion of Theorem 4 is given in Section 4.

3. Overview of the argument

We now give an overview of our argument. As the structure is similar to that of the minimum modulus problem, we refer the proofs of some background statements to [8].

We assume that we are given a congruence system with a finite set of moduli

$$\mathcal{M} \subset \big\{m > 1, (m, 6) = 1\big\},$$

together with a residue class $a_m \bmod m$ for each $m \in \mathcal{M}$. We let

$$Q = \mathrm{LCM}(m : m \in \mathcal{M}),$$

and set

$$R = \mathbb{Z} \setminus \bigcup_{m\in\mathcal{M}} (a_m \bmod m)$$

for the set left uncovered by the congruence system. Theorem 1 follows by showing that the density of R is positive.

To estimate the density of R we appeal to the Lovász local lemma-type arguments of the previous section. These arguments, however, only apply to estimate the density of sets left uncovered by congruence systems whose moduli are composed of a limited number of primes, and so we break the estimate for the density of R into stages.

Let $P_0 = 4 < P_1 < P_2 < \cdots$ be a sequence of real numbers (not equal to prime integers). Let $Q_0 = 1$, and, for $i \geq 1$, let

$$Q_i = \prod_{p^j \| Q, p < P_i} p^j$$

be the part of Q composed of primes less than P_i. We let $\mathcal{M}_i = \{m \in \mathcal{M} : m \mid Q_i\}$ be the P_i-smooth moduli in $\mathcal{M}$, and we let the set of "new factors" be

$$\mathcal{N}_i = \left\{n > 1 : n \mid Q_i, p \mid n \Rightarrow P_{i-1} < p \leq P_i\right\}.$$

Notice that each $m \in \mathcal{M}_{i+1} \setminus \mathcal{M}_i$ has a unique factorization as $m = m_0 n$ with $m_0 \mid Q_i$ and $n \in \mathcal{N}_{i+1}$.

We consider the sequence of sets $\mathbb{Z} = R_0 \supset R_1 \supset \cdots$,

$$\forall i \geq 1, \quad R_i = \mathbb{Z} \setminus \bigcup_{m \in \mathcal{M}_i} (a_m \bmod m).$$

Since $R_i = R$ eventually, it will suffice to show that R_i is nonempty for each i.

The set R_i is defined modulo Q_i. Viewing $\mathbb{Z}/Q_{i+1}\mathbb{Z}$ as fibered over $\mathbb{Z}/Q_i\mathbb{Z}$ we note that

$$R_{i+1} = R_i \setminus \bigcup_{m \in \mathcal{M}_{i+1} \setminus \mathcal{M}_i} (a_m \bmod m),$$

so that we may view R_{i+1} as cut out from the fibers $(r \bmod Q_i)$, $r \in R_i$, by congruences to moduli in $\mathcal{M}_{i+1} \setminus \mathcal{M}_i$. Given $r \in R_i$ and $m \in \mathcal{M}_{i+1} \setminus \mathcal{M}_i$, factor $m = m_0 n$ with $m_0 | Q_i$ and $n \in \mathcal{N}_{i+1}$. Then the congruence $(a_m \bmod m)$ meets $(r \bmod Q_i)$ if and only if $r \equiv a_{m_0 n} \bmod m_0$, and when it does so, it intersects in a single residue class modulo nQ_i. Thus, grouping together moduli according to a common new factor $n \in \mathcal{N}_{i+1}$ we find

$$R_{i+1} \cap (r \bmod Q_i) = (r \bmod Q_i) \setminus \bigcup_{n \in \mathcal{N}_{i+1}} A_{n,r},$$

with

$$A_{n,r} = (r \bmod Q_i) \cap \bigcup_{\substack{m_0 | Q_i \\ m_0 n \in \mathcal{M}_{i+1}}} (a_{m_0 n} \bmod m_0 n).$$

After translating and dilating $(r \bmod Q_i)$ to coincide with the integers, the set $A_{n,r}$ is composed of some residue classes modulo n, a set which we call $\mathbf{a}_{n,r}$. Thus, we can understand the problem of estimating the density of R_{i+1} within $(r \bmod Q_i)$ as sieving the integers by multiple residue classes to moduli in $\mathcal{N}_{i+1}$, a set of moduli whose prime factors are constrained to lie in $(P_i, P_{i+1}]$. This is the situation treated by the Lovász-type result, Theorem 4 above, and so, if we are able to solve the relevant fixed-point problem, then we obtain that the fiber is nonempty. Note that in the initial stage, all of the sieving sets have size 0 or 1, so that in this stage we can appeal to the optimal Shearer-type theorem, Theorem 2.

In practice we will not estimate the density of R_{i+1} over all of R_i, but only within certain "good" fibers above a subset $R_i^* \subset R_i \bmod Q_i$. We will be deliberately vague at this point about the requirements of a good fiber. Roughly speaking, they ensure that the corresponding fixed-point problem has a favorable solution. Also, we require that $R_i^* \subset R_{i-1}^* \cap R_i$ so that the good sets are nested. We let $R_0^* = R_0 = \mathbb{Z}$.

For $i \geq 1$, we give weights to the set $\mathbb{Z}/Q_i\mathbb{Z}$ according to the probability measure μ_i supported on $R_{i-1}^* \cap R_i$, chosen so as to guarantee that a large proportion of the fibers are good. The measure μ_1 is uniform on the set $R_0^* \cap R_1 = R_1 \subset \mathbb{Z}/Q_1\mathbb{Z}$, and

$$\forall r \in R_0^* \cap R_1 \bmod Q_1, \quad \mu_1(r) = \frac{1}{|R_1 \bmod Q_1|}.$$

Taking the measure μ_i as given, define, for $i \geq 1$,

$$\pi_{\text{good}}(i) = \frac{\mu_i(R_i^*)}{\mu_i(R_{i-1}^* \cap R_i)}$$

to be the proportion of good fibers. For $i \geq 1$ and $r \in R_i^* \cap R_{i+1} \bmod Q_{i+1}$, we set

$$\mu_{i+1}(r) = \frac{\mu_i(r \bmod Q_i)}{\pi_{\text{good}}(i)|R_{i+1} \cap (r \bmod Q_i) \bmod Q_{i+1}|}.$$

Thus, for a fixed $r \in R_i$, μ_{i+1} is constant on $R_{i+1} \cap (r \bmod Q_i)$. That μ_i is a sequence of probability measures follows from [8, Lemma 2], although, note that the factor of $\frac{1}{\pi_{\text{good}}(i)}$ is not included in the definition of μ_i in [8], so that the measures there do not have mass 1. Throughout, when we write $\mathbf{E}_{r \in R_{i-1}^* \cap R_i}$ we mean expectation with respect to the measure μ_i.

Along with the measure μ_i we track some bias statistics of $R_{i-1}^* \cap R_i$. Let $\ell_k(m)$ be the multiplicative function given at prime powers by

$$\ell_k(p^j) = (j+1)^k - j^k.$$

For $i \geq 1$, the kth bias statistic of $R_{i-1}^* \cap R_i$ is defined to be

$$\beta_k^k(i) = \sum_{m \mid Q_i} \ell_k(m) \max_{b \bmod m} \mu_i\big((b \bmod m)\big).$$

The importance of the bias statistics is that they control moments of (mixtures of) the sizes of the sets $\mathbf{a}_{n,r}$ as r varies in $R_{i-1}^* \cap R_i$.

LEMMA 5
Let $i \geq 1$. Let $\{w_n : n \in \mathcal{N}_{i+1}\}$ be any collection of nonnegative weights, not all of which are zero. For each $k \geq 1$, we have

$$\mathbf{E}_{r \in R_{i-1}^* \cap R_i} \Big(\sum_{n \in \mathcal{N}_{i+1}} w_n |\mathbf{a}_{n,r} \bmod n| \Big)^k \leq \Big(\sum_{n \in \mathcal{N}_{i+1}} w_n \Big)^k \beta_k^k(i).$$

Proof
See Lemmas 4 and 5 of [8]. □

In addition to the bias statistics, it will be useful for us to track maximum biases among the various good fibers. Let $i \geq 0$, and let $n \in \mathcal{N}_{i+1}$. We define the maximum bias at n to be

$$b_n = \max_{r \in R_i^*} \max_{b \bmod n} \frac{n |R_{i+1} \cap (r \bmod Q_i) \cap (b \bmod n) \bmod Q_{i+1}|}{|R_{i+1} \cap (r \bmod Q_i) \bmod Q_{i+1}|}.$$

Note that these appeared only implicitly in [8], but to get a better quantitative bound it will be useful for us to track them more carefully here.

The iterative growth of the bias statistics $\beta_k(i)$ to $\beta_k(i+1)$ is controlled by the proportion of good fibers $\pi_{\text{good}}(i)$ and the maximal biases at $n \in \mathcal{N}_{i+1}$.

LEMMA 6
Let $i \geq 1$. For each $k \geq 1$, we have the bound

$$\beta_k^k(i+1) \leq \frac{\beta_k^k(i)}{\pi_{\text{good}}(i)} \Big(1 + \sum_{n \in \mathcal{N}_{i+1}} \frac{\ell_k(n) b_n}{n} \Big).$$

Proof
This follows by tracing the proof of Proposition 3 of [8]. □

We now turn to giving a detailed account of Theorem 4.

4. The local lemma and good fibers

Our Theorem 4, which is used to estimate the density of good fibers, is derived from the following improved version of the Lovász local lemma due to [1] (see also [13]).

THEOREM 7 (Clique Lovász local lemma)
Suppose that $G = (V, E)$ is a dependency graph for a family of events $\{A_v\}_{v \in V}$, each

with probability $\mathbf{P}(A_v) \leq \pi_v$. *Let* N_v *be the neighborhood of* $v \in V$. *Suppose that there exists a sequence* $\underline{\mu} = \{\mu_v\}_{v\in V}$ *of reals in* $[0, \infty)$ *such that, for each* $v \in V$,

$$\mu_v \geq \pi_v \phi_v(\underline{\mu}), \tag{7}$$

where

$$\phi_v(\underline{\mu}) = \sum_{\substack{R \subset \{v\} \cup N_v \\ R \text{ indep. in } G}} \prod_{v' \in R} \mu_{v'}.$$

Then

$$\mathbf{P}\Big(\bigcap_{v\in V} A_v^c\Big) \geq \exp\Big(-\sum_{v\in V} \mu_v\Big) \tag{8}$$

and, for all $U \subset V$,

$$\frac{\mathbf{P}(\bigcap_{v\in V} A_v^c)}{\mathbf{P}(\bigcap_{u\in U} A_u^c)} \geq \exp\Big(-\sum_{v \in V\setminus U} \mu_v\Big). \tag{9}$$

Remark
In the definition of ϕ_v, $R = \emptyset$ is to be included, with associated product equal to 1.

