An algebraic inverse theorem for the quadratic Littlewood–Offord problem, and an application to Ramsey graphs

Matthew Kwan * Lisa Sauermann†

Abstract

Consider a quadratic polynomial $f(\xi_1, \dots, \xi_n)$ of independent Bernoulli random variables. What can be said about the concentration of f on any single value? This generalises the classical Littlewood–Offord problem, which asks the same question for linear polynomials. As in the linear case, it is known that the point probabilities of f can be as large as about $1/\sqrt{n}$, but still poorly understood is the "inverse" question of characterising the algebraic and arithmetic features f must have if it has point probabilities comparable to this bound. In this paper we prove some results of an algebraic flavour, showing that if f has point probabilities much larger than $1/n$ then it must be close to a quadratic form with low rank. We also give an application to Ramsey graphs, asymptotically answering a question of Kwan, Sudakov and Tran.

1 Introduction

Consider a random variable of the form $X = a_1\xi_1 + \cdots + a_n\xi_n$, where $(a_1, \dots, a_n) \in \mathbb{R}^n$ is a sequence of real numbers and $\xi = (\xi_1, \dots, \xi_n) \sim \mathrm{Rad}^n$ is a sequence of independent Rademacher random variables (meaning that $\Pr(\xi_i = 1) = \Pr(\xi_i = -1) = 1/2$ for each i). Broadly speaking, the classical Littlewood–Offord problem asks for *anti-concentration* estimates for random variables of this type: what can we say about the maximum probability that X is equal to a single value, or falls in an interval of prescribed length?

In connection with their work on random polynomials, Littlewood and Offord [29] first proved that if each $|a_i| \geq 1$, then the small-ball probabilities $\Pr(|X - x| \leq 1)$, for $x \in \mathbb{R}$, are bounded by $O(\log n/\sqrt{n})$. (here, and for the rest of the paper, all asymptotics are as $n \to \infty$). Erdős [13] later proved the optimal upper bound $\binom{n}{\lfloor n/2 \rfloor}/2^n = O(1/\sqrt{n})$, and further work by Halász [24], Tao and Vu [43, 45], Rudelson and Vershynin [40] and Nguyen and Vu [34] explored the relationship between the concentration behaviour of X and the arithmetic structure of the coefficients $(a_1, \dots, a_n)$. This work has had far-reaching consequences: in particular, these Littlewood–Offord-type theorems were essential tools in some of the landmark results in random matrix theory (see for example [42, 45, 47]).

Observe that $a_1\xi_1 + \cdots + a_n\xi_n$ is a linear polynomial in $\xi = (\xi_1, \dots, \xi_n)$, so a natural variation on the Littlewood–Offord problem is to consider *quadratic* polynomials in ξ. This direction of research was popularised by Costello, Tao and Vu [12] in connection with their proof of Weiss' conjecture that a random symmetric ± 1 matrix typically has full rank, and was further explored by Costello [11] and Nguyen [35] (higher-degree polynomials were also considered by Rosiński and Samorodnitsky [39], Razborov and Viola [38], Meka, Nguyen and Vu [32], and Fox, Kwan and Sauermann [21]). Specifically, Costello [11] proved that if f is a quadratic polynomial in n variables with $\Theta(n^2)$ nonzero coefficients, then $\Pr(f(\xi) = x) \leq n^{o(1)-1/2}$.

The exponent of $1/2$ in Costello's result is best-possible, as can be seen by considering the polynomial $(\xi_1 + \cdots + \xi_n)^2$. However, it seems that for a "typical" quadratic polynomial f we should expect a much stronger bound. For example, if all the coefficients of f are integers of bounded size (and $\Theta(n^2)$ of them are non-zero), then the standard deviation of $f(\xi)$ is of order $\Theta(n)$. In this case it seems reasonable to assume that "typically" the probability mass is roughly evenly distributed over the integer points

*Department of Mathematics, Stanford University, Stanford, CA 94305. Email: mattkwan@stanford.edu. Research supported in part by SNSF project 178493.

†Department of Mathematics, Stanford University, Stanford, CA 94305. Email: lsauerma@stanford.edu.

in a standard-deviation-sized interval around the mean, yielding a bound of about $1/n$ for the point probabilities. Costello made a conjecture (see [11, Conjecture 3]) to this effect, suggesting that the only way $f(\xi)$ can have point probabilities greater than $n^{\varepsilon-1}$, for any constant $\varepsilon > 0$, is if f "differs in only a few coefficients" from a polynomial which splits into two linear factors. We remark that Costello's paper was about polynomials with *complex* coefficients, and splitting over $\mathbb{C}$ is a weaker property than splitting over $\mathbb{R}$. Nevertheless, Costello's conjecture appears to be plausible over both $\mathbb{R}$ and $\mathbb{C}$.

To put Costello's conjecture in a wider context, an important goal for Littlewood–Offord-type problems is to prove *inverse theorems*: in addition to proving general bounds on the maximum point probability, we would also like to understand the structural features exhibited whenever the random variable has a point probability close to this maximum. In the linear case, as previously mentioned, the point probabilities are only affected by the arithmetic structure of the multiset of coefficients $(a_1, \dots, a_n)$, and state-of-the-art inverse theorems due to Rudelson and Vershynin [40] and Nguyen and Vu [34] give a very refined understanding of the way this arithmetic structure is influenced by the maximum point probability (these results build on an earlier, coarser, inverse theorem due to Tao and Vu [45]). However, in the quadratic case the point probabilities are influenced not only by the values of the coefficients, but also by how the different coefficients are arranged in the polynomial (for example, even the case where all the coefficients lie in $\{0, 1\}$ is not well understood). Nguyen [35] proved a coarse inverse theorem (whose exact statement is too technical to reproduce here) showing that if, for a quadratic polynomial f, the maximum point probability of $f(\xi)$ is only polynomially small (that is, $\Pr(f(\xi) = x) \geq n^{-O(1)}$ for some x), then f enjoys some algebraic and arithmetic structure. One can interpret Costello's conjecture as asking for a much more refined inverse theorem, albeit one that only takes algebraic structure into account.

In this paper, we prove some inverse theorems of a similar flavour to Costello's conjecture, giving a connection between anti-concentration of $f(\xi)$ and algebraic properties of f. Roughly speaking, we prove that if f has concentration probability much larger than $1/n$ then it must be close to a quadratic form with low rank[1]. Our first result is in terms of "coefficient L_1 distance".

Theorem 1.1. *Let $\mathbb{F} \in \{\mathbb{C}, \mathbb{R}, \mathbb{Q}\}$. For any integer $r \geq 3$, and any $0 < \varepsilon \leq 1$, there is a constant $C = C(r, \varepsilon)$ such that the following holds. Let $f \in \mathbb{F}[x_1, \dots, x_n]$ be a quadratic polynomial all of whose coefficients have absolute value at most 1. Let $\xi = (\xi_1, \dots, \xi_n) \in \mathrm{Rad}^n$, and suppose that we have*

$$\sup_{x \in \mathbb{F}} \Pr(f(\xi) = x) \geq C \cdot \frac{(\log n)^{r/2}}{n^{1-2/(r+2)}}.$$

Then there is a quadratic form $h \in \mathbb{F}[x_1, \dots, x_n]$ of rank strictly less than r such that the sum of the absolute values of the coefficients of $f - h$ is at most εn^2.

Note that if r is large, then the bound on the point probabilities in Theorem 1.1 is close to $1/n$. Also, note that we can rescale any quadratic polynomial $f \in \mathbb{F}[x_1, \dots, x_n]$ so that all of its coefficients have absolute value at most 1, so Theorem 1.1 can be interpreted as giving a bound in terms of the largest coefficient of f. We remark that for the linear Littlewood–Offord problem, it essentially suffices to consider the case where the coefficients are of bounded size, because if there are many coefficients with dramatically different orders of magnitude, the point probabilities are small for trivial reasons (Littlewood and Offord's original work [29] proceeded along these lines). Unfortunately we were not able to find such a reduction for the quadratic Littlewood–Offord problem.

In certain combinatorial applications, the coefficients of f are integers of bounded size, or lie in some other bounded set of "allowed coefficients". For such polynomials, our next result gives a bound analogous to Theorem 1.1 where the quadratic form h differs in only few coefficients from f (so this version is in terms of "coefficient Hamming distance" as in Costello's conjecture, instead of "coefficient L_1 distance" as in Theorem 1.1).

Theorem 1.2. *Let $\mathbb{F} \in \{\mathbb{C}, \mathbb{R}, \mathbb{Q}\}$. For any integer $r \geq 3$, any $0 < \varepsilon \leq 1$, and any finite set $S \subseteq \mathbb{F}$, there is a constant $C = C(r, \varepsilon, S)$ such that the following holds. Let $f \in \mathbb{F}[x_1, \dots, x_n]$ be a quadratic polynomial*

[1]Recall that an n-variable *quadratic form* over a field $\mathbb{F}$ is a homogeneous quadratic polynomial $h \in \mathbb{F}[x_1, \dots, x_n]$. If $\mathbb{F}$ has characteristic not equal to 2, there is a unique representation $h(x) = x^T Q x$ with a symmetric matrix $Q \in \mathbb{F}^{n \times n}$ (where x denotes the column vector with entries $x_1, \dots, x_n$). The *rank* of h is defined to be the rank of this matrix Q. Equivalently, the rank of h is the minimum r such that there is a representation $h = \lambda_1 h_1^2 + \cdots + \lambda_r h_r^2$ as a linear combination of squares of homogeneous linear polynomials $h_1, \dots, h_r \in \mathbb{F}[x_1, \dots, x_n]$.

all of whose degree-2 coefficients are elements of the set S. Let $\xi = (\xi_1, \dots, \xi_n) \in \mathrm{Rad}^n$, and suppose that we have

$$\sup_{x \in \mathbb{F}} \Pr(f(\xi) = x) \geq C \cdot \frac{(\log n)^{r/2}}{n^{1-2/(r+2)}}.$$

Then there is a quadratic form $h \in \mathbb{F}[x_1, \dots, x_n]$ of rank strictly less than r such that f and h differ in at most εn^2 coefficients.

We remark that Theorems 1.1 and 1.2 can in fact be used to give anti-concentration estimates substantially stronger than $1/\sqrt{n}$ for polynomials that are not close to factorising over $\mathbb{C}$, as in Costello's conjecture. If a complex quadratic form has rank at most 2, then it is a sum of two squares of linear forms. Over $\mathbb{C}$, such forms always split into linear factors. Therefore, applying Theorems 1.1 and 1.2 with $r = 3$, we see that if the point probabilities of $f(\xi)$ are much larger than $n^{-3/5}$, then there is a quadratic form h, close to f, which splits into linear factors over the complex numbers.

The proofs of Theorems 1.1 and 1.2 involve a number of ideas and ingredients that may be independently interesting. In Subsection 1.2 we will outline the proofs and discuss these ideas, but first we describe an application of Theorem 1.2 to anti-concentration of edge-statistics in Ramsey graphs, asymptotically answering a question of Kwan, Sudakov and Tran [27].

1.1 Ramsey graphs

An induced subgraph of a graph is said to be *homogeneous* if it is a clique or independent set. A classical result in Ramsey theory, proved in 1935 by Erdős and Szekeres [18], is that every n-vertex graph has a homogeneous subgraph with at least $\frac{1}{2}\log_2 n$ vertices. On the other hand, Erdős [14] famously used the probabilistic method to prove that, for all n, there exists an n-vertex graph with no homogeneous subgraph on $2\log_2 n$ vertices. Despite significant effort (see for example [5, 8, 9, 23, 28]), there are no known non-probabilistic constructions of graphs whose largest homogeneous subgraphs are of a comparable size.

Say an n-vertex graph is *C-Ramsey* if it has no homogeneous subgraph of size $C\log_2 n$. It is widely believed that for any fixed constant C all C-Ramsey graphs must in some sense resemble random graphs, and this belief has been supported by a number of theorems showing that certain "richness" properties characteristic of random graphs hold for all C-Ramsey graphs. The first result of this type was due to Erdős and Szemerédi [19], who showed that every C-Ramsey graph G has edge-density bounded away from zero and one. Note that this implies fairly strong information about the edge distribution on induced subgraphs of G, because any n^α-vertex induced subgraph of an n-vertex C-Ramsey graph is itself (C/α)-Ramsey.

This basic result was the foundation for a large amount of further research on Ramsey graphs; over the years many conjectures have been proposed and resolved (see [1, 3, 4, 7, 15, 16, 17, 25, 26, 33, 37, 41]). In particular, we mention two results regarding the edge distribution on induced subgraphs. First, solving a conjecture of Narayanan, Sahasrabudhe and Tomon [33] (inspired by an old question of Erdős and McKay [15, 16]), Kwan and Sudakov [26] proved that for any n-vertex C-Ramsey graph there are induced subgraphs with $\Omega(n^2)$ different numbers of edges. Second, resolving a conjecture of Erdős, Faudree and Sós [15, 16] (improving results of Alon and Kostochka [3] and Alon, Balogh, Kostochka and Samotij [1]), Kwan and Sudakov [25] also proved that every n-vertex C-Ramsey graph has the property that for $\Omega(n)$ of the choices $\ell \in \{0, \dots, n\}$, there are ℓ-vertex induced subgraphs with $\Omega(n^{3/2})$ different numbers of edges.

The aforementioned Erdős–Szemerédi theorem can be interpreted as a (weak) "concentration" theorem: the numbers of edges in induced subgraphs cannot be "too extreme". On the other hand, the two results in the last paragraph point in the opposite direction: there are many different possibilities for the numbers of edges in induced subgraphs. In connection with some recent work on anti-concentration of edge-statistics (see [2, 22, 27, 30]), Kwan, Sudakov and Tran [27] asked about anti-concentration of the edge distribution in Ramsey graphs. Specifically, for an n-vertex C-Ramsey graph, let X be the number of edges induced by a uniformly random set of (say) $n/2$ vertices. Is it true that $\Pr(X = x) = O(1/n)$ for all $x \in \mathbb{N}$? If true, this would be best-possible, as can be seen by considering a random graph $\mathbb{G}(n, 1/2)$. One of the motivations for this question was that better understanding of edge-statistics in Ramsey graphs could lead to a unified and more conceptual proof of the conjectures of Narayanan–Sahasrabudhe–Tomon and Erdős–Faudree–Sós concerning induced subgraphs of Ramsey graphs with different numbers of edges. We discuss this further in Section 8.

As an application of Theorem 1.2, we answer Kwan, Sudakov and Tran's question asymptotically. Roughly speaking, we express X as a quadratic polynomial and show that Ramsey graphs are too disordered for this polynomial to be close to a low-rank quadratic form.

Theorem 1.3. *The following holds for any fixed constants $C, c > 0$. Let G be an n-vertex C-Ramsey graph, and, for some $cn \leq k \leq (1-c)n$, let X be the number of edges induced by a uniformly random subset of k vertices of G. Then for any $x \in \mathbb{Z}$, we have*

$$\Pr(X = x) \leq n^{o(1)-1}.$$

In Section 8 we discuss a further related conjecture, and some connections between this line of research and some older conjectures about Ramsey graphs.

1.2 Outline of the paper and the proofs

The structure of the paper is as follows. First, in Section 2 we present the deduction of Theorem 1.3 (our result about Ramsey graphs) from Theorem 1.2. This illustrates some ideas that might be more generally useful in other applications of Theorems 1.1 and 1.2. Via a coupling trick (Lemma 2.3), our random variable X can be represented as a certain quadratic polynomial. To apply Theorem 1.2 we then need to show that a certain $n \times n$ matrix M corresponding to this quadratic polynomial (defined in terms of a Ramsey graph) is far from being low-rank, in the sense that for any fixed $r \in \mathbb{N}$, changing any $o(n^2)$ entries of M results in a matrix with rank at least r. We observe that, to this end, it suffices to show that for any $r \in \mathbb{N}$, our matrix M contains $\Omega(n^{2r})$ invertible $r \times r$ submatrices. This will follow from a simple generalisation (Lemma 2.2) of an old result due to Erdős and Hajnal: for fixed h and C, every C-Ramsey graph contains $\Omega(n^h)$ copies of every possible h-vertex induced subgraph.

Next, in Section 3 we state and prove an anti-concentration inequality for real quadratic polynomials satisfying a certain technical non-degeneracy condition (Lemma 3.2). This will be a key ingredient in the proofs of Theorems 1.1 and 1.2. The main idea that allows us to prove bounds stronger than $1/\sqrt{n}$ is a decoupling trick applied to the characteristic function (Fourier transform) of our random variable of interest. In circumstances where the characteristic function decays in a "Gaussian-like" way, this allows us to reduce from the quadratic case to the linear case without incurring the square-root loss that is usually associated with decoupling tricks of this type.

In Section 4 we state and prove a lemma concerning "real projections" of complex matrices (Lemma 4.1), which essentially allows us to deduce the complex cases of Theorems 1.1 and 1.2 from the real cases. This lemma may also be of independent interest. To illustrate, a special case is the fact that for any nonsingular complex matrix A, there is a phase $\theta \in [-\pi, \pi]$ such that $\Re(e^{i\theta}A)$ is nonsingular.

