Resolution of the quadratic Littlewood–Offord problem

Matthew Kwan[*]   Lisa Sauermann[†]

December 21, 2023

Abstract

Consider a quadratic polynomial $Q(\xi_1, \ldots, \xi_n)$ of independent Rademacher random variables $\xi_1, \ldots, \xi_n$. To what extent can $Q(\xi_1, \ldots, \xi_n)$ concentrate on a single value? This quadratic version of the classical Littlewood–Offord problem was popularised by Costello, Tao and Vu in their study of symmetric random matrices. In this paper, we obtain an essentially optimal bound for this problem, as conjectured by Nguyen and Vu.

Specifically, if $Q(\xi_1, \ldots, \xi_n)$ "robustly depends on at least m of the ξ_i" in the sense that there is no way to pin down the value of $Q(\xi_1, \ldots, \xi_n)$ by fixing values for fewer than m of the variables ξ_i, then we have $\Pr[Q(\xi_1, \ldots, \xi_n) = 0] \le O(1/\sqrt{m})$. This also implies a similar result in the case where $\xi_1, \ldots, \xi_n$ have arbitrary distributions.

Our proof combines a number of ideas that may be of independent interest, including an *inductive decoupling scheme* that reduces quadratic anticoncentration problems to high-dimensional linear anticoncentration problems. Also, one application of our main result is the resolution of a conjecture of Alon, Hefetz, Krivelevich and Tyomkyn related to *graph inducibility*.

1 Introduction

Some of the most fundamental theorems in probability theory (and its applications) are *concentration* inequalities, which show that certain types of random variables are likely to lie in a small interval around their mean. In the other direction, *anticoncentration* inequalities give *upper* bounds on the probability that a random variable falls into a small interval or is equal to a particular value. In this area, one of the most important directions is the (polynomial) *Littlewood–Offord problem*. Roughly speaking, the problem is as follows[1]. Consider an n-variable polynomial $P \in \mathbb{R}[x_1, \ldots, x_n]$, and consider independent Rademacher random variables $\xi_1, \ldots, \xi_n \in \{-1, 1\}$ (meaning that $\Pr[\xi_i = -1] = \Pr[\xi_i = 1] = 1/2$ for each i). What upper bounds can we prove on the *point probabilities* of the form $\Pr[P(\xi_1, \ldots, \xi_n) = z]$? More specifically, how large can the maximum point probability $\sup_{z \in \mathbb{R}} \Pr[P(\xi_1, \ldots, \xi_n) = z]$ be, without making any strong assumptions about the polynomial P?

Historically, the starting point for this problem was the *linear* case, where P is a degree-1 polynomial. Indeed, consider a random variable $X = a_1 \xi_1 + \cdots + a_n \xi_n$, where $a_1, \ldots, a_n \in \mathbb{R}$ are nonzero real numbers and $\xi_1, \ldots, \xi_n \in \{-1, 1\}$ are independent Rademacher random variables. As part of their study of random polynomials, in 1943 Littlewood and Offord [35] proved that

$$\sup_{z \in \mathbb{R}} \Pr[X = z] \le O\left(\frac{\log n}{\sqrt{n}}\right).$$

---

[1]This is the "discrete" form of the polynomial Littlewood–Offord problem; one can also consider the "continuous" form, where we are interested in the maximum probability that $P(\xi_1, \ldots, \xi_n)$ lies in an interval of given length. In the following historical discussion, we specialise all results to this discrete setting.

Littlewood and Offord's result was famously sharpened in 1945 by Erdős [13], who found a purely combinatorial proof of what is now usually called the *Erdős–Littlewood–Offord theorem*: under the same assumptions,

$$\sup_{z \in \mathbb{R}} \Pr[X = z] \leq \binom{n}{\lfloor n/2 \rfloor} \cdot 2^{-n} = O\left(\frac{1}{\sqrt{n}}\right). \tag{1.1}$$

This result is best-possible, as can be seen by considering the case where the coefficients a_i are all equal.

Remark 1.1. It is worth noting that the general topic of anticoncentration of sums of independent random variables was first considered in 1936 by Doeblin and Lévy [12], and this led to a parallel line of research with many similar results (the two lines of research seem to have not been aware of each others' existence until quite recently). In particular, in the above setting, the bound $\sup_{z \in \mathbb{R}} \Pr[X = z] \leq O(1/\sqrt{n})$ follows from a general result claimed in a 1939 paper of Doeblin [11], preceding Littlewood and Offord by several years. However, this general result is also the subject of a 1958 paper of Kolmogorov [30], which claims that Doeblin's paper did not provide a full proof.

By now, the linear Littlewood–Offord problem is very well understood, and many variations and strengthenings are available. For example, there are very powerful *inverse theorems* that relate the maximum concentration probability to the arithmetic structure of the coefficients $a_1, \ldots, a_n$, and these theorems have had a huge impact in random matrix theory (see for example the survey [40]). Also, there are versions of the Erdős–Littlewood–Offord theorem for general distributions (i.e., allowing the variables $\xi_1, \ldots, \xi_n$ to take non-Rademacher distributions), assuming that the distributions of the variables ξ_i do not themselves concentrate too strongly. (Approximate results along these lines follow directly from the Lévy–Doeblin–Kolmogorov theorem mentioned in Remark 1.1, or can be deduced from the Erdős–Littlewood–Offord theorem; in some sense the Rademacher case is the "hardest case". For exact results see [25, 26, 34].)

After the linear case, the next case to consider is the *quadratic* case: what bounds can we give on the point concentration of a quadratic polynomial $Q(\xi_1, \ldots, \xi_n)$ of independent Rademacher random variables $\xi_1, \ldots, \xi_n$? This question came to the forefront in a 2005 paper by Costello, Tao and Vu [6], when they used such a bound in their proof of Weiss' conjecture on singularity of random symmetric matrices. Specifically, they proved that if a quadratic polynomial $Q \in \mathbb{R}[x_1, \ldots, x_n]$ has at least cn^2 nonzero coefficients[2] (for some constant $c > 0$), then $X = Q(\xi_1, \ldots, \xi_n)$ (for independent Rademacher random variables $\xi_1, \ldots, \xi_n$) satisfies

$$\sup_{z \in \mathbb{R}} \Pr[X = z] \leq O\left(\frac{1}{n^{1/8}}\right). \tag{1.2}$$

Remark 1.2. Parallelling the situation described in Remark 1.1, we remark that Costello, Tao and Vu were actually not the first to prove this inequality: in 1996, Rosiński and Samorodnitsky [44] had proved essentially the same result (in fact, a generalisation of it to higher degree polynomials), in their study of zero-one laws for Lévy chaos (*chaoses* are polynomials of independent random variables; they are classical and very well-studied objects in probability theory, statistics and applied mathematics). Seemingly unaware of Rosiński and Samorodnitsky's work, in 2013 Razborov and Viola [43] considered a similar higher-degree generalisation (for applications in the theory of Boolean functions).

The authors of [6, 44] already recognised that (1.2) was likely not optimal. Indeed, just as for the linear Littlewood–Offord problem, one expects a bound of the form $\sup_{z \in \mathbb{R}} \Pr[X = z] \leq O(1/\sqrt{n})$ (this bound is attained in the case $Q(x_1, \ldots, x_n) = (x_1 + \cdots + x_n)^2$, for example). A conjecture to this effect has been attributed to Nguyen and Vu (see [37, 43]).