Proof
This theorem with conclusion

$$\mathbf{P}\Big(\bigcap_{v\in V} A_v^c\Big) \geq \prod_{v\in V} (1-\pi_v)^{\phi_v(\underline{\mu})-\mu_v} \tag{10}$$

is proved in [1], and the corresponding relative conclusion

$$\frac{\mathbf{P}(\bigcap_{v\in V} A_v^c)}{\mathbf{P}(\bigcap_{u\in U} A_u^c)} \geq \prod_{v\in V\setminus U} (1-\pi_v)^{\phi_v(\underline{\mu})-\mu_v} \tag{11}$$

follows directly from the argument there. To deduce (8) and (9), observe that

$$\phi_v(\underline{\mu}) - \mu_v \leq (1-\pi_v)\phi_v(\underline{\mu}),$$

so that

$$\begin{aligned}(1-\pi_v)^{\phi_v(\underline{\mu})-\mu_v} &\geq \exp\big(\phi_v(\underline{\mu})(1-\pi_v)\log(1-\pi_v)\big) \\ &\geq \exp\big(-\phi_v(\underline{\mu})\pi_v\big) \geq \exp(-\mu_v). \qquad \square\end{aligned}$$

Recall that Theorem 4 applies in the context of a congruence system to moduli in a set $\mathcal{N}$, whose prime factors lie in a set $\mathcal{P}$. Each modulus $n \in \mathcal{N}$ has a set of residues $\mathbf{a}_n$, considered to be a probabilistic event with probability $\frac{|\mathbf{a}_n|}{n}$. We require a system of nonnegative weights $\{x_p\}_{p\in\mathcal{P}}$ satisfying

$$x_p \geq \sum_{n\in\mathcal{N}: p|n} \frac{|\mathbf{a}_n \bmod n| \prod_{p'|n}(1+x_{p'})}{n},$$

and the conclusion is that the uncovered set R has density at least

$$\mathbf{P}(R) \geq \exp\Big(-\sum_{n\in\mathcal{N}} \frac{|\mathbf{a}_n \bmod n| \prod_{p|n}(1+x_p)}{n}\Big)$$

and that, for any $n \in \mathcal{N}$, for any $b \bmod n$,

$$\frac{\mathbf{P}(R\cap(b \bmod n))}{\mathbf{P}(R)} \leq \frac{\exp(\sum_{p|n} x_p)}{n}.$$

Deduction of Theorem 4
To deduce Theorem 4 from Theorem 7, we take V to be the set of nontrivial square-free products of primes in $\mathcal{P}$,

$$V = \{v > 1, \text{square-free}, p \mid v \Rightarrow p \in \mathcal{P}\}.$$

The event associated to $v \in V$ is the union of congruences $(\mathbf{a}_n \bmod n)$ for which $\mathrm{sqf}(n) = v$, and this event has probability

$$\pi_v \leq \sum_{n:\mathrm{sqf}(n)=v} \frac{|\mathbf{a}_n|}{n}.$$

The dependency graph connects v_1 and v_2 if and only if $\mathrm{GCD}(v_1, v_2) > 1$.

Define $\mu_v = \pi_v \prod_{p|v}(1+x_p)$. This has the effect of reducing (7) at v to the constraint

$$\prod_{p|v}(1+x_p) \geq \phi_v(\underline{\mu}). \tag{12}$$

Notice that

$$\phi_v(\underline{\mu}) = \sum_{\substack{R\subset\{v\}\cup N_v\\ \text{independent}}} \prod_{v'\in R} \mu_{v'} \leq \prod_{p|v}\Big(1+\sum_{v':p|v'} \mu_{v'}\Big)$$

since each term in the sum on the left appears in the expansion of the product on the right. Thus, if we make the condition that for each $p \mid Q$,

$$x_p \geq \sum_{v':p|v'} \mu_{v'},$$

which is the condition (12) in the case $v = p$, then (12) holds automatically for all v. In this way we have reduced to guaranteeing the system of prime constraints

$$\forall p|Q, \quad x_p \geq \sum_{v':p|v'} \pi_{v'} \prod_{p':p'|v'} (1 + x_{p'}), \tag{13}$$

which is the constraint of Theorem 4.

The first conclusion, (8) of Theorem 7 now gives that

$$\begin{aligned}\mathbf{P}\Big(\bigcap_{n\in\mathcal{N}} (\mathbf{a}_n \bmod n)^c\Big) &\geq \exp\Big(-\sum_{v\in V} \pi_v \prod_{p|v}(1+x_p)\Big)\\ &= \exp\Big(-\sum_{n\in\mathcal{N}} \frac{|\mathbf{a}_n \bmod n| \prod_{p|n}(1+x_p)}{n}\Big),\end{aligned}$$

which is the first conclusion of Theorem 4. To get the second conclusion, use

$$\begin{aligned}\frac{\mathbf{P}(R \cap (b \bmod n))}{\mathbf{P}(R)} &\leq \frac{\mathbf{P}((b \bmod n) \cap \bigcap_{n'\in\mathcal{N},(n,n')=1} (\mathbf{a}_{n'} \bmod n')^c)}{\mathbf{P}(\bigcap_{n'\in\mathcal{N}} (\mathbf{a}_{n'} \bmod n')^c)}\\ &= \frac{1}{n} \frac{\mathbf{P}(\bigcap_{n'\in\mathcal{N},(n,n')=1} (\mathbf{a}_{n'} \bmod n')^c)}{\mathbf{P}(\bigcap_{n'\in\mathcal{N}} (\mathbf{a}_{n'} \bmod n')^c)}\\ &\leq \frac{1}{n} \exp\Big(\sum_{n'\in\mathcal{N}:(n',n)>1} \frac{|\mathbf{a}_{n'} \bmod n'| \prod_{p|n'}(1+x_p)}{n'}\Big).\end{aligned}$$

The last term is bounded by

$$\frac{1}{n} \exp\Big(\sum_{p|n} \sum_{n':p|n'} \frac{|\mathbf{a}_{n'} \bmod n'| \prod_{p'|n'}(1+x_{p'})}{n'}\Big) \leq \frac{1}{n} \exp\Big(\sum_{p|n} x_p\Big). \qquad \square$$

We now give a sufficient criterion to guarantee a good solution to the fixed-point equation governing existence of weights in Theorem 4. Recall that we define

$$G_p(\underline{x}) = \sum_{n\in\mathcal{N}:p|n} \frac{|\mathbf{a}_n \bmod n| \prod_{p'|n}(1+x_{p'})}{n}.$$

A trivial lower bound for a fixed point $\underline{G}(\underline{x}^{\mathrm{fix}}) = \underline{x}^{\mathrm{fix}}$ is

$$\underline{x}^0, \quad x_p^0 = \frac{G_p(\underline{0})}{1 - G_p(\underline{0})},$$

and we wish to say that a fixed point lies near $\underline{x}^0$. The nth derivative $D^n\underline{G}(\underline{0})$ is a multilinear map $\bigotimes^n \ell^2(\mathcal{P}) \to \ell^2(\mathcal{P})$. Give it the usual operator norm,

$$\left\| D^n \underline{G}(\underline{0}) \right\|_{\mathrm{op}} = \sup_{\|v_1\|_{\ell^2} = \cdots = \|v_n\|_{\ell^2} = 1} \left\| D^n \underline{G}(\underline{0})(v_1, \ldots, v_n) \right\|_{\ell^2}.$$

The following theorem guarantees that there exists such a fixed point $\underline{x}^{\mathrm{fix}}$ close to $\underline{x}^0$ when there is good control of the operator norms of the derivatives of $D^n(\underline{G})(\underline{0})$ of $\underline{G}$ at $\underline{0}$. The theorem was motivated by the series of approximations made in "Newton's method."

THEOREM 8
With the notation as above, let $M > 0$ be a parameter. Assume that

$$B_\infty = \left\| \underline{G}(\underline{0}) \right\|_{\ell^\infty} < 1,$$

and set $B_{2,0} = \|\underline{x}^0\|_{\ell^2}$ and

$$B_{\mathrm{op}}(M) = \left\| D\underline{G}(\underline{0}) - \mathrm{diag}\big(D\underline{G}(\underline{0})\big) \right\|_{\mathrm{op}} + \sum_{n=2}^{\infty} \frac{M^{n-1}}{(n-1)!} \left\| D^n \underline{G}(\underline{0}) \right\|_{\mathrm{op}} < \infty.$$

Suppose that $\theta = \frac{B_{\mathrm{op}}}{1 - B_\infty} < 1$ and that $\frac{B_{2,0}}{1-\theta} \le M$. Then there exists $\underline{x}^{\mathrm{fix}} = \underline{x}^0 + \underline{\epsilon}$, $\underline{\epsilon} \ge \underline{0}$ solving the fixed-point equation $\underline{G}(\underline{x}^{\mathrm{fix}}) = \underline{x}^{\mathrm{fix}}$, such that

$$\|\underline{\epsilon}\|_{\ell^2} \le \frac{B_{2,0}\theta}{1-\theta}.$$

Proof
Let $F(\underline{x}) = \underline{G}(\underline{x}) - \underline{x}$, so that we seek to solve $F(\underline{x}^{\mathrm{fix}}) = \underline{0}$. This we can attempt via "Newton's method."