Then, in Section 5 we outline how to deduce Theorems 1.1 and 1.2 from Lemmas 3.2 and 4.1. The main step of the deduction is to show that quadratic polynomials which are far from a low-rank quadratic form satisfy the technical non-degeneracy condition of Lemma 3.2. As it happens, this step is more challenging than it may first seem; it basically amounts to proving that if a matrix is close to being symmetric, and close to having low rank, then it is close to a matrix that is simultaneously symmetric and has low rank. This fact about "symmetric low-rank approximation" is encapsulated in Lemmas 5.5 and 5.7 (there are slightly different versions for the proofs of Theorems 1.1 and 1.2), and most of the rest of the paper is devoted to proving these lemmas. Indeed, in Section 6 we prove some general-purpose lemmas about a certain notion of "robust linear independence", and in Section 7 we use these lemmas to prove Lemmas 5.5 and 5.7.

Finally, Section 8 contains some concluding remarks. In particular, we present a new conjecture about edge-statistics in Ramsey graphs, generalising Theorem 1.3, which would imply the conjectures of Erdős–Faudree–Sós and Narayanan–Sahasrabudhe–Tomon regarding subgraphs of Ramsey graphs with different numbers of edges.

1.3 Notation

We use standard asymptotic notation throughout. For functions $f = f(n)$ and $g = g(n)$ we write $f = O(g)$ to mean there is a constant C such that $|f| \leq C|g|$, we write $f = \Omega(g)$ to mean there is a

constant $c > 0$ such that $f \geq c|g|$ for sufficiently large n, we write $f = \Theta(g)$ to mean that $f = O(g)$ and $f = \Omega(g)$, and we write $f = o(g)$ or $g = \omega(f)$ to mean that $f/g \to 0$ as $n \to \infty$. All asymptotics are as $n \to \infty$ unless specified otherwise.

For a non-negative integer n we define $[n] = \{1, \dots, n\}$, and for a real number x, the floor function is denoted $\lfloor x \rfloor = \max\{i \in \mathbb{Z} : i \leq x\}$. For a vector $v \in \mathbb{C}^n$ or a matrix $A \in \mathbb{C}^{m \times n}$, we let $\Re(v) \in \mathbb{R}^n$ and $\Re(A) \in \mathbb{R}^{m \times n}$ denote the vector or matrix obtained by taking the real part of each entry. We adopt the convention that the determinant of the 0×0 empty matrix is 1. All logarithms are in base e, unless explicitly noted otherwise.

We also use standard graph-theoretic notation. Given a graph G, we denote its vertex set by $V(G)$. For subsets $X, Y \subseteq V(G)$, let $e(X)$ denote the number of edges inside X, and let $e(X, Y)$ denote the number of edges between X and Y. Let $d(X) = e(X)/\binom{|X|}{2}$ and $d(X, Y) = e(X, Y)/(|X||Y|)$ denote the density of edges inside X, and between X and Y, respectively. Abusing notation, we write $e(x, Y)$ or $e(x, y)$ to denote $e(\{x\}, Y)$ or $e(\{x\}, \{y\})$, respectively.

Given a field $\mathbb{F}$ and non-negative integers m and n, let $\mathbb{F}^{m \times n}$ denote the set of all $m \times n$ matrices with entries in $\mathbb{F}$. For a matrix $A \in \mathbb{F}^{m \times n}$, for $i = 1, \dots, m$ we write $\mathrm{row}_i(A) \in \mathbb{F}^n$ for the vector corresponding to the i-th row of A, and for $j = 1, \dots, n$ we write $\mathrm{col}_j(A) \in \mathbb{F}^m$ for the vector corresponding to the j-th column. Given a matrix $A \in \mathbb{F}^{m \times n}$ and a subset $I \subseteq [n]$, let A_I be the $m \times |I|$ submatrix of A consisting of the columns with indices in I. Similarly, for a vector $v \in \mathbb{F}^n$ and a subset $I \subseteq [n]$, let $v_I \in \mathbb{F}^I$ be the vector consisting of the entries of v with indices in I.

For a matrix $A \in \mathbb{C}^{m \times n}$, we denote the sum of the absolute values of the entries of A by $\|A\|_1$, and we denote the maximum of the absolute values of the entries by $\|A\|_\infty$ (these are entrywise norms of A, not to be confused with the more common operator norms). For a vector $v \in \mathbb{C}^n$, we denote the usual L_p-norm by $\|v\|_p$, and for vectors $v, w \in \mathbb{R}^n$ we write $\langle v, w \rangle$ for the standard inner product of v and w.

2 Anti-concentration in Ramsey graphs

In this section we deduce Theorem 1.3 from Theorem 1.2. Our plan will be to express the random variable X in Theorem 1.3 as a quadratic polynomial of independent Rademacher random variables. We can obtain an upper bound on the point probabilities of this polynomial from Theorem 1.2 if we can show that our quadratic polynomial is not close to a low-rank quadratic form. In order to do so, we will apply the following simple lemma to the matrix associated with the homogeneous degree-2 part of the polynomial.

Lemma 2.1. *Let r be a positive integer, let $\delta \geq 0$, and let M be an $m \times m$ matrix over any field which has more than δm^{2r} full-rank $r \times r$ submatrices. Then, if we change up to δm^2 entries of M, the resulting matrix has rank at least r.*

Proof. If we change at most δm^2 entries, we can affect at most $\delta m^2 m^{2(r-1)} = \delta m^{2r}$ of the $r \times r$ submatrices of M, so a full-rank $r \times r$ submatrix remains. □

Next, the only fact about Ramsey graphs we will need is the fact that they have many copies of every possible induced subgraph on a small number of vertices. This generalises an old result of Erdős and Hajnal [17], which asserts the existence of at least one copy of each such subgraph.

Lemma 2.2. *For any fixed $h \geq 1$ and any fixed constant $C > 0$, there is $\delta = \delta(h, C) > 0$ such that the following holds for sufficiently large n. Every n-vertex C-Ramsey graph G contains at least δn^h induced copies of every graph H on h vertices.*

Proof. A graph is said to be ε-regular if for all subsets $X, Y \subseteq V(G)$ with $|X| \geq \varepsilon|V(G)|$ and $|Y| \geq \varepsilon|V(G)|$, we have $|d(X, Y) - d(V(G))| \leq \varepsilon$. It is a consequence of Szemerédi's regularity lemma that for any fixed $\varepsilon > 0$, every n-vertex graph G contains an ε-regular induced subgraph $G[U]$ on $m = \Omega(n)$ vertices (see also [10, Lemma 5.2] for a version of this fact with better dependence on ε). Now, if G is C-Ramsey, then $G[U]$ is still $(C + o(1))$-Ramsey, so by the theorem of Erdős and Szemerédi [19] mentioned in the introduction, the density of $G[U]$ is bounded away from zero and one: that is, there is $\eta > 0$ depending only on C such that $\eta \leq d(U) \leq 1 - \eta$ for sufficiently large n.

Then, provided ε is sufficiently small with respect to η, we can conclude the proof by applying a counting lemma to $G[U]$ (see for example [10, Lemma 5.12]): if an m-vertex graph has density bounded away from zero and one, and it is ε-regular for sufficiently small ε, then it has $\Omega(m^h)$ induced copies of every graph H on h vertices. Thus, we obtain $\Omega(m^h) = \Omega(n^h)$ induced copies of H in $G[U]$ and therefore in G. □

Another crucial ingredient is a variant of [27, Lemma 2.8], to express the random variable X in Theorem 1.3 as a quadratic polynomial of independent random variables. Consider any graph G with vertex set $[n]$ and any $0 \le k \le n$, and let $m = \min(k, n-k)$. First, we want a way to generate a random k-vertex subset of G in a way which involves independent random choices. Let π be a uniformly random permutation of $[n]$ and let $\xi = (\xi_1, \dots, \xi_m) \sim \mathrm{Rad}^m$ be a sequence of independent Rademacher random variables (also independent from π). Note that

$$U_{\pi,\xi} = \{\pi(i) : i \in [m], \xi_i = 1\} \cup \{\pi(i+m) : i \in [m], \xi_i = -1\} \cup \{\pi(i) : 2m+1 \le i \le m+k\} \tag{2.1}$$

is a uniformly random subset of k vertices of G. Indeed, the union of the first two sets on the right-hand side has size m, the third set has size $k - m \ge 0$, and all three sets are disjoint.

Now, recall that for two vertices $v, w \in V(G)$, we defined $e(v,w) = 1$ if there is an edge between v and w and $e(v,w) = 0$ otherwise. For any fixed outcome of π, the number of edges in $U = U_{\pi,\xi}$ is

$$e(U_{\pi,\xi}) = \sum_{1 \le i < j \le n} \mathbf{1}_{\pi(i) \in U} \mathbf{1}_{\pi(j) \in U} e(\pi(i), \pi(j)). \tag{2.2}$$

Note that by the definition of $U = U_{\pi,\xi}$, we have

$$\mathbf{1}_{\pi(i) \in U} = \begin{cases} \frac{1}{2}(1+\xi_i) & \text{if } 1 \le i \le m \\ \frac{1}{2}(1-\xi_{i-m}) & \text{if } m+1 \le i \le 2m \\ 1 & \text{if } 2m+1 \le i \le m+\ell \\ 0 & \text{if } m+\ell+1 \le i \le n \end{cases}.$$

Plugging this into (2.2), and using that $\xi_i^2 = 1$ for all i, we obtain the following lemma.

Lemma 2.3. *Let G be a graph with vertex set $[n]$. Furthermore, let $0 \le k \le n$ and $m = \min(k, n-k)$. Let π be a random permutation of $[n]$, and $\xi = (\xi_1, \dots, \xi_m) \sim \mathrm{Rad}^m$ be a sequence of independent Rademacher random variables, and define $U_{\pi,\xi} \subseteq V(G)$ as in (2.1). Then $U_{\pi,\xi}$ is a uniformly random subset of k vertices of G. Furthermore we can write*

$$e(U_{\pi,\xi}) = f_\pi(\xi) = \sum_{1 \le i < j \le m} a_{ij} \xi_i \xi_j + \sum_{1 \le i \le m} a_i \xi_i + a_0,$$

where the coefficients a_{ij}, a_i and a_0 of the quadratic polynomial f_π only depend on π (and not on ξ). In addition, for $1 \le i < j \le m$ we have

$$a_{ij} = \frac{1}{4} e(\pi(i), \pi(j)) - \frac{1}{4} e(\pi(i), \pi(j+m)) - \frac{1}{4} e(\pi(i+m), \pi(j)) + \frac{1}{4} e(\pi(i+m), \pi(j+m)). \tag{2.3}$$

Note that (2.3) in particular implies that $a_{ij} \in \{-\frac{1}{2}, -\frac{1}{4}, 0, \frac{1}{4}, \frac{1}{2}\}$ for all $i < j$. One can also give explicit formulas for the other coefficients a_i and a_0 of f_π, but this is not necessary for our argument.

Now we put everything together to prove Theorem 1.3.

Proof of Theorem 1.3. Fix some $r \ge 3$, which we treat as a constant in all asymptotic notation. We will prove that $\Pr(X = x) \le n^{-1+2/(r+2)+o(1)}$. Since r was arbitrary, this suffices to prove Theorem 1.3.

Let G be an n-vertex C-Ramsey graph, and let $cn \le k \le (1-c)n$. As before, define $m = \min(k, n-k)$ and note that $cn \le m \le n/2$. As in Lemma 2.3, we can model the random variable X as $X = e(U_{\pi,\xi}) = f_\pi(\xi)$, where π is a random permutation of $[n]$, and $\xi = (\xi_1, \dots, \xi_m) \sim \mathrm{Rad}^m$ is a sequence of independent Rademacher random variables.

Let us say that a $(2r)$-tuple $(i_1, \dots, i_r, j_1, \dots, j_r) \in [m]^{2r}$ is *strong* if $i_1 < \dots < i_r < j_1 < \dots < j_r$ and if we have $a_{i_\ell j_\ell} = 1/2$ for $\ell = 1, \dots, r$, but $a_{i_\ell j_q} = 0$ whenever $\ell \ne q$ (note that this definition depends on the permutation π; recall (2.3)). We first use Lemma 2.2 to show that there are likely to be many strong $(2r)$-tuples.

Claim 2.4. *Subject to the randomness of the random permutation π, with probability $1 - e^{-\Omega(n)}$ there are $\Omega(m^{2r})$ strong $(2r)$-tuples.*

Proof of Claim 2.4. Since G is a C-Ramsey graph, by Lemma 2.2, it has $\Omega(n^{4r})$ induced copies of a perfect matching on $4r$ vertices (consisting of $2r$ edges). That is to say, there are $\Omega(n^{4r})$ sequences of distinct vertices $(v_1, \dots, v_{4r}) \in V(G)^{4r}$ such that for $i = 1, \dots, 2r$ there is an edge between v_i and v_{i+2r} and there are no other edges between $v_1, \dots, v_{4r}$.

There are $\binom{m}{2r} = \Omega(m^{2r})$ different $(2r)$-tuples $(i_1, \dots, i_r, j_1, \dots, j_r) \in [m]^{2r}$ with $i_1 < \dots < i_r < j_1 < \dots < j_r$. For each such $(2r)$-tuple,

$$(\pi(i_1), \dots, \pi(i_r), \pi(i_1 + m), \dots, \pi(i_r + m), \pi(j_1), \dots, \pi(j_r), \pi(j_1 + m), \dots, \pi(j_r + m))$$

is a uniformly random sequence of $4r$ distinct vertices of G. Thus, with probability $\Omega(1)$, it is one of the sequences $(v_1, \dots, v_{4r})$ considered above. But if that is the case, then $(i_1, \dots, i_r, j_1, \dots, j_r)$ is strong: by (2.3), $a_{i_\ell j_\ell} = \frac{1}{4} - 0 - 0 + \frac{1}{4} = \frac{1}{2}$ for $\ell = 1, \dots, r$, but $a_{i_\ell j_q} = 0 - 0 - 0 + 0 = 0$ whenever $\ell \neq q$.

Let Z be the number of strong $(2r)$-tuples: we have just proved that $\mathbb{E}Z = \Omega(m^{2r})$. Now we can conclude the proof with a concentration inequality. Note that changing π by a transposition (swapping some $\pi(t)$ and $\pi(t')$) changes the number of strong $(2r)$-tuples by at most $8r \cdot m^{2r-1}$. Indeed, there are at most $8r \cdot m^{2r-1}$ different $(2r)$-tuples $(i_1, \dots, i_r, j_1, \dots, j_r) \in [m]^{2r}$ such that t or t' occur among $i_1, \dots, i_r, i_1 + m, \dots, i_r + m, j_1, \dots, j_r, j_1 + m, \dots, j_r + m$ (which are the only places where the value of π affects whether $(i_1, \dots, i_r, j_1, \dots, j_r)$ is strong). Thus, by a McDiarmid-type concentration inequality for random permutations (see for example [31, Section 3.2]), we have

$$\Pr(Z < \mathbb{E}Z/2) \leq \exp\left(-\Omega\left(\frac{(\mathbb{E}Z/2)^2}{n \cdot (8r \cdot m^{2r-1})^2}\right)\right) = \exp\left(-\Omega\left(\frac{(m^{2r})^2}{n \cdot (8r \cdot m^{2r-1})^2}\right)\right) = \exp(-\Omega(n)),$$

recalling that $m \geq cn$. □

Now, condition on an outcome of π satisfying the conclusion of Claim 2.4. Conditionally, X can be represented as a quadratic polynomial $f_\pi(\xi)$ in $\xi = (\xi_1, \dots, \xi_m) \sim \mathrm{Rad}^m$. We can express the homogeneous degree-2 part of $f_\pi(\xi)$ as $\xi^T Q_\pi \xi$ for some symmetric $m \times m$ matrix, and note that for $i < j$ the (i, j)-entry of Q_π equals $a_{ij}/2$. Claim 2.4 implies that the matrix Q_π has $\Omega(m^{2r})$ full-rank $r \times r$ submatrices (note that for every strong $(2r)$-tuple $(i_1, \dots, i_r, j_1, \dots, j_r)$ the submatrix with rows $i_1, \dots, i_r$ and columns $j_1, \dots, j_r$ is a diagonal matrix with entries $1/4$ on the diagonal). Hence by Lemma 2.1 there is $\varepsilon = \Omega(1)$ such that whenever we change up to $2\varepsilon m^2$ entries of Q_π, the resulting matrix has rank at least r. Now, if $h(\xi)$ is a quadratic form differing from f in at most εm^2 coefficients, then h is of the form $\xi^T Q'_\pi \xi$ for a symmetric $m \times m$ matrix Q'_π which differs from Q_π in at most $2\varepsilon m^2$ entries and consequently has rank at least r. Thus, using that the degree-2 coefficients a_{ij} of f_π all lie in the set $S = \{-\frac{1}{2}, -\frac{1}{4}, 0, \frac{1}{4}, \frac{1}{2}\}$, applying Theorem 1.2 yields

$$\sup_{x \in \mathbb{Q}} \Pr(f_\pi(\xi) = x) < C(r, \varepsilon, S) \cdot \frac{(\log n)^{r/2}}{(cn)^{1-2/(r+2)}} \leq \frac{n^{o(1)}}{n^{1-2/(r+2)}}.$$

Recalling that we have been conditioning on an event that holds with probability $1 - e^{-\Omega(n)}$, it follows that

$$\sup_{x \in \mathbb{Q}} \Pr(X = x) = \frac{n^{o(1)}}{n^{1-2/(r+2)}} + e^{-\Omega(n)} = n^{-1+2/(r+2)+o(1)},$$

as desired. □

3 A technical anti-concentration inequality for real polynomials

In this section, we will prove an anti-concentration bound for real quadratic polynomials satisfying a certain technical non-degeneracy condition. This will be one of the key ingredients for the proofs of Theorems 1.1 and 1.2.

To cleanly state our anti-concentration inequality, we first make some simple definitions.