The first improvement on (1.2) was by Costello and Vu [8], who showed how to adapt the arguments in [6] to prove a bound of the form $O(n^{-1/4})$. Introducing several new ideas, Costello [5] then managed to obtain the nearly optimal bound $O(n^{-1/2+\varepsilon})$ (for any constant $\varepsilon > 0$). Via a completely different approach, this bound was further refined to $\exp(O((\log \log n)^2))/\sqrt{n}$ by Meka, Nguyen and Vu[3] (see arXiv version v4 of [37]), before it was observed that the bound $(\log n)^{O(1)}/\sqrt{n}$ follows from a powerful general result of Kane [28] (see the journal version of [37] for more discussion.)

---

[2]This is not exactly the assumption that appeared in [6], but it is a simple assumption which is sufficient for essentially all results in this area; see Remark 1.3 for more discussion.

[3]The results in [37] are more general, holding for polynomials of any bounded degree.

In this paper we finally resolve the quadratic Littlewood–Offord problem (up to constant factors), obtaining an optimal bound of $O(1/\sqrt{n})$. We are also able to make a weaker assumption on Q than in previous work.

Theorem 1.1. *Let $Q \in \mathbb{R}[x_1, \ldots, x_n]$ be a polynomial of degree at most 2, and let $\xi_1, \ldots, \xi_n \in \{-1, 1\}$ be independent Rademacher random variables. For some positive integer m, suppose that $Q(\xi_1, \ldots, \xi_n)$ "robustly depends on at least m of the variables ξ_i" in the sense that the value of $Q(\xi_1, \ldots, \xi_n)$ cannot be determined by specifying any outcomes of any $m-1$ of the variables $\xi_1, \ldots, \xi_n \in \{-1, 1\}$. Then,*

$$\sup_{z \in \mathbb{R}} \Pr[Q(\xi_1, \ldots, \xi_n) = z] \leq \frac{C}{\sqrt{m}},$$

for some absolute constant C.

Remark 1.3. Recall that the Erdős–Littlewood–Offord theorem has an assumption that each linear coefficient a_i is nonzero. Of course, zero coefficients can be ignored, so without such an assumption one immediately obtains a bound of the form $\sup_{z \in \mathbb{R}} \Pr[a_1\xi_1 + \cdots + a_n\xi_n = z] \leq O(1/\sqrt{m})$ where m is the number of nonzero a_i. Unfortunately, the situation is not so simple in the quadratic case. One could make the very strong assumption that *every* degree-2 coefficient is nonzero, but this is too strong of an assumption for most applications[4]. Alternatively, one might wish to consider the very weak assumption that every variable x_i features in at least one nonzero term of $Q(x_i)$, but unfortunately this is too weak to get a sensible bound: indeed, consider for example the case where $Q(x_1, \ldots, x_n) = (1 + \xi_1)(\xi_1 + \cdots + \xi_n)$, which is zero whenever $\xi_1 = -1$, and therefore $\Pr[Q(\xi_1, \ldots, \xi_n) = 0] \geq 1/2$. More generally, if it is possible to determine the value of $Q(\xi_1, \ldots, \xi_n)$ by fixing the outcomes of a small number of variables to certain values, then this automatically leads to a large point probability for $Q(\xi_1, \ldots, \xi_n)$. So it is necessary to make an assumption guaranteeing that Q "robustly depends on many of the ξ_i".

We believe that our assumption in Theorem 1.1 captures this "robust dependence on many of the ξ_i" in a natural way. To compare to previous work: our assumption is slightly weaker than the assumption in [37] (which says that $Q(x_1, \ldots, x_n)$ has many quadratic terms featuring disjoint variables), and is much weaker than the assumptions in [5, 6] (which say, in slightly different ways, that $Q(x_1, \ldots, x_n)$ has a huge number of nonzero coefficients).

Remark 1.4. Of course, we could ask for the optimal constant factor C in Theorem 1.1. It is not necessarily clear what to expect: one may guess that the polynomial $Q(x_1, \ldots, x_m) = (x_1 + \cdots + x_m)(x_1 + \cdots + x_m + 2)$ or $Q(x_1, \ldots, x_m) = (x_1 + \cdots + x_m + 1)(x_1 + \cdots + x_m - 1)$ (depending on whether m is odd or even) is the worst case, but recent developments on the so-called *Gotsman–Linial conjecture* (see Section 12) suggest that this might be too naïve. It is also quite possible that the optimal value for the constant factor in the bound on $\sup_{z \in \mathbb{R}} \Pr[Q(\xi_1, \ldots, \xi_n) = z]$ is sensitive to the precise assumption one makes on the quadratic polynomial Q (to ensure that it "robustly depends on many of the variables ξ_i"; see Remark 1.3).

In Section 2 we give a brief summary of the methods that had previously been applied to the (quadratic) Littlewood–Offord problem, their limitations, and the new ideas in our proof of Theorem 1.1. In particular, our key contribution is a new *inductive decoupling scheme*: we take the well-known technique of *decoupling*, usually viewed as a tool to inefficiently "reduce from a quadratic problem to a linear problem", and reinterpret it as a tool to efficiently "reduce the relative dimension of the quadratic part of a problem". We also develop a new way to study anticoncentration of random vectors, via a technique we call *witness-counting*. We believe both these aspects of our proof may have broader applications.

We also remark that, while the details of the proof of Theorem 1.1 are rather involved, there is a certain special case of Theorem 1.1 (the case where the quadratic part of Q has "bounded rank") which permits a much simpler proof. This case is already interesting, and we present its proof in Section 4 to serve as an accessible illustration of our inductive decoupling scheme (and the results of Section 4 will also be used later in the paper).

Finally, we remark that just as for the linear Littlewood–Offord problem, one can deduce a version of Theorem 1.1 in which the variables ξ_i are allowed to take essentially any discrete distribution (not just the Rademacher distribution), as follows.

---

[4]We also remark that this strong assumption doesn't seem to make the problem much easier.

Theorem 1.2. *Fix $0 < \delta < 1$. Let $Q \in \mathbb{R}[x_1, \ldots, x_n]$ be a polynomial of degree at most 2, and let $\zeta_1, \ldots, \zeta_n \in \mathbb{R}$ be independent discrete random variables.*

For nonempty subsets $R_1, \ldots, R_n$ of the supports of $\zeta_1, \ldots, \zeta_n$, respectively, say that the product $R_1 \times \cdots \times R_n$ is a fixing box *for Q if the polynomial Q is constant on $R_1 \times \cdots \times R_n$. For some positive integer m, suppose that for any fixing box $R_1 \times \cdots \times R_n$ there are at least m indices $i \in \{1, \ldots, n\}$ such that $\Pr[\zeta_i \in R_i] \le 1 - \delta$. Then, we have*

$$\sup_{z \in \mathbb{R}} \Pr[Q(\zeta_1, \ldots, \zeta_n) = z] \le \frac{C_\delta}{\sqrt{m}},$$

for some constant C_δ only depending on δ.

