under certain "non-triviality" assumptions on $\mathbf{v}$ (basically, that the set of signs $(\eta_1, \ldots, \eta_n) \in \{-1, 1\}^n$ for which $\eta_1 v_1 + \cdots + \eta_n v_n = 0$ has to span the hyperplane) one can also place the v_i in an *arithmetic* progression of length $n^{o(n)}$. For more precise statements and proofs see [365]. The main point is to inspect the use of the Cauchy–Davenport inequality in the proof of Lemma 7.14, and observe that this inequality is only efficient when sets such as $\{\xi \in \mathbf{Z}_p : F(\xi) > \alpha\}$ have small doubling constant. This in turn can be used (via some duality arguments) to place the $v_1, \ldots, v_n$ in a "Bohr set" of small doubling constant, at which point one can apply a Freiman-type theorem (e.g. Theorem 5.44) to place the v_j in a progression. This result played an essential role in establishing the bound $\mathbf{P}(\det(M_n) = 0) = (\frac{3}{4} + o(1))^n$ for $n \times n$ random Bernoulli matrices; see Section 7.5 for further discussion.

Exercise

7.4.1 Let the notation and hypotheses be as in Proposition 7.21, and let $1 \leq m \leq k$. Show that either

$$\mathbf{P}\big(X_{\mathbf{v}}^{(1)} = x\big) = O_d\big(mk^{-d/2}\big)$$

or there exists a progression $P = [-k, k]^{d-1} \cdot (w_1, \ldots, w_{d-1})$ in Z such that for all but at most k^2 exceptional values of $j \in [1, n]$, there exist at least k/m values $a_0 \in [1, k]$ such that $a_0 v_j \in P$. (Hint: argue as in Proposition 7.21, but work with $k/2$-dissociated tuples instead of k-dissociated ones, and add one extra copy of $\mathbf{v}$ in (7.7). Then if the latter conclusion fails, use Corollary 7.12 one final time to exploit the sparseness of the a_0 for which $a_0 v_j \in P$ and thence obtain the former conclusion.)

7.5 Random Bernoulli matrices

Let M_n be the random $n \times n$ matrix whose entries are independent uniformly distributed signs ± 1 (M_n is often referred to as the *random Bernoulli matrix*). The distribution of several quantities relating to M_n, such as its determinant and singular values, is of interest to a number of fields, including theoretical physics, combinatorics and theoretical computer science. It turns out that the tools developed in earlier sections are very well adapted for the study of M_n.

In this section we focus on a specific problem, namely to understand the singularity probability $\mathbf{P}(\det(M_n) = 0)$. An equivalent formulation is: given n vectors $X_1, \ldots, X_n$ chosen uniformly at random from the unit cube $\{-1, 1\}^n \in \mathbf{R}^n$, what is the probability that these vectors are linearly independent?

This simple-sounding problem has turned out to be surprisingly non-trivial. It is easy enough to show that

$$\mathbf{P}(X_i = \pm X_j \text{ for some } 1 \leq i < j \leq n \text{ and sign } \pm) = (1 + o(1))n^2 2^{-n}. \quad (7.9)$$

A similar argument (taking into account both the rows and columns of M_n) gives

$$\mathbf{P}(\det(M_n) = 0) \geq (2 + o(1))n^2 2^{-n}. \quad (7.10)$$

It is conjectured that this is sharp; thus

Conjecture 7.24 $\mathbf{P}(\det(M_n) = 0) = (2 + o(1))n^2 2^{-n}$. *In particular,* $\mathbf{P}(\det(M_n) = 0) = (\frac{1}{2} + o(1))^n$.

This conjecture remains open, although we will discuss some progress on this problem in this section. Notice that M_n is singular if and only there is a non-zero vector $v \in \mathbf{R}^n$ such that $M_n v = 0$. By restricting v to some special sets of vectors, we can obtain the conjectured bound $(1/2 + o(1))^n$. The following result is due to Komlós.

Theorem 7.25 *Let $n \geq 3$, and let Ω_1 be the set of vectors in $\mathbf{R}^n$ with at least $3n/\log_2 n$ coordinates. The probability that $M_n v = 0$ for some non-zero $v \in \Omega_1$ is $(1 + o(1))n^2 2^{-n}$.*

By considering the transpose of M_n, one can see that this theorem is equivalent to the following lemma.