Definition 3.1. For an $n \times n$ matrix M and a tuple $(i_1, \dots, i_r) \in [n]^r$, let $M(i_1, \dots, i_r)$ be the $r \times n$ matrix whose rows are $\mathrm{row}_{i_1}(M), \dots, \mathrm{row}_{i_r}(M)$. For $\delta > 0$, let us say that a $r \times n$ matrix M' is *δ-non-degenerate* if for any unit vector $e \in \mathbb{R}^r$, there are at least δn columns w of M' satisfying $|\langle w, e\rangle| \geq \delta$.

Now, our anti-concentration inequality is as follows.

Lemma 3.2. *For any integer $r \geq 3$ and any $\delta > 0$ there is a constant $C = C(r, \delta)$ such that the following holds. Consider a real quadratic polynomial $f(x) = \sum_{1 \leq i \leq j \leq n} a_{ij} x_i x_j + \sum_{1 \leq i \leq n} a_i x_i + a_0$, let $a_{ji} = a_{ij}$ for $i > j$, and let $M = (a_{ij})_{i,j} \in \mathbb{R}^{n \times n}$. Suppose that each $|a_{ij}| \leq 1$, and suppose that there is a set $T \subseteq [n]^r$ of δn disjoint r-tuples such that for each $(i_1, \dots, i_r) \in T$, the matrix $M(i_1, \dots, i_r)$ is δ-non-degenerate. Then, for $\xi \in \mathrm{Rad}^n$ and any $x \in \mathbb{R}$, we have*

$$\Pr\Big(|f(\xi) - x| \leq n^{2/(r+2)}\Big) \leq C \cdot \frac{(\log n)^{r/2}}{n^{1-2/(r+2)}}.$$

The notion of being δ-non-degenerate is closely related to the condition in an anti-concentration theorem due to Halász [24] (stated as Theorem 3.5 below), which will be used in the proof of Lemma 3.2. A matrix M' being δ-non-degenerate can be interpreted as a robust version of M' having full row rank, so if an $n \times n$ matrix M has many $r \times n$ submatrices that are δ-non-degenerate, then there is some sense in which M robustly has rank at least r. The reader may wish to compare the statement of Lemma 3.2 with the statements of Theorems 1.1 and 1.2 in the introduction.

Before proving Lemma 3.2, we discuss some of the main ideas and ingredients. The most crucial idea is a variant of a *decoupling* trick due to Costello, Tao and Vu [12], as follows. If $[n] = I \cup J$ is a partition of the index set into two subsets, then we can break $\xi = (\xi_1, \dots, \xi_n)$ into two subsequences ξ_I and ξ_J. The quadratic polynomial $f(\xi)$ can then be written as $f(\xi) = f(\xi_I, \xi_J)$. The crucial observation is that if ξ'_J is an independent copy of ξ_J then it is possible to relate the anti-concentration of $f(\xi)$ to the anti-concentration of $Y := f(\xi_I, \xi_J) - f(\xi_I, \xi'_J)$: for example, one can use the Cauchy–Schwarz inequality to prove that

$$\Pr(f(\xi) = x) \leq \Pr\big(f(\xi_I, \xi_J) = x \text{ and } f\big(\xi_I, \xi'_J\big) = x\big)^{1/2} \leq \Pr(Y = 0)^{1/2}. \tag{3.1}$$

After conditioning on an outcome of (ξ_J, ξ'_J), the random variable Y then becomes a *linear* polynomial in ξ_I, which is much easier to study.

This approach results in a square-root loss, and therefore seems unsuitable to prove Lemma 3.2. However, a variation on this approach is to instead use decoupling to study the modulus of the *characteristic function* (Fourier transform) $t \mapsto \mathbb{E}e^{2\pi i t f(\xi)}$ of $f(\xi)$. Specifically, we will use the following simple observation.

Lemma 3.3. *Let ξ_I and ξ_J be independent random vectors, and let $f(\xi_I, \xi_J)$ be a real-valued random variable defined in terms of these random vectors. Let ξ'_J be an independent copy of ξ_J. Then for any $t \in \mathbb{R}$,*

$$\Big|\mathbb{E}e^{2\pi i t f(\xi_I, \xi_J)}\Big|^2 \leq \mathbb{E}\Big[\Big|\mathbb{E}[e^{2\pi i t \left(f(\xi_I, \xi_J) - f(\xi_I, \xi'_J)\right)} \mid \xi_J, \xi_{J'}]\Big|\Big].$$

Proof. First, by convexity we have

$$\Big|\mathbb{E}e^{2\pi i t f(\xi_I, \xi_J)}\Big|^2 = \Big|\mathbb{E}\Big[\mathbb{E}[e^{2\pi i t f(\xi_I, \xi_J)} | \xi_I]\Big]\Big|^2 \leq \mathbb{E}\Big[\Big|\mathbb{E}[e^{2\pi i t f(\xi_I, \xi_J)} | \xi_I]\Big|^2\Big].$$

Then, observe that for independent identically distributed complex-valued random variables Z, Z', we have

$$|\mathbb{E}Z|^2 = \mathbb{E}Z\overline{\mathbb{E}Z} = \mathbb{E}Z\overline{\mathbb{E}Z'} = \mathbb{E}\big[Z\overline{Z'}\big].$$

In particular, we obtain

$$\Big|\mathbb{E}[e^{2\pi i t f(\xi_I, \xi_J)} \mid \xi_I]\Big|^2 = \mathbb{E}[e^{2\pi i t \left(f(\xi_I, \xi_J) - f(\xi_I, \xi'_J)\right)} \mid \xi_I].$$

It follows that

$$\Big|\mathbb{E}e^{2\pi i t f(\xi_I, \xi_J)}\Big|^2 \leq \mathbb{E}\Big[\mathbb{E}[e^{2\pi i t \left(f(\xi_I, \xi_J) - f(\xi_I, \xi'_J)\right)} \mid \xi_I]\Big] = \mathbb{E}\Big[\mathbb{E}[e^{2\pi i t \left(f(\xi_I, \xi_J) - f(\xi_I, \xi'_J)\right)} \mid \xi_J, \xi_{J'}]\Big],$$

from which we can conclude the desired result. □

We remark that while we were working on this paper, some similar decoupling tricks were independently developed by Berkowitz [6] in connection with his work on local central limit theorems for clique counts in random graphs. We also remark that a very similar argument appears implicitly in a paper of Nguyen [35].

Next, the following result is called *Esséen's concentration inequality* [20]. It gives a way to prove anti-concentration bounds by integrating bounds on the characteristic function of a random variable. This particular statement is a special case of [46, Lemma 7.17].

Lemma 3.4 ([46])**.** *There is a constant $C > 0$ such that the following holds. Let X be a real-valued random variable which takes only a finite number of values. Then for any $\varepsilon > 0$, any $s > 0$ and any $x \in \mathbb{R}$, we have*

$$\Pr(|X - x| \leq s) \leq C(s + 1/\varepsilon) \int_{-\varepsilon}^{\varepsilon} \left|\mathbb{E}e^{2\pi i t X}\right| \mathrm{d}\, t.$$

It may not be immediately obvious how one can benefit from using a decoupling trick for characteristic functions (as in Lemma 3.3) instead of using a decoupling trick for point probabilities directly (as in (3.1)). Indeed, Lemma 3.3 also involves a square-root loss when studying $f(\xi_I, \xi_J)$ via $f(\xi_I, \xi_J) - f(\xi_I, \xi'_J)$. The key is that the square-root loss is "inside the integral". It turns out that in the setting of Lemma 3.2, the characteristic function has "sharp threshold" behaviour: if $|t|$ is much smaller than $1/n$, then $|\mathbb{E}e^{2\pi i t f(\xi)}|$ is very close to one, whereas if $|t|$ is much larger than $1/n$ then $|\mathbb{E}e^{2\pi i t f(\xi)}|$ is very close to zero. Therefore taking the square root of the modulus of the characteristic function has a relatively small effect on its integral.

In order to effectively apply Lemma 3.3, we will need some understanding of the typical structure of $f(\xi_I, \xi_J) - f(\xi_I, \xi'_J)$ as a linear polynomial in ξ_I, subject to the randomness of ξ_J and ξ'_J. In particular, we need to show that this polynomial is unlikely to have many coefficients that are close to zero. To accomplish this, we use the following multi-dimensional extension of the (linear) Littlewood–Offord theorem due to Halász [24][2].

Theorem 3.5. *For any integer $d \geq 1$ and any $\delta > 0$, there is $C = C(d, \delta) > 0$ such that the following holds. Let $a_1, \dots, a_n$ be a collection of vectors in $\mathbb{R}^d$ and let $s > 0$. Suppose that for any unit vector $e \in \mathbb{R}^d$, there are at least δn vectors a_i with $|\langle a_i, e\rangle| \geq s$. Then for $\xi = (\xi_1, \dots, \xi_n) \in \mathrm{Rad}^n$ we have*

$$\sup_{u \in \mathbb{R}^d} \Pr\left(\left\|\sum_{i=1}^n \xi_i a_i - u\right\|_2 < s\right) \leq C n^{-d/2}.$$

We have still not yet described how to choose the partition $[n] = I \cup J$ for decoupling. We will actually just choose the partition randomly; we will then need the fact that a random submatrix of a non-degenerate matrix is typically still non-degenerate, as follows. Recall that for a matrix $M \in \mathbb{R}^{r \times n}$ and a subset $I \subseteq [n]$, we defined M_I to be the $r \times |I|$ submatrix of M consisting of the columns with indices in I. Also recall that matrix norms in this paper are entrywise.

Lemma 3.6. *Fix an integer $r \geq 1$ and fix $\delta > 0$. Suppose $M \in \mathbb{R}^{r \times n}$ is a δ-non-degenerate matrix with $\|M\|_\infty \leq 1$. Then, for a uniformly random subset $I \subseteq [n]$, with probability $1 - e^{-\Omega(n)}$, the matrix M_I is $(\delta/3)$-non-degenerate. (Here, the implicit constant in the Ω-term may depend on r and δ.)*

Proof. Let $\varepsilon = \delta/(2r) > 0$, and fix a finite set $E \subseteq \{e \in \mathbb{R}^r : \|e\|_2 = 1\}$ such that for any unit vector $e \in \mathbb{R}^r$ we can find $e' \in E$ with $\|e - e'\|_2 \leq \varepsilon$ (that is, E is an ε-net of the unit sphere in $\mathbb{R}^r$).

Note that every column $w \in \mathbb{R}^r$ of M satisfies $\|w\|_2 \leq r$. Thus, whenever $e' \in E$ and w is a column of M with $|\langle w, e'\rangle| \geq \delta$, then all unit vectors $e \in \mathbb{R}^r$ with $\|e - e'\|_2 \leq \varepsilon$ satisfy

$$|\langle w, e\rangle| \geq |\langle w, e'\rangle| - |\langle w, e' - e\rangle| \geq \delta - \|w\|_2 \|e - e'\|_2 \geq \delta - r\varepsilon \geq \delta/3.$$

For every $e' \in E$, the δ-non-degenerate matrix M has at least δn columns w such that $|\langle w, e'\rangle| \geq \delta$. By a Chernoff bound, with probability $1 - e^{-\Omega(n)}$, at least $(\delta/3)n$ of these columns are still in M_I. By taking the union bound over all $e' \in E$, we see that with probability $1 - e^{-\Omega(n)}$, for every $e' \in E$ the matrix M_I has $(\delta/3)n$ columns w with $|\langle w, e'\rangle| \geq \delta$. Whenever this happens, for every unit vector $e \in \mathbb{R}^r$, the matrix M_I has $(\delta/3)n$ columns w with $|\langle w, e\rangle| \geq \delta/3$. As M_I has at most n columns in total, this means that M_I is $(\delta/3)$-non-degenerate. □

[2] We remark that a very similar inequality was also proved by Tao and Vu [44, Theorem 1.4], and that there is a large body of work generalising the Erdős–Littlewood–Offord theorem to higher dimensions without this non-degeneracy condition (in which case the bounds are much weaker; see for example the survey [36, Section 2]).

We are now ready to prove Lemma 3.2.

Proof of Lemma 3.2. Fix an integer $r \geq 3$ and fix $\delta > 0$. For all asymptotic notation in this proof, we treat r and δ as constants. Let $f \in \mathbb{R}[x_1, \dots, x_n]$ and $T \subseteq [n]^r$ be as in the lemma statement. For any subset $J \subseteq [n]$ and tuple $(i_1, \dots, i_r) \in T$, let $M_J(i_1, \dots, i_r)$ be the submatrix of $M(i_1, \dots, i_r)$ consisting of the columns with indices $j \in J$.

Now, consider a uniformly random subset $J \subseteq [n]$, and let $I = [n] \setminus J$. By Lemma 3.6 and the union bound, with probability $1 - e^{-\Omega(n)}$ each $M_J(i_1, \dots, i_r)$, for $(i_1, \dots, i_r) \in T$, is $(\delta/3)$-non-degenerate. Also, by a Chernoff bound, with probability $1 - e^{-\Omega(n)}$ we have $|J| \geq n/3$ and $|I^r \cap T| \geq 2^{-r-1}|T| \geq 2^{-r-1}\delta n$.

Thus, we can fix a partition $[n] = I \cup J$ such that $|J| \geq n/3$ and such that there exists a set $T_I \subseteq I^r \cap T$ of at least $2^{-r-1}\delta n = \Omega(n)$ disjoint r-tuples, with the property that for each $(i_1, \dots, i_r) \in T_I$, the matrix $M_J(i_1, \dots, i_r)$ is $(\delta/3)$-non-degenerate.

Now, let $\xi_I = (\xi_\ell)_{\ell \in I}$ and $\xi_J = (\xi_j)_{j \in J}$. We write $f(\xi) = f(\xi_I, \xi_J)$. Let ξ'_J be an independent copy of ξ_J, so by Lemma 3.3, for any $t \in \mathbb{R}$ we have

$$\left|\mathbb{E}e^{2\pi i t f(\xi)}\right|^2 \leq \mathbb{E}\Big[\Big|\mathbb{E}[e^{2\pi i t\left(f(\xi_I,\xi_J)-f(\xi_I,\xi'_J)\right)} \mid \xi_J, \xi_{J'}]\Big|\Big]. \tag{3.2}$$

Note that $f(\xi_I, \xi_J) - f(\xi_I, \xi'_J) = \sum_{\ell \in I} A_\ell \xi_\ell + A$, where $A_\ell = \sum_{j \in J} a_{\ell j}(\xi_j - \xi'_j)$ for all $\ell \in I$ and

$$A = \sum_{j,j' \in J,\, j \leq j'} a_{jj'}(\xi_j \xi_{j'} - \xi'_j \xi'_{j'}) + \sum_{j \in J} a_j(\xi_j - \xi'_j).$$

Thus, when conditioning on any fixed outcome of $\xi_J, \xi_{J'}$, we can interpret $f(\xi_I, \xi_J) - f(\xi_I, \xi'_J)$ as a linear function in ξ_I with coefficients A_ℓ. Hence, for any $t \in \mathbb{R}$, we obtain

$$\left|\mathbb{E}[e^{2\pi i t\left(f(\xi_I,\xi_J)-f(\xi_I,\xi'_J)\right)} \mid \xi_J, \xi_{J'}]\right| = \prod_{\ell \in I}\left|\frac{e^{2\pi i t A_\ell} + e^{-2\pi i t A_\ell}}{2}\right| = \prod_{\ell \in I}|\cos(2\pi t A_\ell)|. \tag{3.3}$$

Now, for real $s \geq 0$, $t \in [-1, 1] \setminus \{0\}$ and $\ell \in I$, let $\mathcal{E}_\ell^{s,t}$ be the event that $|2A_\ell - k/t| \leq s$ for some integer multiple k/t of $1/t$. Note that if $s|t| \leq 1$ and $\mathcal{E}_\ell^{s,t}$ does not hold then $|\cos(2\pi t A_\ell)| = e^{-\Omega(t^2 s^2)}$. We will now use a concentration inequality and Halász' inequality (Theorem 3.5) to find an upper bound for the probability of the event that $\mathcal{E}_\ell^{s,t}$ holds for many different ℓ.

Claim 3.7. *Consider some r-tuple $(i_1, \dots, i_r) \in T_I \subseteq I^r$, and let $\mathcal{E}_{i_1,\dots,i_r}^{s,t} = \mathcal{E}_{i_1}^{s,t} \cap \dots \cap \mathcal{E}_{i_r}^{s,t}$. Then $\Pr(\mathcal{E}_{i_1,\dots,i_r}^{s,t}) = O(p(s,t))$, where $p(s,t) = (|t| \log n + 1/\sqrt{n})^r (s+1)^r$.*

Proof. We may assume that $s|t| \leq 1$ as otherwise the claim is trivial. Let J^* be the subset of indices $j \in J$ such that $(\xi_j - \xi'_j) \neq 0$. This is a uniformly random subset of J, so by the Chernoff bound and Lemma 3.6, with probability $1 - e^{-\Omega(n)}$, we have $|J^*| \geq |J|/3 \geq n/9$ and $M_{J^*}(i_1, \dots, i_r)$ is $(\delta/9)$-non-degenerate. Condition on such an outcome of J^*.