Note that if $\zeta_1, \ldots, \zeta_n$ are independent uniformly random integers in $\{-B, -(B-1), \ldots, B-1, B\}$, then $\Pr[Q(\zeta_1, \ldots, \zeta_n) = z] \cdot (2B+1)^n$ is the number of integer solutions to $Q(x_1, \ldots, x_n) = z$ among integers $x_1, \ldots, x_n$ with absolute value ("height") at most B. We remark that quantities of this form have been extensively studied in analytic number theory, in the regime where n is constant and B is large (in contrast, our estimates are effective in the regime where B is constant and n is large); see for example [22, 42] and [3, Theorem 1.11].

1.1 An application to edge-statistics

Apart from the intrinsic value of Theorem 1.1, of course it also enables us to improve bounds in any place where quadratic Littlewood–Offord inequalities had previously been applied (see for example [1, 6, 7, 8, 15, 19, 33]). Here we highlight one particular application: we can resolve a conjecture of Alon, Hefetz, Krivelevich and Tyomkyn [2] related to the so-called *graph inducibility problem*. Specifically, let $\mathcal{G}_n$ be the set of n-vertex graphs, and for a graph G let $N_G(k, \ell)$ be the number of sets of k vertices of G inducing exactly ℓ edges. Then, define the *edge-inducibility* (with parameters k and ℓ) by

$$\mathrm{ind}(k, \ell) = \lim_{n \to \infty} \max_{G \in \mathcal{G}_n} \frac{N_G(k, \ell)}{\binom{n}{k}}.$$

This parameter measures, for large graphs, the maximum possible fraction of k-vertex subsets which induce ℓ edges. By considering complete or empty graphs we have

$$\mathrm{ind}(k, 0) = \mathrm{ind}\left(k, \binom{k}{2}\right) = 1,$$

but for $0 < \ell < \binom{k}{2}$ we have $\mathrm{ind}(k, \ell) < 1$ (this follows easily from *Ramsey's theorem*, which says that large graphs must have large complete or empty subgraphs). In fact, for $0 < \ell < \binom{k}{2}$, and large k, we now know that $\mathrm{ind}(k, \ell)$ cannot be much larger than $1/e$; this was the content of the *Edge-Statistics conjecture*, proved in a combination of papers by Kwan, Sudakov and Tran [33], Fox and Sauermann [18], and Martinsson, Mousset, Noever and Trujić [36]. One can use Theorem 1.1 to prove the following much stronger bound when ℓ is far from zero and far from $\binom{k}{2}$, which was conjectured by Alon, Hefetz, Krivelevich and Tyomkyn (see [2, Conjecture 6.2]):

Theorem 1.3. *For $0 < \ell < \binom{k}{2}$ we have*

$$\mathrm{ind}(k, \ell) = O\left(\frac{1}{\sqrt{\min\left(\ell, \binom{k}{2} - \ell\right)/k}}\right).$$

The deduction of Theorem 1.3 from Theorem 1.1 is exactly as in the proof of [33, Theorem 1.1] (which uses the weaker quadratic Littlewood–Offord inequality from [37]), so we do not include it here. The idea is that for any large graph G, the number of edges in a random set of k vertices of G can be interpreted as a quadratic polynomial of independent Rademacher random variables (via a certain coupling).

4

Notation. As, usual, for a nonnegative integer n, we write $[n] := \{1, \ldots, n\}$ (note that for $n = 0$, this means $[n] = \emptyset$). For an $n \times n$ matrix A and subsets $I, J \subseteq [n]$, we write $A[I \times J]$ for the $|I| \times |J|$ submatrix of A consisting of the rows with indices in I and the columns with indices in J. Similarly for a vector $\vec{v} \in \mathbb{R}^n$ and a subset $I \subseteq [n]$ we write $\vec{v}[I] \in \mathbb{R}^I$ for the vector obtained from $\vec{v}$ by only taking the coordinates with indices in I. Finally, for a vector $\vec{v} \in \mathbb{R}^n$ and $i \in [n]$, we write $\vec{v}[i] \in \mathbb{R}$ for the i-th entry of $\vec{v}$.

In this paper, we say that $Q \in \mathbb{R}[x_1, \ldots, x_n]$ is a quadratic polynomial if its degree is *at most* 2. For $t > 0$, we write $\log(t)$ for the base-2 logarithm of t.

2 Key ideas, in comparison with previous work

By now there are several different proofs of the $O(1/\sqrt{n})$ bound in the (linear) Erdős–Littlewood–Offord theorem. As far as we know, they all take advantage of at least one of two very special properties of random variables of the form $X = a_1\xi_1 + \cdots + a_n\xi_n$. First, Erdős' original proof [13] used the *monotonicity* of X (for every index i, changing ξ_i from -1 to 1 will always make the value of X increase, or always make the value of X decrease). Second, one can take advantage of the fact that X is a *sum of independent random variables* (each with a very simple distribution), so its Fourier transform is very well-behaved. (The first Fourier-analytic proof seems to have been by Halász [21]; see also the very simple proof of a $O(1/\sqrt{n})$ bound due to Croot [9]).

Unfortunately, in the quadratic setting (where $X = Q(\xi_1, \ldots, \xi_n)$ for some quadratic polynomial Q), both of the above properties of X may fail in a very strong way. There are two general approaches that have been most successful so far: Gaussian approximation, and a technique called *decoupling*. We briefly discuss both these approaches and their limitations, before describing our new ideas.

2.1 Gaussian approximation and combinatorial partitioning

Whether one is interested in the linear or the quadratic cases of the Littlewood–Offord problem, perhaps the most natural starting point is to try to leverage some of the vast literature in probability theory on distributional approximation: if one can approximate the entire distribution of a random variable X, then anticoncentration should be an easy corollary.

This angle of attack is especially compelling in the linear case, since $X = a_1\xi_1 + \cdots + a_n\xi_n$ is a sum of independent random variables; it is tempting to try to apply a central limit theorem. One cannot be too naïve here, as the limiting distribution of X could actually be very far from Gaussian (consider for example the case where $a_i = 2^i$ for each i). However, in their foundational paper, Littlewood and Offord [35] were in fact able to prove their $O(\log n / \sqrt{n})$ bound via Gaussian approximation. The key idea was to partition the coefficients a_i into $O(\log n)$ "buckets" according to their orders of magnitude, in such a way that Gaussian approximation is effective within each bucket.

It is far from obvious how to extend this type of strategy to the higher-degree case (when $X = Q(\xi_1, \ldots, \xi_n)$ for a bounded-degree polynomial Q), but this is more or less what was accomplished by Meka, Nguyen and Vu [37] and by Kane [28] in their bounds of the form $\exp(O((\log\log n)^2))/\sqrt{n}$ and $(\log n)^{O(1)}/\sqrt{n}$, respectively. Instead of a central limit theorem, one needs a *Gaussian invariance principle* (which provides sufficient conditions under which one can approximate a polynomial of independent Rademacher random variables by a polynomial of independent Gaussian random variables), and instead of a simple "bucketing" argument one needs a *regularity lemma* to describe $Q(\xi_1, \ldots, \xi_n)$ in terms of a "low-complexity decision tree".