Let

$$\mathcal{D} = \mathrm{diag}\big(D\underline{G}(\underline{0})\big) = \mathrm{diag}\big(G_p(\underline{0})\big), \qquad \mathcal{O}\mathcal{D} = D\underline{G}(\underline{0}) - \mathrm{diag}\big(D\underline{G}(\underline{0})\big).$$

Starting from the initial guess $\underline{x}^0$ as above, set $\underline{x}^{i+1} = \underline{x}^i + (I - \mathcal{D})^{-1} F(\underline{x}^i)$. We may also set $\underline{x}^{-1} = \underline{0}$, which is consistent with this definition. A moment's thought shows that the sequence $\underline{x}^i$ is increasing, so that if it is bounded it converges to the desired fixed point, and $\underline{\epsilon} = \sum_{i=0}^{\infty} (\underline{x}^{i+1} - \underline{x}^i)$. Plainly $\|(I - \mathcal{D})^{-1}\|_{\mathrm{op}} = \frac{1}{1 - B_\infty}$, so that

$$\|\underline{x}^{i+1} - \underline{x}^i\|_{\ell^2} \leq \frac{1}{1 - B_\infty} \left\| F(\underline{x}^i) \right\|_{\ell^2}.$$

Note that F is a polynomial. Thus,

$$F(\underline{x}^i) = \sum_{k=0}^{\infty} \frac{D^k F(\underline{0})(\underline{x}^i, \dots, \underline{x}^i)}{k!}.$$

In this sum, write $DF(\underline{0}) = (\mathscr{D} - I) + \mathscr{O}\mathscr{D}$, and recall that $\underline{x}^i = \underline{x}^{i-1} + (I - \mathscr{D})^{-1} F(\underline{x}^{i-1})$, so that

$$DF(\underline{0})\underline{x}^i = \mathscr{O}\mathscr{D}\underline{x}^i + (\mathscr{D} - I)\underline{x}^{i-1} - F(\underline{x}^{i-1}).$$

On Taylor expanding $F(\underline{x}^{i-1})$, we find

$$\begin{aligned} F(\underline{x}^i) &= \mathscr{O}\mathscr{D}(\underline{x}^i - \underline{x}^{i-1}) \\ &\quad + \sum_{k=2}^{\infty} \frac{1}{k!} \big(D^k F(\underline{0})(\underline{x}^i, \dots, \underline{x}^i) - D^k F(\underline{0})(\underline{x}^{i-1}, \dots, \underline{x}^{i-1}) \big). \end{aligned} \tag{14}$$

Now we impose the constraint $\|\underline{x}^j\|_{\ell^2} \leq M$, which holds for $j = 0$ due to the condition on $B_{2,0}$, and which we will verify for all j by induction. With this assumption, by the usual trick with the triangle inequality in which we change one coordinate at a time,

$$\begin{aligned} &\left\| D^k F(\underline{0})(\underline{x}^i, \dots, \underline{x}^i) - D^k F(\underline{0})(\underline{x}^{i-1}, \dots, \underline{x}^{i-1}) \right\|_{\ell^2} \\ &\qquad \leq k M^{k-1} \left\| D^k F(\underline{0}) \right\|_{\mathrm{op}} \|\underline{x}^i - \underline{x}^{i-1}\|_{\ell^2}, \end{aligned}$$

so that $\|F(\underline{x}^i)\|_{\ell^2} \leq B_{\mathrm{op}} \|\underline{x}^i - \underline{x}^{i-1}\|$ and

$$\|\underline{x}^{i+1} - \underline{x}^i\|_{\ell^2} \leq \frac{B_{\mathrm{op}}}{1 - B_\infty} \|\underline{x}^i - \underline{x}^{i-1}\|_{\ell^2} = \theta \|\underline{x}^i - \underline{x}^{i-1}\|_{\ell^2}.$$

Since

$$\|\underline{x}^0 - \underline{x}^{-1}\|_{\ell^2} = \|\underline{x}^0\|_{\ell^2} = B_{2,0},$$

we have $\|\underline{x}^i\| \leq \frac{B_{2,0}}{1-\theta} \leq M$ for all i, which verifies the condition above. It follows that $\|\underline{\epsilon}\|_{\ell^2} \leq \frac{B_{2,0}\theta}{1-\theta}$. □

4.1. Random Lovász weights

For each $i = 1, 2, \ldots$ on (a subset of good) fibers above $R^*_{i-1} \cap R_i$, we apply Theorem 4 with moduli $\mathcal{N} = \mathcal{N}_{i+1}$ and residues $\mathbf{a}_n = \mathbf{a}_{n,r}$. Thus, we think of the quantities from Theorems 4 and 8 as depending upon the random variable r, for example, $\underline{G}(\underline{0}) = \underline{G}(\underline{0}, r)$, $B_\infty = B_{\infty,r}$, $\underline{x}^0 = \underline{x}^0(r)$. We wish to understand properties of the distribution of $\underline{x}^{\text{fix}}(r)$, but will instead define good fibers in terms of ℓ^p control of $\underline{G}(\underline{0}, r)$, and control of $B_{\text{op}}(M, r)$, the other quantities of interest being controlled in terms of these. We now work to control B_{op}.

We directly verify that the partial derivatives of $D^k \underline{G}$ are given by

$$D_{p_1} \cdots D_{p_k} G_p = \begin{cases} \sum_{\substack{n \in \mathcal{N} \\ p, p_1 \cdots p_k | n}} \frac{|\mathbf{a}_n \bmod n|}{n} \prod_{\substack{p' | n \\ p' \notin \{p_1, \ldots, p_k\}}} (1 + x_{p'}) \\ \quad \text{if } p_1, \ldots, p_k \text{ distinct}, \\ 0 \quad \text{otherwise.} \end{cases}$$

A simple bound for the operator norm of $D^k \underline{G}(\underline{0})$ is

$$\left\| D^k \underline{G}(\underline{0}) \right\|_{\text{op}} \le \left\| D^k \underline{G}(\underline{0}) \right\|_{\ell^2} = \Big(\sum_{\substack{p_1, \ldots, p_k \\ \text{distinct}}} \left\| D_{p_1} \cdots D_{p_k} \underline{G}(\underline{0}) \right\|_{\ell^2}^2 \Big)^{\frac{1}{2}},$$

which, in view of the evaluation of $D^k \underline{G}$, is given by, for $k \ge 2$,

$$\begin{aligned} \left\| D^k \underline{G}(\underline{0}) \right\|_{\text{op}}^2 &\le S_k, \\ S_k := \sum_{p_1, \ldots, p_k \text{ distinct}} \Big(& k \Big(\sum_{\substack{n \in \mathcal{N} \\ p_1 \cdots p_k | n}} \frac{|\mathbf{a}_n \bmod n|}{n} \Big)^2 \\ & + \sum_{p \notin \{p_1, \ldots, p_k\}} \Big(\sum_{\substack{n \in \mathcal{N} \\ p p_1 \cdots p_k | n}} \frac{|\mathbf{a}_n \bmod n|}{n} \Big)^2 \Big). \end{aligned}$$

In $\mathcal{OD} = D\underline{G}(\underline{0}) - \operatorname{diag}(D\underline{G}(\underline{0}))$ the diagonal terms $p = p_1$ are missing, so that we recover the bound

$$\|\mathcal{OD}\|_{\text{op}}^2 \le \sum_{p \ne p_1} \Big(\sum_{\substack{n \in \mathcal{N} \\ p p_1 | n}} \frac{|\mathbf{a}_n \bmod n|}{n} \Big)^2 =: S_1.$$

By Cauchy–Schwarz, for positive weights $W_1, W_2, W_3, \ldots$,

$$B_{\text{op}}(M)^2 \le \Big(\sum_{k=1}^{\infty} \frac{W_k M^{k-1}}{(k-1)!} \Big) \Big(\sum_{k=1}^{\infty} \frac{M^{k-1}}{(k-1)!} \frac{S_k}{W_k} \Big).$$

Let $\min(\mathcal{P}) \geq P + 1$. We choose

$$W_1^2 = \frac{1}{(P \log P)^2}, \quad \forall k \geq 2, \qquad W_k^2 = \frac{k}{(P \log P)^k},$$

from which it follows that

$$\begin{aligned} B_{\mathrm{op}}(M)^2 &\leq \Big(\frac{1}{P \log P} + \sum_{k=2}^{\infty} \frac{k^{\frac{1}{2}} M^{k-1}}{(k-1)!(P \log P)^{\frac{k}{2}}}\Big) \\ &\quad \times \Big((P \log P) S_1 + \sum_{k=2}^{\infty} \frac{M^{k-1}(P \log P)^{\frac{k}{2}}}{k^{\frac{1}{2}}(k-1)!} S_k\Big) \\ &=: \mathcal{C} \times \mathcal{S}. \end{aligned} \tag{15}$$

We record bounds for $\|\underline{G}(\underline{0})\|_{\ell^2}$ and $S_1, S_2, \dots$ averaged over $r \in R_{i-1}^* \cap R_i$.

LEMMA 9

Let $i \geq 1$. For $r \in R_{i-1}^* \cap R_i$, consider $\mathcal{N} = \mathcal{N}_{i+1}$ and $\mathbf{a}_n = \mathbf{a}_{n,r}$ as in the discussion above. We have the bounds

$$\mathbf{E}_{r \in R_{i-1}^* \cap R_i} \big\|\underline{G}(\underline{0}, r)\big\|_{\ell^2}^2 \leq \beta_2^2(i) \prod_{P_i \leq p < P_{i+1}} \Big(1 + \frac{1}{p-1}\Big)^2 \sum_{P_i \leq p < P_{i+1}} \frac{1}{(p-1)^2},$$

$$\mathbf{E}_{r \in R_{i-1}^* \cap R_i} \big\|\underline{G}(\underline{0}, r)\big\|_{\ell^3}^3 \leq \beta_3^3(i) \prod_{P_i \leq p < P_{i+1}} \Big(1 + \frac{1}{p-1}\Big)^3 \sum_{P_i \leq p < P_{i+1}} \frac{1}{(p-1)^3},$$

$$\begin{aligned} \mathbf{E}_{r \in R_{i-1}^* \cap R_i} S_1(r) &\leq \beta_2^2(i) \prod_{P_i \leq p < P_{i+1}} \Big(1 + \frac{1}{p-1}\Big)^2 \\ &\quad \times \Big(\sum_{P_i \leq p < P_{i+1}} \frac{1}{(p-1)^2}\Big)^2, \end{aligned}$$

and, for $k \geq 2$,

$$\begin{aligned} &\mathbf{E}_{r \in R_{i-1}^* \cap R_i} S_k(r) \\ &\quad \leq \beta_2^2(i) \prod_{P_i \leq p < P_{i+1}} \Big(1 + \frac{1}{p-1}\Big)^2 \\ &\quad\quad \times \Big[k\Big(\sum_{P_i \leq p < P_{i+1}} \frac{1}{(p-1)^2}\Big)^k \Big(1 + \frac{1}{k} \sum_{P_i \leq p < P_{i+1}} \frac{1}{(p-1)^2}\Big)\Big]. \end{aligned}$$

Proof

We use the bound, for distinct $P_i \le p_1, \dots, p_k < P_{i+1}$,

$$\sum_{\substack{n \in \mathcal{N}_{i+1} \\ p_1 \cdots p_k | n}} \frac{1}{n} < \frac{1}{(p_1 - 1) \cdots (p_k - 1)} \sum_{n \in \mathcal{N}_{i+1} \cup \{1\}} \frac{1}{n}$$
$$< \frac{1}{(p_1 - 1) \cdots (p_k - 1)} \prod_{P_i \le p < P_{i+1}} \Big(1 + \frac{1}{p-1}\Big).$$

Since

$$\|\underline{G}(\underline{0})\|_{\ell^2}^2 = \sum_{P_i \le p < P_{i+1}} \Big(\sum_{n \in \mathcal{N}_{i+1} : p | n} \frac{|\mathbf{a}_n \bmod n|}{n} \Big)^2, \tag{16}$$

by the convexity lemma, Lemma 5,

$$\mathbf{E}_{r \in R_{i-1}^* \cap R_i} \|\underline{G}(\underline{0}, r)\|_{\ell^2}^2 \le \beta_2^2(i) \sum_{P_i \le p < P_{i+1}} \Big(\sum_{n \in \mathcal{N}_{i+1} : p | n} \frac{1}{n} \Big)^2$$
$$\le \beta_2^2(i) \prod_{P_i \le p < P_{i+1}} \Big(1 + \frac{1}{p-1}\Big)^2 \sum_{P_i \le p < P_{i+1}} \frac{1}{(p-1)^2}.$$

The proof for $\|\underline{G}(\underline{0})\|_{\ell^3}^3$ is similar.