Now, conditionally, the random variables $\xi_j^* := (\xi_j - \xi'_j)/2$, for $j \in J^*$, are Rademacher distributed and mutually independent. For $j \in J^*$ let $b_j = (a_{i_1 j}, \dots, a_{i_r j})$ be the column of $M_{J^*}(i_1, \dots, i_r)$ indexed by $j \in J^*$. Recall that for $q = 1, \dots, r$, we have $A_{i_q} = \sum_{j \in J} a_{i_q j}(\xi_j - \xi'_j) = 2\sum_{j \in J} a_{i_q j}\xi_j^*$. Hence the vector $(A_{i_1}, \dots, A_{i_q}) \in \mathbb{R}^r$ equals $2\sum_{j \in J} \xi_j^* b_j$. Thus, by Halász' inequality (Theorem 3.5), still conditioning on our outcome of J^*, for each $v = (k_1, \dots, k_r) \in \mathbb{Z}^r$ we have

$$\Pr\big(|2A_{i_q} - k_q/t| \leq s \text{ for all } q = 1, \dots, r\big) = \Pr\left(\left\|\sum_{j \in J^*} \xi_j^* b_j - \frac{v}{4t}\right\|_\infty \leq \frac{s}{4}\right) = O\Big(\frac{(s+1)^r}{n^{r/2}}\Big).$$

(Note that the above equation features the norm $\|\cdot\|_\infty$, while Theorem 3.5 concerns the norm $\|\cdot\|_2$. We can cover a r-dimensional box of side-length $s/2$ with $O(s+1)^r$ balls of radius $\delta/9$, and then we can apply Theorem 3.5 to each of these balls using that the matrix $M_{J^*}(i_1, \dots, i_r)$ is $(\delta/9)$-non-degenerate.) Also, by the Azuma–Hoeffding inequality and the union bound,

$$\Pr\big(|2A_{i_q}| \geq \sqrt{n}\log n \text{ for some } q \in [r]\big) = e^{-\Omega((\log n)^2)}.$$

Finally, note that there are $O\big((|t|(\sqrt{n}\log n + s) + 1)^r\big)$ choices $(k_1, \dots, k_r) \in \mathbb{Z}^r$ such that each $|k_q/t| \le \sqrt{n}\log n + s$, so we can conclude that

$$\Pr\big(\mathcal{E}^{s,t}_{i_1,\dots,i_r}\big) = O\bigg(\big(|t|(\sqrt{n}\log n + s) + 1\big)^r \frac{(s+1)^r}{n^{r/2}} + e^{-\Omega((\log n)^2)} + e^{-\Omega(n)}\bigg) = O(p(s,t)). \qquad \square$$

Now, let $W^{s,t}$ be the number of r-tuples $(i_1, \dots, i_r) \in T_I$ satisfying $\mathcal{E}^{s,t}_{i_1,\dots,i_r}$. By Claim 3.7, $\mathbb{E}W^{s,t} = O(p(s,t)|T_I|)$, so $\Pr(W^{s,t} \ge |T_I|/2) = O(p(s,t))$ by Markov's inequality. But observe that if $W^{s,t} < |T_I|/2$, then at least $|T_I|/2$ different r-tuples $(i_1, \dots, i_r) \in T_I$ contain some index i_q such that $\mathcal{E}^{s,t}_{i_q}$ does not hold. As the r-tuples in T_I are all disjoint, this means that there are at least $|T_I|/2 = \Omega(n)$ indices $\ell \in I$ such that $\mathcal{E}^{s,t}_{\ell}$ does not hold, so $\prod_{\ell\in I}|\cos(2\pi t A_\ell)| = e^{-\Omega(t^2 s^2 n)}$.

We have proved that for any $x \in (0,1)$, there is $s = O\Big(\sqrt{-\log(x)/(t^2 n)}\Big)$ such that

$$\Pr\bigg(\prod_{\ell\in I}|\cos(2\pi t A_\ell)| \ge x\bigg) \le \Pr\big(W^{s,t} \ge |T_I|/2\big) = O(p(s,t)). \tag{3.4}$$

For this value of s, we compute

$$\begin{aligned} p(s,t) &= O\Big(\big(|t|\log n + 1/\sqrt{n}\big)\Big(\sqrt{\log(1/x)/(t^2 n)} + 1\Big)\Big)^r \\ &= \begin{cases} O\bigg(\Big(\dfrac{\log n}{\sqrt{n}}\Big)^r + (\log(1/x))^{r/2}\Big(\dfrac{\log n}{|t|n}\Big)^r\bigg) & \text{for } |t| \le \dfrac{1}{\sqrt{n}}, \\ O\bigg((|t|\log n)^r + (\log(1/x))^{r/2}\Big(\dfrac{\log n}{\sqrt{n}}\Big)^r\bigg) & \text{for } |t| \ge \dfrac{1}{\sqrt{n}}. \end{cases} \end{aligned} \tag{3.5}$$

Note that

$$\int_0^1 (\log(1/x))^{r/2}\,\mathrm{d}\,x \le \sum_{j=1}^{\infty}\int_{2^{-j}}^{2^{-j+1}} (\log_2(1/x))^{r/2}\,\mathrm{d}\,x \le \sum_{j=1}^{\infty} 2^{-j} j^{r/2} = O(1). \tag{3.6}$$

Combining Equations (3.2) to (3.6), we have

$$\begin{aligned} \Big|\mathbb{E}e^{2\pi i t f(\xi)}\Big|^2 &\le \mathbb{E}\bigg[\prod_{\ell\in I}|\cos(2\pi t A_\ell)|\bigg] \\ &= \int_0^1 \Pr\bigg(\prod_{\ell\in I}|\cos(t A_\ell)| \ge x\bigg)\,\mathrm{d}\,x \\ &= \begin{cases} O\bigg(\Big(\dfrac{\log n}{\sqrt{n}}\Big)^r + \Big(\dfrac{\log n}{|t|n}\Big)^r\bigg) = O\bigg(\Big(\dfrac{\log n}{|t|n}\Big)^r\bigg) & \text{for } |t| \le \dfrac{1}{\sqrt{n}}, \\ O\bigg((|t|\log n)^r + \Big(\dfrac{\log n}{\sqrt{n}}\Big)^r\bigg) = O\Big((|t|\log n)^r\Big) & \text{for } |t| \ge \dfrac{1}{\sqrt{n}}. \end{cases} \end{aligned} \tag{3.7}$$

Finally, we apply Esséen's concentration inequality (Lemma 3.4) with $s = n^{2/(r+2)}$ and $\varepsilon = 1/s$ (note that we need to take the square root of the bound in (3.7)). This yields

$$\begin{aligned} \Pr\Big(|f(\xi) - x| \le n^{2/(r+2)}\Big) &\le O(s)\int_{-1/s}^{1/s} |\mathbb{E}e^{2\pi i t f(\xi)}|\,\mathrm{d}\,t \\ &= O\bigg(\frac{s}{n} + s\int_{1/n}^{1/\sqrt{n}}\Big(\frac{\log n}{tn}\Big)^{r/2}\,\mathrm{d}\,t + s\int_{1/\sqrt{n}}^{1/s}(\log n)^{r/2}t^{r/2}\,\mathrm{d}\,t\bigg) \\ &= O\bigg(\frac{s}{n} + s\Big(\frac{\log n}{n}\Big)^{r/2}\cdot(1/n)^{-r/2+1} + s(\log n)^{r/2}\cdot(1/s)^{r/2+1}\bigg) \\ &= O\Big((\log n)^{r/2}\Big(\frac{s}{n} + s^{-r/2}\Big)\Big) = O\Big((\log n)^{r/2}n^{-r/(r+2)}\Big). \qquad \square \end{aligned}$$

4 Real projections of complex nonsingular matrices

The anti-concentration inequality in the last section (Lemma 3.2) was only for quadratic polynomials with real coefficients, whereas in Theorems 1.1 and 1.2 we wish to consider complex polynomials as well. The following lemma will be an important tool to reduce from the complex case to the real case, and may be of independent interest. Recall that matrix norms in this paper are entrywise.

Lemma 4.1. *For every integer $r \geq 1$ and any $\varepsilon > 0$, there exists $c = c(r, \varepsilon) > 0$ such that the following holds. Let A be a complex $r \times r$ matrix with $|\det A| \geq \varepsilon$ and $\|A\|_\infty \leq 1$. Let $\theta \in [-\pi, \pi]$ be a uniformly random phase. Then with probability at least c, the matrix $\Re\big(e^{i\theta}A\big)$ satisfies $|\det \Re\big(e^{i\theta}A\big)| \geq c$.*

Proof. Define the polynomial $p(z) = \det\big(z^2 A + \overline{A}\big)$, so that

$$\det\big(\Re\big(e^{i\theta}A\big)\big) = \det\left(\frac{e^{i\theta}A + e^{-i\theta}\overline{A}}{2}\right) = \frac{e^{-ir\theta}}{2^r} p\big(e^{i\theta}\big).$$

Let ϕ be the phase of $p(0) = \det \overline{A}$. We have $\mathbb{E}\Re\big(e^{-i\phi + ik\theta}\big) = 0$ for each positive integer k, so

$$\mathbb{E}\big|\det\big(\Re\big(e^{i\theta}A\big)\big)\big| = 2^{-r}\mathbb{E}\big|p\big(e^{i\theta}\big)\big| \geq 2^{-r}\mathbb{E}\Re\big(e^{-i\phi}p\big(e^{i\theta}\big)\big) = 2^{-r}\Re\big(e^{-i\phi}p(0)\big) \geq \varepsilon 2^{-r}.$$

On the other hand, since the entries of $\big|\Re\big(e^{i\theta}A\big)\big|$ each have absolute value at most 1, we have $\big|\det\big(\Re\big(e^{i\theta}A\big)\big)\big| \leq r!$. So, by Markov's inequality,

$$\Pr\big(\big|\det\big(\Re\big(e^{i\theta}A\big)\big)\big| \geq c\big) = 1 - \Pr\big(r! - \big|\det\big(\Re\big(e^{i\theta}A\big)\big)\big| \geq r! - c\big) \geq 1 - \frac{r! - c}{r! - \varepsilon 2^{-r}}.$$

For sufficiently small $c > 0$, this probability is at least c. □

In this section we also prove that if a small matrix has bounded entries and determinant bounded away from zero, then its least singular value is bounded away from zero as well.

Lemma 4.2. *For some integer $q \geq 1$, let B be a complex nonsingular $q \times q$ matrix with $\|B\|_\infty \leq 1$. Then for any unit vector $e \in \mathbb{C}^q$, the vector Be satisfies $\|Be\|_1 \geq (q!)^{-1} \cdot |\det B|$.*

Proof. Let $v = Be$. First, we have $\|B^{-1}v\|_1 = \|e\|_1 \geq \|e\|_2 = 1$. On the other hand, observe that B^{-1} can be calculated from the determinant of B and the adjugate matrix of B. All entries of the adjugate matrix of B have absolute value at most $(q-1)!$, and therefore all entries of B^{-1} all have absolute value at most $(q-1)! \cdot |\det B|^{-1}$. Thus, each entry of the vector $e = B^{-1}v$ has absolute value at most $(q-1)! \cdot |\det B|^{-1} \cdot \|v\|_1$. It follows that

$$1 \leq \|B^{-1}v\|_1 \leq q \cdot (q-1)! \cdot |\det B|^{-1} \cdot \|v\|_1,$$

from which the desired result immediately follows. □

5 Deducing the main theorems

In this section we explain how to prove Theorems 1.1 and 1.2. Before getting into the details, we first observe that Theorem 1.2 actually follows from a slight variant of Theorem 1.1, where we control "coefficient-L_1 norm" but we demand that certain coefficients lie in a certain finite set.

Theorem 5.1. *Let $\mathbb{F} \in \{\mathbb{C}, \mathbb{R}, \mathbb{Q}\}$. For any integer $r \geq 3$, any $0 < \varepsilon \leq 1$, and any finite set $S \subseteq \mathbb{F}$ with $|s| \leq 1$ for all $s \in S$, there is a constant $C = C(r, \varepsilon, S)$ and a finite set $S^* = S^*(r, S) \subseteq \mathbb{F}$ such that the following holds. Let $f \in \mathbb{F}[x_1, \ldots, x_n]$ be a quadratic polynomial, let $\tilde{f}$ be the homogeneous degree-2 part of f and assume that the coefficients of $\tilde{f}$ are elements of the set S. Let $\xi = (\xi_1, \ldots, \xi_n) \in \mathrm{Rad}^n$, and suppose that we have*

$$\sup_{x \in \mathbb{F}} \Pr(f(\xi) = x) \geq C \cdot \frac{(\log n)^{r/2}}{n^{1-2/(r+2)}}.$$

Then there is a quadratic form $h \in \mathbb{F}[x_1, \ldots, x_n]$ of rank strictly less than r such that the sum of the absolute values of the coefficients of $\tilde{f} - h$ is at most εn^2, and such that all coefficients of h are elements of the set S^*.

Proof of Theorem 1.2 given Theorem 5.1. First of all, by rescaling we may assume that the set S in Theorem 1.2 satisfies $|s| \leq 1$ for all $s \in S$. Now, let $S^* = S^*(r, S)$ be the finite set in Theorem 5.1, and let Δ be the minimum distance between two elements of $S^* \cup S$.

Let $f \in \mathbb{F}[x_1, \ldots, x_n]$ be as in Theorem 1.2, and note that we may assume that n is large with respect to ε. Let $\tilde{f}$ be the homogeneous degree-2 part of f. By Theorem 5.1 applied with the error parameter $\varepsilon \cdot \Delta/2$, we either have the desired inequality on point probabilities of $f(\xi)$, or there is a quadratic form $h \in \mathbb{F}[x_1, \ldots, x_n]$ of rank less than r with coefficients in S^* such that the sum of absolute values of $\tilde{f} - h$ is at most $\varepsilon \cdot (\Delta/2) \cdot n^2$. By the choice of Δ, this implies that h and $\tilde{f}$ differ in at most $(\varepsilon/2) \cdot n^2$ coefficients. Thus, h and f differ in at most $(\varepsilon/2) \cdot n^2 + n + 1 \leq \varepsilon n^2$ coefficients (if n is sufficiently large). □

Now, for the proofs of Theorems 1.1 and 5.1 we will need a robust version of linear independence.

Definition 5.2. Consider $\mathbb{F} \in \{\mathbb{C}, \mathbb{R}, \mathbb{Q}\}$. For any $0 \leq \varepsilon \leq 1$, let us say that vectors $v_1, \ldots, v_q \in \mathbb{F}^n$ are ε-dependent (over $\mathbb{F}$) if there are linearly dependent vectors $v_1', \ldots, v_q' \in \mathbb{F}^n$ with

$$\|v_1 - v_1'\|_1 + \cdots + \|v_q - v_q'\|_1 \leq \varepsilon n.$$

Otherwise, say that $v_1, \ldots, v_q$ are ε-independent.

Note that the usual notion of being linearly independent corresponds to being 0-independent. Also note that the empty collection of vectors (taking $q = 0$) is ε-independent for any $0 \leq \varepsilon \leq 1$. In Section 6 we will observe some more basic properties of ε-independence, and prove some analogues of simple facts about ordinary linear independence.

Recall that we already introduced the notion of being δ-non-degenerate in Definition 3.1, which can also be interpreted as a type of robust linear independence. The following lemma connects these two notions.

Lemma 5.3. *For any integer $r \geq 1$ and any $0 < \varepsilon \leq 1$, there is a constant $c = c(r, \varepsilon) > 0$ such that the following holds. Suppose that $v_1, \ldots, v_r \in \mathbb{R}^n$ are ε-independent vectors with $\|v_i\|_\infty \leq 1$ for each i. Then the $r \times n$ matrix with rows $v_1, \ldots, v_r$ is c-non-degenerate.*

We will also need a variant of Lemma 5.3 for complex vectors, incorporating our lemma concerning real projections of complex matrices (Lemma 4.1).

Lemma 5.4. *For any integer $r \geq 1$ and any $0 < \varepsilon \leq 1$, there is a constant $c = c(r, \varepsilon) > 0$ such that the following holds. Suppose $v_1, \ldots, v_r \in \mathbb{C}^n$ are ε-independent vectors with $\|v_i\|_\infty \leq 1$ for each i. Furthermore, let $\theta \in [-\pi, \pi]$ be a uniformly random phase. Then with probability at least c, the $r \times n$ matrix with rows $\Re(e^{i\theta} v_1), \ldots, \Re(e^{i\theta} v_r) \in \mathbb{R}^n$ is c-non-degenerate.*

We defer the proofs of Lemmas 5.3 and 5.4 to Section 6.

Now, the remaining ingredient for the proofs of Theorems 1.1 and 5.1 is the following lemma. It states that if a symmetric matrix does not "robustly" have rank at least r, then it must be close (in terms of entrywise L_1 norm) to a symmetric matrix of rank less than r.

Lemma 5.5. *Fix $\mathbb{F} \in \{\mathbb{C}, \mathbb{R}, \mathbb{Q}\}$ and an integer $r \geq 1$. Let $0 < \alpha \leq 1$ and $0 < \delta < 1/r$ and let $A \in \mathbb{F}^{n \times n}$ be a symmetric matrix with $\|A\|_\infty \leq 1$. Suppose that there do not exist δn disjoint r-tuples of α-independent rows of A. Then there exists a symmetric matrix $H \in \mathbb{F}^{n \times n}$ of rank less than r such that $\|A - H\|_1 \leq O(\delta + \alpha^{(6r)^{-r}}) \cdot n^2$.*

Here, the implicit constant in the O-term may depend on r. We defer the proof of Lemma 5.5 to Section 7.

We remark that Lemma 5.5 also implies the following corollary, which may be of independent interest: if a symmetric matrix is close to a matrix which has rank less than r, then it is close to a symmetric matrix with rank less than r.