Lemma 7.26 *Let $n \geq 3$, and let E denote the event that $a_1 X_1 + \cdots + a_n X_n = 0$ for some non-zero $(a_1, \ldots, a_n) \in \Omega_1$. Then $\mathbf{P}(E) = (1 + o(1))n^2 2^{-n}$.*

Proof To establish the upper bound, we use the union bound to give

$$\mathbf{P}(E) = \sum_{2 \leq k \leq n - 3n/\log_2 n} \mathbf{P}(E_k \backslash E_{k-1})$$

where E_k is the event that $a_1 X_1 + \cdots + a_n X_n = 0$ for some $(a_1, \ldots, a_n) \in \mathbf{R}^n$ with exactly k of the a_j being non-zero. (Note that the event E_1 is vacuous.) From (7.9) we easily see that $\mathbf{P}(E_2) = (1 + o(1))n^2 2^{-n}$, so it will suffice to show that

$$\sum_{3 \leq k \leq n - 3n/\log_2 n} \mathbf{P}(E_k \backslash E_{k-1}) = o(n^2 2^{-n}).$$

From symmetry we have $\mathbf{P}(E_k) \leq \binom{n}{k} \mathbf{P}(F_k \backslash E_{k-1})$, where F_k is the event that $a_1 X_1 + \cdots + a_k X_k = 0$ for some non-zero $a_1, \ldots, a_k$. If $F_k \backslash E_{k-1}$ occurs, then the $n \times k$ matrix whose columns are $X_1, \ldots, X_k$ has rank exactly $k - 1$, and so $(a_1, \ldots, a_k)$ is essentially the wedge product of $k - 1$ of the rows of this matrix. There are $\binom{n}{k-1}$ ways to choose these rows, and then, on fixing all the entries of

those rows (and hence fixing $a_1, \ldots, a_k$), we see from Corollary 7.4 that each of the other $n-k+1$ rows will be consistent with the equation $a_1X_1 + \cdots + a_kX_k = 0$ with probability $\binom{k}{\lfloor k/2 \rfloor}/2^k$. We conclude that

$$\sum_{3 \le k \le n-3\log_2 n} \mathbf{P}(E_k \backslash E_{k-1}) \le \sum_{3 \le k \le n-3n/\log_2 n} \binom{n}{k}\binom{n}{k-1}\left(\binom{k}{\lfloor k/2 \rfloor}/2^k\right)^{n-k+1}.$$

The claim then follows by direct computation (estimating $\binom{k}{\lfloor k/2 \rfloor}/2^k$ by $O(1/\sqrt{n})$ when $k = \Theta(n)$). □

Let us consider another restricted class. Let Ω_2 be the set of integer vectors in $\mathbf{R}^n$ where the coordinates have absolute values at most n^C, for some positive constant C.

Theorem 7.27 *The probability that $M_n v = 0$ for some non-zero $v \in \Omega_2$ is $(1/2 + o(1))^n$. (The error term $o(1)$ depends of course on C.)*

Proof The lower bound is trivial so we focus on the upper. For each non-zero vector v, let $p(v)$ be the probability that $X \cdot v = 0$, where X is a random Bernoulli vector. It is trivial that $\mathbf{P}(M_n v = 0) = p(v)^n$. Since a hyperplane can contain at most 2^{n-1} ± 1 vectors, $p(v)$ is at most $1/2$. For $j = 1, 2, \ldots$ let S_j be the number of non-zero vectors v in Ω_2 such that $2^{-j-1} < p(v) \le 2^{-j}$. Then the probability that $M_n v = 0$ for some non-zero $v \in \Omega_2$ is at most

$$\sum_{j=1}^{n} (2^{-j})^n S_j.$$

Let us now restrict the range of j. Notice that if $p(v) \ge n^{-1/3}$, then by Corollary 7.4 most of the coordinates of v are zero and then by Theorem 7.25 the contribution from these v is at most $(1/2 + o(1))^n$. Next, since the number of vectors in Ω_2 is at most $(2n^C + 1)^n \le n^{(C+1)n}$, we can ignore those j where $2^{-j} \le n^{-C-2}$. Now it suffices to show

