The Full Rank Condition for Sparse Random Matrices

Amin Coja-Oghlan
Department of Computer Science, TU Dortmund, Germany

Jane Gao
Department of Combinatorics and Optimization, University of Waterloo, Canada

Max Hahn-Klimroth
Department of Computer Science, TU Dortmund, Germany

Joon Lee
Communication Theory Laboratory, École Polytechnique Fédérale de Lausanne, Switzerland

Noela Müller
Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands

Maurice Rolvien
Department of Computer Science, TU Dortmund, Germany

Abstract

We derive a sufficient condition for a sparse random matrix with given numbers of non-zero entries in the rows and columns having full row rank. Inspired by low-density parity check codes, the family of random matrices that we investigate is very general and encompasses both matrices over finite fields and $\{0,1\}$-matrices over the rationals. The proof combines statistical physics-inspired coupling techniques with local limit arguments.

2012 ACM Subject Classification Mathematics of computing; Theory of computation → Error-correcting codes

Keywords and phrases random matrices, rank, finite fields, rationals

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2023.54

Category RANDOM

Related Version *Full Version*: https://arxiv.org/abs/2112.14090

Funding *Amin Coja-Oghlan*: DFG CO 646/3 and DFG CO 646/5
Max Hahn-Klimroth: DFG FOR 2975
Noela Müller: NWO Gravitation grant NETWORKS-024.002.003

1 Introduction

1.1 Background and motivation

While "continuous" random matrices such as, for example, a random $n \times n$-matrix with independent Gaussian entries have full rank almost surely for trivial reasons, the rank problem for random combinatorial matrices with entries drawn from discrete distributions poses deep mathematical challenges. In the 1960s Komlós, among the first to study this type of problem, proved that a random $n \times n$-matrix with independent ± 1-entries has full rational rank with high probability [24]. An obvious lower bound on the singularity probability is the probability $2^{-n+o(n)}$ that two rows or columns coincide. The conjecture that this lower bound is tight inspired an impressive line of work (e.g., [23, 33]), which culminated in Tikhomirov's proof of the conjecture [34].

By comparison to the case of dense random matrices, relatively little is known about the sparse case where the average number of non-zero entries per row or column is bounded. Moreover, techniques developed for dense random matrices such as large deviations inequalities or Littlewood-Offord arguments do not easily carry over to the sparse case. Yet sparse random matrices are of key interest in computer science. Prominent applications include low-density parity check codes [32], data compression [1, 36] and hashing [16].

One of the first full rank theorems for sparse random matrices came in the guise of a random constraint satisfaction problem. Specifically, in the random k-XORSAT problem we form a random Boolean formula over n variables with m independent XOR-clauses of length k. The question is for what clause densities m/n such a random formula admits an (XOR-)satisfying assignment. Because Boolean XOR is equivalent to addition over the field $\mathbb{F}_2$, this question boils down to determining the threshold m/n up to which a random $m \times n$-matrix over $\mathbb{F}_2$ with precisely k ones per row has full row rank. For the case $k = 3$ this problem was solved by Dubois and Mandler [18] in 2002. Remarkably, the case of general k was solved only more than ten years later by Pittel and Sorkin [31]. Both proofs depend on delicate and technically demanding moment computations.

The contribution of the present paper is a sufficient condition for a sparse random combinatorial matrix to have full row rank. We derive this sufficient condition within the framework of a very general model of random matrices that hails from coding theory [32]. As a very special case this general result encompasses the random k-XORSAT problem. But in addition, we obtain a range of other important special cases such as matrices in which the number of non-zero entries per row or column follows a power law. In fact, the sufficient condition that we obtain is essentially necessary, too. The proof of the main result is based on a novel combination of statistical physics-inspired coupling arguments and local limit theorem techniques. These methods are conceptually more powerful than the method of moments as there exist concrete instances of the present random matrix model where the matrix provably has full rank even though the method of moments fails spectacularly.

1.2 Results

The random matrix model that we investigate allows to control the number of non-zero entries in the rows and columns. The model is identical to the type of model used to construct low-density parity check codes [9, 32]. Specifically, let $\boldsymbol{d} \geq 0$, $\boldsymbol{k} \geq 3$ be independent integer-valued random variables such that $\mathbb{E}[\boldsymbol{d}^{2+\eta}] + \mathbb{E}\left[\boldsymbol{k}^{2+\eta}\right] < \infty$ for an arbitrarily small $\eta > 0$. Let $(\boldsymbol{d}_i, \boldsymbol{k}_i)_{i\geq 1}$ be independent copies of $(\boldsymbol{d}, \boldsymbol{k})$ and set $d = \mathbb{E}[\boldsymbol{d}], k = \mathbb{E}[\boldsymbol{k}]$. Furthermore, let $\mathfrak{d}$ and $\mathfrak{k}$ be the greatest common divisors of the support of $\boldsymbol{d}$ and $\boldsymbol{k}$, respectively. Finally, let $n > 0$ be a large integer divisible by $\mathfrak{k}$ and let $\boldsymbol{m} \sim \mathrm{Po}(dn/k)$ be independent of $(\boldsymbol{d}_i, \boldsymbol{k}_i)_{i\geq 1}$. These definitions ensure that the event

$$\sum_{i=1}^{n} \boldsymbol{d}_i = \sum_{j=1}^{\boldsymbol{m}} \boldsymbol{k}_j \tag{1.1}$$

occurs with probability $\Omega(n^{-1/2})$ [9, Proposition 1.7]. Hence, assuming (1.1) occurs, let $\mathbb{G} = \mathbb{G}_n(\boldsymbol{d}, \boldsymbol{k})$ be a simple random bipartite graph on a set $\{a_1 \ldots, a_{\boldsymbol{m}}\}$ of *check nodes* and a set $\{x_1, \ldots, x_n\}$ of *variable nodes* such that the degree of a_i equals $\boldsymbol{k}_i$ and the degree of x_j equals $\boldsymbol{d}_j$ for all i, j. Adopting coding terminology, we refer to $\mathbb{G}$ as the *Tanner graph*. We need to assume that the second moment is bounded so that the *Tanner graph* is locally finite. The random graph $\mathbb{G}$ naturally induces a $\{0, 1\}$-matrix, namely the $\boldsymbol{m} \times n$-biadjacency matrix $\mathbb{B} = \mathbb{B}(\mathbb{G})$ of the bipartite graph. Explicitly,

$$\mathbb{B}_{ij} = \mathbb{1}\{a_i x_j \in E(\mathbb{G})\} \qquad (1 \le i \le \boldsymbol{m},\, 1 \le i \le n).$$

Let

$$D(z) = \sum_{\ell=0}^{\infty} \mathbb{P}\left[\boldsymbol{d} = \ell\right] z^{\ell} \qquad \text{and} \qquad K(z) = \sum_{\ell=0}^{\infty} \mathbb{P}\left[\boldsymbol{k} = \ell\right] z^{\ell}$$

be the probability generating functions of $\boldsymbol{d}$ and $\boldsymbol{k}$. Since $\mathbb{E}[\boldsymbol{d}^2] + \mathbb{E}[\boldsymbol{k}^2] < \infty$, the function

$$\Phi : [0,1] \to \mathbb{R}, \qquad z \mapsto D\left(1 - K'(z)/k\right) - \frac{d}{k}\left(1 - K(z) - (1-z)K'(z)\right) \tag{1.2}$$

is well-defined. The following result renders a sufficient condition for $\mathbb{B}$ to have full row rank.

▶ **Theorem 1.** *If*

$$\Phi(z) < \Phi(0) \qquad \textit{for all } 0 < z \le 1, \tag{1.3}$$

then $\mathbb{B}$ has full rational row rank w.h.p.