Corollary 5.6. *Fix $\mathbb{F} \in \{\mathbb{C}, \mathbb{R}, \mathbb{Q}\}$ and an integer $r \geq 1$. Let $0 < \alpha \leq 1$ and let $A \in \mathbb{F}^{n \times n}$ be a symmetric matrix with $\|A\|_\infty \leq 1$. Suppose that there is a matrix $A' \in \mathbb{F}^{n \times n}$ of rank less than r such that $\|A - A'\|_1 \leq \alpha n^2$. Then there exists a symmetric matrix $H \in \mathbb{F}^{n \times n}$ of rank less than r such that $\|A - H\|_1 \leq O(\alpha^{1/(2 \cdot (6r)^r)}) \cdot n^2$.*

Proof. It suffices to prove that A cannot have $\alpha^{1/2}n$ disjoint r-tuples of $\alpha^{1/2}$-independent rows; we would then be able to apply Lemma 5.5. So, suppose there were $\alpha^{1/2}n$ disjoint r-tuples $(i_1, \dots, i_r) \in [n]^r$ such that the rows $\mathrm{row}_{i_1}(A), \dots, \mathrm{row}_{i_r}(A)$ are $\alpha^{1/2}$-independent.

As in Definition 3.1, for each of our r-tuples $(i_1, \dots, i_r)$ we use notation like $A(i_1, \dots, i_r)$ to represent the $r \times n$ submatrix of A with these rows. Since our r-tuples are disjoint, for at least one of our r-tuples $(i_1, \dots, i_r)$ we have $\|A(i_1, \dots, i_r) - A'(i_1, \dots, i_r)\|_1 \le \|A - A'\|_1/(\alpha^{1/2}n) \le \alpha^{1/2}n^2$. Since A' has rank less than r, the rows of $A'(i_1, \dots, i_r)$ are linearly dependent, so the rows of $A(i_1, \dots, i_r)$ are $\alpha^{1/2}$-dependent, a contradiction. □

In order to prove Theorem 5.1, we also need the following variant of Lemma 5.5, stating that if the entries of A lie in a finite set S, then the matrix H can be chosen such that its entries lie in a finite set S' (depending only on S and r).

Lemma 5.7. *For $\mathbb{F} \in \{\mathbb{C}, \mathbb{R}, \mathbb{Q}\}$, an integer $r \ge 1$, and a finite set $S \subseteq \mathbb{F}$ with $|s| \le 1$ for all $s \in S$, there is a finite set $S' = S'(r, S) \subseteq \mathbb{F}$ such that the following holds: let $0 < \alpha \le 1$ and $0 < \delta < 1/r$ and let $A \in \mathbb{F}^{n \times n}$ be a symmetric matrix all of whose entries are in S. Suppose that there do not exist δn disjoint r-tuples of α-independent rows of A. Then there exists a symmetric matrix $H \in \mathbb{F}^{n \times n}$ of rank less than r such that $\|A - H\|_1 \le O(\delta + \alpha^{(6r)^{-r}}) \cdot n^2$ and such that all entries of H are elements of S'.*

Now, we can deduce Theorems 1.1 and 5.1. The proofs are virtually the same (the only difference is whether we use Lemma 5.7 or Lemma 5.5), so we present both proofs together.

Proof of Theorems 1.1 and 5.1. First, choose some small $0 < \alpha \le 1$ and $0 < \delta < 1/r$ such that the $O(\delta + \alpha^{(6r)^{-r}})$-term in Lemmas 5.5 and 5.7 is at most ε. Also, note that we may assume n is sufficiently large with respect to ε. For each i, j, let a_{ij} be the coefficient of $x_i x_j$ and in f (so a_{ii} is the coefficient of x_i^2), and define the (symmetric) "coefficient matrix" $A = (a_{ij})_{i,j}$. Note that by our assumptions on f, $\|A\|_\infty \le 1$. We consider two cases.

Case 1: A does not have δn disjoint r-tuples of α-independent rows

By Lemma 5.5 (and our choice of δ and α), there exists a symmetric matrix $H = (h_{ij})_{1 \le i,j \le n} \in \mathbb{F}^{n \times n}$ of rank less than r such that $\|A - H\|_1 \le \varepsilon n^2$. In the setting of Theorem 5.1, we can instead apply Lemma 5.7 to get the same conclusion, with the additional property that all entries of H lie in some fixed finite set S' that only depends on r and S.

Now, let $h \in \mathbb{F}[x_1, \dots, x_n]$ be the quadratic form defined by $h(x) = \frac{1}{2}x^T H x = \sum_{i<j} h_{ij}x_i x_j + \sum_i (h_{ii}/2)x_i^2$, which also has rank less than r (and in the setting of Theorem 5.1, the coefficients of h lie in the finite set $S^* = S' \cup \{s/2 : s \in S'\}$). Let $\tilde{f}$ be the homogeneous degree-2 part of f. Then, the sum of the absolute values of the coefficients of $\tilde{f} - h$ is

$$\sum_{i<j} |a_{ij} - h_{ij}| + \sum_i |a_{ii} - h_{ii}/2| \le \frac{1}{2}\|A - H\|_1 + \sum_i |a_{ii}/2| \le (\varepsilon/2)n^2 + n/2.$$

For the proof of Theorem 5.1, this already gives the desired conclusion (for sufficiently large n). In the setting of Theorem 1.1, we additionally note that the sum of the absolute values of the coefficients of $\tilde{f} - f$ is at most $n + 1 = o(n^2)$.

Case 2: A has δn disjoint r-tuples of α-independent rows

In this case, we use Lemma 3.2 to prove that

$$\sup_{x \in \mathbb{F}} \Pr(f(\xi) = x) = O\left(\frac{(\log n)^{r/2}}{n^{1-2/(r+2)}}\right),$$

showing that the assumptions of Theorems 1.1 and 5.1 cannot hold for large C. (For the rest of the proof, all asymptotic notation treats r, δ and α as fixed constants.) Let $T_A \subseteq [n]^r$ be a collection of δn disjoint r-tuples, such that for any $(i_1, \dots, i_r) \in T_A$ the corresponding rows of A are α-independent.

If $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{Q}$, then by Lemma 5.3, for each $(i_1, \dots, i_r) \in T_A$, the $r \times n$ matrix of $A(i_1, \dots, i_r)$, defined as in Definition 3.1, is c-non-degenerate for $c = \Omega(1)$. Thus, Lemma 3.2 implies that for random $\xi = (\xi_1, \dots, \xi_n) \in \mathrm{Rad}^n$, we have

$$\sup_{x \in \mathbb{F}} \Pr(f(\xi) = x) \leq \sup_{x \in \mathbb{F}} \Pr\Big(|f(\xi) - x| \leq n^{2/(r+2)}\Big) = O\left(\frac{(\log n)^{r/2}}{n^{1-2/(r+2)}}\right),$$

as desired.

For the case $\mathbb{F} = \mathbb{C}$, choose $c = \Omega(1)$ such that if $\theta \in [-\pi, \pi]$ is a uniformly random phase then for each $(i_1, \dots, i_r) \in T_A$, the matrix $\Re(e^{i\theta} A)(i_1, \dots, i_r)$ is c-degenerate with probability at least c. Such a c exists by Lemma 5.4. Then, let T_θ be the set of all $(i_1, \dots, i_r) \in T_A$ such that $\Re(e^{i\theta} A)(i_1, \dots, i_r)$ is c-degenerate. For random θ we have $\mathbb{E}|T_\theta| \geq c\delta n$, so we can fix θ such that $|T_\theta| \geq c\delta n$. Then, let f^* be the polynomial obtained from $e^{i\theta} f$ by taking the real part of each coefficient, so for $\xi \in \mathrm{Rad}^n$ Lemma 3.2 implies

$$\sup_{x \in \mathbb{C}} \Pr(f(\xi) = x) \leq \sup_{x \in \mathbb{C}} \Pr\big(\Re(e^{i\theta} f(\xi)) = \Re(e^{i\theta} x)\big) = \sup_{x \in \mathbb{R}} \Pr(f^*(\xi) = x) = O\left(\frac{(\log n)^{r/2}}{n^{1-2/(r+2)}}\right),$$

as desired. □

6 Lemmas on robust linear independence

In this section we prove Lemmas 5.3 and 5.4, and several other auxiliary lemmas concerning ε-independence (defined in Definition 5.2). Throughout this section we fix $\mathbb{F} \in \{\mathbb{C}, \mathbb{R}, \mathbb{Q}\}$.

First, with a view towards proving Lemmas 5.3 and 5.4, we start with the fact that if the rows of a $q \times n$ matrix are ε-independent then there is a $q \times q$ submatrix with large determinant.

Lemma 6.1. *For any $0 \leq \varepsilon \leq 1$ and any ε-independent vectors $v_1, \dots, v_q \in \mathbb{F}^n$, the $q \times n$ matrix with rows $v_1, \dots, v_q$ has a $q \times q$ submatrix whose determinant has absolute value at least ε^q.*

Lemma 6.1 will be an immediate consequence of a more general lemma (Lemma 6.3) that we prove later in this section. Now we prove Lemma 5.3.

Proof of Lemma 5.3. We will take $c = \varepsilon^r/(2^r r! r)$. Let M be the $r \times n$ matrix with rows $v_1, \dots, v_r$.

First, we claim that M has $m \geq \varepsilon/(2r^2) \cdot n$ disjoint $r \times r$ submatrices $B_1, \dots, B_m$ whose determinants have absolute value at least $(\varepsilon/2)^r$. Indeed, consider a maximal collection such disjoint submatrices and suppose that this collection consists of fewer than $\varepsilon/(2r^2) \cdot n$ submatrices. Let $M' \in \mathbb{R}^{r \times n}$ be obtained from M by setting all the entries in our identified submatrices to zero. By maximality, every $r \times r$ submatrix of M' has determinant bounded in absolute value by $(\varepsilon/2)^r$. On the other hand, since $\|M\|_\infty \leq 1$, we have $\|M - M'\|_1 \leq \varepsilon/(2r^2) \cdot n \cdot r^2 \leq (\varepsilon/2)n$. Since the rows of M are ε-independent, this shows that the rows of M' are $(\varepsilon/2)$-independent. But then, by Lemma 6.1, the matrix M' has a $r \times r$ submatrix whose determinant has absolute value at least $(\varepsilon/2)^r$, which is a contradiction.

Now, in order to show that M is c-non-degenerate, we need to show that for every unit vector $e \in \mathbb{R}^r$, there are at least cn columns w of M such that $|\langle w, e\rangle| \geq c$. As $m \geq \varepsilon/(2r^2) \cdot n \geq cn$, it suffices to show that for each $j = 1, \dots, m$ there is a column w of B_j with $|\langle w, e\rangle| \geq c$. This is equivalent to showing that the vector $B_j^T e \in \mathbb{R}^r$ has at least one entry with absolute value at least c. However, since $|\det B_j^T| \geq (\varepsilon/2)^r$, Lemma 4.2 implies that $\|B_j^T e\|_1 \geq (r!)^{-1} \cdot (\varepsilon/2)^r = r \cdot c$. Therefore one of the r entries of $B_j^T e$ must indeed have absolute value at least c. This finishes the proof of Lemma 5.3. □

Next, to prove Lemma 5.4, we modify the proof Lemma 5.3 to incorporate our lemma concerning real projections of complex matrices (Lemma 4.1).

Proof of Lemma 5.4. Let M be the $r \times n$ matrix with rows $v_1, \dots, v_r$. As in the proof of Lemma 5.3, in M we can find $m \geq \varepsilon/(2r^2) \cdot n$ disjoint $r \times r$ submatrices $B_1, \dots, B_m$ whose determinants have absolute value at least $(\varepsilon/2)^r$.

Consider a random phase $\theta \in [-\pi, \pi]$. By Lemma 4.1, for some $0 < c' < 1$ only depending on r and ε, for each $1 \le j \le m$, with probability at least c' we have $|\det \Re(e^{i\theta} B_j)| \ge c'$. Let J be the random set of j such that this holds, and observe that $\mathbb{E}|J| \ge c'm$. Then, since $|J| - c'm/2 \le m$ we have $m \Pr(|J| - c'm/2 \ge 0) \ge \mathbb{E}[|J| - c'm/2]$, so

$$\Pr(|J| \ge c'm/2) = \Pr(|J| - c'm/2 \ge 0) \ge \frac{\mathbb{E}[|J| - c'm/2]}{m} \ge \frac{c'm - c'm/2}{m} = \frac{c'}{2}.$$

But, if $|J| \ge c'm/2$ then $\Re(e^{i\theta} M)$ has $c'm/2 \ge c'\varepsilon/(4r^2) \cdot n$ disjoint $r \times r$ submatrices whose determinants have absolute value at least c'. As in the proof of Lemma 5.3, it follows from Lemma 4.2 that $\Re(e^{i\theta} M)$ is c-non-degenerate for some $0 < c < c'/2$ depending only on ε and r (via c'). The desired result follows. □

In the remainder of this section, we prove some simple facts about ε-independence that will be useful for the proof of Lemma 5.5 in Section 7. From now on, fix any $\mathbb{F} \in \{\mathbb{C}, \mathbb{R}, \mathbb{Q}\}$.

The following lemma says that for ε-dependent vectors $v_1, \dots, v_q$ with entries of absolute value at most 1, the vectors $v'_1, \dots, v'_q$ in Definition 5.2 can be chosen in such a way that their entries have absolute value at most $q + 1$.

Lemma 6.2. *Let $0 \le \varepsilon \le 1$ and let $v_1, \dots, v_q \in \mathbb{F}^n$ be ε-dependent such that $\|v_i\|_\infty \le 1$ for each i. Then there are linearly dependent vectors $v'_1, \dots, v'_q \in \mathbb{F}^n$ with*

$$\|v_1 - v'_1\|_1 + \cdots + \|v_q - v'_q\|_1 \le \varepsilon n$$

and such that all entries of the vectors $v'_1, \dots, v'_q$ have absolute value at most $q + 1$.

Proof. By the definition of $v_1, \dots, v_q \in \mathbb{F}^n$ being ε-dependent, there are linearly dependent vectors $v'_1, \dots, v'_q \in \mathbb{F}^n$ with

$$\|v_1 - v'_1\|_1 + \cdots + \|v_q - v'_q\|_1 \le \varepsilon n.$$

It may be the case that for one or more indices i, the i-th entry of one of the vectors $v'_1, \dots, v'_q$ has absolute value larger than $q + 1$. For all such i, let us change the i-th entry of each of the vectors $v'_1, \dots, v'_q$ to zero. It is not hard to see that this does not increase the value of $\|v_1 - v'_1\|_1 + \cdots + \|v_q - v'_q\|_1$, and that the new vectors $v'_1, \dots, v'_q$ are still linearly dependent. □

The next lemma is a generalisation of Lemma 6.1 where we can specify some forbidden pairs of vectors.

Lemma 6.3. *Let $0 \le \varepsilon \le 1$ and $0 \le \delta \le 1$. For some integer $q \ge 0$, let M be a $q \times n$ matrix whose rows are $(\varepsilon + q(q-1)\delta)$-independent and whose entries have absolute value at most 1. Consider a graph G on the vertex set $[n]$ in which every vertex has degree at most δn. Then there exists a q-vertex independent set $I \subseteq [n]$ of G, such that the $q \times q$ matrix M_I satisfies $|\det M_I| \ge \varepsilon^q$.*

Note that Lemma 6.1 follows from Lemma 6.3 by taking $\delta = 0$ and the graph G with no edges.

Proof of Lemma 6.3. Let $\varepsilon' = \varepsilon + q(q-1)\delta$. We prove the lemma by induction on q. Note that the base case $q = 0$ is trivial (since we adopted the convention that the determinant of the 0×0 empty matrix is 1). So assume that $q \ge 1$ and that we have already proved the lemma for $q - 1$. Let $M = (a_{ij})_{i,j}$ be the $q \times n$ matrix with rows $v_1, \dots, v_q$. As $v_1, \dots, v_q$ are ε'-independent, the vectors $v_1, \dots, v_{q-1}$ are ε'-independent as well and in particular $(\varepsilon + (q-1)(q-2)\delta)$-independent. Let M' be the matrix obtained from M by removing the last row. Then, by the induction hypothesis, there is a $(q-1)$-vertex independent set $I' \subseteq [n]$ of G such that $|\det M'_{I'}| \ge \varepsilon^{q-1}$. Let us assume without loss of generality that $I' = [q-1]$.

Now, let $J \subseteq \{q, \dots, n\}$ be the set of those vertices which have a neighbour in I' in the graph G. By the degree assumption, we have $|J| \le (q-1)\delta n$. Then, consider all the $q \times q$ submatrices of M formed by the first $q - 1$ columns together with a column which has index in $\{q, \dots, n\} \setminus J$. Assume for contradiction that for all of these submatrices the absolute value of their determinant is smaller than ε^q.

We want to modify M to create a matrix M' contradicting the definition of ε'-independence. First, for $j \in \{q, \dots, n\} \setminus J$, let us change the entry a_{qj} in such a way that the determinant of the $q \times q$ submatrix of M formed by the first $q - 1$ columns together with the j-th column becomes zero. By the assumption in the last paragraph, we need to change a_{qj} by at most $\varepsilon^q / |\det M'_{I'}| \le \varepsilon^q / \varepsilon^{q-1} = \varepsilon$ in order to achieve this. Second, for every $j \in J$ change all the entries in column j to zero. The resulting matrix M' then satisfies $\|M - M'\|_1 \le \varepsilon(n - (q-1)) + q(q-1)\delta n \le \varepsilon' n$. But by construction, the first $q - 1$ columns of M' span its entire column space, so M' has rank less than q and its rows are not linearly independent. This is in contradiction to the rows of M being ε'-independent. □

Next, the following lemma states that for robustly independent vectors $v_1, \dots, v_q \in \mathbb{F}^n$ and a vector $v \in \mathbb{F}^n$, either $v_1, \dots, v_q$ and v are robustly independent together, or v is close to a linear combination of $v_1, \dots, v_q$.