Since

$$S_1 = \sum_{P_i \le p \ne p_1 \le P_{i+1}} \Big(\sum_{\substack{n \in \mathcal{N}_{i+1} \\ p p_1 | n}} \frac{|\mathbf{a}_n \bmod n|}{n} \Big)^2, \tag{17}$$

the convexity lemma implies that

$$\mathbf{E}_{r \in R_{i-1}^* \cap R_i} S_1(r) \le \beta_2^2(i) \sum_{P_i \le p \ne p_1 < P_{i+1}} \Big(\sum_{n \in \mathcal{N}_{i+1}, p p_1 | n} \frac{1}{n} \Big)^2$$
$$\le \beta_2^2(i) \prod_{P_i \le p < P_{i+1}} \Big(1 + \frac{1}{p-1}\Big)^2 \Big(\sum_{P_i \le p < P_{i+1}} \frac{1}{(p-1)^2} \Big)^2.$$

Since, for $k \ge 2$,

$$S_k = \sum_{\substack{P_i \le p_1, \dots, p_k < P_{i+1} \\ \text{distinct}}} \Big(k \Big(\sum_{\substack{n \in \mathcal{N}_{i+1} \\ p_1 \cdots p_k | n}} \frac{|\mathbf{a}_n \bmod n|}{n} \Big)^2$$
$$+ \sum_{p \notin \{p_1, \dots, p_k\}} \Big(\sum_{\substack{n \in \mathcal{N}_{i+1} \\ p p_1 \cdots p_k | n}} \frac{|\mathbf{a}_n \bmod n|}{n} \Big)^2 \Big), \tag{18}$$

the convexity lemma implies that

$$\begin{aligned}
&\mathbf{E}_{r\in R_{i-1}^*\cap R_i} S_k(r)\\
&\quad\le \beta_2^2(i) \sum_{\substack{P_i\le p_1,\dots,p_k<P_{i+1}\\ \text{distinct}}} \Bigl(k\Bigl(\sum_{\substack{n\in\mathcal{N}_{i+1}\\ p_1\cdots p_k|n}} \frac{1}{n}\Bigr)^2 + \sum_{p\notin\{p_1,\dots,p_k\}} \Bigl(\sum_{\substack{n\in\mathcal{N}_{i+1}\\ pp_1\cdots p_k|n}} \frac{1}{n}\Bigr)^2\Bigr)\\
&\quad\le \beta_2^2(i) \prod_{P_i\le p<P_{i+1}} \Bigl(1+\frac{1}{p-1}\Bigr)^2\\
&\qquad\times\Bigl[k\Bigl(\sum_{P_i\le p<P_{i+1}} \frac{1}{(p-1)^2}\Bigr)^k \Bigl(1+\frac{1}{k}\sum_{P_i\le p<P_{i+1}} \frac{1}{(p-1)^2}\Bigr)\Bigr]. \qquad\square
\end{aligned}$$

Inserting (15) in the last lemma, we conclude the following bound.

LEMMA 10
Let B_{op} be the constant from Theorem 8. Averaged over $R_{i-1}^*\cap R_i$, we have the bound

$$\begin{aligned}
&\mathbf{E}_{r\in R_{i-1}^*\cap R_i} B_{\mathrm{op}}(M)^2\\
&\quad\le \mathcal{C}_i\beta_2^2(i) \prod_{P_i\le p<P_{i+1}} \Bigl(1+\frac{1}{p-1}\Bigr)^2\\
&\qquad\times\Bigl(P_i\log P_i\Bigl(\sum_{P_i\le p<P_{i+1}} \frac{1}{(p-1)^2}\Bigr)^2\\
&\qquad+\Bigl(1+\frac{1}{P_i}\Bigr)\sum_{n=2}^{\infty} \frac{n^{\frac12}M^{n-1}(P_i\log P_i)^{\frac{n}{2}}}{(n-1)!}\Bigl(\sum_{P_i\le p<P_{i+1}} \frac{1}{(p-1)^2}\Bigr)^n\Bigr)
\end{aligned}$$

with $\mathcal{C}_i$ given as above by

$$\mathcal{C}_i=\Bigl(\frac{1}{P_i\log P_i}+\sum_{n=2}^{\infty}\frac{n^{\frac12}M^{n-1}}{(n-1)!(P_i\log P_i)^{\frac{n}{2}}}\Bigr).$$

We conclude this section with a brief discussion of how we apply Theorem 4. Beyond demonstrating that fibers above a good set R_i^* are nonempty, the information that we wish to obtain from Theorem 4 is a bound for the bias statistics $\beta_k(i+1)$ in the next stage of iteration. Lemma 6 reduces this problem to bounding the individual biases b_n of $R_i^*\cap R_{i+1}$ at $n\in\mathcal{N}_{i+1}$, and Theorem 4 demonstrates that this bias is bounded by

$$b_n\le\max_{r\in R_i^*}\exp\Bigl(\sum_{p|n}x_p^{\mathrm{fix}}(r)\Bigr). \tag{19}$$

We bound this quantity in terms of the number of prime factors $\omega = \omega(n)$ of n. Thinking of $\underline{\epsilon} = \underline{\epsilon}(r)$ as a small error, we have

$$\sum_{p|n} x_p^{\mathrm{fix}}(r) \le \|\underline{x}^{\mathrm{fix}}\|_{\infty,\omega} \le \|\underline{x}^0\|_{\infty,\omega} + \sqrt{\omega}\|\underline{\epsilon}\|_{\ell^2},$$

where $\|\cdot\|_{\infty,k}$ denotes the norm

$$\|\underline{x}\|_{\infty,\omega} = \max_{i_1<i_2<\cdots<i_\omega} \big(|x_{i_1}| + \cdots + |x_{i_\omega}|\big).$$

LEMMA 11
Let $q = 2$ or 3. Given the optimization problem

$$\begin{aligned} &\textit{given:} && 0 < B_q < 1, \quad 1 \le \omega, \\ &\textit{maximize:} && \|\underline{x}^0\|_{\infty,\omega}, \\ &\textit{subject to:} && \big\|\underline{G}(\underline{0})\big\|_{\ell^q} \le B_q, \end{aligned} \tag{20}$$

when $q = 2$, the optimizing solution $\underline{G}(\underline{0})$ has coordinates that take at most three values, $0 = c_1 < c_2 \le \frac{1}{3} \le c_3 \le B_2$, subject to $c_2(1-c_2)^2 = c_3(1-c_3)^2$. When there are two nonzero values, c_2 is constrained by $c_2(1-c_2)^2 \ge B_2(1-B_2)^2$, which is only possible for a bounded number of nonzero entries. When there is only one nonzero value, the optimum is $\frac{\sqrt{\omega}B_2}{1-\frac{B_2}{\sqrt{\omega}}}$.

When $q = 3$, the optimizing solution satisfies $\underline{G}(\underline{0})$ has at most two nonzero values, and they necessarily satisfy $c_2 = 1 - c_3$.

Proof
Without loss of generality, $\underline{y} = \underline{G}(\underline{0})$ has at most ω nonzero entries, and all of these may be assumed to be nonnegative, since replacing a negative entry y with its absolute value increases $|\frac{y}{1-y}|$, without changing the norm.

Applying Lagrange multipliers in the case $q = 2$ with constraint $\|\underline{y}\|_{\ell^2}^2 \le B_2^2$ and objective $\sum_i \frac{y_i}{1-y_i}$ obtains that those nonzero coordinates y_i satisfy $\frac{1}{(1-y_i)^2} = \lambda y_i$ or $y_i(1-y_i)^2 = c$ with $c = \frac{1}{\lambda}$. Notice that the constraint implies that $y_i \le B_2 < 1$, and on $[0,1]$, $\frac{d}{dy}(y(1-y)^2) = (3y-1)(y-1)$ so that $y(1-y)^2$ is increasing on $[0, \frac{1}{3})$ and decreasing on $(\frac{1}{3}, 1]$. Hence, there are at most two nonzero solutions to $y(1-y)^2 = c$, and if there are 2, then $c \ge B_2(1-B_2)^2$. When there is only one nonzero value, by Lagrange multipliers each of the ω variables appearing in the objective function takes the value $\frac{B_2}{\sqrt{\omega}}$, which is as large as possible, since the objective function is increasing in each variable.

When $q = 3$, applying Lagrange multipliers with constraint $\|\underline{y}\|_{\ell^3}^3 \le B_3^3$, $y_i \ge 0$ and objective function $\sum_i \frac{y_i}{1-y_i}$ implies that those positive coordinates satisfy

$\frac{1}{(1-y_i)^2} = \lambda y_i^2$ or $y_i^2(1-y_i)^2 = c$. Since $0 \le y_i, 1-y_i \le 1$, this has two solutions, c_2 and $1-c_2$. □

5. Explicit calculation in initial stages

In the initial stage, we appeal to the Shearer-type theorem, Theorem 3, with the primes in the range $P_0 = 4 < p < P_1 = 222$, and we verify numerically that the condition of the theorem holds. We also calculate the bound for bias statistics

$$\beta_2(1) \le 12.25, \qquad \beta_3(1) \le 25.$$

The method of performing these explicit computations is described in Appendix A. Empirically, the barrier to ruling out an odd covering using the current method is that the optimal application in the initial stage can only accommodate a few primes, so that the resulting bounds for moments do not permit the process to continue.