Theorem 1 is a direct consequence of a significantly stronger result on matrices over finite fields. Specifically, suppose that $q \ge 2$ is a prime power, let $\mathbb{F}_q$ signify the field with q elements and let $\boldsymbol{\chi}$ be a random variable that takes values in $\mathbb{F}_q^* = \mathbb{F}_q \setminus \{0\}$. Let $(\boldsymbol{\chi}_{i,j})_{i,j\ge 1}$ be independent copies of $\boldsymbol{\chi}$. Finally, let $\mathbb{A} = \mathbb{A}_n(\boldsymbol{d}, \boldsymbol{k}, \boldsymbol{\chi})$ be the $\boldsymbol{m} \times n$-matrix with entries

$$\mathbb{A}_{i,j} = \mathbb{1}\left\{a_i x_j \in E(\mathbb{G})\right\} \cdot \boldsymbol{\chi}_{i,j} \in \mathbb{F}_q.$$

Hence, the i-th row of $\mathbb{A}$ contains $\boldsymbol{k}_i$ non-zero entries and the j-th column contains $\boldsymbol{d}_j$ non-zero entries.

▶ **Theorem 2.** *If q and $\mathfrak{d}$ are coprime and* (1.3) *is satisfied, then $\mathbb{A}$ has full row rank over $\mathbb{F}_q$ w.h.p.*

Theorem 1 follows from Theorem 2 and a few lines of linear algebra.

The sufficient condition (1.3) is quite close to being necessary, too. Indeed, the *normalised* rank of $\mathbb{A}$ (and $\mathbb{B}$) can be expressed in terms of the function Φ as follows [9, Theorem 1.1]:

$$\frac{\mathrm{rk}(\mathbb{A})}{n} \xrightarrow{n\to\infty} 1 - \max_{z\in[0,1]} \Phi(z) \qquad \text{in probability.} \tag{1.4}$$

Since $\boldsymbol{k} \ge 3$, the definition (1.2) ensures that $\Phi(0) = 1 - d/k$ and thus $n\Phi(0) \sim n - \boldsymbol{m}$ w.h.p. Hence, (1.4) implies that $\mathrm{rk}(\mathbb{A}) \le \boldsymbol{m} - \Omega(n)$ w.h.p. unless $\Phi(z)$ attains its maximum at $z = 0$. In other words, $\mathbb{A}$ has full row rank *only if* $\Phi(z) \le \Phi(0)$ for all $0 < z \le 1$. Indeed, in Section 1.3 we will discover examples that require a strict inequality as in (1.3). The condition that q and $\mathfrak{d}$ be coprime is generally necessary as well, as we will see in Example 8 below.

▶ Remark 3. We emphasise that (1.4) does not guarantee that $\mathbb{A}$ has full row rank w.h.p. even if (1.3) is satisfied. In fact, due to the normalisation on the l.h.s., (1.4) only implies that $\mathrm{rk}(\mathbb{A}) = \boldsymbol{m} - o(n)$ w.h.p., rather than the far stronger statement that $\mathrm{rk}(\mathbb{A}) = \boldsymbol{m}$ w.h.p. delivered by Theorem 2.

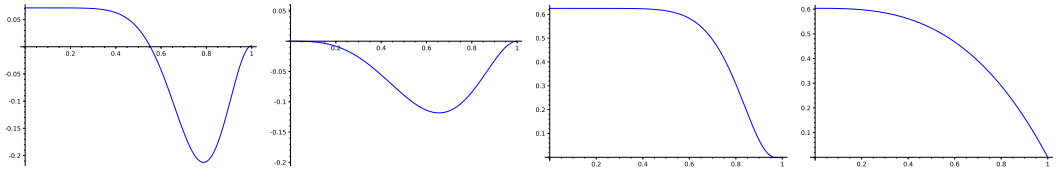

Figure 1 From left to right: the shape of Φ for Examples 4–7.

1.3 Examples

To illustrate the power of Theorems 1 and 2 we consider a few instructive special cases of distributions $\boldsymbol{d}, \boldsymbol{k}, \boldsymbol{\chi}$.

▶ **Example 4** (random k-XORSAT)**.** In random k-XORSAT we are handed a number of independent random constraints c_i of the type $c_i = y_{i1}$ XOR $\cdots$ XOR y_{ik} where each literal y_{ij} is either one of n available Boolean variables $x_1, \ldots, x_n$ or a negation $\neg x_1, \ldots, \neg x_n$. The goal to determine the maximum number of random constraints can be satisfied simultaneously w.h.p. Because Boolean XOR comes down to addition over $\mathbb{F}_2$ and since the clauses are drawn independently, XOR-satisfiability can be rephrased as the full rank problem for the random matrix $\mathbb{A}$ over $\mathbb{F}_q$ with $q = 2$, $\boldsymbol{k} = k$ fixed to a deterministic value and $\boldsymbol{d} \sim \mathrm{Po}(d)$ a Poisson variable. Hence, the generating functions of $\boldsymbol{d}, \boldsymbol{k}$ read $D(z) = \exp(d(z-1))$ and $K(z) = z^k$ and $\Phi_{d,k}(z) = \exp(-dz^{k-1}) - \frac{d}{k}\left(1 - kz^{k-1} + (k-1)z^k\right)$. Thus, Theorem 2 implies that for a given $k \geq 3$ the threshold of d up to which random k-XORSAT is satisfiable w.h.p. equals the largest d such that

$$\Phi_{d,k}(z) < \Phi_{d,k}(0) = 1 - d/k \qquad \text{for all } 0 < z \leq 1. \tag{1.5}$$

A few lines of calculus verify that (1.5) matches the formulas for the k-XORSAT threshold derived by combinatorial methods tailored to this specific case [18, 31]. Theorem 2 also encompasses the generalisations of XORSAT to other finite fields $\mathbb{F}_q$ from [5, 21].

▶ **Example 5** (identical distributions)**.** An interesting scenario arises when $\boldsymbol{d}, \boldsymbol{k}$ are identically distributed. For example, suppose that $\mathbb{P}[\boldsymbol{d} = 3] = \mathbb{P}[\boldsymbol{d} = 4] = \mathbb{P}[\boldsymbol{k} = 3] = \mathbb{P}[\boldsymbol{k} = 4] = 1/2$. Thus, $D(z) = K(z) = (z^3 + z^4)/2$. The resulting $\Phi(z)$ attains two identical maxima, namely $\Phi(0) = \Phi(1) = 0$. Since $\boldsymbol{k}_i, \boldsymbol{d}_i$ are chosen independently subject only to (1.1), the probability that $\mathbb{A}$ has more rows than columns works out to be $1/2 + o(1)$. As a consequence, $\mathbb{A}$ cannot have full row rank w.h.p. This shows that the condition that 0 be the *unique* maximiser of $\Phi(x)$ is generally necessary.

▶ **Example 6** (fixed $\boldsymbol{d}, \boldsymbol{k}$)**.** Suppose that both $\boldsymbol{d} = d, \boldsymbol{k} = k \geq 3$ are constants rather than genuinely random. Then $\Phi(z) = (1 - z^{k-1})^d - \frac{d}{k}(1 - kz^{k-1} + (k-1)z^k)$. Clearly, $\mathbb{A}$ cannot have full row rank unless $d \leq k$, while Theorem 2 implies that $\mathbb{A}$ has full row rank w.h.p. if $d < k$. This result was previously established via the second moment method [30]. But in the critical case $d = k$ the function $\Phi(z)$ attains its identical maxima at $z = 0$ and $z = 1$. Specifically, $0 = \Phi(0) = \Phi(1) > \Phi(z)$ for all $0 < z < 1$. Hence, Theorem 2 does not cover this special case. Nonetheless, Huang [22] and Mészáros [29] proved that the random $\{0,1\}$-matrix $\mathbb{B}$ has full rational rank w.h.p. The proof is based on a delicate moment computation in combination with a precise local expansion via the Laplace method. However, numerical evidence indicates that the corresponding "d-regular" random matrix $\mathbb{A}$ over a finite field fails to have full rank w.h.p.