These types of methods are very powerful and very flexible, but unfortunately it seems that one inevitably needs to "pay logarithmic factors" in order to deconstruct an arbitrary[5] quadratic polynomial into "well-behaved pieces" for which Gaussian approximation is effective. Even in the linear case, we are not aware of a way to prove an optimal $O(1/\sqrt{n})$ bound via Gaussian approximation. Perhaps this is not surprising: in

---

[5]If one makes strong assumptions about the quadratic polynomial, it is sometimes possible to prove exact bounds via Gaussian approximation; see for example [31].

some sense the entire philosophy of the Littlewood–Offord problem is that anticoncentration is an extremely robust phenomenon that holds under much weaker assumptions than central limit theorems, and we should not expect to be able to *use* central limit theorems to prove optimal anticoncentration.

2.2 Decoupling

A completely different technique was employed in the papers of Costello, Tao and Vu [6] and Rosiński and Samorodnitsky [44] which first studied the quadratic Littlewood–Offord problem. This technique is now usually called *decoupling*, following [6]. Roughly speaking, decoupling is a general technique to "reduce a quadratic anticoncentration problem to a linear one"[6]. Below we sketch the basic idea, incorporating an improvement due to Costello and Vu [8][7].

Let $Q \in \mathbb{R}[x_1, \ldots, x_n]$ be a quadratic polynomial and let $\vec{\xi} = (\xi_1, \ldots, \xi_n)$ be a sequence of independent Rademacher random variables. Given a partition of the index set $\{1, \ldots, n\}$ into two subsets I and J, we can break the random vector $\vec{\xi}$ into two parts $\vec{\xi}[I] \in \{-1, 1\}^I$ and $\vec{\xi}[J] \in \{-1, 1\}^J$. Then, a simple application of the Cauchy–Schwarz inequality (see Lemma 3.6) shows that if $\vec{\xi}'[I]$ is an independent copy of $\vec{\xi}[I]$, we have

$$\Pr\Big[Q(\vec{\xi}) = z\Big] = \Pr\Big[Q(\vec{\xi}[I], \vec{\xi}[J]) = z\Big] \leq \Pr\Big[Q(\vec{\xi}[I], \vec{\xi}[J]) = z \text{ and } Q(\vec{\xi}'[I], \vec{\xi}[J]) = z\Big]^{1/2} \qquad (2.1)$$

$$\leq \Pr\Big[Q(\vec{\xi}[I], \vec{\xi}[J]) - Q(\vec{\xi}'[I], \vec{\xi}[J]) = 0\Big]^{1/2}. \qquad (2.2)$$

Now, $Q(\vec{\xi}[I], \vec{\xi}[J]) - Q(\vec{\xi}'[I], \vec{\xi}[J])$ can be interpreted as a *linear* function of $\vec{\xi}[J]$, with coefficients that depend on $(\vec{\xi}[I], \vec{\xi}'[I])$ (since the terms that are quadratic in $\vec{\xi}[J]$ cancel out). Furthermore, this linear function typically has many nonzero coefficients (since it is unlikely that most of the "cross terms" between $\vec{\xi}[I]$ and $\vec{\xi}'[I]$ cancel out). So, after conditioning on a typical outcome of $(\vec{\xi}[I], \vec{\xi}'[I])$, one can apply the Erdős–Littlewood–Offord theorem, to obtain a bound of the form $\Pr[Q(\vec{\xi}) = z] \leq (O(1/\sqrt{n}))^{1/2} = O(n^{-1/4})$.

The great advantage of decoupling is that the resulting *linear* anticoncentration problem is much easier: one has access to the much wider range of tools available to study sums of independent random variables. However, the inequality between (2.1) and (2.2) is rather lossy, and one tends to end up with bounds that are at least "a square root away" from best-possible[8].

Costello's paper [5], proving the first bound of the form $n^{-1/2+o(1)}$ for the quadratic Littlewood–Offord problem, combined decoupling with a structural dichotomy. Namely, Costello discovered that if the quadratic polynomial Q is in a certain sense "robustly irreducible", then number-theoretic arguments give the stronger bound $\Pr[Q(\vec{\xi}[I], \vec{\xi}[J]) - Q(\vec{\xi}'[I], \vec{\xi}[J]) = 0] \leq n^{-1+o(1)}$, and so even after the "square-root loss" of decoupling one has $\Pr[Q(\vec{\xi}[I], \vec{\xi}[J]) = z] \leq n^{-1/2+o(1)}$. He then gave a separate argument to handle the case where Q does not satisfy the robust irreducibility condition (i.e., when Q essentially splits into linear factors), based on the Szemerédi–Trotter theorem from discrete geometry.

We cannot conclusively rule out the possibility that one could prove an optimal $O(1/\sqrt{n})$ bound with a similar kind of case analysis, but this seems to be extremely difficult. In particular, we are not aware of any suitable candidate for a condition on Q which ensures that $\Pr[Q(\vec{\xi}[I], \vec{\xi}[J]) - Q(\vec{\xi}'[I], \vec{\xi}[J]) = 0] \leq O(1/n)$ (Costello's notion of robust irreducibility, and similar notions of "robust rank" or "robust sum-of-squares complexity", are only suitable for an $n^{-1+o(1)}$ bound on such probabilities). Also, Costello's Szemerédi–Trotter-based proof for the nearly-reducible case does not seem to easily generalise to other simple-looking families of quadratic polynomials (for example, polynomials which can be written as the sum of four squares; see the discussion in [17, Section 10]).

---

[6]The term "decoupling" refers more generally to a class of techniques used to reduce from dependent situations to independent situations, see for example the book-length treatment in [10].

[7][6, 44] featured a "more symmetric" decoupling argument, which gives worse bounds.

[8]But we remark that there are some circumstances where one can "do decoupling in Fourier space" in such a way that the resulting square-root loss in Fourier space corresponds to a much smaller loss in physical space, see [18, 31].

2.3 A geometric point of view, and an inductive decoupling scheme

In this paper, we consider a different perspective on decoupling: instead of using decoupling to immediately reduce from a quadratic problem to a linear one, we reinterpret decoupling as a tool to obtain a problem with a linear and a quadratic part, and to inductively "reduce the proportion of our problem that is quadratic". To explain this, it is helpful to take a geometric perspective.

Specifically, for a quadratic polynomial $Q \in \mathbb{R}[x_1, \ldots, x_n]$ and a random vector $\vec{\xi} \in \{1, -1\}^n$, note that $\Pr[Q(\vec{\xi}) = z]$ can be interpreted as the probability that $\vec{\xi}$ lies in the quadric (quadratic variety) $\mathcal{Z}$ given by $\mathcal{Z} = \{\vec{x} \in \mathbb{R}^n : Q(\vec{x}) = z\} \subseteq \mathbb{R}^n$. Similarly, the expression $\Pr[Q(\vec{\xi}[I], \vec{\xi}[J]) = z$ and $Q(\vec{\xi}'[I], \vec{\xi}[J]) = z]$ appearing in (2.1) can be interpreted as the probability that $\vec{\xi}[J]$ lies in the variety

$$\mathcal{Z}^{(1)} = \left\{\vec{x} \in \mathbb{R}^J : Q(\vec{\xi}[I], \vec{x}) = z \text{ and } Q(\vec{\xi}'[I], \vec{x}) = z\right\} = \mathcal{Z}_{\vec{\xi}[I]} \cap \mathcal{Z}_{\vec{\xi}'[I]} \subseteq \mathbb{R}^J,$$

where, for $\vec{u} \in \mathbb{R}^I$, we write $\mathcal{Z}_{\vec{u}}$ for the set of all $\vec{x} \in \mathbb{R}^J$ with $(\vec{u}, \vec{x}) \in \mathcal{Z}$ (typically, this is a quadric in $\mathbb{R}^J$). Using this language, (2.1) can be restated as

$$\Pr\left[Q(\vec{\xi}) = z\right] = \Pr\left[\vec{\xi} \in \mathcal{Z}\right] \leq \Pr\left[\vec{\xi}[J] \in \mathcal{Z}^{(1)}\right]^{1/2}. \tag{2.3}$$