Let $P_2 = 4000$. In order to choose the good set $R_1^* \subset R_1 (= R_0^* \cap R_1)$, we appeal to Lemmas 9 and 10 to calculate, for any $C_2, C_{\mathrm{op}} > 0$, and for $M = 1.769746269$,

$$\begin{aligned}
&\mathbf{E}_{r \in R_1}\left(C_2 \left\|\underline{G}(\underline{0}, r)\right\|_{\ell^2}^2 + C_{\mathrm{op}} B_{\mathrm{op}}(1.769746269, r)^2\right) \\
&\le C_2 \beta_2^2(1) \prod_{222 \le p < 4000} \left(1 + \frac{1}{p-1}\right)^2 \sum_{222 \le p < 4000} \frac{1}{(p-1)^2} \\
&\quad + C_{\mathrm{op}} \mathscr{C}_1 \beta_2^2(1) \prod_{222 \le p < 4000} \left(1 + \frac{1}{p-1}\right)^2 \\
&\quad \times \Bigg[(222 \log 222) \sum_{222 \le p < 4000} \frac{1}{(p-1)^2} \\
&\quad + \frac{223}{222} \sum_{n=2}^{\infty} \frac{n^{\frac{1}{2}} 1.769746269^{n-1} (2 \cdot 222^2 \log 222)^{\frac{n}{2}}}{(n-1)!} \Bigg(\sum_{222 \le p < 4000} \frac{1}{(p-1)^2} \Bigg)^n \Bigg]
\end{aligned}$$

and

$$\mathscr{C}_1 = \frac{1}{222(\log 222)} + \sum_{n=2}^{\infty} \frac{n^{\frac{1}{2}} 1.769746269^{n-1}}{(n-1)!(222 \log 222)^{\frac{n}{2}}}.$$

We calculate numerically that

$$\mathbf{E}_{r \in R_1}\left(C_2 \left\|\underline{G}(\underline{0}, r)\right\|_{\ell^2}^2 + C_{\mathrm{op}} B_{\mathrm{op}}(M, r)^2\right) < C_2 \cdot 0.246514091 + C_{\mathrm{op}} \cdot 0.002220166.$$

We choose $C_2 = \frac{0.9}{0.246514091}$ and $C_{\mathrm{op}} = \frac{0.1}{0.002220166}$, so that the above inequality reads

$$\mathbf{E}_{r \in R_1}\left(C_2 \left\|\underline{G}(\underline{0}, r)\right\|_{\ell^2}^2 + C_{\mathrm{op}} B_{\mathrm{op}}(1.769746269, r)^2\right) < 1.$$

We say that $r \in R_1$ is *good* if

$$\left\|C_2\underline{G}(\underline{0},r)\right\|_{\ell^2}^2 + C_{\mathrm{op}}B_{\mathrm{op}}(1.769746269,r)^2 \le \frac{1}{1-0.3}.$$

By Markov's inequality, $\pi_{\mathrm{good}}(1) \ge 0.3$.

Evidently, for all r,

$$\left\|\underline{G}(\underline{0},r)\right\|_{\ell^2}^2 \le \frac{1}{C_2(1-0.3)} < 0.391292208.$$

We save a little extra ground by conditioning on the actual size of $\|\underline{G}(\underline{0},r)\|_{\ell^2}$. Let $K = 100$ be a parameter. For $1 \le j \le K$, we say that $r \in R_1^*$ is in bin $\mathscr{B}_j$ if

$$\left\|\underline{G}(\underline{0},r)\right\|_{\ell^2}^2 \in \left(\frac{j-1}{K}, \frac{j}{K}\right] \cdot \frac{1}{C_2(1-0.3)}.$$

For $r \in \mathscr{B}_j$, we have

$$B_{\mathrm{op}}(1.769746269,r)^2 \le \frac{K-j+1}{K} \cdot \frac{1}{C_{\mathrm{op}}(1-0.3)}.$$

Abusing notation, for $r \in \mathscr{B}_j$ we write quantities depending upon r as depending upon j instead, so we use $\underline{G}(\underline{0},j)$, $B_2(j)$, $B_{\mathrm{op}}(j)$, and so forth.

In each bin we update

$$B_\infty(j) = \left\|\underline{G}(\underline{0},j)\right\|_{\ell^\infty} \le B_2(j)$$

and thus

$$B_{2,0}(j) = \|\underline{x}^0\|_{\ell^2} \le \frac{B_2(j)}{1-B_2(j)},$$

and $\theta(j) \le \frac{B_{\mathrm{op}}(j)}{1-B_2(j)}$. We check numerically, bin-by-bin, that for all bins,

$$\frac{B_{2,0}(j)}{1-\theta(j)} < 1.769746269 = M,$$

so that the condition of Theorem 8 is met. In particular, each good fiber is nonempty.

Again, we apply Theorem 8 bin-by-bin so that, in each bin, we obtain a bound of

$$\left\|\underline{\epsilon}(j)\right\|_{\ell^2} \le \frac{B_{2,0}(j)\theta(j)}{1-\theta(j)}.$$

Beginning from the information $B_2^2(j) \le \frac{j}{K}0.625533539$, we solve the optimization problem (20) for $\omega = 1,2,3,\ldots$. As we have already commented, for each ω the

Table 1.

ω	$\|\underline{x}^{\text{fix}}\|_{\infty,\omega}$
1	1.769746269
2	1.900670975
3	2.033321919
4	2.184489901
5	2.363269323
6	2.530235874
7	2.686345986
8	2.833661687
9	2.973253326
10	3.106051540
$\omega > 10$	$\omega\frac{0.625533539}{\sqrt{\omega}-0.625533539} + 0.292129153\sqrt{\omega}$

optimal solution has no more than two nonzero values among the x_p's. When it has the two values c_1, c_2 these satisfy for some positive integers $a \le b$, $a + b \le \omega$,

$$ac_1^2 + bc_2^2 \le \frac{j}{K}0.391292208, \qquad c_1(1-c_1)^2 = c_2(1-c_2)^2.$$

It transpires that this possibility occurs only for $\omega \le 4$ and when $a = 1$. For $\omega \ge 5$, the optimum in each bin is given by

$$\left\|\underline{x}^0(j)\right\|_{\infty,\omega} \le \omega\frac{B_2(j)}{\sqrt{\omega}-B_2(j)}, \qquad \left\|\underline{x}^{\text{fix}}(j)\right\| \le \omega\frac{B_2(j)}{\sqrt{\omega}-B_2(j)} + \sqrt{\omega}\left\|\underline{\epsilon}(j)\right\|_{\ell^2}.$$

Obviously $B_2(j) \le B_2(K) \le 0.391292208^{\frac{1}{2}} = 0.625533539$, and we find

$$\sup_j \left\|\underline{\epsilon}(j)\right\|_{\ell_2} \le 0.292129153.$$

Resulting bounds for $\sup_j \|\underline{x}^{\text{fix}}\|_{\infty,\omega}$ are recorded in Table 1.

We can thus update the bound for bias statistics $\beta_k(2)$ according to Lemma 6. Write, for $k = 1, 2, 3, \ldots,$

$$\tau_k(p) = \sum_{i=1}^{\infty} \frac{(i+1)^k - i^k}{p^i},$$

for the local factor at p that occurs at the kth bias statistic. Then the new bound becomes

$$\beta_k^k(2) \le \frac{\beta_k^k(1)}{\pi_{\text{good}}(1)}\Big(1 + \sum_{j=1}^{\infty} \exp\big(\|\underline{x}^{\text{fix}}\|_{\infty,j}\big)e_j\big(\tau_k(p) : P_1 \le p < P_2\big)\Big),$$

where e_j indicates the jth elementary symmetric function. For large j, we use the bound $e_j(\underline{\tau}) \le \frac{e_1(\underline{\tau})^j}{j!}$. In this way we calculate that

$$\beta_2(2) < 94.66051416, \qquad \beta_3(2) < 199.2834489.$$

6. Asymptotic estimates

Recall that $P_2 = 4000$. For all $i \geq 2$, we let $P_{i+1} = P_i^{1.5}$. In this section we use the following explicit estimates for sums and products over primes, which hold for $i \geq 2$:

$$\begin{aligned}\prod_{P_i \leq p < P_{i+1}} \Big(\frac{p}{p-1}\Big) &< (1.004212)(1.5)\\ &= 1.506318,\\ \sum_{P_i \leq p < P_{i+1}} \frac{1}{(p-1)^2} &< \frac{1.002631}{P_i \log P_i},\\ \sum_{P_i \leq p < P_{i+1}} \frac{1}{(p-1)^3} &< \frac{1.004382}{2P_i^2 \log P_i}.\end{aligned}$$

These are verified in Appendix B.

For the remainder of the argument our inductive assumption is, for $i \geq 2$,

$$\begin{aligned}\beta_2(i) &\leq 0.5197033883 \cdot (P_i \log P_i)^{\frac{1}{2}},\\ \beta_3(i) &\leq 0.3100980448 \cdot (2P_i^2 \log P_i)^{\frac{1}{3}}.\end{aligned} \tag{21}$$

Note that both of these hold at $i = 2$.

Setting $M = 2.949873427$ in Theorem 8, we estimate

$$\mathbf{E}_{r \in R_{i-1}^* \cap R_i} \big(C_3 \big\| \underline{G}(\underline{0}, r)\big\|_{\ell^3}^3 + C_2 \big\| \underline{G}(\underline{0}, r)\big\|_{\ell^2}^2 + C_{\mathrm{op}} B_{\mathrm{op}}(2.949873427, r)^2\big).$$

Appealing to Lemma 10, we bound the sums in $\mathscr{C}$ and $\mathbf{E}\mathscr{S}$, implicitly defined in (15), by

$$\begin{aligned}\mathscr{C}_i \leq \mathscr{C}_2 &= \Big(\frac{1}{4000 \log 4000} + \sum_{n=2}^{\infty} \frac{n^{\frac{1}{2}} M^{n-1}}{(n-1)!(4000 \log 4000)^{\frac{n}{2}}}\Big)\\ &\leq 0.0001571422884,\\ \mathbf{E}\mathscr{S}_i &\leq (1.506318)^2 \beta_2^2(i)\\ &\quad \times \Big(\frac{(1.002631)^2}{P_i \log P_i} + \Big(1 + \frac{1}{P_i}\Big) \sum_{n=2}^{\infty} \frac{n^{\frac{1}{2}} M^{n-1} (1.002631)^n}{(n-1)!(P_i \log P_i)^{\frac{n}{2}}}\Big)\\ &\leq (1.506318)^2 (0.5197033883)^2\\ &\quad \times \Big((1.002631)^2 + \frac{4001}{4000} \sum_{n=2}^{\infty} \frac{n^{\frac{1}{2}} M^{n-1} (1.002631)^n}{(n-1)!(4000 \log 4000)^{\frac{n}{2}-1}}\Big)\\ &\leq 3.212501212.\end{aligned}$$