▶ **Example 7** (power laws). Let $\mathbb{P}(\boldsymbol{d} = \ell) \propto \ell^{-\alpha}$ for some $\alpha > 3$ and $\boldsymbol{k} = k \geq 3$. Thus,

$$D(z) = \frac{1}{\zeta(\alpha)} \sum_{\ell=1}^{\infty} \frac{z^\ell}{\ell^\alpha}, \quad K(z) = z^k,$$
$$\Phi(z) = D\left(1 - z^{k-1}\right) - \frac{\zeta^{-1}(\alpha)\zeta(\alpha-1)}{k}\left(1 - kz^{k-1} + (k-1)z^k\right)$$

and it can be verified that $\Phi'(z) < 0$ for all $z \in (0,1)$. Hence, (1.3) is always satisfied and Theorems 1 and 2 show that $\mathbb{A}, \mathbb{B}$ have full row rank.

▶ **Example 8** (zero row sums). Theorem 2 requires the assumption that q and the g.c.d. $\mathfrak{d}$ of the support of $\boldsymbol{d}$ be coprime. This assumption is indeed necessary. To see this, consider the case that $q = 2$, $\boldsymbol{\chi} = 1$, $\boldsymbol{d} = 4$ and $\boldsymbol{k} = 8$ deterministically. Then the rows of $\mathbb{A}$ always sum to zero. Hence, $\mathbb{A}$ cannot have full row rank.

2 Proof Strategy

The proof of the main theorem (Theorem 2) substantially extends the techniques behind the asymptotic rank formula (1.4) from [9]. As one key additional ingredient we require a new method to count "equitable" vectors in the kernel of $\mathbb{A}$, i.e., vectors in which each element of $\mathbb{F}_q$ occurs with frequency $1/q + o(1)$. This part of the proof, which involves the asymptotic enumeration of lattice points that satisfy certain arithmetic conditions, hinges on local limit techniques and algebraic arguments, specifically the identification of suitable bases of $\mathbb{Z}$-modules generate by lattice points. This argument generalises techniques that were also used in the study of adjacency matrices of random d-regular graphs [22, 29].

To motivate the proof strategy we first go down the "classical" path of the method of moments. We will discover where this proof strategy gets stuck and then work our way around the obstacle by means of physics-inspired coupling arguments. In statistical physics jargon, the moment calculation amounts to an "annealed" analysis. In a nutshell, the issue with such analyses is that unlikely events can render outsized contributions to moments of exponentially large random variables such as the number of vectors in the kernel of $\mathbb{A}$. Once we see where such large deviations hazards lurk, we will be able to replace the "annealed" strategy by a "quenched" approach that sidesteps these large deviations effects.

2.1 The method of moments

By extension of random k-XORSAT from Example 4, the full rank problem for the random matrix $\mathbb{A}$ over $\mathbb{F}_q$ can be viewed as a random constraint satisfaction problem. Specifically, choose a vector $\boldsymbol{y} \in \mathbb{F}_q^{\boldsymbol{m}}$ uniformly independently of $\mathbb{A}$. Then a solution to our random CSP is just a solution $x \in \mathbb{F}_q^n$ to the linear system $\mathbb{A}x = \boldsymbol{y}$. Thus, together with the corresponding entry of $\boldsymbol{y}$ each of the $\boldsymbol{m}$ rows of $\mathbb{A}$ induces a constraint.

Since the early 2000s the default method for approaching random CSPs has been the second moment method [2, 3]. Indeed, one of the first contributions to this line of work was the aforementioned work of Dubois and Mandler on random 3-XORSAT [18], which corresponds to the special case $q = 2$, $\boldsymbol{k} = 3$, $\boldsymbol{d} = \mathrm{Po}(d)$. A natural first stab at the full rank problem therefore appears to be to run the second moment routine on the number $\boldsymbol{Z} = \boldsymbol{Z}(\mathbb{A}, \boldsymbol{y})$ of solutions to $\mathbb{A}x = \boldsymbol{y}$. But clearly, to have any chance of success we need to condition on the degrees of the variable and check nodes, and a few more innocent pieces of information. Formally, let $\mathfrak{A}$ be the σ-algebra generated by $\boldsymbol{m}, (\boldsymbol{k}_i)_{i\geq 1}, (\boldsymbol{d}_i)_{i\geq 1}$ and by the numbers $\boldsymbol{m}(\chi_1, \ldots, \chi_\ell)$ of rows with non-zero entries $\chi_1, \ldots, \chi_\ell \in \mathbb{F}_q^*$. Let us write $\mathbb{P}_{\mathfrak{A}} = \mathbb{P}[\cdot \mid \mathfrak{A}]$ and $\mathbb{E}_{\mathfrak{A}} = \mathbb{E}[\cdot \mid \mathfrak{A}]$ for the conditional probability and expectation given $\mathfrak{A}$.

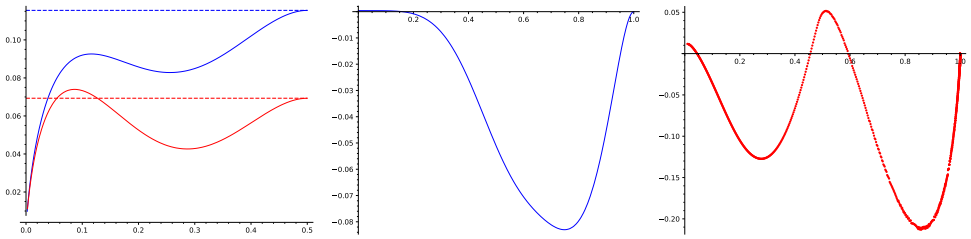

Figure 2 *Left:* the r.h.s. of (2.6) for $d = 2.5$ (blue) and $d = 2.7$ (red) in the interval $[0, \frac{1}{2}]$. *Middle:* the function $\Phi(z)$ from Example 9. *Right:* numerical lower bound on the moment formula from Example 9.

Since $\boldsymbol{y}$ is independent of $\mathbb{A}$, for any fixed $x \in \mathbb{F}_q^n$ the event $\mathbb{A}x = \boldsymbol{y}$ has probability $q^{-\boldsymbol{m}}$. As there are q^n choices of x, linearity of expectation yields

$$\mathbb{E}_{\mathfrak{A}}[\boldsymbol{Z}] = q^{n-\boldsymbol{m}}. \tag{2.1}$$

For the second moment method to succeed we need to verify that $\mathbb{E}_{\mathfrak{A}}[\boldsymbol{Z}^2] \sim \mathbb{E}_{\mathfrak{A}}[\boldsymbol{Z}]^2$. Then Chebyshev's inequality yields $\boldsymbol{Z} \sim \mathbb{E}_{\mathfrak{A}}[\boldsymbol{Z}]$ w.h.p., and thus $\mathbb{A}x = \boldsymbol{y}$ has a solution w.h.p., provided that $n \leq \boldsymbol{m}$. This fact, in turn, would imply that $\mathbb{A}$ has full row rank w.h.p.; for if it were the case that $\operatorname{rk}\mathbb{A} < \boldsymbol{m}$, then the linear system $\mathbb{A}x = \boldsymbol{y}$ would fail to have a solution with probability at least $1/q$.