Lemma 6.4. *Fix an integer $q \geq 0$. Let $0 < \varepsilon \leq 1$ and let $v_1, \dots, v_q \in \mathbb{F}^n$ be ε-independent vectors such that $\|v_i\|_\infty \leq 1$ for each i. Furthermore let $0 \leq \delta < \varepsilon/2$ and let $v \in \mathbb{F}^n$ be a vector with $\|v\|_\infty \leq 1$. Then at least one of the following two conditions holds:*

(a) $v_1, \dots, v_q, v$ are δ-independent, or
(b) there is a vector $v^* \in \operatorname{span}(v_1, \dots, v_q)$ with $\|v - v^*\|_1 \leq O(\varepsilon^{-q}\delta) \cdot n$.

Here the implicit constant in the O-term may depend on q.

Proof. Suppose (a) is not satisfied, so $v_1, \dots, v_q, v$ are δ-dependent. Thus, by Lemma 6.2 there are linearly dependent vectors $v_1', \dots, v_q', v' \in \mathbb{F}^n$ with

$$\|v_1 - v_1'\|_1 + \cdots + \|v_q - v_q'\|_1 + \|v - v'\|_1 \leq \delta n, \tag{6.1}$$

such that all entries of $v_1', \dots, v_q', v'$ have absolute value at most $q+1$. Now, since $v_1, \dots, v_q \in \mathbb{F}^n$ are ε-independent, and $\delta < \varepsilon/2$, the vectors $v_1', \dots, v_q'$ are $(\varepsilon/2)$-independent, and in particular linearly independent. Thus, as $v_1', \dots, v_q', v'$ are linearly dependent, we can write v' as $v' = a_1 v_1' + \cdots + a_q v_q'$ for some $a_1, \dots, a_q \in \mathbb{F}$.

We claim that $|a_i| = O(\varepsilon^{-q})$ for all i. Indeed, since $v_1', \dots, v_q'$ are $(\varepsilon/2)$-independent, by Lemma 6.1 the matrix with rows $v_1', \dots, v_q'$ has a $q \times q$ submatrix A whose determinant has absolute value $\Omega(\varepsilon^q)$. Let $I \subseteq [n]$ be the set of the indices of the columns contained in this submatrix. Consider the row vector $(v')_I \in \mathbb{F}^q$ obtained from v' by taking the coordinates indexed by I. Now, since $v' = a_1 v_1' + \cdots + a_q v_q'$ we have $(a_1, \dots, a_q)A = (v')_I$, so $(a_1, \dots, a_q) = (v')_I A^{-1}$. Since $\|A\|_\infty \leq q+1 = O(1)$, and $|\det A| = \Omega(\varepsilon^q)$, we can see (from the formula for A^{-1} in terms of the adjugate of A) that all entries of the matrix A^{-1} are of the form $O(\varepsilon^{-q})$. Since $\|(v')_I\|_\infty \leq q+1 = O(1)$ as well, we can conclude that the absolute values of $a_1, \dots, a_q$ are of the form $O(\varepsilon^{-q})$, as claimed.

Now, define the linear combination $v^* = a_1 v_1 + \cdots + a_q v_q$. We have

$$\|v - v^*\|_1 \leq \|v - v'\|_1 + \|v' - v^*\|_1 = \|v - v'\|_1 + \big\|a_1(v_1' - v_1) + \cdots + a_q(v_q' - v_q)\big\|_1.$$

Recalling (6.1) and that $|a_i| = O(\varepsilon^{-q})$ for all i, we deduce that $\|v - v^*\|_1 \leq O(\varepsilon^{-q}) \cdot \delta n$, so (b) holds. □

We then deduce the following lemma. It is a "robust version" of the fact that if a list of vectors does not contain r linearly independent vectors, then among the vectors on this list we can find a linearly independent set of size less than r whose span contains the entire list.

Lemma 6.5. *Fix a positive integer r. Consider some $0 < \varepsilon \leq 1/2$, and consider a list of vectors $v_1, \dots, v_k \in \mathbb{F}^n$ such that $\|v_i\|_\infty \leq 1$ for all i. Suppose that there is no subset of r vectors from this list which are $\varepsilon^{(6r)^r}$-independent. Then for some $0 \leq q \leq r-1$, we can choose vectors $w_1, \dots, w_q$ among the list $v_1, \dots, v_k$ such that both of the following conditions are satisfied:*

(i) $w_1, \dots, w_q$ are $\varepsilon^{(6r)^q}$-independent, and
(ii) for each $i = 1, \dots, k$, there is a vector $v_i' \in \operatorname{span}(w_1, \dots, w_q)$ with $\|v_i - v_i'\|_1 \leq O(\varepsilon^{5r \cdot (6r)^q}) \cdot n$.

Here the implicit constant in the O-term may depend on r.

Proof. Let us choose a collection of $\varepsilon^{(6r)^q}$-independent vectors $w_1, \dots, w_q$ among the list $v_1, \dots, v_k$, with $q \in \{0, \dots, r\}$ as large as possible (this is well-defined, since $q = 0$ is definitely possible). By our assumption on $v_1, \dots, v_k$ we must have $q \leq r-1$.

We need to check condition (ii). Note that for all i for which v_i is one of the vectors $w_1, \dots, w_q$, condition (ii) holds trivially with $v_i' = v_i$. For all other $1 \leq i \leq k$, the $q+1$ vectors $w_1, \dots, w_q, v_i$ cannot be $\varepsilon^{(6r)^{q+1}}$-independent by maximality of q, so by Lemma 6.4 there is $v_i' \in \operatorname{span}(w_1, \dots, w_q)$ with

$$\|v_i - v_i'\|_1 \leq O\left(\left(\varepsilon^{(6r)^q}\right)^{-q} \cdot \varepsilon^{(6r)^{q+1}}\right) \cdot n = O\left(\varepsilon^{5r \cdot (6r)^q}\right) \cdot n. \qquad \square$$

Finally, the following lemma is not strictly about ε-independence but will be used several times in the proofs of Lemmas 5.5 and 5.7. For given vectors $w_1, \dots, w_q$, the lemma is about finding a vector $v \in \operatorname{span}(w_1, \dots, w_q)$ with certain prescribed coordinates, and controlling the distance of v to another given vector $\tilde{v}$ in this span.

Lemma 6.6. *Fix a non-negative integer q. Let $w_1, \dots, w_q \in \mathbb{F}^n$ be vectors satisfying $\|w_i\|_\infty \le 1$ for each i, and let M be the $q \times n$ matrix whose rows are the vectors $w_1, \dots, w_q$. Consider some subset $I \subseteq [n]$ of size q, and suppose that the $q \times q$ matrix M_I satisfies $\det M_I \neq 0$.*

Now, for each $i \in I$ let us specify a value $v^{(i)} \in \mathbb{F}$. Then there is a unique vector $v \in \operatorname{span}(w_1, \dots, w_q)$ having the prescribed values $v^{(i)}$ in the coordinates $i \in I$. Furthermore, for any vector $\tilde{v} \in \operatorname{span}(w_1, \dots, w_q)$ we have

$$\|v - \tilde{v}\|_1 \le O(|\det M_I|^{-1}) \cdot n \cdot \|\tilde{v}_I - v_I\|_1.$$

Here the implicit constant in the O-term may depend on q.

Proof. The existence and uniqueness of v follow directly from the fact that $\det M_I \neq 0$. Then, consider some vector $\tilde{v} \in \operatorname{span}(w_1, \dots, w_q)$, and write $v - \tilde{v} = a_1 w_1 + \cdots + a_q w_q$ for some $a_1, \dots, a_q \in \mathbb{F}$.

Note that we can determine the coefficients $a_1, \dots, a_q$ by the equation $(a_1, \dots, a_q) M_I = (v - \tilde{v})_I$, where we interpret $(v - \tilde{v})_I \in \mathbb{F}^I$ as a row vector. In other words, we have $(a_1, \dots, a_q) = (v - \tilde{v})_I M_I^{-1}$. The entries of M_I^{-1} have absolute value $O(|\det M_I|^{-1})$ (by the formula for the inverse of a matrix in terms of its determinant and its adjugate). Thus, the absolute values of $a_1, \dots, a_q$ are of the form $O(|\det M_I|^{-1}) \cdot \|\tilde{v}_I - v_I\|_1$. Since $\|w_i\|_\infty \le 1$ for each i, we conclude

$$\|v - \tilde{v}\|_1 = \|a_1 w_1 + \cdots + a_q w_q\|_1 \le (|a_1| + \cdots + |a_q|) \cdot n \le O(|\det M_I|^{-1}) \cdot n \cdot \|\tilde{v}_I - v_I\|_1. \qquad \square$$

7 Proving closeness to a low-rank symmetric matrix

7.1 Proof of Lemma 5.5

In this section, we finally prove Lemma 5.5: given a symmetric matrix A which does not have δn disjoint r-tuples of α-independent rows, we show that there is a symmetric matrix H that is close to A (in terms of the entrywise L_1 norm) and has rank less than r. We outline the approach with a sequence of claims, whose proofs we will provide afterwards. We assume that α is sufficiently small, and for all asymptotic notation we treat r as a constant.

First, we find a symmetric matrix $A^* \in \mathbb{F}^{n \times n}$ which is close to A and does not have any r-tuple of α-independent rows, as in the following claim.

Claim 7.1. *Consider the setting of Lemma 5.5. Then we can find a symmetric matrix $A^* \in \mathbb{F}^{n \times n}$ with $\|A^*\|_\infty \le 1$, such that $\|A^* - A\|_1 \le O(\delta) \cdot n^2$ and such that the matrix A^* does not have any r-tuple of α-independent rows.*

Second, we use Lemma 6.5 to identify a subset of rows $w_1, \dots, w_q$ of A^*, where $0 \le q \le r - 1$, such that every row of A^* can be approximated by a linear combination of $w_1, \dots, w_q$. We then form a matrix B^* by replacing the rows of A^* by these approximations. This will give the following.

Claim 7.2. *Let the matrix A^* be as in Claim 7.1. For some $0 \le q \le r - 1$ and $0 < \tilde{\alpha} \le \alpha^{(6r)^{-r}}$ we can find $\tilde{\alpha}$-independent rows $w_1, \dots, w_q$ of A^*, and a matrix $B^* \in \mathbb{F}^{n \times n}$, such that each row of B^* lies in $\operatorname{span}(w_1, \dots, w_q)$ and such that $\|\operatorname{row}_i(A^*) - \operatorname{row}_i(B^*)\|_1 \le O(\tilde{\alpha}^{4r}) \cdot n$ for all $i \in [n]$.*

Claim 7.2 ensures that each row of A^* is close to the corresponding row of B^*. However, we have no control over the columns of A^* and B^*. In the next step, we "zero out" some rows and columns of A^* and B^* to obtain matrices A' and B' such that every row and column of A' is close to the corresponding row or column of B'.

Claim 7.3. *Consider A^*, B^*, $\tilde{\alpha}$ and $w_1, \dots, w_q$ as in Claim 7.2. We can find a symmetric matrix $A' \in \mathbb{F}^{n \times n}$, a matrix $B' \in \mathbb{F}^{n \times n}$, and $(\tilde{\alpha}/2)$-independent vectors $w'_1, \dots, w'_q \in \mathbb{F}^n$ such that each row of B' lies in $\operatorname{span}(w'_1, \dots, w'_q)$, such that each $\|w'_i\|_\infty \leq 1$, and such that we have*

$$\|A' - A^*\|_1 \leq O(\tilde{\alpha}^r) \cdot n^2$$

and, for each $i \in [n]$,

$$\|\operatorname{row}_i(A') - \operatorname{row}_i(B')\|_1 \leq O(\tilde{\alpha}^{3r}) \cdot n \quad \text{and} \quad \|\operatorname{col}_i(A') - \operatorname{col}_i(B')\|_1 \leq O(\tilde{\alpha}^{3r}) \cdot n.$$

While Claim 7.3 ensures that each row or column of A' is close to the corresponding row or column of B', it does not give control over individual entries of A'. However, the following claim (proved using Lemma 6.3) states that we can find a subset $I \subseteq [n]$ such that for all distinct $i, j \in I$, the (i,j)-entry of A' is close to the (i,j)-entry of B'.

Claim 7.4. *Let $A' = (a'_{ij})_{i,j}$, $B' = (b'_{ij})_{i,j}$ and $w'_1, \dots, w'_q$ be as in Claim 7.3, and let M be the $q \times n$ matrix whose rows are $w'_1, \dots, w'_q$. There is a subset $I \subseteq [n]$ of size q such that for all distinct $i, j \in I$ we have $|a'_{ij} - b'_{ij}| < \tilde{\alpha}^{2r}$, and such that the $q \times q$ matrix M_I satisfies $|\det M_I| \geq \Omega(\tilde{\alpha}^{r-1})$.*

We can then use this subset $I \subseteq [n]$ to construct our final matrix H, in the following claim.

Claim 7.5. *Let $A' = (a'_{ij})_{i,j}$, $B' = (b'_{ij})_{i,j}$ and $w'_1, \dots, w'_q$ be as Claim 7.3 and let $I \subseteq [n]$ be as in Claim 7.4. Define $h_{ii} = b'_{ii}$ for all $i \in I$ and $h_{ij} = a'_{ij}$ for all distinct $i, j \in I$. Then we can extend these values to a symmetric matrix $H = (h_{ij}) \in \mathbb{F}^{n \times n}$ such that every row of H is in $\operatorname{span}(w'_1, \dots, w'_q)$ and such that $\|H - B'\| \leq O(\tilde{\alpha}) \cdot n^2$.*

The proof of Claim 7.5 will require Lemma 6.6 and the following technical lemma.

Lemma 7.6. *Consider vectors $v_1, \dots, v_q \in \mathbb{F}^n$, and let M be the matrix with these vectors as rows. Consider a subset $I \subseteq [n]$ with size q, such that the $q \times q$ matrix M_I is invertible. Let $H = (h_{ij})_{i,j} \in \mathbb{F}^{n \times n}$ be a matrix each of whose rows is in $\operatorname{span}(v_1, \dots, v_q)$, such that $h_{ij} = h_{ji}$ for all $i \in I$ and all $j \in [n]$. Then H is symmetric.*

It is not hard to see that Claims 7.1 to 7.5 imply Lemma 5.5. Indeed, the symmetric matrix H in Claim 7.5 clearly has rank at most $q \leq r - 1$. Furthermore, using that $\|A' - B'\|_1 = \sum_{i=1}^n \|\operatorname{row}_i(A') - \operatorname{row}_i(B')\|_1 \leq O(\tilde{\alpha}^{3r}) \cdot n^2$ by Claim 7.3, we obtain

$$\|H - A\|_1 \leq \|H - B'\|_1 + \|B' - A'\|_1 + \|A' - A^*\|_1 + \|A^* - A\|_1 \leq O(\tilde{\alpha} + \delta) \cdot n^2 = O(\alpha^{(6r)^{-r}} + \delta) \cdot n^2.$$

It remains to prove Claims 7.1 to 7.5 and Lemma 7.6.

Proof of Claim 7.1. By assumption, there do not exist δn disjoint r-tuples of α-independent rows of A. Choose a maximal collection of such r-tuples and let $J \subseteq [n]$ be the set of all rows involved (so $|J| \leq r\delta n$). Let $A^* = (a^*_{ij})_{i,j} \in \mathbb{F}^{n \times n}$ be the symmetric matrix obtained from A by setting to zero all rows and all columns with indices in J. As $\|A\|_\infty \leq 1$, we have $\|A^* - A\|_1 \leq 2|J|n \leq 2r\delta n^2 = O(\delta) \cdot n^2$. We claim that A^* does not have any r-tuple of α-independent rows. Clearly, no r-tuple containing a zero row can be α-independent. For any r-tuple of rows of A^* with indices in $[n] \setminus J$, the corresponding r-tuple of rows in A must be α-dependent (by the maximality of the collection chosen in the beginning). It is not hard to see that this r-tuple stays α-dependent when setting to zero the columns with indices in J. □

Proof of Claim 7.2. We can apply Lemma 6.5 with $\varepsilon = \alpha^{(6r)^{-r}}$ and the list of rows $v_1, \dots, v_n$ of A^*, to obtain $\alpha^{(6r)^{q-r}}$-independent rows $w_1, \dots, w_q$ of A^* for some $0 \leq q \leq r - 1$, as well as a list of vectors $v'_1, \dots, v'_n \in \operatorname{span}(w_1, \dots, w_q)$ satisfying $\|v_i - v'_i\|_1 \leq O(\alpha^{5r(6r)^{q-r}})n \leq O(\alpha^{4r(6r)^{q-r}})n$ for all $i \in [n]$ (the wasteful second inequality here will make it easier to explain how to adapt this proof to prove Lemma 5.7 in Subsection 7.2). Then define $\tilde{\alpha} = \alpha^{(6r)^{q-r}} \leq \alpha^{(6r)^{-r}}$ and let $B^* \in \mathbb{F}^{n \times n}$ be the matrix with rows $v'_1, \dots, v'_n$. □

Proof of Claim 7.3. Note that by the assumptions on A^* and B^*, we have

$$\sum_{i=1}^{n} \|\operatorname{col}_i(A^*) - \operatorname{col}_i(B^*)\|_1 = \|A^* - B^*\|_1 = \sum_{i=1}^{n} \|\operatorname{row}_i(A^*) - \operatorname{row}_i(B^*)\|_1 \leq O(\tilde{\alpha}^{4r}) \cdot n^2.$$

Defining $J \subseteq [n]$ be the set of those indices $j \in [n]$ such that $\|\operatorname{col}_j(A^*) - \operatorname{col}_j(B)\|_1 \geq \tilde{\alpha}^{3r} \cdot n$, we obtain that $|J| \leq O(\tilde{\alpha}^r) \cdot n$. Let the matrices A' and B' be obtained from A^* and B^* by setting to zero all the entries in all rows and columns with indices in J. Similarly, let the vectors $w'_1, \dots w'_q$ be obtained from $w_1, \dots, w_q$ by setting the entries with indices in J to zero. Since each row of B^* lies in $\operatorname{span}(w_1, \dots, w_q)$, each row of B' lies in $\operatorname{span}(w'_1, \dots, w'_q)$ (note that this is trivially true for the all-zero rows with indices in J).