Combined with the asymptotics of $\mathbf{E}\|\underline{G}(\underline{0})\|_{\ell^p}^p$ from Lemma 9,

$$\begin{aligned}\mathbf{E}_{r\in R_{i-1}^*\cap R_i}\big\|\underline{G}(\underline{0})\big\|_{\ell^3}^3 &\le \prod_{P_i\le p<P_{i+1}}\Big(\frac{p}{p-1}\Big)^3\beta_3^3(i)\sum_{P_i\le p<P_{i+1}}\frac{1}{(p-1)^3}\\ &< (1.506318)^3(1.004382)(0.3100980448)^3\\ &< 0.1023637064,\\ \mathbf{E}_{r\in R_{i-1}^*\cap R_i}\big\|\underline{G}(\underline{0})\big\|_{\ell^2}^2 &\le \prod_{P_i\le p<P_{i+1}}\Big(\frac{p}{p-1}\Big)^2\beta_2^2(i)\sum_{P_i\le p<P_{i+1}}\frac{1}{(p-1)^2}\\ &< (1.506318)^2(1.002631)(0.5197033883)^2\\ &< 0.6144485964,\end{aligned}$$

we deduce

$$\begin{aligned}&\mathbf{E}_{r\in R_{i-1}^*\cap R_i}\big(C_3\big\|\underline{G}(\underline{0})\big\|_{\ell^3}^3 + C_2\big\|\underline{G}(\underline{0})\big\|_{\ell^2}^2 + C_{\mathrm{op}}B_{\mathrm{op}}(2.949873427, r)^2\big)\\ &\qquad\le 0.1023637064C_3 + 0.6144485964C_2 + 0.0005048197920C_{\mathrm{op}}.\end{aligned}$$

Choose $C_3 = \frac{0.7}{0.1023637064}$, $C_2 = \frac{0.2}{0.6144485964}$, $C_{\mathrm{op}} = \frac{0.1}{0.0005048197920}$ so that the expectation is bounded by 1. As before, declare $r \in R_{i-1}^* \cap R_i$ to be good if

$$C_3\big\|\underline{G}(\underline{0}, r)\big\|_{\ell^3}^3 + C_2\big\|\underline{G}(\underline{0}, r)\big\|_{\ell^2}^2 + C_{\mathrm{op}}B_{\mathrm{op}}(2.949873427, r)^2 \le \frac{1}{1-0.3}.$$

Evidently $\pi_{\mathrm{good}}(i) \ge 0.3$, and for good r,

$$\begin{aligned}\big\|\underline{G}(\underline{0}, r)\big\|_{\ell^\infty}^3 \le \big\|\underline{G}(\underline{0}, r)\big\|_{\ell^3}^3 &\le \frac{1}{C_3(1-0.3)} < 0.2089055233,\\ \big\|\underline{G}(\underline{0}, r)\big\|_{\ell^2}^2 &\le \frac{1}{C_2(1-0.3)} < 4.388918546,\end{aligned}$$

but, again, we bin to get a stronger result.

For $K = 100$ and integers $0 < i, j, i + j \le K + 1$, let the bin $\mathscr{B}_{i,j}$ be those r for which

$$\begin{aligned}\big\|\underline{G}(\underline{0}, r)\big\|_{\ell^3}^3 &\in \Big(\frac{i-1}{K}, \frac{i}{K}\Big]\frac{1}{C_3(1-0.3)},\\ \big\|\underline{G}(\underline{0}, r)\big\|_{\ell^2}^2 &\in \Big(\frac{j-1}{K}, \frac{j}{K}\Big]\frac{1}{C_2(1-0.3)}.\end{aligned}$$

For $r \in \mathscr{B}_{i,j}$, we have

$$B_{\mathrm{op}}^2(r) \le \frac{K-i-j+1}{K}\frac{1}{C_{\mathrm{op}}(1-0.3)}.$$

We proceed much as before (again replacing r with i, j in each argument) updating bin-by-bin

$$B_{2,0}(i,j) \le \frac{\|\underline{G}(\underline{0},i,j)\|_{\ell^2}}{1-\|\underline{G}(\underline{0},i,j)\|_{\ell^3}},$$

and

$$\theta(i,j) \le \frac{B_{\mathrm{op}}(2.949873427,i,j)}{1-\|\underline{G}(\underline{0},i,j)\|_{\ell^3}}.$$

We check bin-by-bin that

$$\frac{B_{2,0}(i,j)}{1-\theta(i,j)} < 2.949873427 = M$$

so that our choice of $M = 2.949873427$ in Theorem 8 is valid.

In each bin we solve the optimization problem (20) with $p = 3$, and we find that for all $\omega \ge 1$ and for all bins the optimum is

$$\left\|\underline{x}^0(i,j)\right\|_{\infty,\omega} \le \omega^{\frac{2}{3}} \frac{\frac{i}{K}0.5933577790}{1-\frac{\frac{i}{K}0.5933577790}{\omega^{\frac{1}{3}}}},$$

so that we guarantee

$$\left\|\underline{x}^{\mathrm{fix}}(i,j)\right\|_{\infty,\omega} \le \frac{\frac{i}{K}0.5933577790\omega^{\frac{2}{3}}}{1-\frac{\frac{i}{K}0.5933577790}{\omega^{\frac{1}{3}}}} + \left\|\underline{\epsilon}(i,j)\right\|_{\ell^2}\omega^{\frac{1}{2}}.$$

We calculate

$$\max_{i,j}\left\|\underline{\epsilon}(i,j)\right\|_{\ell^2} \le 0.190000303.$$

Thus, we find the bounds in Table 2.

In Appendix B we verify that, for $i \ge 2$,

$$e_1\big(\tau_2(p)\big) \le 3\log 1.5 + 0.00334 < 1.21974,$$
$$e_1\big(\tau_3(p)\big) \le 7\log 1.5 + 0.00779 < 2.84605.$$

Hence,

$$e_j\big(\tau_2(p) : P_i \le p < P_{i+1}\big) \le \frac{e_1(\tau_2(p))^j}{j!} < \frac{(1.21974)^j}{j!},$$
$$e_j\big(\tau_3(p)\big) < \frac{(2.84605)^j}{j!}$$

Table 2.

ω	$\Vert\underline{x}^{\text{fix}}\Vert_{\infty,\omega}$
1	1.459164221
2	1.780349459
3	2.096937862
4	2.387653719
5	2.656941273
6	2.909180305
7	3.147611526
8	3.374605257
9	3.591932780
10	3.800951606
$\omega > 10$	$\frac{0.5933577790\omega}{\omega^{\frac{1}{3}}-0.5933577790} + 0.190000303\sqrt{\omega}$

and we find

$$\frac{\beta_2^2(i+1)}{\beta_2^2(i)} \leq \frac{1}{0.3}\Big(1 + \sum_{\omega=1}^{\infty} \exp\big(\max_{i,j}\Vert\underline{x}^{\text{fix}}(i,j)\Vert_{\infty,\omega}\big)\frac{(1.21974)^{\omega}}{\omega!}\Big) < 48.515$$

and

$$\frac{\beta_3^3(i+1)}{\beta_3^3(i)} \leq \frac{1}{0.3}\Big(1 + \sum_{\omega=1}^{\infty} \exp\big(\max_{i,j}\Vert\underline{x}^{\text{fix}}(i,j)\Vert_{\infty,\omega}\big)\frac{(2.84605)^{\omega}}{\omega!}\Big) < 487.17.$$

On the other hand,

$$\frac{P_{i+1}\log P_{i+1}}{P_i \log P_i} \geq 1.5 \cdot P_i^{\frac{1}{2}} \geq 1.5 \cdot 4000^{\frac{1}{2}} > 94$$

and

$$\frac{P_{i+1}^2 \log P_{i+1}}{P_i^2 \log P_i} \geq 1.5 \cdot P_i \geq 6000,$$

so that (21) is preserved, which completes the proof by induction.

Appendices

A. Symmetric functions

We briefly describe how we performed the calculations in the initial stage of the argument (see Section 5). There we appealed to Theorem 3, which is Theorem 2 with set $[n]$ identified with $\{p : 4 < p < 222\} = p_1 < p_2 < \cdots < p_n$ and weights $\pi_p = \frac{1}{p-1}$. We identify square-free number m with the set of its prime factors. The Shearer functions

$$\rho(p_1) > \rho(p_1 p_2) > \cdots > \rho(p_1 \cdots p_n)$$

are easily computed via

$$\rho(p_1 \cdots p_j) = \sum_{i=0}^{j} X(i) e_i(\pi_{p_1}, \dots, \pi_{p_j}),$$

with the e_i elementary symmetric functions (take $e_0 = 1$) (see [16]).

The bias statistics are also not difficult to bound. Recall that $Q = \mathrm{LCM}(m : m \in \mathcal{M})$ and that Q_1 is the part of Q composed of primes less than $P_1 = 222$. The kth bias statistic is

$$\beta_k^k(1) = \sum_{m|Q_1} \ell_k(m) \max_{b \bmod m} \frac{|R_1 \cap (b \bmod m) \bmod Q_1|}{|R_1 \bmod Q_1|}.$$

Let $\mathrm{sqf}(m) = \prod_{p \in S} p =: m_S$. Appealing to (4) of Theorem 3 we have

$$\frac{|R_1 \cap (b \bmod m) \bmod Q_1|}{|R_1 \bmod Q_1|} \leq \frac{1}{m} \frac{\rho([n] \setminus S)}{\rho([n])},$$

so that

$$\begin{aligned}\beta_k^k(1) &\leq \frac{1}{\rho([n])} \sum_{S \subset [n]} \rho\big([n] \setminus S\big) \sum_{m:\mathrm{sqf}(m) = m_S} \frac{\ell_k(m)}{m} \\ &\leq \frac{1}{\rho([n])} \sum_{S \subset [n]} \rho\big([n] \setminus S\big) \prod_{s \in S} \Big(\sum_{j=1}^{\infty} \frac{\ell_k(p_s^j)}{p_s^j} \Big). \qquad \text{(A.1)}\end{aligned}$$

Recall that we define

$$\begin{aligned}\tau_{k,i} = \tau_k(p_i) &= \sum_{j=1}^{\infty} \frac{\ell_k(p_i^j)}{p_i^j} = \sum_{j=1}^{\infty} \frac{(j+1)^k - j^k}{p_i^k} \\ &= \Big[\Big(\frac{1}{x} - 1\Big) \Big(x \frac{\partial}{\partial x}\Big)^k \frac{1}{1-x} - 1 \Big]_{x = \frac{1}{p_i}}.\end{aligned}$$

Define for $i + j \leq n$ the mixed symmetric functions $f_{i,j}(\underline{\pi}, \underline{\tau}_k)$ by

$$f_{i,j}(\underline{\pi}, \underline{\tau}_k) = \frac{i!j!(n-i-j)!}{n!} \sum_{\sigma \in \mathrm{Sym}([n])} \underline{\pi}_{\sigma(1)} \cdots \underline{\pi}_{\sigma(i)} \underline{\tau}_{k,\sigma(i+1)} \cdots \underline{\tau}_{k,\sigma(i+j)}$$

or, equivalently, by

$$\sum_{0\le i+j\le n} f_{i,j}(\underline{\pi}, \underline{\tau}_k)x^i y^j = \prod_{i=1}^{n}(1 + x\pi_i + y\tau_{k,i}).$$

The sum of (A.1) is a linear combination of the mixed symmetric functions $f_{i,j}(\underline{\pi}, \underline{\tau}_k)$

$$\sum_{S\subset[n]} \rho\big([n]\setminus S\big)\prod_{s\in S}\tau_{k,s} = \sum_{0\le i+j\le n} X(i) f_{i,j}(\underline{\pi}, \underline{\tau}_k),$$

and so is rapidly computable.