The next step leading to the traditional decoupling inequality (2.2) can geometrically be phrased as observing that $\mathcal{Z}^{(1)}$ lies inside the affine-linear subspace

$$\mathcal{W}^{(1)} = \left\{\vec{x} \in \mathbb{R}^J : Q(\vec{\xi}[I], \vec{x}) - Q(\vec{\xi}'[I], \vec{x}) = 0\right\}.$$

Indeed, this yields the probability bound

$$\Pr\left[Q(\vec{\xi}) = z\right] \leq \Pr\left[\vec{\xi}[J] \in \mathcal{Z}^{(1)}\right]^{1/2} \leq \Pr\left[\vec{\xi}[J] \in \mathcal{W}^{(1)}\right]^{1/2}$$

stated in (2.2). One can then forget about $\mathcal{Z}^{(1)}$ and restrict one's attention to the affine-linear subspace $\mathcal{W}^{(1)}$, where the relevant probabilities are easier to analyse (under suitable assumptions on Q, one can show that $\Pr[\vec{\xi}[J] \in \mathcal{W}^{(1)}] \leq O(1/\sqrt{n})$, leading to the bound $\Pr[Q(\vec{\xi}) = z] \leq O(n^{-1/4})$ described in Section 2.2).

However, it turns out that this "forgetting" of $\mathcal{Z}^{(1)}$ is precisely the cause of the square-root loss usually associated with decoupling. Indeed, $\mathcal{W}^{(1)}$ is a variety with codimension 1 (being defined by a single equation), while $\mathcal{Z}^{(1)}$ is typically[9] a variety with codimension (at least) 2 (being defined by two equations), so intuitively we should be much less likely to have $\vec{\xi}[J] \in \mathcal{Z}^{(1)}$ than $\vec{\xi}[J] \in \mathcal{W}^{(1)}$. More specifically, in order to have $\vec{\xi}[J] \in \mathcal{Z}^{(1)}$, we need $\vec{\xi}[J]$ to satisfy two different equations simultaneously, and we might expect each of these to be satisfied with probability $O(1/\sqrt{n})$. So for typical outcomes of $\vec{\xi}[I], \vec{\xi}'[I]$ (which determine $\mathcal{Z}^{(1)}$) we might expect $\Pr[\vec{\xi}[J] \in \mathcal{Z}^{(1)} \mid \vec{\xi}[I], \vec{\xi}'[I]] \leq (O(1/\sqrt{n}))^2$. If we could prove this, we would be able to deduce Theorem 1.1 (recalling (2.3)).

So, we choose not to "forget" $\mathcal{Z}^{(1)}$, and our task is to show that $\Pr[\vec{\xi}[J] \in \mathcal{Z}^{(1)}] \leq (O(1/\sqrt{n}))^2$. While this new task may seem harder than the previous one (as $\mathcal{Z}^{(1)}$ seems like a more complicated object than $\mathcal{Z}$), the key observation is that we have "reduced the relative proportion of the quadratic part of our problem". Indeed, at the start we were interested in a quadric $\mathcal{Z}$ described by a single quadratic equation, but now we are interested in $\mathcal{Z}^{(1)}$, which can be interpreted as a quadric *inside* the affine-linear subspace $\mathcal{W}^{(1)} \subseteq \mathbb{R}^J$. That is to say, $\mathcal{Z}^{(1)}$ is described by one linear and one quadratic equation, so now "only half of our problem is quadratic".

Crucially, it is possible to iterate this entire procedure: we next fix a partition $I^{(2)} \cup J^{(2)}$ of J, and consider the variety

$$\mathcal{Z}^{(2)} = \mathcal{Z}^{(1)}_{\vec{\xi}[I^{(2)}]} \cap \mathcal{Z}^{(1)}_{\vec{\xi}'[I^{(2)}]} \subseteq \mathbb{R}^{J^{(2)}}$$

---

[9]There are certain degenerate situations where $\mathcal{Z}^{(1)}$ does not have codimension at least 2. We will ignore degeneracies of this type throughout this outline, but they do cause challenges in the actual proof.

(where, for $\vec{u} \in \mathbb{R}^{I^{(2)}}$, we write $\mathcal{Z}^{(1)}_{\vec{u}}$ for the set of all $\vec{x} \in \mathbb{R}^{J^{(2)}}$ with $(\vec{u}, \vec{x}) \in \mathcal{Z}^{(1)}$). Now, decoupling, analogously to the inequality in (2.3), yields

$$\Pr\!\Big[\vec{\xi}[J] \in \mathcal{Z}^{(1)} \,\Big|\, \vec{\xi}[I], \vec{\xi}'[I]\Big] \le \Pr\!\Big[\vec{\xi}[J^{(2)}] \in \mathcal{Z}^{(2)} \,\Big|\, \vec{\xi}[I], \vec{\xi}'[I]\Big]^{1/2}. \tag{2.4}$$

So, it suffices to show that $\Pr\!\big[\vec{\xi}[J^{(2)}] \in \mathcal{Z}^{(2)} \,\big|\, \vec{\xi}[I], \vec{\xi}'[I]\big] \le (O(1/\sqrt{n}))^4$ for typical outcomes of $\vec{\xi}[I]$ and $\vec{\xi}'[I]$. Now, for $\mathcal{Z}^{(2)}$ to be nonempty, it must be the case that $\mathcal{W}^{(1)}_{\vec{\xi}[I^{(2)}]} \cap \mathcal{W}^{(1)}_{\vec{\xi}'[I^{(2)}]}$ is nonempty, or equivalently that $\mathcal{W}^{(1)}_{\vec{\xi}[I^{(2)}]} = \mathcal{W}^{(1)}_{\vec{\xi}'[I^{(2)}]}$ (it is not hard to see that $\mathcal{W}^{(1)}_{\vec{\xi}[I^{(2)}]}$ and $\mathcal{W}^{(1)}_{\vec{\xi}'[I^{(2)}]}$ are parallel translates of each other). That is to say, $\vec{\xi}[I^{(2)}] - \vec{\xi}'[I^{(2)}]$ must lie in a certain affine-linear subspace which typically has codimension 1; this happens with probability $O(1/\sqrt{n})$ by the (linear) Erdős–Littlewood–Offord theorem.