Concerning the computation of $\mathbb{E}_{\mathfrak{A}}[\boldsymbol{Z}^2]$, because the set of solutions is either empty or a translation of the kernel, we obtain

$$\begin{aligned}\mathbb{E}_{\mathfrak{A}}[\boldsymbol{Z}^2] &= \sum_{\sigma,\tau\in\mathbb{F}_q^n} \mathbb{P}_{\mathfrak{A}}\left[\mathbb{A}\sigma = \mathbb{A}\tau = \boldsymbol{y}\right] = \sum_{\sigma,\tau\in\mathbb{F}_q^n} \mathbb{P}_{\mathfrak{A}}\left[\mathbb{A}\sigma = \boldsymbol{y}\right]\mathbb{P}_{\mathfrak{A}}\left[\sigma-\tau\in\ker\mathbb{A}\right] \\ &= \mathbb{E}_{\mathfrak{A}}\left[\boldsymbol{Z}\right]\mathbb{E}_{\mathfrak{A}}\left[|\ker\mathbb{A}|\right].\end{aligned} \tag{2.2}$$

Hence, we are left to calculate $\mathbb{E}_{\mathfrak{A}}\left[|\ker\mathbb{A}|\right]$.

Unlike in the case (2.1) of the first moment of $\boldsymbol{Z}$, the probability of belonging to the kernel of $\mathbb{A}$ is not the same for all $x \in \mathbb{F}_q^n$. Indeed, as an extreme example, the zero vector always belongs to the kernel. By contrast, depending on $\boldsymbol{d}, \boldsymbol{k}$ and q there may be vectors that cannot possibly belong to the kernel for divisibility reasons; e.g., if $\boldsymbol{k} = 3$ and $q = 2$, then the all-ones vector cannot lie in $\ker\mathbb{A}$.

Hence, we need to tread carefully. In order to calculate the expected size of the kernel we need to discriminate vectors x according to the number $n_\ell(s)$ of variable nodes of a given degree ℓ that take a specific value $s \in \mathbb{F}_q$. Further, given the $n_\ell(s)$ we need to know the numbers $m_{\chi_1,\ldots,\chi_\ell}(s_1,\ldots,s_\ell)$ of rows with non-zero entries $\chi_1,\ldots,\chi_\ell$ whose neighbouring variable nodes in $\mathbb{G}$ receive values $s_1,\ldots,s_\ell$. Since given $\mathfrak{A}$ the matching of the variable and check nodes remains random given their degrees, the ensuing contribution to the first moment works out to be

$$\begin{aligned}\Xi(n_\ell(s), m_{\chi_1,\ldots,\chi_\ell}(s_1,\ldots,s_\ell))_{\ell,s,s_1,\ldots,s_\ell} &= \sum_{s\in\mathbb{F}_q}\mathbb{E}\left[(\boldsymbol{d}-1)n_{\boldsymbol{d}}(s)\log\frac{n_{\boldsymbol{d}}(s)}{n}\right] \\ -\frac{d}{k}\mathbb{E}\left[\sum_{\sigma_1,\ldots,\sigma_{\boldsymbol{k}}\in\mathbb{F}_q}\mathbb{1}\left\{\boldsymbol{\chi}\perp\sigma\right\}m_{\boldsymbol{\chi}_{1,1},\ldots,\boldsymbol{\chi}_{1,\boldsymbol{k}}}(\sigma_1,\ldots,\sigma_{\boldsymbol{k}})\log\frac{m_{\boldsymbol{\chi}_{1,1},\ldots,\boldsymbol{\chi}_{1,\boldsymbol{k}}}(\sigma_1,\ldots,\sigma_{\boldsymbol{k}})}{\boldsymbol{m}}\right].\end{aligned} \tag{2.3}$$

Then

$$\mathbb{E}_{\mathfrak{A}}|\ker\mathbb{A}| = \exp\left[\max_{n_\ell(s),m_{\chi_1,\ldots,\chi_\ell}(s_1,\ldots,s_\ell)}\Xi(n_\ell(s),m_{\chi_1,\ldots,\chi_\ell}(s_1,\ldots,s_\ell))+o(n)\right]. \tag{2.4}$$

Hence, in order to compute the expected kernel size we should maximise the fairly impressive formula (2.3) over a potentially *very* large range of parameters $n_\ell(s)$, $m_{\chi_1,\ldots,\chi_\ell}(s_1,\ldots,s_\ell)$. The choice of these parameters is subject to the constraint that for every value $s \in \mathbb{F}_q$ the number of occurrences of s counted from the side of the check nodes must equal the number of occurrences viewd from the variable side. This leads to the equations

$$\mathbb{E}[\boldsymbol{d}n_{\boldsymbol{d}}(s)] = \mathbb{E}\left[\sum_{\sigma_1,\ldots,\sigma_{\boldsymbol{k}}\in\mathbb{F}_q} \boldsymbol{k}\mathbb{1}\left\{\sigma_1 = s\right\}\mathbb{1}\left\{\boldsymbol{\chi} \perp \sigma\right\} m_{\boldsymbol{\chi}_{1,1},\ldots,\boldsymbol{\chi}_{1,\boldsymbol{k}}}(\sigma_1,\ldots,\sigma_{\boldsymbol{k}})\right] \forall s \in \mathbb{F}_q.$$

Taking these constraints into account, we can transform (2.4) into a Lagrangian optimisation problem whose only variables are the $n_\ell(s)$, $s \in \mathbb{F}_q$, $\ell \in \mathrm{supp}\boldsymbol{d}$. A somewhat delicate application of the Laplace method then shows that $\mathbb{E}_{\mathfrak{A}}[|\ker\mathbb{A}|] = O(q^{n-\boldsymbol{m}})$, i.e., that the second moment method "nearly works", if and only if the maximum of (2.3) is attained at the "equitable" solution

$$n_\ell(s) \sim n\mathbb{P}\left[\boldsymbol{d} = \ell\right]/q \qquad \text{for all } s \in \mathbb{F}_q,\ \ell \in \mathrm{supp}\boldsymbol{d}. \tag{2.5}$$

Unfortunately, we have no idea how to solve the optimisation problem (2.4) in any generality. Worse, even if we knew how to tackle this optimisation task, that would still not suffice to prove Theorem 2. Indeed, the plain moment calculation fails even for random 3-XORSAT, i.e., the case $q = 2$, $\boldsymbol{k} = 3$ constant and $\boldsymbol{d} = \mathrm{Po}(d)$. In this case the second moment calculation reduces to the one-dimensional optimisation problem

$$\log \mathbb{E}_{\mathfrak{A}}|\ker\mathbb{A}| \sim n \cdot \max_{z\in[0,1]} -z\log z - (1-z)\log(1-z) + \frac{\boldsymbol{m}}{n}\log\frac{1+(1-2z)^3}{2} \quad \text{(cf. [18])}. \tag{2.6}$$

At $z = 1/2$ the r.h.s. of (2.6) simplifies to $(n-\boldsymbol{m})\log 2$, and thus $\mathbb{E}_{\mathfrak{A}}|\ker\mathbb{A}| = 2^{n-\boldsymbol{m}}$ matches the first moment (2.1). But if the maximum (2.6) is attained at $z \neq 1/2$, then $\mathbb{E}_{\mathfrak{A}}|\ker\mathbb{A}| \gg 2^{n-\boldsymbol{m}}$ and the second moment method fails. Figure 2 displays (2.6) for $d = 2.5$ and $d = 2.7$. While for $d = 2.5$ the function takes its maximum at $z = 1/2$, for $d = 2.7$ the maximum is attained at $z \approx 0.085$. However, the actual random 3-XORSAT threshold is $d \approx 2.75$ [18]. Thus, method of moments fails short of the real threshold.