$$\sum_{n^{-C-2} \le 2^{-j} \le n^{-1/3}} (2^{-j})^n S_j = o((1/2)^n).$$

Let ϵ be a small positive constant (say .001). As we have $j = \Theta(\log n)$ for all relevant j, we can find an integer $d = O(1)$ such that

$$n^{-(d-1+1/3)\epsilon} > 2^{-j} \ge n^{-(d+1/3)\epsilon}.$$

(The value of d depends on j, but is bounded from above by a constant.) Set $k = n^\epsilon$. Thus $2^{-j} \gg k^{-d}$ and we can use Proposition 7.21 to estimate S_j. Indeed, by invoking this theorem, we see that there are at most $\binom{n}{k^2}(2n^C + 1)^{k^2} = n^{O(k^2)} = n^{o(n)}$ ways to choose the positions and values of exceptional coordinates of v. There

are only $(2n^C + 1)^{d-1} = n^{O(1)}$ ways to fix the generalized progression P. Once P is fixed, the number of ways to set the rest of the coordinates of v is at most $|P|^n = (2k + 1)^{(d-1)n}$. Putting these together,

$$S_j \leq O(1)^n n^{O(k^2)} k^{(d-1)n}.$$

Since $k = n^\varepsilon$ and $2^{-j} \leq n^{-(d-1+1/3)\epsilon}$, it follows that

$$2^{-jn} S_j \leq O(1)^n n^{o(n)} n^{-\epsilon n/3}.$$

As the number of js is only $O(\log n)$, and $n^{-\Omega(n)} \log n = o((1/2)^n)$, we are done. □

By combining Theorem 7.25 with Corollary 7.13, we have the following consequence.

Corollary 7.28 *[215] Let $n \geq 3$. Then for any $1 \leq i \leq n$ we have*

$$\mathbf{P}(X_i \text{ is a linear combination of } X_1, \ldots, X_{i-1}) \leq \min\left(2^{i-n-1}, O\left(\frac{1}{\sqrt{n}}\right)\right).$$

Proof Let us first prove the upper bound of 2^{i-n-1}. Note that $X_1, \ldots, X_{i-1}$ span a space of dimension at most $i - 1$, and so there exist $i - 1$ coordinates which determine all the other coordinates of the space. But if one fixes $i - 1$ coordinates of X_{i-1} then X_{i-1} is still uniformly distributed among 2^{n-i+1} remaining points, and the claim follows. Now we prove the bound of $O(\frac{1}{\sqrt{n}})$. We may assume n is large and i is close to n (say $i > .9n$). The vectors $X_1, \ldots, X_{i-1}$ will be contained in at least one hyperplane $\{(x_1, \ldots, x_n) \in \mathbf{R}^n : a_1 x_1 + \cdots + a_n x_n = 0\}$; choose one arbitrarily. By Corollary 7.25, we certainly will have $\Theta(n)$ of the coordinates non-zero with probability $1 - O(\frac{1}{\sqrt{n}})$ (in fact, we can have much higher probability here). By Corollary 7.13, the probability that $X_i \cdot (a_1, \ldots, a_n) = 0$ is at most $O(\frac{1}{\sqrt{n}})$. Since this event is necessary in order for X_i to be a linear combination of $X_1, \ldots, X_{i-1}$, the claim follows. □

From this corollary, Bayes' identity, and independence, one easily verifies that

$$\begin{aligned}\mathbf{P}(\det(M_n) = 0) &\leq \sum_{i=2}^{n} \mathbf{P}(X_i \text{ is a linear combination of } X_1, \ldots, X_{i-1}) \\ &= O\left(\frac{\log n}{\sqrt{n}}\right)\end{aligned}$$

for large n. This bound was sharpened slightly to $O(\frac{1}{\sqrt{n}})$ in [215], [216] by a variant of this method.

Using a refinement of this argument, one can infact obtain the following estimate for the determinant [364]

$$\mathbf{P}\big(|\det(M_n)| = \sqrt{n!}\exp\big(O\big(n^{1/2}\log^{1/2} n\big)\big)\big) = 1 - o(1).$$

The right-hand side is nearly optimal (see Exercises 7.5.3 and 7.5.4). With the help of recent results from [366], one can have $o(1) = 1/n^C$ for any fix C, at the cost of changing the hidden constant in the O on the left-hand side. It is not clear, however, that one can have $o(1) = \exp(-\Omega(n))$.