Note that A' is symmetric, since A^* is symmetric. Furthermore, $\|A^*\|_\infty \leq 1$, so $\|w_i\|_\infty \leq 1$ and $\|w'_i\|_\infty \leq 1$ for each i. We claim that $w'_1, \dots, w'_q$ are $(\tilde{\alpha}/2)$-independent. For $q = 0$ this is trivially true, so we may assume that $q \geq 1$ and therefore $r \geq 2$. Recall that $w_1, \dots, w_q$ are $\tilde{\alpha}$-independent, and that we changed only $r|J| = O(\tilde{\alpha}^r) \cdot n \leq (\tilde{\alpha}/2) \cdot n$ entries of $w_1, \dots, w_q$ to zero to obtain $w'_1, \dots, w'_q$. Thus, the vectors $w'_1, \dots, w'_q$ are indeed $(\tilde{\alpha}/2)$-independent.

We have $\|A' - A^*\|_1 \leq 2|J|n \leq O(\tilde{\alpha}^r) \cdot n^2$. For each $i \in J$, we have $\operatorname{row}_i(A^*) = 0 = \operatorname{row}_i(B^*)$ and $\operatorname{col}_i(A^*) = 0 = \operatorname{col}_i(B^*)$. On the other hand, for $i \in [n] \setminus J$, we have $\|\operatorname{row}_i(A') - \operatorname{row}_i(B')\|_1 \leq \|\operatorname{row}_i(A^*) - \operatorname{row}_i(B^*)\|_1 \leq O(\tilde{\alpha}^{4r}) \cdot n$ by the properties in Claim 7.2 and $\|\operatorname{col}_i(A') - \operatorname{col}_i(B')\|_1 \leq \|\operatorname{col}_i(A^*) - \operatorname{col}_i(B^*)\|_1 < \tilde{\alpha}^{3r} \cdot n$ by the definition of J. □

Proof of Claim 7.4. Note that the case $q = 0$ is trivial, so we may assume that $q \geq 1$ and therefore $r \geq 2$. Consider the graph G on the vertex set $[n]$ where for any $1 \leq i < j \leq n$ we draw an edge between the vertices i and j if $|a'_{ij} - b'_{ij}| \geq \tilde{\alpha}^{2r}$ or if $|a'_{ji} - b'_{ji}| \geq \tilde{\alpha}^{2r}$. By the last part of Claim 7.3, this graph has maximum degree $O(\tilde{\alpha}^r) \cdot n$, so the desired result follows from Lemma 6.3 with $\varepsilon = \tilde{\alpha}/4$ and $\delta = O(\tilde{\alpha}^r)$ (using that $\varepsilon + q(q-1)\delta \leq \tilde{\alpha}/4 + O(\tilde{\alpha}^r) \leq \tilde{\alpha}/2$ as $r \geq 2$, and also recalling that $q \leq r - 1$). □

Proof of Claim 7.5. Recall that for $i \in I$, we defined $h_{ii} = b'_{ii}$, and for distinct $i, j \in I$ we defined $h_{ij} = a'_{ij}$. For every $i \in I$, we can uniquely extend the vector $(h_{ij})_{j \in I} \in \mathbb{F}^I$ to a vector $\operatorname{row}_i(H) = (h_{ij})_{1 \leq j \leq n} \in \operatorname{span}(w'_1, \dots, w'_q)$, using Lemma 6.6. Then, using that every row of B' is also in $\operatorname{span}(w'_1, \dots, w'_q)$, by the second part of Lemma 6.6 we have (recalling the defining properties of I in Claim 7.4)

$$\|\operatorname{row}_i(H) - \operatorname{row}_i(B')\|_1 \leq O(|\det M_I|^{-1}) \cdot n \cdot \sum_{j \in I \setminus \{i\}} |a'_{ij} - b'_{ij}| \leq O(\tilde{\alpha}^{1-r}) \cdot n \cdot q \cdot \tilde{\alpha}^{2r} = O(\tilde{\alpha}^{r+1}) \cdot n. \quad (7.1)$$

for every $i \in I$. So far we have defined h_{ij} for $i \in I$ and $j \in [n]$. Since A' is symmetric, we have $h_{ij} = a'_{ij} = a'_{ji} = h_{ji}$ for distinct $i, j \in I$. Now, for $j \in I$ and $i \in [n] \setminus I$, let us define $h_{ij} = h_{ji}$. Then, for $i \in [n] \setminus I$, we proceed very similarly to before: we can uniquely extend the vector $(h_{ij})_{j \in I} \in \mathbb{F}^I$ to a vector $\operatorname{row}_i(H) = (h_{ij})_{1 \leq j \leq n} \in \operatorname{span}(w'_1, \dots, w'_q)$. The resulting matrix $H = (h_{ij})_{i,j}$ satisfies $h_{ij} = h_{ji}$ for all $i \in I$ and $j \in [n]$ and all of its rows lie in $\operatorname{span}(w'_1, \dots, w'_q)$. So by Lemma 7.6, H is therefore symmetric. Furthermore, for all $i \in [n]$ by the second part of Lemma 6.6 we have

$$\|\operatorname{row}_i(H) - \operatorname{row}_i(B')\|_1 \leq O(\tilde{\alpha}^{1-r}) \cdot n \cdot \sum_{j \in I} |h_{ij} - b'_{ij}|$$

and therefore we obtain

$$\|H - B'\|_1 = \sum_{i=1}^{n} \|\operatorname{row}_i(H) - \operatorname{row}_i(B')\|_1 \leq O(\tilde{\alpha}^{1-r}) \cdot n \cdot \sum_{i=1}^{n} \sum_{j \in I} |h_{ij} - b'_{ij}| = O(\tilde{\alpha}^{1-r}) \cdot n \cdot \|(H - B')_I\|_1.$$

For an $n \times n$ matrix B, let B^I denote the submatrix consisting of the rows indexed by I. By the properties of A' and B' in Claim 7.3 we have

$$\|(A' - B')^I\|_1 \leq O(\tilde{\alpha}^{3r}) \cdot n \quad \text{and} \quad \|(A' - B')_I\|_1 \leq O(\tilde{\alpha}^{3r}) \cdot n,$$

and by symmetry of H and A', we have $\|(H - A')_I\|_1 = \|(H - A')^I\|_1$, so

$$\|(H - B')_I\|_1 \leq \|(H - A')_I\|_1 + \|(A' - B)_I\|_1$$

$$\leq \|(H-A')^I\|_1 + O(\tilde{\alpha}^{3r})\cdot n \leq \|(H-B')^I\|_1 + O(\tilde{\alpha}^{3r})\cdot n.$$

On the other hand, from (7.1) we obtain $\|(H-B')^I\|_1 \leq O(\tilde{\alpha}^{r+1})\cdot n$, so it follows that $\|(H-B')_I\|_1 \leq O(\tilde{\alpha}^{r+1})\cdot n$ and therefore $\|H-B'\|_1 \leq O(\tilde{\alpha}^{1-r})\cdot n \cdot O(\tilde{\alpha}^{r+1})\cdot n = O(\tilde{\alpha}^2)\cdot n^2$. □

Proof of Lemma 7.6. Write v_{ij} for the jth component of v_i. Without loss of generality we may assume that $I=[q]\subseteq[n]$. Also, we may assume that M_I is the $q\times q$ identity matrix (we can replace $v_1,\ldots,v_q$ by different vectors with the same span).

Now, each row of H is a linear combination of $v_1,\ldots,v_q$, and given the above assumptions it is easy to read off the coefficients: $\mathrm{row}_i(H) = h_{i1}v_1+\cdots+h_{iq}v_q$ for all $i\in[n]$. So, using the assumption that $h_{ik}=h_{ki}$ for all $k\in I=[q]$ and all $i\in[n]$, we have

$$h_{ij} = \sum_{k=1}^{q} h_{ik}v_{kj} = \sum_{k=1}^{q} h_{ki}v_{kj} = \sum_{k=1}^{q}\left(\sum_{\ell=1}^{q} h_{k\ell}v_{\ell i}\right)v_{kj} = \sum_{k,\ell\in[q]} h_{k\ell}v_{\ell i}v_{kj}$$

for all $i,j\in[n]$. (For the third equality, we read off the coefficients for $\mathrm{row}_k(H)$ in the same way we read off the coefficients for $\mathrm{row}_i(H)$.) This expression is symmetric in i and j, since $h_{k\ell}=h_{\ell k}$ for all $k,\ell\in[q]$. □

7.2 Adapting the proof for Lemma 5.7

In this section we describe how to prove Lemma 5.7, by slightly modifying the proof of Lemma 5.5 in the previous subsection. Specifically, given a finite set $S\subseteq\mathbb{F}$ we define a finite set $S'=S'(r,S)$, and given a matrix A with entries in S, we describe how to adapt the proof of Lemma 5.5 to ensure that H has entries in S'.

Definition 7.7. Given a non-negative integer q and a finite set $S\subseteq\mathbb{F}$ with $0\in S$, define $T_q(S)\subseteq\mathbb{F}^r$ to be the set of those vectors $v\in\mathbb{F}^r$ which are the solution to an equation of the form $Mv=z$ for some invertible matrix $M\in S^{q\times q}$ and some vector $z\in S^q$. Also, define $\tau_q(S)=\{v^Tw : v\in T_q(S), w\in S^q\}$. Finally, for a positive integer r, define $\sigma_r(S)=S\cup\tau_1(S)\cup\cdots\cup\tau_{r-1}(S)$ and $S'(r,S)=\sigma_r(\sigma_r(\sigma_r(S)))$.

Note that the sets $T_q(S)$, $\tau_q(S)$, $\sigma_r(S)$ and $S'(r,S)$ in Definition 7.7 are finite.

Recall that in the proof of Lemma 5.5 (more specifically, in the proof of Claim 7.5) all rows of the matrix H were chosen by applying Lemma 6.6 to find a vector in $\mathrm{span}(w_1',\ldots,w_q')$ with certain prescribed entries. In order to ensure that the entries of H are in S', we need a way to control the entries of the vectors found when applying Lemma 6.6. The following lemma gives such control.

Lemma 7.8. *Fix a non-negative integer q and a finite set $S\subseteq\mathbb{F}$ with $0\in S$. Consider vectors $w_1,\ldots,w_q\in S^n$ and a subset $I\subseteq[n]$ of size q. Let M be the $q\times n$ matrix whose rows are the vectors $w_1,\ldots,w_q$, and suppose that its $q\times q$ submatrix M_I is invertible. Finally, consider $v=(v^{(1)},\ldots,v^{(n)})\in\mathrm{span}(w_1,\ldots,w_q)$ such that $v^{(i)}\in S$ for all $i\in I$. Then $v\in(\tau_q(S))^n$. In particular, if r is a positive integer such that $q\leq r-1$, we have $v\in(\sigma_r(S))^n$.*

Note that Lemma 7.8 implies the following: if we apply Lemma 6.6 to vectors $w_1,\ldots,w_q$ with entries in S, and the prescribed values $v^{(i)}$ also all lie in S, then the resulting vector v in Lemma 6.6 satisfies $v\in(\tau_q(S))^n\subseteq(\sigma_r(S))^n$.

Proof of Lemma 7.8. Write $v=\lambda_1w_1+\cdots+\lambda_qw_q$ for some $\lambda=(\lambda_1,\ldots,\lambda_q)\in\mathbb{F}^q$. Only considering the coordinates with indices in I, this implies $v_I=M_I^T\lambda$. We have $v_I\in S^q$, and $M_I^T\in S^{q\times q}$ is an invertible matrix, so $\lambda\in T_q(S)$. Now, for each $i\in[n]$, the entry $v^{(i)}$ is the product of the row vector $\lambda=(\lambda_1,\ldots,\lambda_n)$ with the column vector formed by the i-th coordinates of $w_1,\ldots,w_q$ (this column vector is in S^q). Hence $v^{(i)}\in\tau_q(S)\subseteq\sigma_r(S)$ for all $i\in[n]$. □

In order to control the entries of the rows of H when applying Lemma 6.6 together with Lemma 7.8, we clearly also need to control the entries we prescribe. These prescribed entries ultimately depend on the entries of A' and B' (see our definition of h_{ij} for $i,j\in I$ in Claim 7.5). Each entry of A' is also an entry of A or equals zero (by the way we constructed A' and A^* in the proofs of Claims 7.1 and 7.2), so we

can easily control the entries of A' (and similarly A^*). However, in order to control the entries of B', we need to control the entries of B^*, which were obtained by applying Lemma 6.5 to the list of rows of A^*.

The following lemma will be used in combination with Lemma 6.5. While it does not give direct control over the vectors $v_i' \in \mathrm{span}(w_1, \dots, w_q)$ obtained from Lemma 6.5 as approximations of the rows of A^*, it states that we can find slightly weaker approximations ensuring that the entries of these approximating vectors lie in a certain set.

Lemma 7.9. *Fix a non-negative integer q and a finite set $S \subseteq \mathbb{F}$ with $0 \in S$ and $|s| \leq 1$ for all $s \in S$. Consider $0 < \alpha \leq 1$ and $0 < \eta \leq 1$, as well as α-independent vectors $w_1, \dots, w_q \in S^n$ and a vector $v \in S^n$, such that there is some vector $\tilde{v} \in \mathrm{span}(w_1, \dots, w_q)$ with $\|v - \tilde{v}\|_1 \leq \eta \cdot n$. Then there is a vector $v^* \in (\tau_q(S))^n \cap \mathrm{span}(w_1, \dots, w_q)$ with $\|v - v^*\|_1 \leq O(\alpha^{-(q+1)}) \cdot \eta \cdot n$.*

Proof. Note that for $q = 0$ the statement is trivial: Indeed, we must have $\tilde{v} = 0$ and can therefore take $v^* = \tilde{v} = 0 \in (\tau_0(S))^n \cap \mathrm{span}(w_1, \dots, w_q)$. So let us from now on assume that $q \geq 1$.

Let $v = (v^{(1)}, \dots, v^{(n)}) \in S^n$ and $\tilde{v} = (\tilde{v}^{(1)}, \dots, \tilde{v}^{(n)}) \in \mathbb{F}^n$. Since $\|v - \tilde{v}\|_1 \leq \eta \cdot n$, there are at most $(\alpha/(2q)) \cdot n$ indices $i \in [n]$ with $|\tilde{v}^{(i)} - v^{(i)}| \geq 2q\alpha^{-1}\eta$. Let J be the set of those indices $i \in [n]$, then $|J| \leq (\alpha/(2q)) \cdot n$. Then, let $w_1', \dots, w_q'$ be obtained from $w_1, \dots, w_q$ by setting to zero all entries with coordinates in J. Since all entries of the α-independent vectors $w_1, \dots, w_q$ have absolute value at most 1, and $|J| \cdot q \leq (\alpha/2) \cdot n$, these new vectors $w_1', \dots, w_q'$ are $(\alpha/2)$-independent.

Let M (respectively M') be the $q \times n$ matrix with $w_1, \dots, w_q$ (respectively $w_1', \dots, w_q'$) as rows. By Lemma 6.1, there exists a subset $I \subseteq [n]$ of size q such that $|\det M_I'| \geq (\alpha/2)^q$. Note that $I \cap J = \emptyset$, since if M_I' had a zero column it would have zero determinant. Hence $M_I = M_I'$ and $|\det M_I| \geq (\alpha/2)^q$.

Now we can apply Lemma 6.6 to obtain a vector $v^* \in \mathrm{span}(w_1, \dots, w_q)$ such that for each $i \in I$ the i-th entry of v^* equals $v^{(i)} \in S$. By Lemma 7.8 we have $v^* \in (\tau_q(S))^n$, and by the second part of Lemma 6.6 we have (using that $|\tilde{v}^{(i)} - v^{(i)}| < 2q\alpha^{-1}\eta$ for all $i \in I \subseteq [n] \setminus J$)

$$\|v^* - \tilde{v}\|_1 \leq O(|\det M_I|^{-1}) \cdot n \cdot \sum_{i \in I} |\tilde{v}^{(i)} - v^{(i)}| \leq O(\alpha^{-q}) \cdot n \cdot q \cdot 2q\alpha^{-1}\eta = O(\alpha^{-(q+1)}) \cdot \eta \cdot n.$$

Hence $\|v - v^*\|_1 \leq \|v - \tilde{v}\|_1 + \|\tilde{v} - v^*\|_1 \leq O(\alpha^{-(q+1)}) \cdot \eta \cdot n$. □

Now, using Lemma 7.8 and Lemma 7.9, it is not hard to modify the proof of Lemma 5.5 to prove Lemma 5.7.