The reason for this is that rare events are apt to boost the *expected* number of vectors in the kernel. This is precisely what happens in random k-XORSAT. The rare event in question is a fluctuation of the density of the *2-core* $\mathbb{G}^{(2)}$ of the Tanner graph, which is obtained by iteratively removing any variable nodes of degree at most one along with their unique adjacent check node if the variable degree equals one. Dubois and Mandler therefore pinpointed the 3-XORSAT threshold by applying the second moment method to the minor $\mathbb{A}^{(2)}$ induced by $\mathbb{G}^{(2)}$ while conditioning on the 2-core having its typical size and density. However, even in random k-XORSAT with $k > 3$ the ensuing optimisation problem (2.3) is anything but straightforward, as witnessed by the work of Pittel and Sorkin [31]. Furthermore, increasing the size of the field to $q > 2$ boosts the number of variables involved, which adds further significant challenges to the optimisation problem; even the case $q = 3$ turns out to be essentially intractable [21]. Finally, for general $\boldsymbol{d}, \boldsymbol{k}$ and q it is far from clear what the "relevant" variables would be that are responsible for any large deviations effects. Inspecting a few examples of degree distributions $\boldsymbol{d}, \boldsymbol{k}$ reveals that conditioning on the size and density of the 2-core will not generally suffice.

The upshot is that the second moment method hardly seems like a promising path towards Theorem 2. But we learned that we basically need to get a handle on the typical size of the kernel of $\mathbb{A}$. Specifically, if we could prove that typical vectors in the kernel are nearly equitable in the sense that all elements $s \in \mathbb{F}_q$ occur about n/q times, then we could conceivably derive the desired bound $|\ker\mathbb{A}| \leq q^{n-\boldsymbol{m}}$ w.h.p.

▶ **Example 9** (failure of the moment method)**.** To underscore the issue with the method of moments, consider the random variables $\boldsymbol{d}, \boldsymbol{k}$ with generating functions $D(z) = 0.889z^3 + 0.111z^{21}$ and $K(z) = z^5$ and set $q = 101$. The resulting function $\Phi(z)$ (just barely) attains its unique maximum at $z = 1$. Hence, Theorem 2 shows that $\mathbb{A}$ has full row rank w.h.p. However, the moment formula (2.4) fails to attain its global maximum at the uniform solution; hence, the method of moments provably fails on this example, even though the 2-core $\mathbb{G}^{(2)}$ coincides with the entire original Tanner graph $\mathbb{G}$. Indeed, Figure 2 displays $\Phi(z)$ (middle) along with a numerical *lower* bound on the moment formula (right). The parameter on the horizontal axis of the right plot corresponds to the fraction variable occurrences set to zero. Hence, a necessary condition for the method of moments to succeed is that the maximum value be attained at $1/q$, which clearly is not the case.

2.2 Quenched analysis

Informed by this discussion, we are thus going to seize upon a different set of techniques to show that typical kernel vectors are essentially equitable. To be precise, let $\boldsymbol{x}_{\mathbb{A}} = (\boldsymbol{x}_{\mathbb{A},i})_{i\in[n]} \in \mathbb{F}_q^n$ be a random vector from the kernel of $\mathbb{A}$. We would like to show that for a given random matrix $\mathbb{A}$, such a random vector $\boldsymbol{x}_{\mathbb{A}} \in \ker \mathbb{A}$ is equitable w.h.p. In physics jargon, such a direct investigation of random solutions to a typical random combinatorial problem instance (in contrast to a moment calculation) is termed a *quenched analysis*. The fundamental merit of such a conditional (or quenched) analysis is that we may condition on the matrix $\mathbb{A}$ being *typical*; hence, we do not need to take very unlikely outcomes of $\mathbb{A}$ into consideration. By contrast, in the moment computations that we sketched in Section 2.1 we average over *all* possible outcomes of $\mathbb{A}$, including pathological cases that for some reason possess excessively large kernels.

The cornerstone of the quenched analysis will be to prove that w.h.p. over the choice of $\mathbb{A}$ the event

$$\mathfrak{O} = \left\{ \sum_{\sigma,\tau\in\mathbb{F}_q} \sum_{i,j=1}^{n} \left| \mathbb{P}\left[\boldsymbol{x}_{\mathbb{A},i} = \sigma,\ \boldsymbol{x}_{\mathbb{A},j} = \tau \mid \mathbb{A}\right] - q^{-2} \right| = o(n^2) \right\} \tag{2.7}$$

occurs. In words, $\mathfrak{O}$ asks that for any two field elements $\sigma, \tau \in \mathbb{F}_q$ for most pairs $1 \le i, j \le n$ the probability that the i-th entry $\boldsymbol{x}_{\mathbb{A},i}$ of a random kernel vector $\boldsymbol{x}_{\mathbb{A}}$ equals σ while the j-th entry $\boldsymbol{x}_{\mathbb{A},j}$ equals τ is about q^{-2}. Thus, for most choices of the indices i, j the pair $(\boldsymbol{x}_{\mathbb{A},i}, \boldsymbol{x}_{\mathbb{A},j}) \in \mathbb{F}_q^2$ is approximately uniformly distributed. Together with Chebyshev's inequality, this implies that a random vector $\boldsymbol{x}_{\mathbb{A}} \in \ker \mathbb{A}$ is equitable w.h.p. In fact, if $\mathfrak{O}$ occurs then even the degree-weighted empirical distribution of the entries of a typical $\boldsymbol{x}_{\mathbb{A}}$ is asymptotically uniform w.h.p., i.e., w.h.p. over the choice of $\boldsymbol{x}_{\mathbb{A}}$ we have

$$\sum_{i=1}^{n} \boldsymbol{d}_i \mathbb{1}\{\boldsymbol{x}_{\mathbb{A}} = \tau\} \sim q^{-1} \sum_{i=1}^{n} \boldsymbol{d}_i \qquad \text{for all } \tau \in \mathbb{F}_q. \tag{2.8}$$

Thus, the thrust behind considering the event $\mathfrak{O}$ is to accomplish just what we failed to accomplish via the moment computation: to show that the dominant contribution to the kernel comes from approximately equitable vectors.

Apart from showing that $\mathbb{A} \in \mathfrak{O}$ w.h.p., the following proposition also shows that the first moment formula (2.1) remains true on $\mathfrak{O}$.

▶ **Proposition 10.** *Under the assumptions of Theorem 2 we have* $\mathbb{P}[\mathbb{A} \in \mathfrak{O}] \sim 1$ *and*

$$\mathbb{E}_{\mathfrak{A}}\left[\boldsymbol{Z} \cdot \mathbb{1}\left\{\mathbb{A} \in \mathfrak{O}\right\}\right] \sim \mathbb{E}_{\mathfrak{A}}\left[\boldsymbol{Z}\right] \sim q^{n-\boldsymbol{m}}. \tag{2.9}$$

Before we elaborate on the proof of Proposition 10 in Section 2.3, we remark that the second moment method "works" once we condition on the event $\mathfrak{O}$. Indeed, the estimate (2.8), which is valid on $\mathfrak{O}$ w.h.p., demonstrates that once we condition on $\mathfrak{O}$, the dominant contribution to (2.3) comes from approximately uniform choices of $n_{\boldsymbol{d}}(s)$ as in (2.5). Due to the concavity of the entropy function, (2.5) implies that the optimal choices of the check variables $m_{\chi_1,\ldots,\chi_\ell}$ are asymptotically uniform as well, subject to the obvious linear constraint. Explicitly, the optimal $m_{\chi_1,\ldots,\chi_\ell}$ read

$$m_{\chi_1,\ldots,\chi_\ell}(s_1,\ldots,s_\ell) \sim \mathbb{1}\{s_1\chi_1+\cdots+s_\ell\chi_\ell=0\}q^{1-\ell}\boldsymbol{m}\mathbb{P}\left[\boldsymbol{k}=\ell\right]\prod_{i=1}^{\ell}\mathbb{P}\left[\boldsymbol{\chi}=\chi_i\right]. \tag{2.10}$$

Expanding (2.3) around (2.5) and (2.10), one could derive the bound $\mathbb{E}_{\mathfrak{A}}\left[\boldsymbol{Z}^2\cdot\mathbb{1}\left\{\mathbb{A}\in\mathfrak{O}\right\}\right] = O(\mathbb{E}_{\mathfrak{A}}\left[\boldsymbol{Z}\right])^2$ via a routine application of the Laplace method. However, to prove Theorem 2 we actually require the following more precise estimate.