Now let us present a breakthrough result of Kahn, Komlós, and Szemerédi [195], which established an exponential bound without any restriction.

Theorem 7.29 *[195] There is a positive constant ε such that* $\mathbf{P}(\det(M_n) = 0) \leq (1-\varepsilon)^n$.

In fact the explicit value $\varepsilon = 0.001$ was obtained in [195]. This was improved to roughly $\varepsilon = 0.042$ in [364], and then to $\varepsilon = \frac{1}{4} + o(1)$ in [365]. Conjecture 7.24 asserts that one can take $\varepsilon = \frac{1}{2} + o(1)$, which would be best possible.

We now sketch the proof of Theorem 7.29. It is convenient to rephrase the problem using the following lemma:

Lemma 7.30 *[195],[374],[364] We have*

$$\mathbf{P}(\det(M_n) = 0) = 2^{o(n)}\mathbf{P}(X_1, \ldots, X_n \textit{ span a hyperplane}).$$

Proof We already know that

$$\mathbf{P}(\det(M_n) = 0) = \mathbf{P}(X_1, \ldots, X_n \text{ linearly dependent}).$$

Thus the lower bound is obvious, and we need only to establish the upper. If $X_1, \ldots, X_n$ are linearly dependent, then there must exist $0 \leq d \leq n-1$ such that $X_1, \ldots, X_{d+1}$ span a d-dimensional subspace. Fixing d and conditioning on this event, we see from repeated application of Corollary 7.28 that $X_1, \ldots, X_n$ will span a hyperplane with probability $2^{-o(n)}$. The claim follows. □

Using this lemma followed by the union bound, it thus suffices to show

$$\sum_V \mathbf{P}(X_1, \ldots, X_n \text{ span } V) \leq (1 - \varepsilon + o(1))^n$$

where V ranges over all hyperplanes. Note that we can restrict our attention to the hyperplanes V which are spanned by their intersection with $\{-1, 1\}^n$; it is easy to see that this is a finite set. Let us call such hyperplanes *non-trivial*. An important quantity associated to a non-trivial hyperplane is its *density*

$$\mathbf{P}(X \in V) = \frac{|V \cap \{-1, 1\}^n|}{|\{-1, 1\}|^n}$$

where we think of X as a random element of $\{-1,1\}^n$. Note that $\mathbf{P}(X \in V) = \mathbf{P}(X_{\mathbf{v}}^{(1)} = 0)$ whenever $\mathbf{v}$ is a normal vector to V. We can exclude the contribution of all the hyperplanes of low density by the following lemma:

Lemma 7.31 *[195] For any $0 < \alpha < 1$, we have*

$$\sum_{V:\mathbf{P}(X\in V)\leq\alpha} \mathbf{P}(X_1,\ldots,X_n \text{ span } V) \leq n\alpha.$$

Proof If $X_1,\ldots,X_n$ span the hyperplane V, then there exists $1 \leq i \leq n$ such that the $n-1$ vectors formed by omitting X_i from $X_1,\ldots,X_n$ still span V. Fixing i and conditioning on this event, we see that V is determined by all the vectors other than X_i, and then X_i has a probability of at most α of also lying in V. The claim follows. □

Thus to establish the claim, it suffices to consider only the high-density hyperplanes for which $\mathbf{P}(X \in V) \geq (1-\varepsilon)^n$. On the other hand, from Lemma 7.26 and Corollary 7.13 we can control the extremely high-density hyperplanes for which $\mathbf{P}(X \in V) \gg \frac{1}{\sqrt{n}}$. So in fact we only need to deal with the range where $(1-\varepsilon)^n \leq \mathbf{P}(X \in V) \leq O(\frac{1}{\sqrt{n}})$.