B. Explicit prime number estimates

In this appendix we sketch proofs for explicit bounds on well-known prime sums and products. Recall that $P_2 = 4000$ and, for $i \ge 2$, $P_{i+1} = P_i^{1.5}$. In particular, no P_i is prime. Let γ denote the Euler–Mascheroni constant. Dusart [2, Theorem 5.9] proves the following estimate.

THEOREM B.1
For $x > 2278382$, we have

$$e^{\gamma}(\log x)\Big(1 - \frac{0.2}{(\log x)^3}\Big) < \prod_{p\le x}\frac{p}{p-1}$$

and

$$\prod_{p\le x}\frac{p}{p-1} < e^{\gamma}(\log x)\Big(1 + \frac{0.2}{(\log x)^3}\Big).$$

As a consequence, we obtain the following.

COROLLARY B.2
For $i \ge 2$, we have

$$\prod_{P_i\le p<P_{i+1}}\Big(\frac{p}{p-1}\Big) < (1.5)(1.004212).$$

For the sums of reciprocals of squares of primes, we have the following estimate.

PROPOSITION B.3
For $i \ge 2$,

$$\sum_{P_i\le p<P_{i+1}}\frac{1}{(p-1)^2} < \frac{1.002631}{P_i\log P_i}$$

and

$$\sum_{P_i \le p < P_{i+1}} \frac{1}{(p-1)^3} < \frac{1.004382}{2P_i^2 \log P_i}.$$

Proof
We prove only the first inequality, as the second is similar. One easily checks that

$$\sum_{P_2 \le p < P_3} \frac{1}{(p-1)^2} < \frac{1}{P_2 \log P_2}, \qquad \sum_{P_3 \le p < P_4} \frac{1}{(p-1)^2} < \frac{1}{P_3 \log P_3}.$$

For $i \ge 4$, use $\frac{0.99999997}{(p-1)^2} < \frac{1}{p^2}$ so that, for $x \ge P_4$,

$$0.99999997 \sum_{p \ge x} \frac{1}{(p-1)^2} < \sum_{p \ge x} \frac{1}{p^2} = -\frac{\theta(x)}{x^2 \log x} + \int_x^\infty \frac{\theta(y)}{y^3} \frac{1 + 2\log y}{(\log y)^2}\, dy.$$

By [2], we have the inequality $|\theta(x) - x| < 0.2\frac{x}{(\log x)^2}$ for $x \ge 3594641$. In particular, $0.99913x < \theta(x) < 1.00088x$ in this range. Also, Lemma 9 of [12] yields

$$\int_x^\infty \frac{1 + 2\log y}{y^2(\log y)^2}\, dy < \frac{2}{x \log x}.$$

Combined, these estimates give the claim. □

Recall that we define $\tau_k(x) = \sum_{i=1}^\infty \frac{(i+1)^k - i^k}{x^i}$. We have

$$\tau_2(x) = \frac{3x-1}{(x-1)^2}, \qquad \tau_3(x) = \frac{7x^2 - 2x + 1}{(x-1)^3}.$$

PROPOSITION B.4
For $i \ge 2$,

$$\sum_{P_i \le p < P_{i+1}} \tau_2(p) < 3 \log 1.5 + 0.00334$$

and

$$\sum_{P_i \le p < P_{i+1}} \tau_3(p) < 7 \log 1.5 + 0.00779.$$

Proof
For $i = 2, 3$, this is verified directly. For $x \ge P_4$, this is a consequence of the following estimate of [2, Theorem 5.6]. □

THEOREM B.5
There is a constant B, such that, for any $x \geq 2278383$,

$$\Big|\sum_{p\leq x}\frac{1}{p} - \log\log x - B\Big| \leq \frac{0.2}{(\log x)^3}.$$

C. Theorems of Lovász and Shearer-type

The Lovász local lemma considers the following scenario. In a probability space $\mathcal{X}$ there are events $\{A_v\}_{v\in V}$ with dependency graph $G = (V, E)$, that is, A_v is independent of the σ-algebra $\sigma(A_w : (v, w) \notin E)$. One seeks a positive lower bound for $\mathbf{P}(\bigcap_{v\in V}\overline{A_v})$. The local lemma guarantees that if there exist weights $1 > x_v \geq \mathbf{P}(A_v)$ satisfying

$$\forall v \in V, \quad x_v \prod_{w:(v,w)\in E}(1 - x_w) \geq \mathbf{P}(A_v),$$

then

$$\mathbf{P}\Big(\bigcap_{v\in V}\overline{A}_v\Big) \geq \prod_{v\in V}(1 - x_v).$$

Shearer [14] gives an optimal bound of the above type via the independent set polynomial

$$\Xi(z_v : v \in V) = 1 + \sum_{n=1}^{\infty}\frac{1}{n!}\sum_{\substack{(v_1,\dots,v_n)\in V^n \\ \forall i\neq j,\, v_i\sim v_j}} z_{v_1}\cdots z_{v_n},$$

where $(v_i \sim v_j)$ means $v_i \neq v_j$ and $(v_i, v_j) \notin E$.

THEOREM (Shearer's theorem; [14, Theorem 1])
Given $S \subset V$, let $\Xi_S(z_v : v \in V)$ denote Ξ with arguments $z_v : v \in S$ replaced by 0. Subject to

$$\forall S \subset V, \quad \Xi_S\big(-\mathbf{P}(A_v) : v \in V\big) \geq 0,$$

it holds that

$$\mathbf{P}\Big(\bigcap_{v\in V}\overline{A_v}\Big) \geq \Xi\big(-\mathbf{P}(A_v) : v \in V\big).$$

Although Shearer's theorem is tight, evaluating the independent set polynomial is difficult and so there remains interest in finding statements of a similar type to the local lemma, which is more easily applied.

One way to reduce the complexity of Shearer's theorem is to organize the events into collection of cliques. We consider the scenario in which graph $G = (V, E)$ is covered by a collection of cliques $\mathcal{K}$, that is, $E = \bigcup_{K\in\mathcal{K}} E_K$. For $v \in V$, let

$$p(v) = \{K : v \in K\}. \tag{C.1}$$

We make the assumption that $p(v)$ uniquely determines v and we assume that all vertices have self-loops. Moreover, we take $V = \mathcal{P}(\mathcal{K}) \setminus \{\emptyset\}$ to be the collection of all nonempty subsets of $\mathcal{K}$, and, for $S_1, S_2 \in V$, set $(S_1, S_2) \in E$ if and only if $S_1 \cap S_2 \neq \emptyset$. Consider vertex variables $(z_v)_{v\in V}$ and clique variables $(\theta_K)_{K\in\mathcal{K}}$. For $v \in V$, set also $\theta_v = \prod_{K:v\in K} \theta_K$. The clique partition function is defined to be

$$\Xi(\mathbf{v}, \boldsymbol{\theta}) = 1 + \sum_{n\geq 1} \frac{1}{n!} \sum_{\substack{v_1,\dots,v_n \\ \text{indep. in } G}} z_{v_1} \cdots z_{v_n} \theta_{v_1} \cdots \theta_{v_n}.$$

Evidently, $\Xi(\mathbf{v}, \boldsymbol{\theta})$ specializes to $\Xi(\mathbf{v})$ at $\boldsymbol{\theta} = 1$. Using this, we prove a clique version of Shearer's theorem.

THEOREM C.1 (Clique Shearer theorem)
Let events $\{A_v : v \in V\}$ in probability space $\mathcal{X}$ have dependency graph $G = (V, E)$ covered by cliques $\mathcal{K}$ as above. For $S \subset \mathcal{K}$, define event $B_S = \bigcup_{v:p(v)\subset S} A_v$. Subject to the condition

$$\forall S \subset \mathcal{K}, \quad \Xi\big(-\mathbf{P}(A_v), 1_S\big) > 0,$$

we have for all $\emptyset \subset S \subset T \subset \mathcal{K}$,

$$\mathbf{P}(\overline{B_T} | \overline{B_S}) \geq \frac{\Xi(-\mathbf{P}(A_v), 1_T)}{\Xi(-\mathbf{P}(A_v), 1_S)}.$$

Remark
As compared to Shearer's theorem, the clique Shearer theorem has the advantage that the number of conditions which must be checked is exponential in the number of cliques, rather than in the number of vertices.

Proof
The proof is by induction. Let $S \subset \mathcal{K}$, and suppose that the conclusion holds for subsets $T \subset S$. Let $K \in \mathcal{K} \setminus S$. Then

$$\begin{aligned}
\frac{\mathbf{P}(\overline{B_{S\cup\{K\}}})}{\mathbf{P}(\overline{B_S})} &\geq 1 - \sum_{\substack{w:K\in p(w) \\ p(w)\subset S\cup\{K\}}} \frac{\mathbf{P}(A_w \cap \overline{B_S})}{\mathbf{P}(\overline{B_S})} \\
&\geq 1 - \sum_{\substack{w:K\in p(w) \\ p(w)\subset S\cup\{K\}}} \frac{\mathbf{P}(A_w)\mathbf{P}(\overline{B_{S\setminus p(w)}})}{\mathbf{P}(\overline{B_S})}
\end{aligned}$$

$$\geq 1 - \sum_{\substack{w:K\in p(w)\\ p(w)\subset S\cup\{K\}}} \mathbf{P}(A_w)\frac{\Xi(-\mathbf{P}(A_v),1_{S\setminus p(w)})}{\Xi(-\mathbf{P}(A_v),1_S)}$$

$$= \frac{\Xi(-\mathbf{P}(A_v),1_{S\cup\{K\}})}{\Xi(-\mathbf{P}(A_v),1_S)}. \qquad \square$$

As a consequence, we obtain a proof of a generalization of Theorem 3.