▶ **Proposition 11.** *Under the assumptions of Theorem 2 we have*

$$\mathbb{E}_{\mathfrak{A}}\left[\boldsymbol{Z}^2\cdot\mathbb{1}\left\{\mathbb{A}\in\mathfrak{O}\right\}\right]\sim\mathbb{E}_{\mathfrak{A}}\left[\boldsymbol{Z}\right]^2. \tag{2.11}$$

The key challenge towards the proof the (2.11) is to obtain asymptotic equality, rather than the weaker bound $\mathbb{E}_{\mathfrak{A}}\left[\boldsymbol{Z}^2\cdot\mathbb{1}\left\{\mathbb{A}\in\mathfrak{O}\right\}\right] = O(\mathbb{E}_{\mathfrak{A}}\left[\boldsymbol{Z}\right]^2)$. This requires a meticulous expansion of the second moment around the equitable solution, which involves the detailed analysis of the lattices generated by integer vectors that encode conceivable values of the variables from (2.3). We are going to outline this analysis in Section 2.4. But first let us observe that Theorem 2 follows from Propositions 10 and 11 easily.

Proof of Theorem 2. The assumption (1.3) implies that $1-d/k=\Phi(0)>\Phi(1)=0$. Since $\boldsymbol{m}=\mathrm{Po}(dn/k)$, we thus obtain $n-\boldsymbol{m}=\Omega(n)$ w.h.p. Therefore, (2.9) implies that $\mathbb{E}_{\mathfrak{A}}\left[\boldsymbol{Z}\cdot\mathbb{1}\left\{\mathbb{A}\in\mathfrak{O}\right\}\right]\sim q^{n-\boldsymbol{m}}=q^{\Omega(n)}$ w.h.p. Hence, (2.11) implies together with Chebyshev's inequality that $\boldsymbol{Z}\geq\boldsymbol{Z}\mathbb{1}\{\mathbb{A}\in\mathfrak{O}\}=q^{\Omega(n)}$ w.h.p. Consequently, the random linear system $\mathbb{A}x=\boldsymbol{y}$ has a solution w.h.p., which implies that $\operatorname{rk}\mathbb{A}=\boldsymbol{m}$ w.h.p. ◀

2.3 Proof of Proposition 10: typical kernel vectors

The asymptotic rank formula (1.4) provides our point of departure toward the proof of Proposition 10. The basic idea is to show that (1.4) could not possibly be correct unless $\mathbb{A}\in\mathfrak{O}$ w.h.p. However, at closer inspection it turns out we cannot just apply (1.4) as is. Instead, we need to derive the analogue of (1.4) for a slightly enhanced random matrix from scratch.

Specifically, for an integer $t\geq 0$ obtain $\mathbb{A}_{[t]}$ from $\mathbb{A}$ by adding t more rows that each contain precisely three non-zero entries. The positions of these non-zero entries are chosen uniformly, mutually independently and independently of $\mathbb{A}$. The non-zero entries themselves are independent copies of $\boldsymbol{\chi}$. For this enhanced matrix we derive the following upper bound on its asymptotic rank.

▶ **Proposition 12.** *If* (1.3) *is satisfied then there exists* $\delta_0=\delta_0(\boldsymbol{d},\boldsymbol{k})>0$ *such that for all* $0<\delta<\delta_0$ *we have*

$$\limsup_{n\to\infty}\frac{1}{n}\mathbb{E}[\operatorname{nul}\mathbb{A}_{[\lfloor\delta n\rfloor]}]\leq 1-\frac{d}{k}-\delta. \tag{2.12}$$

The proof of Proposition 12 relies on the so-called "Aizenman-Sims-Starr scheme" [4], a coupling argument inspired by spin glass theory that also constituted the cornerstone of the derivation of (1.4) in [9]. That said, a subtle modification of this argument is necessary to accommodate the additional ternary equations. A vital assumption towards the proof of Proposition 12 is that the function Φ from (1.2) attains its *unique* global max at $z = 0$. In fact, the proof of Proposition 12 is the only place where the uniqueness of the maximiser is required.

How does Proposition 10 follow from Proposition 12? Assuming (1.3), we obtain from (1.4) that

$$\frac{1}{n}\operatorname{nul}\mathbb{A} \sim 1 - \frac{d}{k} \qquad \text{w.h.p.} \tag{2.13}$$

Now suppose that we add $\lfloor \delta n \rfloor$ extra ternary rows to $\mathbb{A}$ to obtain $\mathbb{A}_{[\lfloor \delta n \rfloor]}$. Comparing (2.12) and (2.13), we conclude that all but $o(n)$ of these extra rows decrease the nullity by one. Indeed, adding a single row cannot decrease the nullity by more than one, and routine arguments show that $\operatorname{nul}\mathbb{A}_{[\lfloor \delta n \rfloor]}$ concentrates about its expectation.

But a drop in nullity of $\delta n + o(n)$ w.h.p. is conceivable only if $\mathbb{A} \in \mathfrak{O}$ w.h.p. To see this, let us contemplate the kernel of a general $M \times N$ matrix A over $\mathbb{F}_q$ for a brief moment. Draw $\boldsymbol{x}_A = (\boldsymbol{x}_{A,i})_{i\in[N]} \in \ker A$ uniformly at random. For any given coordinate $\boldsymbol{x}_{A,i}$, $i \in [N]$ there are two possible scenarios: either $\boldsymbol{x}_{A,i} = 0$ with probability one, or $\boldsymbol{x}_{A,i}$ is uniformly distributed over $\mathbb{F}_q$. To see this, consider a basis $\zeta_1, \ldots, \zeta_h$ of the kernel of A. Then we can sample $\boldsymbol{x}_A$ by just multiplying each ζ_j with a random scalar $\boldsymbol{z}_j \in \mathbb{F}_q$ and summing up: $\boldsymbol{x}_A = \boldsymbol{z}_1\zeta_1 + \cdots + \boldsymbol{z}_h\zeta_h$. If the i-th coordinate of all ζ_j is zero, then $\boldsymbol{x}_{A,i} = 0$ deterministically; otherwise $\boldsymbol{x}_{A,i}$ is a sum of uniformly random elements of $\mathbb{F}_q$, and thus uniformly random itself. It therefore makes sense to call coordinate i *frozen* if $x_i = 0$ for all $x \in \ker A$, and unfrozen otherwise. Let $\mathfrak{F}(A)$ be the set of frozen coordinates.

If $\mathbb{A}$ had many frozen coordinates then adding an extra random row with three non-zero entries could hardly decrease the nullity w.h.p. For if all three non-zero coordinates fall into the frozen set, then we get the new equation "for free", i.e., $\operatorname{nul}\mathbb{A}_{[1]} = \operatorname{nul}\mathbb{A}$. Thus, Proposition 12 implies that $|\mathfrak{F}(\mathbb{A})| = o(n)$ w.h.p. We conclude that $\boldsymbol{x}_{\mathbb{A},i}$ is uniformly distributed over $\mathbb{F}_q$ for all but $o(n)$ coordinates $i \in [n]$. However, this does not yet imply that $\boldsymbol{x}_{\mathbb{A},i}$, $\boldsymbol{x}_{\mathbb{A},j}$ are independent for most i, j, as required by $\mathfrak{O}$. Yet a slightly more careful deliberation based on linear algebra and the "pinning lemma" [9, Proposition 2.4] shows that $\mathbb{A} \in \mathfrak{O}$ w.h.p.