We now crucially exploit the relative Halász inequality, Lemma 7.14. Let $0 < \mu \ll 1$ be a small parameter (independent of n), and let $Y \in \{-1,0,1\}^n$ be the random variable $Y = (\eta_1^{(\mu)},\ldots,\eta_n^{(\mu)})$. Lemma 7.14 implies (if n is large enough) that Y concentrates on the above hyperplanes V more strongly than X does, if μ is sufficiently small:

$$\mathbf{P}(Y \in V) = O(\sqrt{\mu})\mathbf{P}(X \in V). \tag{7.11}$$

If we use the informal heuristic

$$\mathbf{P}(X_1,\ldots,X_n \text{ span } V) \approx \mathbf{P}(X \in V)^n$$

then we thus expect

$$\mathbf{P}(X_1,\ldots,X_n \text{ span } V) \leq O(\sqrt{\mu})^n \mathbf{P}(Y_1,\ldots,Y_n \text{ span } V)$$

where $Y_1,\ldots,Y_n$ are identical independent copies of Y. Summing this in V, and using the trivial fact that each $Y_1,\ldots,Y_n$ can span at most one hyperplane V we thus expect

$$\mathbf{P}(M_n = 0) \leq O(\sqrt{\mu})^n$$

which certainly gives Theorem 7.29 by setting μ small enough.

The above strategy almost works, except for a slight problem in that the $Y_1,\ldots,Y_n$ may be so linearly dependent that they will only span a subspace of V rather than V itself. The simplest way to solve this problem is to use only a small

number of Y, say $Y_1, \ldots, Y_{\delta n}$ for some small[1] δ. If V is sufficiently high-density and δ is small enough, we can ensure that $Y_1, \ldots, Y_{\delta n}$ will remain linearly independent in V. This reduces the potential gain in this argument from $O(\sqrt{\mu})^n$ to only $O(\sqrt{\mu})^{\delta n}$, but this is still enough to establish Theorem 7.29.

More rigorously, we introduce $Y_1, \ldots, Y_{\delta n}$ independently of $X_1, \ldots, X_n$. Fix a density $(1-\varepsilon)^n \leq \sigma \leq O(\frac{1}{\sqrt{n}})$, and let V be such that $\mathbf{P}(X \in V) = (1 + O(\frac{1}{n}))\sigma$:

$$\mathbf{P}(Y_1, \ldots, Y_{\delta n} \in V) \geq \Omega\left(\frac{1}{\sqrt{\mu}}\right)^{\delta n} \sigma^{\delta n}.$$

If δ is sufficiently small depending on μ, and ε is sufficiently small depending on δ and μ, then one can modify Corollary 7.28 to refine this to

$$\mathbf{P}(Y_1, \ldots, Y_{\delta n} \text{ linearly dependent in } V) \geq \Omega\left(\frac{1}{\sqrt{\mu}}\right)^{\delta n} \sigma^{\delta n}; \tag{7.12}$$

we leave this as an exercise. From independence we thus have

$$\mathbf{P}(X_1, \ldots, X_n \text{ span } V) \leq O(\sqrt{\mu})^{\delta n} \sigma^{-\delta n} \mathbf{P}(E_V)$$

where E_V is the event that $X_1, \ldots, X_n$ span V and $Y_1, \ldots, Y_{\delta n}$ are linearly independent in V. But if this event occurs, then there exist $n - \delta n$ vectors in $X_1, \ldots, X_n$ which, together with $Y_1, \ldots, Y_{\delta n}$, span V. If we fix all these vectors then V is also fixed, and the remaining δn vectors in $X_1, \ldots, X_n$ have a probability of $\Theta(\sigma^{\delta n})$ of lying in V. We thus conclude that

$$\sum_{V: \mathbf{P}(X \in V) = (1+O(\frac{1}{n}))\sigma} \mathbf{P}(E_V) \leq \binom{n}{\delta n} \Theta(\sigma^{\delta n})$$

which, when combined with the preceding estimates, give

$$\sum_{V: \mathbf{P}(X \in V) = (1+O(\frac{1}{n}))\sigma} \leq \mathbf{P}(X_1, \ldots, X_n \text{ span } V) \leq O(\sqrt{\mu})^{\delta n} \binom{n}{\delta n}.$$

If we choose δ sufficiently small depending on μ, and ε sufficiently small depending on δ, μ, we can make the right-hand side $(1 - \varepsilon + o(1))^n$. Summing over all relevant σ (there are only about $O(n^2)$ such σ to sum over) we obtain Theorem 7.29 as desired.

By using Theorem 7.23 one can boost ε to be as large as $\frac{1}{4} + o(1)$. The basic point is that Theorem 7.23 allows one to improve (7.11) significantly unless the hyperplane V has an exceptional form (in particular, the coordinates of its normal

[1] Strictly speaking we should use $\lfloor \delta n \rfloor$ instead of δn but we shall omit this inessential detail for ease of exposition.