If we also condition on outcomes of $\vec{\xi}[I^{(2)}]$ and $\vec{\xi}'[I^{(2)}]$ such that $\mathcal{W}^{(1)}_{\vec{\xi}[I^{(2)}]} = \mathcal{W}^{(1)}_{\vec{\xi}'[I^{(2)}]}$, it is not hard to see that $\mathcal{Z}^{(2)}$ is typically a quadric inside the affine-linear subspace $\mathcal{W}^{(2)}$ of $\mathcal{W}^{(1)}_{\vec{\xi}[I^{(2)}]} = \mathcal{W}^{(1)}_{\vec{\xi}'[I^{(2)}]} \subseteq \mathbb{R}^{J^{(2)}}$ given by the linear equation $Q(\vec{\xi}[I], \vec{\xi}[I^{(2)}], \vec{x}) - Q(\vec{\xi}[I], \vec{\xi}'[I^{(2)}], \vec{x}) = 0$ (in much the same way that $\mathcal{Z}^{(1)}$ is a quadric inside the affine-linear subspace $\mathcal{W}^{(1)}$). That is to say, $\mathcal{Z}^{(2)}$ is typically a variety of codimension (at least) 3, described by two linear equations and one quadratic equation. So we might expect that (for typical outcomes of $\vec{\xi}[I], \vec{\xi}'[I], \vec{\xi}[I^{(2)}], \vec{\xi}'[I^{(2)}]$ for which $\mathcal{W}^{(1)}_{\vec{\xi}[I^{(2)}]} = \mathcal{W}^{(1)}_{\vec{\xi}'[I^{(2)}]}$)

$$\Pr\!\Big[\vec{\xi}[J^{(2)}] \in \mathcal{Z}^{(2)} \,\Big|\, \vec{\xi}[I], \vec{\xi}'[I], \vec{\xi}[I^{(2)}], \vec{\xi}'[I^{(2)}]\Big] \le \big(O(1/\sqrt{n})\big)^3. \tag{2.5}$$

If we were able to prove (2.5), we would obtain a bound of the form

$$\Pr\!\Big[\vec{\xi}[J^{(2)}] \in \mathcal{Z}^{(2)}\Big] \le O(1/\sqrt{n}) \cdot \big(O(1/\sqrt{n})\big)^3 = \big(O(1/\sqrt{n})\big)^4,$$

which would imply Theorem 1.1, tracing back through our decoupling inequalities (2.3) and (2.4). We have made progress by "reducing the proportion of our problem that is quadratic": If, instead of (2.5), we were only able to prove that

$$\Pr\!\Big[\vec{\xi}[J^{(2)}] \in \mathcal{Z}^{(2)} \,\Big|\, \vec{\xi}[I], \vec{\xi}'[I], \vec{\xi}[I^{(2)}], \vec{\xi}'[I^{(2)}]\Big] \le \Pr\!\Big[\vec{\xi}[J^{(2)}] \in \mathcal{W}^{(2)} \,\Big|\, \vec{\xi}[I], \vec{\xi}'[I], \vec{\xi}[I^{(2)}], \vec{\xi}'[I^{(2)}]\Big] \le \big(O(1/\sqrt{n})\big)^2$$

("forgetting" the quadratic part of the problem and only bounding the probability that $\vec{\xi}[J^{(2)}]$ lies in the affine-linear subspace $\mathcal{W}^{(2)}$ of codimension 2), we would end up with a final bound of the form $\Pr[\vec{\xi} \in \mathcal{Z}] \le O(n^{-3/8})$, which is much better than the $O(n^{-1/4})$ bound we obtained with a single decoupling step.

In general, after k steps of this scheme, we will have considered k "nested" partitions of the form $J^{(i-1)} = I^{(i)} \cup J^{(i)}$, and defined k varieties $\mathcal{Z}^{(1)}, \ldots, \mathcal{Z}^{(k)}$. We will have applied the decoupling inequality k times, and considered various conditional probabilities that the vectors $\vec{\xi}[I^{(i)}] - \vec{\xi}'[I^{(i)}]$ lie in certain affine-linear subspaces (to ensure that certain intersections $\mathcal{W}^{(i-1)}_{\vec{\xi}[I^{(i)}]} \cap \mathcal{W}^{(i-1)}_{\vec{\xi}'[I^{(i)}]}$ are nonempty). After all this, we find ourselves in a position where if we "forget the quadratic part of the problem" we obtain a bound of the form $\Pr[Q(\vec{\xi}) = z] \le O(n^{-1/2+1/2^{k+1}})$. That is to say, if k is at least $\log\log n$, the quadratic part of the problem is so insignificant that "forgetting" it only costs us a constant factor in the final bound.

At an extremely high level, this explains our strategy to prove Theorem 1.1. However, we omitted a number of important details in this outline. In particular, our strategy heavily depends on being able to obtain suitable upper bounds on probabilities that certain random vectors fall into certain affine-linear subspaces, and there are crucial "robust rank" nondegeneracy conditions that must be satisfied in order for such bounds to hold. This is not just a technicality; we require significant new ideas to maintain these robust rank properties during our iterative decoupling scheme, which we discuss next.

Remark 2.1. Theorem 1.1 concerns quadratic polynomials, but one may be interested in a generalisation to cubic polynomials, or to polynomials of any fixed degree. Decoupling still makes sense for general polynomials: for example, if $\mathcal{Z}$ is a cubic (variety), then decoupling yields an inequality of the form

$\Pr[\vec{\xi} \in \mathcal{Z}] \le \Pr[\vec{\xi}[J] \in \mathcal{Z}^{(2)}]^{1/2}$, where $\mathcal{Z}^{(2)}$ is the intersection of a quadric and a cubic. As observed by Rosiński and Samorodnitsky [44] and Razborov and Viola [43], the basic type of decoupling argument described in Section 2.2 generalises quite straightforwardly to higher degrees (but the bounds get worse as the degree increases). However, it is less clear how to generalise the inductive decoupling scheme described in this subsection. Roughly speaking, in the degree-d case, instead of affine-linear subspaces (obtained as intersections of affine-linear hyperplanes), one must work with intersections of degree-$(d-1)$ varieties, which are much more complicated objects. We hope that the relevant complexities can be handled with some kind of multiple-level induction, but so far we were not able to accomplish this.

2.4   High-dimensional anticoncentration inequalities, and witness-counting

Our proof, as outlined in the previous subsection, relies on bounds on probabilities that random vectors lie in certain affine-linear subspaces. More specifically, for a suitably nondegenerate affine-linear subspace $\mathcal{W} \subseteq \mathbb{R}^n$ of codimension k, and a uniformly random vector $\vec{\xi} \in \{1, -1\}^n$, we need a probability bound of the form $\Pr[\vec{\xi} \in \mathcal{W}] \le O(n^{-k/2})$. Intuitively, this is because $\vec{\xi}$ needs to simultaneously satisfy k different linear equations, each of which is satisfied with probability roughly $n^{-1/2}$. More formally, such a bound follows from a high-dimensional version of the Erdős–Littlewood–Offord theorem.