THEOREM (Shearer-type theorem)
Suppose that we have a probability space and a parameter $\theta \geq 1$. Let $[n] = \{1, 2, \ldots, n\}$, and assume that for each $1 \leq i \leq n$ there is a weight π_i assigned, satisfying $\frac{1}{\theta} \geq \pi_1 \geq \pi_2 \geq \cdots \geq \pi_n \geq 0$. Let the sets $\emptyset \neq T \subset [n]$ index events A_T each having probability

$$0 \leq \mathbf{P}(A_T) \leq \theta \prod_{t\in T} \pi_t := \pi_T.$$

Assume that A_T is independent of $\sigma(\{A_S : S \subset [n], S \cap T = \emptyset\})$, so that a valid dependency graph for the events $\{A_T : \emptyset \neq T \subset [n]\}$ has an edge between $S \neq T$ whenever $S \cap T \neq \emptyset$.

Define $\rho_\theta(\emptyset) = 1$, and given $\emptyset \neq T \subset [n]$, set

$$\rho_\theta(T) = 1 - \sum_{\emptyset\neq S_1\subset T} \pi_{S_1} + \sum_{\substack{\emptyset\neq S_1,S_2\subset T\\ S_1<S_2 \text{ disjoint}}} \pi_{S_1}\pi_{S_2}$$

$$- \sum_{\substack{\emptyset\neq S_1,S_2,S_3\subset T\\ S_1<S_2<S_3 \text{ disjoint}}} \pi_{S_1}\pi_{S_2}\pi_{S_3} + \cdots.$$

Suppose that $\rho_\theta([1]) \geq \rho_\theta([2]) \geq \cdots \geq \rho_\theta([n]) > 0$. Then for any $\emptyset \neq T \subset [n]$,

$$\mathbf{P}\Big(\bigcap_{\emptyset\neq S\subset T} A_S^c\Big) \geq \rho_\theta(T) > 0$$

and, for any $T_1 \subset T_2 \subset [n]$,

$$\frac{\mathbf{P}(\bigcap_{\emptyset\neq S\subset T_2} A_S^c)}{\mathbf{P}(\bigcap_{\emptyset\neq S\subset T_1} A_S^c)} \geq \frac{\rho_\theta(T_2)}{\rho_\theta(T_1)}.$$

Proof
It is observed in [16] that $\rho_\theta(T)$ may be expressed as a linear combination of elementary symmetric functions in $\{\pi_t : t \in T\}$. Indeed, if $B(m, j)$ denotes the generalized

Bell number, that is, the number of ways of partitioning a set of size m into j parts, then setting $|T| = M$ and making the convention $e_0(\underline{\pi}) = 1$,

$$\rho_\theta(T) = 1 + \sum_{m=1}^{M} \Big(\sum_{j=1}^{m} (-\theta)^j B(m, j) \Big) e_m(\underline{\pi}) := \sum_{i=0}^{M} X_\theta(i) e_i(\underline{\pi}),$$

where X_θ satisfies the recurrence

$$X_\theta(0) = 1, \quad \forall i \geq 1, \qquad X_\theta(i) = -\theta \sum_{j=0}^{i-1} \binom{i-1}{j} X_\theta(j).$$

In particular, as exploited in [13], $\rho(T)$ is affine linear in each variable π_t. We check that under the given conditions, $\rho_\theta(T) > 0$ for any $T \subset [n]$, which reduces this theorem to the clique Shearer theorem.

Given vectors $\underline{x}, \underline{y} \in \mathbb{R}^m$, say that $\underline{x} \leq \underline{y}$ if $x_i \leq y_i$ for each i. By induction, we show that for any $1 \leq m \leq n$ and for $\underline{0} \leq \underline{x} \leq \underline{\pi}$, $\rho_\theta(\underline{x}) \geq \rho_\theta(\underline{\pi}) > 0$, from which the case for T follows since the π_i's are decreasing.

When $m = 1$, $\rho_\theta(\pi_1) = 1 - \theta\pi_1 \leq 1 - \theta x_1 = \rho_\theta(x_1)$. Given $m > 1$, assume inductively the statement for all $m' < m$.

Note that, by hypothesis, we have $\rho_\theta(\pi_1, \ldots, \pi_{m-1}) \geq \rho_\theta(\pi_1, \ldots, \pi_m) > 0$. We show by an inner induction that for $1 \leq j \leq m$,

$$\rho_\theta(x_1, \ldots, x_j, \pi_{j+1}, \ldots, \pi_m) \geq \rho_\theta(\pi_1, \ldots, \pi_m).$$

When $j = 1$ this holds, since $(\pi_2, \ldots, \pi_m) \leq (\pi_1, \ldots, \pi_{m-1})$ so that, by the inductive assumption

$$\rho_\theta(\pi_2, \ldots, \pi_m) \geq \rho_\theta(\pi_1, \ldots, \pi_{m-1}) \geq \rho_\theta(\pi_1, \ldots, \pi_m),$$

from which

$$\rho_\theta(x_1, \pi_2, \ldots, \pi_m) \geq \rho_\theta(\pi_1, \ldots, \pi_m)$$

follows by affine linearity.

Having shown

$$\rho_\theta(x_1, \ldots, x_{j-1}, \pi_j, \ldots, \pi_m) \geq \rho_\theta(\pi_1, \ldots, \pi_m),$$

the case

$$\rho_\theta(x_1, \ldots, x_j, \pi_{j+1}, \ldots, \pi_m) \geq \rho_\theta(\pi_1, \ldots, \pi_m)$$

again follows by affine linearity from

$$\begin{aligned}\rho_\theta(x_1,\ldots,x_{j-1},0,\pi_{j+1},\ldots,\pi_m) &\geq \rho_\theta(\pi_1,\ldots,\pi_{j-1},0,\pi_{j+1},\ldots,\pi_m)\\ &\geq \rho_\theta(\pi_1,\ldots,\pi_{m-1},0)\\ &\geq \rho_\theta(\pi_1,\ldots,\pi_m). \qquad \square\end{aligned}$$

Acknowledgments. We thank the anonymous referees for comments which improved the paper.

The authors were partially sponsored by National Security Agency grant H98230-16-1-0048 and by National Science Foundation grants DMS-1712682 and DMS-1802336.

References

[1] R. BISSACOT, R. FERNÁNDEZ, A. PROCACCI, and B. SCOPPOLA, *An improvement of the Lovász local lemma via cluster expansion*, Combin. Probab. Comput. **20** (2011), no. 5, 709–719. MR 2825585. DOI 10.1017/S0963548311000253. *(5, 9, 10)*

[2] P. DUSART, *Explicit estimates of some functions over primes*, Ramanujan J. **45** (2018), no. 1, 227–251. MR 3745073. DOI 10.1007/s11139-016-9839-4. *(28, 29)*

[3] P. ERDŐS, *On integers of the form $2^k + p$ and some related problems*, Summa Brasil. Math **2** (1950), 113–123. MR 0044558. *(1)*

[4] M. FILASETA, K. FORD, and S. KONYAGIN, *On an irreducibility theorem of A. Schinzel associated with coverings of the integers*, Illinois J. Math. **44** (2000), no. 3, 633–643. MR 1772434. *(1)*

[5] M. FILASETA, K. FORD, S. KONYAGIN, C. POMERANCE, and G. YU, *Sieving by large integers and covering systems of congruences*, J. Amer. Math. Soc. **20** (2007), no. 2, 495–517. MR 2276778. DOI 10.1090/S0894-0347-06-00549-2. *(1, 2)*

[6] S. GUO and Z. SUN, *On odd covering systems with distinct moduli*, Adv. Appl. Math. **35** (2005), no. 2, 182–187. MR 2152886. DOI 10.1016/j.aam.2005.01.004. *(2)*

[7] R. K. GUY, *Unsolved Problems in Number Theory*, Unsolved Problems in Intuitive Mathematics **1**, Springer, New York, 1981. MR 0656313. *(2)*

[8] B. HOUGH, *Solution of the minimum modulus problem for covering systems*, Ann. of Math. (2) **181** (2015), no. 1, 361–382. MR 3272928. DOI 10.4007/annals.2015.181.1.6. *(1, 6, 8, 9)*

[9] P. P. NIELSEN, *A covering system whose smallest modulus is 40*, J. Number Theory **129** (2009), no. 3, 640–666. MR 2488595. DOI 10.1016/j.jnt.2008.09.016. *(1)*

[10] T. OWENS, *A covering system with minimum modulus 42*, Ph.D. dissertation, Brigham Young University, Provo, UT, 2014. *(1)*

[11] N. P. ROMANOFF, *Über einige Sätze der additiven Zahlentheorie*, Math. Ann. **109** (1934), no. 1, 668–678. MR 1512916. DOI 10.1007/BF01449161. *(1)*

[12] J. B. ROSSER and L. SCHOENFELD, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. **6** (1962), no. 1, 64–94. MR 0137689. *(29)*

[13] A. D. SCOTT and A. D. SOKAL, *The repulsive lattice gas, the independent-set polynomial, and the Lovász local lemma*, J. Stat. Phys. **118** (2005), nos. 5–6, 1151–1261. MR 2130890. DOI 10.1007/s10955-004-2055-4. *(5, 9, 33)*

[14] J. B. SHEARER *On a problem of Spencer*, Combinatorica **5** (1985), no. 3, 241–245. MR 0837067. DOI 10.1007/BF02579368. *(3, 30)*

[15] W. SIERPINSKI, *Sur un problème concernant les nombres* $k \cdot 2^n + 1$, Elem. Math. **15** (1960), 73–74. MR 0117201. *(1)*

[16] R. J. SIMPSON and D. ZEILBERGER, *Necessary conditions for distinct covering systems with square-free moduli*, Acta Arith. **59** (1991), no. 1, 59–70. MR 1133237. DOI 10.4064/aa-59-1-59-70. *(2, 3, 27, 32)*

Hough
Department of Mathematics, Stony Brook University, Stony Brook, New York, USA;
robert.hough@stonybrook.edu

Nielsen
Department of Mathematics, Brigham Young University, Provo, Utah, USA;
pace@math.byu.edu