2.4 Proof of Proposition 11: expansion around the equitable solution

We prove Proposition 11 by way of expanding (2.3) carefully around the uniform distribution (2.5). Recall that once the $n_\ell(s)$ are set to the equitable solution (2.5), the optimal check variables $m_{\chi_1,\ldots,\chi_\ell}(s_1, \ldots, s_\ell)$ are given by (2.10). This observation by itself now suffices to conclude without (much) further ado that

$$\mathbb{E}_{\mathfrak{A}}[\boldsymbol{Z}^2 \cdot \mathbb{1}\{\mathbb{A} \in \mathfrak{O}\}] = O\left(\mathbb{E}_{\mathfrak{A}}[\boldsymbol{Z} \cdot \mathbb{1}\{\mathbb{A} \in \mathfrak{O}\}]^2\right). \tag{2.14}$$

The challenge is to sharpen this estimate so as to obtain the asymptotic equality claimed in (2.11). In his work on adjacency matrices of random regular graphs, Huang [22] actually faced a similar issue (with $\boldsymbol{d} = \boldsymbol{k}$ constant and q a prime number). To prove Proposition 11 we need to cope with the (significantly) more general situation of arbitrary $\boldsymbol{d}, \boldsymbol{k}$ and prime powers q. This improvement actually constitutes one of the main technical obstacles that we need to surmount toward the proof of Theorem 2.

The issue is that in order to eliminate the constant factor hidden in the $O(\cdot)$ in (2.14) we need to carefully consider divisibility properties that make it possible or impossible for a vector $x \in \mathbb{F}_q^n$ to belong to the kernel. These questions depend not only on the degree distributions $\boldsymbol{d}, \boldsymbol{k}$ but also on q and the distribution $\boldsymbol{\chi}$ of the non-zero entries. Hence, to estimate the kernel size precisely we need to crystallise the conceivable frequencies of field elements that may lead to solutions. Specifically, for an integer $\ell \geq 3$ and $\chi_1, \ldots, \chi_\ell \in \mathbb{F}_q \setminus \{0\}$ let

$$\mathcal{S}_q(\chi_1, \ldots, \chi_\ell) = \left\{ \sigma \in \mathbb{F}_q^\ell : \sum_{i=1}^{\ell} \chi_i \sigma_i = 0 \right\} \tag{2.15}$$

comprise all solutions to a linear equation with coefficients $\chi_1, \ldots, \chi_{k_0} \in \mathbb{F}_q$. Furthermore, for each $\sigma \in \mathcal{S}_q(\chi_1, \ldots, \chi_\ell)$ we define the vector

$$\hat{\sigma} = \left(\sum_{i=1}^{\ell} \mathbb{1}\{\sigma_i = s\} \right)_{s \in \mathbb{F}_q \setminus \{0\}} \in \mathbb{Z}^{\mathbb{F}_q \setminus \{0\}} \tag{2.16}$$

to track the frequencies with which the various non-zero field elements appear. Moreover, let

$$\mathfrak{M}_q(\chi_1, \ldots, \chi_\ell) \subseteq \mathbb{Z}^{\mathbb{F}_q \setminus \{0\}}$$

be the $\mathbb{Z}$-module generated by the frequency vectors $\hat{\sigma}$ for $\sigma \in \mathcal{S}_q(\chi_1, \ldots, \chi_\ell)$. Thus, $\mathfrak{M}_q(\chi_1, \ldots, \chi_\ell) \subseteq \mathbb{Z}^{\mathbb{F}_q \setminus \{0\}}$ captures all conceivable frequency vectors of solutions σ to $\sum_{i=1}^{\ell} \chi_i \sigma_i$.

Depending on the coefficients $\chi_1, \ldots, \chi_\ell$, the module $\mathfrak{M}_q(\chi_1, \ldots, \chi_\ell)$ may be a proper sub-module of the integer lattice $\mathbb{Z}^{\mathbb{F}_q \setminus \{0\}}$. For example, in the case $q = \ell = 3$ and $\chi_1 = \chi_2 = \chi_3 = 1$ the module $\mathfrak{M}_3(1,1,1)$ constitutes the sub-lattice spanned by $\binom{1}{1}$ and $\binom{0}{3}$, which is a proper sub-lattice of $\mathbb{Z}^2$. The following proposition characterises the lattice spanned by the frequency vectors for general $\chi_1, \ldots, \chi_\ell$. The determinant formula that the proposition provides shows that $\mathfrak{M}_q(\chi_1, \ldots, \chi_\ell)$ is a proper sub-module iff all the coefficients $\chi_1, \ldots, \chi_\ell$ coincide.

▶ **Proposition 13.** *Let $q \geq 2$ be a prime power, $\ell \geq 3$ and let $\chi_1, \ldots, \chi_\ell \in \mathbb{F}_q \setminus \{0\}$. Then $\mathfrak{M}_q(\chi_1, \ldots, \chi_\ell)$ has a basis $\mathfrak{b}_1, \ldots, \mathfrak{b}_{q-1}$ of non-negative integer vectors with $\|\mathfrak{b}_i\|_1 \leq 3$ for all $1 \leq i \leq q-1$ such that*

$$\det(\mathfrak{b}_1 \ \cdots \ \mathfrak{b}_{q-1}) = q^{\mathbb{1}\{\chi_1 = \cdots = \chi_\ell\}}.$$

A vital feature of Proposition 13 is that the module basis consists of non-negative integer vectors with small ℓ_1-norm. In effect, the basis vectors are "combinatorially meaningful" towards our purpose of counting solutions. Perhaps surprisingly, the proof of Proposition 13 turns out to be rather delicate, with details depending on whether q is a prime or a prime power, among other things.

In addition to the subgrid constraints imposed by the linear equations themselves, we need to take another divisibility condition into account. Indeed, for any assignment $\sigma \in \mathbb{F}_q^n$ of values to variables the frequencies of the various field elements $s \in \mathbb{F}_q$ are divisible by the g.c.d. $\mathfrak{d}$ of $\mathrm{supp}(\boldsymbol{d})$, i.e.

$$\mathfrak{d} \mid \sum_{i=1}^{n} \boldsymbol{d}_i \mathbb{1}\{\sigma_i = s\} \qquad \text{for all } s \in \mathbb{F}_q. \tag{2.17}$$

Thus, to compute the expected kernel size we need to study the intersection of the subgrid (2.17) with the grid spanned by the frequency vectors $\hat{\sigma}$ for $\sigma \in \mathcal{S}_q(\boldsymbol{\chi}_{1,1}, \ldots, \boldsymbol{\chi}_{1,\boldsymbol{k}})$. Specifically, in order to derive Proposition 11 from Proposition 13 we need to estimate the number of vectors $\sigma \in \mathbb{F}_q^n$ represented by each grid point and calculate the ensuing satisfiability probability. This argument combines the Laplace method with local limit techniques.

3 Discussion

While there is a substantial body of work on dense random matrices where the average number of non-zero entries per row/column diverges or even is linear in the size of the matrix (e.g., [6, 7, 14, 15, 23, 24, 33, 34]), far less is known about sparse random matrices. The aim of this paper has been to determine sufficient (as well as necessary) conditions for a sparse random matrix to have *full* row rank. To this end we drew upon some of the elements of prior work on the *asymptotic* rank of random matrices [5, 9], specifically the formula (1.4). In particular, the proof of Proposition 12 adapts and extends the Aizenman-Sims-Starr scheme from [9]. Additionally, the expansion around the centre employs some of the techniques developed in the study of satisfiability thresholds, particularly the extensive use of local limit theorems [12, 11]. These also played a role in prior work on the adjacency matrices of random d-regular graphs [22, 29].

A principal new proof ingredient is the asymptotically precise analysis of the moment formula (2.3) for general $\boldsymbol{d}, \boldsymbol{k}, q$ around the equitable solution by means of the study of the sub-grids of the integer lattice induced by the constraints. This issue that was absent in the prior literature on variations on random k-XORSAT [5, 9, 13] and on other random constraint satisfaction problems [12, 11]. That said, in the study of the random regular matrix from Example 6 Huang [22] faced a similar issue in the special case $\boldsymbol{d} = \boldsymbol{k}$ constant and $\boldsymbol{\chi} = 1$ deterministically. Proposition 13, whose proof is based on a combinatorial investigation of lattices in the general case, constitutes a considerable generalisation of this case. A further new feature of the proof of Proposition 13 is the explicit ℓ_1-bound on the basis vectors, which greatly facilitates the proof of Theorem 2.