The first such high-dimensional version was due to Halász [21]. In linear-algebraic language, it can be phrased as follows: for any fixed k, if $M \in \mathbb{R}^{k \times n}$ is a matrix that "robustly has rank k" in the sense that (for some fixed $\delta > 0$) one cannot delete δn columns of M to obtain a matrix with rank less than k, then for a uniformly random vector $\vec{\xi} \in \{-1, 1\}^n$ we have

$$\sup_{\vec{w} \in \mathbb{R}^k} \Pr[M\vec{\xi} = \vec{w}] \le O(n^{-k/2}).$$

Note that some kind of "robust rank k" condition is necessary here: for example, if n is even and $M \in \mathbb{R}^{2 \times n}$ has rows $(1, \ldots, 1) \in \mathbb{R}^n$ and $(1, \ldots, 1, 0, 0) \in \mathbb{R}^n$, then it is easy to check that $\Pr[M\vec{\xi} = \vec{0}]$ has order of magnitude $1/\sqrt{n}$.

Several extensions and variants of Halasz' inequality have since been proved (see for example [14, 16, 23, 47]); in particular, Ferber, Jain and Zhao [14] proved a version of Halasz' theorem with a much better dependence on k (allowing k to vary with n, instead of viewing it as a constant). We state (a corollary of) this theorem as Theorem 3.3.

Of course, whenever we want to apply any Halász-type theorem, we need a "robust rank" condition to hold. So, in order to execute the strategy described in the last subsection, at each step of the decoupling scheme we need a "robust rank inheritance" lemma, proving that a robust rank condition is likely to hold for the next step, given that it holds for the current step. The key ingredient for our robust rank inheritance lemma is a new high-dimensional anticoncentration inequality for the probability that a random vector falls in a small ball in the *Hamming* norm. We believe this inequality (and the techniques in its proof) to be of independent interest; an important special case is as follows.

Lemma 2.1. *For any fixed positive integer r, there are constants $C_r > 0$ and $c_r > 0$ only depending on r such that the following holds. Consider a matrix $A \in \mathbb{R}^{m \times n}$ which has rank at least r after deletion of any t rows and t columns (for some positive integer t). Then for a sequence $\vec{\xi} = (\xi_1, \ldots, \xi_n) \in \{-1, 1\}^n$ of independent Rademacher random variables, we have*

$$\sup_{\vec{v} \in \mathbb{R}^m} \Pr[A\vec{\xi} \text{ differs from } \vec{v} \text{ in fewer than } c_r t \text{ coordinates}] \le C_r \cdot t^{-r/2}.$$

The assumption in Lemma 2.1 says that A robustly has rank at least r, but in a stronger sense than typical Halász-type theorems: the rank needs to remain at least r after *row* deletion as well as column deletion. As a result, we are able to obtain a stronger conclusion, namely that for any vector $\vec{v}$, it is unlikely that $A\vec{\xi}$ agrees with $\vec{v}$ in almost all its coordinates. (It is not hard to see that for this stronger conclusion, such a row-deletion assumption is indeed necessary).

We prove Lemma 2.1 (and our more general robust rank inheritance lemma) in Section 7 using a *witness-counting* technique, which we outline here. First, note that the most naïve strategy to prove Lemma 2.1 would be to simply take a union bound over all sets I of $m - c_r t$ coordinates in which $A\vec{\xi}$ and $\vec{v}$ could agree. For some specific I, the probability that $A\vec{\xi}$ and $\vec{v}$ agree on the coordinates indexed by I can be easily understood using existing tools (i.e., Halász' inequality and its variants) and is at most on the order of $t^{-r/2}$. However, it is far too wasteful to simply take a union bound summing over all possibilities for I (the number of possibilities is exponential in t).

Instead, for each I we consider small "witness" subsets $I' \subseteq I$, such that the submatrix of A consisting of the rows with indices in I' still has (robustly) high rank. Note that for each $I' \subseteq I$, whenever $A\vec{\xi}$ and $\vec{v}$ agree on the coordinates indexed by I, they also in particular agree on the coordinates indexed by I'. Given the high-rank property, for each "witness" subset I' we can still easily bound the probability that $A\vec{\xi}$ and $\vec{v}$ agree on the coordinates indexed by I' (using Halász' inequality and its variants).

There are still too many possible "witness" subsets I' to be able to simply take a union bound over all possibilities for I', but (roughly speaking) we can show that whenever $A\vec{\xi}$ and $\vec{v}$ agree on a large set of coordinates I, then they must agree on the coordinates of *many* "witness" subsets $I' \subseteq I$. We can show this is unlikely by computing the expected number of "witness" coordinate-subsets on which $A\vec{\xi}$ and $\vec{v}$ agree, and applying Markov's inequality.

Remark 2.2. One might be interested in adapting Theorem 1.1 to study small-ball probabilities of the form $\sup_{z \in \mathbb{R}} \Pr[|Q(\xi_1, \ldots, \xi_n) - z| \leq 1]$ instead of point probabilities $\sup_{z \in \mathbb{R}} \Pr[Q(\xi_1, \ldots, \xi_n) = z]$. The robust rank inheritance lemma seems to be the main point of difficulty for such an adaptation; it is not clear how to extend the witness-counting arguments to the small-ball setting (e.g., in the setting of Lemma 2.1, we would be interested in the event that there are fewer than $c_r t$ coordinates i for which $|(A\vec{\xi} - \vec{v})[i]| \leq 1$).

Before ending this overview section, we remark that there is a way to sidestep the robust rank inheritance issue in the special case where Q has "bounded rank", meaning that the quadratic part of Q can be written as $\vec{x}^\mathsf{T} A \vec{x}$ for some symmetric matrix A of rank $O(1)$. Indeed, in this case we can reduce our entire problem to a certain *bounded-dimensional* geometric anticoncentration problem (involving a quadric), where the robust rank conditions for Halasz' inequality are always automatically satisfied when following the strategy in the previous subsection. In Section 4, we give a simple self-contained proof of an essentially optimal anticoncentration bound in this setting. We believe this to be a good illustration of the basic principles of our inductive decoupling scheme, with minimal technicalities (the results of Section 4 will actually also be used later in the paper).

3 Preliminaries

First, as outlined in Section 2, we need a "high-dimensional" version of the Erdős–Littlewood–Offord theorem. This will require a robust rank condition, defined as follows.

Definition 3.1. *For integers $0 \leq k \leq n$ and a real number $s \geq 0$, let $\mathcal{H}^{k \times n}(s)$ be the set of matrices $M \in \mathbb{R}^{k \times n}$ such that* $\operatorname{rank} M[[k] \times J] = k$ *for all subsets $J \subseteq [n]$ of size $|J| \geq n - s$, i.e., the set of matrices having rank k after any deletion of up to s columns.*

We clearly have $\mathcal{H}^{k \times n}(s) \subseteq \mathcal{H}^{k \times n}(s')$ if $s \geq s'$. Furthermore, for a partition $[n] = S \cup I$ with $|S| \leq s$, for every matrix $M \in \mathcal{H}^{k \times n}(s)$ we also have $M[[k] \times I] \in \mathcal{H}^{k \times I}(s - |S|)$. Also note that in the case $k = 0$, the (unique) empty matrix $M \in \mathbb{R}^{0 \times n}$ is contained in $\mathcal{H}^{0 \times n}(s)$ for all $s \geq 0$.