Proof of Lemma 5.7. Let r, S and A be as in Lemma 5.7. Then A satisfies the assumptions in Lemma 5.5 with the additional property that all entries of A are in the set S. Let $S' = S'(r, S)$ be as in Definition 7.7 (note that we may assume that $0 \in S$, since otherwise we can replace S by $S \cup \{0\}$).

We can now proceed as in the proof of Lemma 5.5 in the previous subsection. First, without making any changes to the proof of Claim 7.1, we observe that each entry of A^* is either zero (which we are assuming is in S), or is the same as the corresponding entry in A (which is also in S). Second, we consider the matrix B^* in Claim 7.2. In the proof of Lemma 5.5 we applied Lemma 6.5 to the rows $v_1, \dots, v_n \in S^n$ of A^* to obtain a list of vectors $v_1', \dots, v_n' \in \mathrm{span}(w_1, \dots, w_q)$ satisfying $\|v_i - v_i'\|_1 \leq O(\alpha^{5r(6r)^{q-r}})n = O(\tilde{\alpha}^{5r})n$ for all $i \in [n]$. We then took these vectors as the rows of B^*. For the proof of Lemma 5.7 we need an additional step: for each i we apply Lemma 7.9 with $v = v_i$ and $\tilde{v} = v_i'$, which gives a vector $v_i^* \in (\tau_q(S))^n \cap \mathrm{span}(w_1, \dots, w_q)$ with $\|v_i - v_i^*\|_1 \leq O(\tilde{\alpha}^{-(q+1)}) \cdot O(\tilde{\alpha}^{5r}) \cdot n \leq O(\tilde{\alpha}^{4r})n$ (since $q \leq r - 1$). We can then take $v_1^*, \dots, v_n^*$ as the rows of B^*, so that all entries of B^* lie in the set $\tau_q(S) \subseteq \sigma_r(S)$.

Next, without making any modifications to the proof of Claim 7.3, observe that the matrices A' and B' have entries are in $\sigma_r(S)$ (the entries of A' and B' were chosen to be zeroes or entries of A^* or B^*). Finally, when constructing the matrix $H = (h_{ij})_{i,j}$ as in the proof of Claim 7.5, observe that the entries of H all lie in $\sigma_r(\sigma_r(\sigma_r(S))) = S'$. Indeed, the entries h_{ij} for $i, j \in I$ are in $\sigma_r(S)$. We obtained $\mathrm{row}_i(H)$ for $i \in I$ by applying Lemma 6.6 to these entries, so by Lemma 7.8 all entries h_{ij} with $i \in I$ and $j \in [n]$ lie in $\sigma_r(\sigma_r(S))$. Then we defined $h_{ij} = h_{ji}$ for $i \in [n] \setminus I$ and $j \in I$, and applied Lemma 6.6 again to obtain $\mathrm{row}_i(H)$ for $i \in [n] \setminus I$. Again by Lemma 7.8 all entries of these rows lie in $\sigma_r(\sigma_r(\sigma_r(S)))$. Thus, all entries of H lie in $\sigma_r(\sigma_r(\sigma_r(S)))$, as desired. □

8 Concluding remarks

There are still a number of future directions of research. Most obviously, Costello's conjecture remains open, and there are several weakenings that we think would already be very interesting to prove. For example, can we remove the restriction on the coefficients in Theorem 1.2? Can we prove any bound of the form $n^{-1/2-\Omega(1)}$ on the point probabilities of a real quadratic polynomial that is not close to splitting into linear factors over the real numbers, even with strong assumptions on the allowed coefficients?

Costello's conjecture also concerned higher-degree polynomials, and it would be interesting to investigate this direction as well. Iterating Lemma 3.3 does give a way to control higher-degree polynomials, but it seems unlikely that one could prove results as strong as Theorems 1.1 and 1.2 without new ideas.

Regarding Ramsey graphs, it would be interesting if one could remove the $o(1)$-term in Theorem 1.3. It would also be very interesting to prove a bound on $\Pr(X = x)$ in terms of the variance of X, as follows. Such a bound would be best-possible due to Chebyshev's inequality.

Conjecture 8.1. *The following holds for any fixed constants $C, c > 0$. Let G be an n-vertex C-Ramsey graph, and, for some $cn \le k \le (1-c)n$, let X be the number of edges induced by a uniformly random subset of k vertices of G. Then for any $x \in \mathbb{Z}$, we have*

$$\Pr(X = x) = O\Big(\frac{1}{\sqrt{\operatorname{Var} X}}\Big).$$

We remark that $\sqrt{\operatorname{Var} X}$ can be as small as $\Theta(n)$ (which is typical for a random graph $\mathbb{G}(n, 1/2)$), and can be as large as $\Theta(n^{3/2})$ (for example, consider a disjoint union of two random graphs $\mathbb{G}(n/2, 1/2)$ on $n/2$ vertices). We find Conjecture 8.1 particularly compelling because, if true, it would give a very simple, unified proof of the conjectures of Narayanan–Sahasrabudhe–Tomon and Erdős–Faudree–Sós stated in the introduction, concerning subgraphs of Ramsey graphs with different numbers of edges. For example, consider an n-vertex C-Ramsey graph G and consider some $k \in \mathbb{N}$ satisfying $k = \Omega(n)$ and $n/2 - k = \Omega(n)$. To prove the Erdős–Faudree–Sós conjecture, it suffices to show that G has $\Omega(n^{3/2})$ induced subgraphs with k vertices and different numbers of edges. This can be shown using Conjecture 8.1, as follows.

Sketch proof of the Erdős–Faudree–Sós conjecture, assuming Conjecture 8.1. Let G and k be as above. Alon and Kostochka (see [3, Equation (2)]) observed that there is a sequence of $2k$-vertex sets $U_0, \dots, U_t \subseteq V(G)$ such that $e(U_t) - e(U_1) = \Omega(n^{3/2})$, but $|e(U_i) - e(U_{i-1})| \le n$ for all $1 \le i \le t$. That is, we can "loosely fill" an interval of length $\Omega(n^{3/2})$ with edge-counts of $2k$-vertex induced subgraphs. This can be shown by using a discrepancy theorem to find two $2k$-vertex induced subgraphs whose numbers of edges differ by $\Omega(n^{3/2})$, and switching vertices one-by-one between these two subgraphs.

Now, let X_i be the number of edges in a random subset of k vertices of $G[U_i]$, and let $\operatorname{supp}(X_i)$ be the support of the random variable X_i. Note that by Chebyshev's inequality, more than half of the probability mass of each X_i falls in the interval $I_i = [e(U_i)/2 - 2\sqrt{\operatorname{Var}(X_i)}, e(U_i)/2 + 2\sqrt{\operatorname{Var}(X_i)}]$, which has length $O(\sqrt{\operatorname{Var}(X_i)})$. Since each $G[U_i]$ is itself an $O(1)$-Ramsey graph, Conjecture 8.1 would say that every individual point in this interval has probability mass only $O(1/\sqrt{\operatorname{Var}(X_i)})$, from which one can show that $|\operatorname{supp}(X_i) \cap I_i| = \Omega(\sqrt{\operatorname{Var}(X_i)}) = \Omega(|I_i|)$. Given that $\sqrt{\operatorname{Var}(X_i)} = \Omega(n)$ for all i, the union $\bigcup_{i=0}^t I_i$ covers an $\Omega(1)$-fraction of the length-$\Omega(n^{3/2})$ interval between $e(U_0)/2$ and $e(U_t)/2$, which means that $|\bigcup_{i=0}^t I_i| = \Omega(n^{3/2})$. By a simple greedy algorithm, we can choose a disjoint collection $(I_i)_{i \in J}$ of intervals covering an $\Omega(1)$-fraction of $\bigcup_{i=0}^t I_i$. Then $\sum_{i \in J} |I_i| = |\bigcup_{i \in J} I_i| = \Omega(n^{3/2})$ and therefore $|\bigcup_{i \in J} \operatorname{supp}(X_i)| \ge \sum_{i \in J} |\operatorname{supp}(X_i) \cap I_i| = \sum_{i \in J} \Omega(|I_i|) = \Omega(n^{3/2})$. This gives $\Omega(n^{3/2})$ different edge-counts of k-vertex induced subgraphs. □

Actually, it is tempting to wonder whether one can strengthen Conjecture 8.1 even further, and prove a local central limit theorem estimating each of the point probabilities $\Pr(X = x)$ in terms of a Gaussian density function. With sufficiently detailed understanding of the local behaviour of "edge-statistic" random variables in Ramsey graphs, it might be possible to prove an old conjecture of Erdős and McKay [15, 16] that there is an interval $I = \{0, 1, 2 \dots, \Omega(n^2)\}$ such that in every $O(1)$-Ramsey graph, for every $m \in I$ there is an induced subgraph with exactly m edges (see [4] for some progress on this question).

Finally, there may also be interesting directions of research concerning some of the auxiliary lemmas in this paper. For example, in Corollary 5.6 we proved that if a matrix with bounded entries is close to being

low-rank, and close to being symmetric, then it is close to being simultaneously low-rank and symmetric. Can we drop the assumption that the entries of the matrix are bounded? There may be a connection between results of this type and the existing literature on low-rank approximation of matrices.

Concluding remarks: We would like to thank Jacob Fox for insightful discussions, and Ragib Zaman for simplifying the proof of Lemma 4.1.

References

[1] N. Alon, J. Balogh, A. Kostochka, and W. Samotij, *Sizes of induced subgraphs of Ramsey graphs*, Combin. Probab. Comput. **18** (2009), no. 4, 459–476.

[2] N. Alon, D. Hefetz, M. Krivelevich, and M. Tyomkyn, *Edge-statistics on large graphs*, Combin. Probab. Comput., to appear, arXiv preprint arXiv:1805.06848 (2018).

[3] N. Alon and A. V. Kostochka, *Induced subgraphs with distinct sizes*, Random Structures Algorithms **34** (2009), no. 1, 45–53.

[4] N. Alon, M. Krivelevich, and B. Sudakov, *Induced subgraphs of prescribed size*, J. Graph Theory **43** (2003), no. 4, 239–251.

[5] B. Barak, A. Rao, R. Shaltiel, and A. Wigderson, *2-source dispersers for $n^{o(1)}$ entropy, and Ramsey graphs beating the Frankl-Wilson construction*, Ann. of Math. (2) **176** (2012), no. 3, 1483–1543.

[6] R. Berkowitz, *A Local Limit Theorem for cliques in $G(n, p)$*, arXiv preprint arXiv:1811.03527 (2018).

[7] B. Bukh and B. Sudakov, *Induced subgraphs of Ramsey graphs with many distinct degrees*, J. Combin. Theory Ser. B **97** (2007), no. 4, 612–619.

[8] E. Chattopadhyay and D. Zuckerman, *Explicit two-source extractors and resilient functions*, STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, ACM, New York, 2016, pp. 670–683.

[9] G. Cohen, *Two-source dispersers for polylogarithmic entropy and improved Ramsey graphs*, STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, ACM, New York, 2016, pp. 278–284.

[10] D. Conlon and J. Fox, *Bounds for graph regularity and removal lemmas*, Geom. Funct. Anal. **22** (2012), no. 5, 1191–1256.

[11] K. P. Costello, *Bilinear and quadratic variants on the Littlewood-Offord problem*, Israel J. Math. **194** (2013), no. 1, 359–394.

[12] K. P. Costello, T. Tao, and V. Vu, *Random symmetric matrices are almost surely nonsingular*, Duke Math. J. **135** (2006), no. 2, 395–413.

[13] P. Erdős, *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc. **51** (1945), 898–902.

[14] P. Erdős, *Some remarks on the theory of graphs*, Bull. Amer. Math. Soc. **53** (1947), 292–294.

[15] P. Erdős, *Some of my favourite problems in various branches of combinatorics*, Matematiche (Catania) **47** (1992), no. 2, 231–240 (1993), Combinatorics 92 (Catania, 1992).

[16] P. Erdős, *Some recent problems and results in graph theory*, Discrete Math. **164** (1997), no. 1-3, 81–85, The Second Krakow Conference on Graph Theory (Zgorzelisko, 1994).

[17] P. Erdős and A. Hajnal, *On spanned subgraphs of graphs*, Contributions to graph theory and its applications (Internat. Colloq., Oberhof, 1977), Tech. Hochschule Ilmenau, Ilmenau, 1977, pp. 80–96.

[18] P. Erdős and G. Szekeres, *A combinatorial problem in geometry*, Compositio Math. **2** (1935), 463–470.

[19] P. Erdős and A. Szemerédi, *On a Ramsey type theorem*, Period. Math. Hungar. **2** (1972), 295–299, Collection of articles dedicated to the memory of Alfréd Rényi, I.

[20] C. G. Esseen, *On the Kolmogorov-Rogozin inequality for the concentration function*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **5** (1966), 210–216.

[21] J. Fox, M. Kwan, and L. Sauermann, *Combinatorial anti-concentration inequalities, with applications*, preprint (2019).

[22] J. Fox and L. Sauermann, *A completion of the proof of the edge-statistics conjecture*, arXiv preprint arXiv:1809.01352 (2018).

[23] P. Frankl and R. M. Wilson, *Intersection theorems with geometric consequences*, Combinatorica **1** (1981), no. 4, 357–368.

[24] G. Halász, *Estimates for the concentration function of combinatorial number theory and probability*, Period. Math. Hungar. **8** (1977), no. 3-4, 197–211.

[25] M. Kwan and B. Sudakov, *Proof of a conjecture on induced subgraphs of Ramsey graphs*, Trans. Amer. Math. Soc., to appear, arXiv preprint arXiv:1712.05656 (2017).

[26] M. Kwan and B. Sudakov, *Ramsey graphs induce subgraphs of quadratically many sizes*, Int. Math. Res. Not. IMRN, to appear, arXiv preprint arXiv:1711.02937 (2017).

[27] M. Kwan, B. Sudakov, and T. Tran, *Anticoncentration for subgraph statistics*, J. London Math. Soc. **99** (2019), no. 3, 757–777.

[28] X. Li, *Non-malleable extractors and non-malleable codes: Partially optimal constructions*, 34th Computational Complexity Conference, to appear, arXiv preprint arXiv:1804.04005 (2018).

[29] J. E. Littlewood and A. C. Offord, *On the number of real roots of a random algebraic equation. III*, Rec. Math. [Mat. Sbornik] N.S. **12(54)** (1943), 277–286.

[30] A. Martinsson, F. Mousset, A. Noever, and M. Trujić, *The edge-statistics conjecture for $\ell \ll k^{6/5}$*, Israel J. Math., to appear, arXiv preprint arXiv:1809.02576 (2018).

[31] C. McDiarmid, *Concentration*, Probabilistic methods for algorithmic discrete mathematics, Algorithms Combin., vol. 16, Springer, Berlin, 1998, pp. 195–248.

[32] R. Meka, O. Nguyen, and V. Vu, *Anti-concentration for polynomials of independent random variables*, Theory Comput. **12** (2016), Paper No. 11, 16 pages.

[33] B. Narayanan, J. Sahasrabudhe, and I. Tomon, *Ramsey graphs induce subgraphs of many different sizes*, Combinatorica **39** (2019), no. 1, 215–237.

[34] H. Nguyen and V. Vu, *Optimal inverse Littlewood-Offord theorems*, Adv. Math. **226** (2011), no. 6, 5298–5319.

[35] H. H. Nguyen, *Inverse Littlewood-Offord problems and the singularity of random symmetric matrices*, Duke Math. J. **161** (2012), no. 4, 545–586.

[36] H. H. Nguyen and V. H. Vu, *Small ball probability, inverse theorems, and applications*, Erdős centennial, Bolyai Soc. Math. Stud., vol. 25, János Bolyai Math. Soc., Budapest, 2013, pp. 409–463.

[37] H. J. Prömel and V. Rödl, *Non-Ramsey graphs are $c \log n$-universal*, J. Combin. Theory Ser. A **88** (1999), no. 2, 379–384.

[38] A. Razborov and E. Viola, *Real advantage*, ACM Trans. Comput. Theory **5** (2013), no. 4, Art. 17.

[39] J. Rosiński and G. Samorodnitsky, *Symmetrization and concentration inequalities for multilinear forms with applications to zero-one laws for Lévy chaos*, Ann. Probab. **24** (1996), no. 1, 422–437.

[40] M. Rudelson and R. Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), no. 2, 600–633.

[41] S. Shelah, *Erdős and Rényi conjecture*, J. Combin. Theory Ser. A **82** (1998), no. 2, 179–185.

[42] T. Tao and V. Vu, *From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices*, Bull. Amer. Math. Soc. (N.S.) **46** (2009), no. 3, 377–396.

[43] T. Tao and V. Vu, *A sharp inverse Littlewood-Offord theorem*, Random Structures Algorithms **37** (2010), no. 4, 525–539.

[44] T. Tao and V. Vu, *The Littlewood-Offord problem in high dimensions and a conjecture of Frankl and Füredi*, Combinatorica **32** (2012), no. 3, 363–372.

[45] T. Tao and V. H. Vu, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, Ann. of Math. (2) **169** (2009), no. 2, 595–632.

[46] T. Tao and V. H. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2010.

[47] K. Tikhomirov, *Singularity of random Bernoulli matrices*, arXiv preprint arXiv:1812.09016 (2018).