Satisfiability thresholds of random constraint satisfaction problems have been studied extensively in the statistical physics literature via a non-rigorous technique called the "cavity method". The cavity method comes in two installments: the simpler "replica symmetric ansatz" associated with the Belief Propagation message passing scheme, and the more intricate "replica symmetry breaking ansatz". The proof of Theorem 2 demonstrates that the former renders the correct prediction as to the satisfiability threshold of random linear equations. By contrast, in quite a few problems, notoriously random k-SAT, replica symmetry breaking occurs [10, 17], requiring a substantially different proof strategy.

A natural question is whether the methods presented in this work can be extended to the adjacency matrices of random graphs. Apart from the aforementioned works regarding the regular case [22, 29] and the work of Bordenave, Lelarge and Salez [8], an exciting recent contribution by Glasgow, Kwan, Sah and Sawhney deals with the precise connection between the matching number and the rank [20]. By contrast to the present work, these contributions rely on local weak convergence and/or Littlewood-Offord techniques; see also [19]. Furthermore, recently the methods from [9] were extended to obtain a rank formula for the adjacency matrices of Erdös-Rényi graphs over arbitrary fields [35]. In fact, the consideration of general fields reveals new phenomena, as was already discovered in some of the earlier literature [6, 7, 25, 26, 27, 28].

References

1 D. Achlioptas and F. McSherry. Fast computation of low-rank matrix approximations. *J. ACM*, 54(2):9, 2007.

2 D. Achlioptas and C. Moore. Random k-sat: Two moments suffice to cross a sharp threshold. *SIAM J. Comput.*, 36(3):740–762, 2006.

3 D. Achlioptas, A. Naor, and Y. Peres. Rigorous location of phase transitions in hard optimization problems. *Nature*, 435(7043):759–764, June 2005.

4 M. Aizenman, R. Sims, and S. L. Starr. Extended variational principle for the sherrington-kirkpatrick spin-glass model. *Phys. Rev. B*, 68:214403, 2003.

5 P. J. Ayre, A. Coja-Oghlan, P. Gao, and N. Müller. The satisfiability threshold for random linear equations. *Comb.*, 40(2):179–235, 2020.

6 G. V. Balakin. The distribution of the rank of random matrices over a finite field. *Theory of Probability & Its Applications*, 13(4):594–605, 1968.

7 J. Blömer, R. Karp, and E. Welzl. The rank of sparse random matrices over finite fields. *Random Struct. Algorithms*, 10(4):407–419, 1997.

8 C. Bordenave, M. Lelarge, and J. Salez. The rank of diluted random graphs. *The Annals of Probability*, 39(3):1097–1121, 2011.

9 A. Coja-Oghlan, A. A. Ergür, P. Gao, S. Hetterich, and Rolvien M. The rank of sparse random matrices. *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020*, pages 579–591, 2020.

10 A. Coja-Oghlan, N. Müller, and J.B. Ravelomanana. Belief propagation on the random k-sat model. *CoRR*, abs/2011.02303, 2020. `arXiv:2011.02303`.

11 A. Coja-Oghlan and K. Panagiotou. Catching the k-naesat threshold. *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 899–908, 2012.

12 A. Coja-Oghlan and K. Panagiotou. The asymptotic k-sat threshold. *Advances in Mathematics*, 288:985–1068, 2016.

13 C. Cooper, A. M. Frieze, and W. Pegden. On the rank of a random binary matrix. *Electron. J. Comb.*, 26(4):4, 2019.

14 K. P. Costello and V. H. Vu. The rank of random graphs. *Random Struct. Algorithms*, 33(3):269–285, 2008.

15 K. P. Costello and V. H. Vu. On the rank of random sparse matrices. *Comb. Probab. Comput.*, 19(3):321–342, 2010.

16 M. Dietzfelbinger, A. Goerdt, M. Mitzenmacher, A. Montanari, R. Pagh, and M. Rink. Tight thresholds for cuckoo hashing via XORSAT. *Proceedings of the 37th International Colloquium on Automata, Languages and Programming (ICALP 2010)*, 6198:213–225, 2010.

17 J. Ding, A. Sly, and N. Sun. Proof of the satisfiability conjecture for large k. *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing (STOC 2015)*, pages 59–68, 2015.

18 O. Dubois and J. Mandler. The 3-xorsat threshold. *43rd Symposium on Foundations of Computer Science (FOCS 2002)*, pages 769–778, 2002.

19 A. Ferber, M. Kwan, A. Sah, and M. Sawhney. Singularity of the k-core of a random graph. *Duke Mathematical Journal*, 172(7):1293–1332, 2023.

20 M. Glasgow, M. Kwan, A. Sah, and M. Sawhney. The exact rank of sparse random graphs. *arXiv preprint*, 2023. `arXiv:2303.05435`.

21 A. Goerdt and L. Falke. Satisfiability thresholds beyond k- xorsat. *Proceedings of the 7th International Computer Science Symposium in Russia (CSR 2012)*, pages 148–159, 2012.

22 J. Huang. Invertibility of adjacency matrices for random d-regular graphs. *Duke Mathematical Journal*, 170(18):3977–4032, 2021.

23 J. Kahn, J. Komlós, and E. Szemerédi. On the probability that a random±1-matrix is singular. *Journal of the American Mathematical Society*, 8(1):223–240, 1995.

24 J. Komlós. On the determinant of (0-1) matrices. *Studia Scientiarium Mathematicarum Hungarica*, 2:7–21, 1967.

25 I. Kovalenko. On the limit distribution of the number of solutions of a random system of linear equations in the class of boolean functions. *Theory of Probability & Its Applications*, 12(1):47–56, 1967.

26 I. Kovalenko, A.A. Levitskaya, and MN Savchuk. Selected problems in probabilistic combinatorics. *Naukova Dumka, Kiev*, 1986.

27 A. A. Levitskaya. Invariance theorems for a system of random linear equations over an arbitrary finite ring. *Doklady Akademii Nauk*, 263(2):289–291, 1982.

28 A. A. Levitskaya. The probability of consistency of a system of random linear equations over an arbitrary finite ring. *Theory of Probability & Its Applications*, 30(2):364–375, 1986.

29 A. Mészáros. The distribution of sandpile groups of random regular graphs. *Transactions of the American Mathematical Society*, 373(9):6529–6594, 2020.

30 G. Miller and G. D. Cohen. The rate of regular LDPC codes. *IEEE Trans. Inf. Theory*, 49(11):2989–2992, 2003.

31 B. Pittel and G. Sorkin. The satisfiability threshold for k-xorsat. *Combinatorics, Probability and Computing*, 25(2):236–268, 2016.

32 T. Richardson and R. Urbanke. *Modern Coding Theory*. Cambridge University Press, 2008.

33 T. Tao and V. H. Vu. On the singularity probability of random bernoulli matrices. *Journal of the American Mathematical Society*, 20, February 2005.

34 K. Tikhomirov. Singularity of random bernoulli matrices. *Annals of Mathematics*, 191:593, March 2020.

35 R. van der Hofstad, N. Müller, and H. Zhu. The rank of sparse symmetric matrices over arbitrary fields. *arXiv preprint*, 2023. `arXiv:2301.12978`.

36 M. J. Wainwright, E. N. Maneva, and E. Martinian. Lossy source compression using low-density generator matrix codes: analysis and algorithms. *IEEE Trans. Inf. Theory*, 56(3):1351–1368, 2010.