For integers $0 \leq k \leq n$ and $t \geq 0$, we say that a matrix $M \in \mathbb{R}^{k \times n}$ contains t disjoint nonsingular $k \times k$ submatrices if there exist disjoint subsets $J_1, \ldots, J_t \subseteq [n]$ of size $|J_1| = \cdots = |J_t| = k$ such that $\operatorname{rank} M[[k] \times J_i] = k$ for $i = 1, \ldots, t$. This is the case if and only if among the column vectors $\vec{a}_1, \ldots, \vec{a}_n$ of M we can form t disjoint bases of $\mathbb{R}^k$ (here, the vectors $\vec{a}_1, \ldots, \vec{a}_n$ are considered with multiplicities, i.e., a vector in $\mathbb{R}^k$ can occur in two different bases if it occurs twice among $\vec{a}_1, \ldots, \vec{a}_n$).

Lemma 3.2. *For integers $0 \leq k \leq n$ and a real number $s \geq 0$, the following statements hold:*

(i) If $M \in \mathbb{R}^{k \times n}$ contains t disjoint nonsingular $k \times k$ submatrices for an integer $t > s$, then $M \in \mathcal{H}^{k \times n}(s)$.

(ii) If $M \in \mathcal{H}^{k \times n}(s)$ and $k \geq 1$, then M contains $\lceil s/k \rceil$ disjoint nonsingular $k \times k$ submatrices.

Proof. For (i), let $J_1, \ldots, J_t \subseteq [n]$ be disjoint subsets of size $|J_1| = \cdots = |J_t| = k$ such that rank $M[[k] \times J_i] = k$ for $i = 1, \ldots, t$. Then for any subsets $J \subseteq [n]$ of size $|J| \geq n - s > n - t$, we have $J_i \subseteq J$ for some $i \in \{1, \ldots, t\}$ and hence $k \geq \text{rank } M[[k] \times J] \geq \text{rank } M[[k] \times J_i] = k$, meaning that rank $M[[k] \times J] = k$.

For (ii), we can greedily find disjoint subsets $J_1, \ldots, J_{\lceil s/k \rceil} \subseteq [n]$ of size $|J_1| = \cdots = |J_{\lceil s/k \rceil}| = k$ such that rank $M[[k] \times J_i] = k$ for $i = 1, \ldots, \lceil s/k \rceil$. Indeed, after having chosen $J_1, \ldots, J_{i-1}$ for some $i \leq \lceil s/k \rceil$, we have $|J_1 \cup \cdots \cup J_{i-1}| = (i-1)k < s$ and hence M still has rank k after deleting the columns with indices in $J_1 \cup \cdots \cup J_{i-1}$. So we can find a subset $J_i \subseteq [n] \setminus (J_1 \cup \cdots \cup J_{i-1})$ of size $|J_i| = k$ with rank $M[[k] \times J_i] = k$. $\square$

Halász [21] proved that for fixed constants $\delta > 0$ and $k \in \mathbb{N}$, if $M \in \mathcal{H}^{k \times n}(\delta n)$ then $\Pr[M\vec{\xi} = \vec{w}] \leq O(n^{-k/2})$. We will use the following quantitative version of Halász' theorem, which is a special case[10] of a result of Ferber, Jain and Zhao [14, Theorem 1.11] (see also [23] for a previous result with a weaker dependence on k).

Theorem 3.3. *Let $0 \leq k \leq n$ and $t \geq 1$ be integers. Consider a matrix $M \in \mathbb{R}^{k \times n}$ containing t disjoint nonsingular $k \times k$ submatrices, and a vector $\vec{w} \in \mathbb{R}^k$. Letting $\vec{\xi} = (\xi_1, \ldots, \xi_n) \in \{-1, 1\}^n$ be a sequence of independent Rademacher random variables, we then have*

$$\mathbb{P}[M\vec{\xi} = \vec{w}] \leq t^{-k/2}.$$

Corollary 3.4. *Let $0 \leq d \leq k \leq n$ and $t \geq 1$ be integers. Consider vectors $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{R}^k$ such that one can form t disjoint bases of $\mathbb{R}^k$ from the vectors $\vec{a}_1, \ldots, \vec{a}_n$, and let $\mathcal{W} \subseteq \mathbb{R}^k$ be a d-dimensional affine-linear subspace. Letting $\vec{\xi} = (\xi_1, \ldots, \xi_n) \in \{-1, 1\}^n$ be a sequence of independent Rademacher random variables, we then have*

$$\Pr[\xi_1 \vec{a}_1 + \cdots + \xi_n \vec{a}_n \in \mathcal{W}] \leq t^{-(k-d)/2}.$$

Proof. Write $\tilde{\mathcal{W}}$ for the d-dimensional linear subspace parallel to the affine-linear subspace $\mathcal{W}$ (i.e., $\mathcal{W} = \tilde{\mathcal{W}} + \vec{v}$ for some $\vec{v} \in \mathbb{R}^k$), and consider a linear map $\phi : \mathbb{R}^k \to \mathbb{R}^{k-d}$ with kernel $\tilde{\mathcal{W}}$ (so $\phi(\mathcal{W})$ consists of a single point $\vec{p} \in \mathbb{R}^{k-d}$). Writing $M \in \mathbb{R}^{(k-d) \times n}$ for the matrix whose columns are the vectors $\phi(\vec{a}_1), \ldots, \phi(\vec{a}_n)$, we have

$$\Pr[\xi_1 \vec{a}_1 + \cdots + \xi_n \vec{a}_n \in \mathcal{W}] = \Pr[\xi_1 \phi(\vec{a}_1) + \cdots + \xi_n \phi(\vec{a}_n) = \vec{p}] = \Pr[M\vec{\xi} = \vec{p}] \leq t^{-(k-d)/2}.$$

by Theorem 3.3 (note that for each of the t disjoint bases formed among the vectors $\vec{a}_1, \ldots, \vec{a}_n$, the image under ϕ is a spanning set of $\mathbb{R}^{k-d}$ and hence contains a basis of $\mathbb{R}^{k-d}$, so we can find t disjoint bases among the columns of M and consequently M contains t disjoint nonsingular $(k-d) \times (k-d)$ submatrices). $\square$

Corollary 3.5. *Let $1 \leq k \leq n$ be integers and $s > 0$ be a real number. Consider a matrix $M \in \mathcal{H}^{k \times n}(s)$ and a vector $\vec{w} \in \mathbb{R}^k$. Letting $\vec{\xi} = (\xi_1, \ldots, \xi_n) \in \{-1, 1\}^n$ be a sequence of independent Rademacher random variables, we then have*

$$\mathbb{P}[M\vec{\xi} = \vec{w}] \leq (s/k)^{-k/2}.$$

Proof. By Lemma 3.2(ii) M contains $\lceil s/k \rceil$ disjoint nonsingular $k \times k$ submatrices, so we can apply Theorem 3.3. $\square$

Next, we recall some basic facts about linear forms. A *linear form* $g \in \mathbb{R}[x_1, \ldots, x_n]$ is a linear polynomial with constant term zero. That is, we can write $g(\vec{x}) = \vec{v} \cdot \vec{x} = \vec{v}^\mathsf{T} \vec{x} = \vec{x}^\mathsf{T} \vec{v}$, where $\vec{v} \in \mathbb{R}^n$ is the *coefficient vector* of g. Note that for two linear forms $g_1, g_2 \in \mathbb{R}[x_1, \ldots, x_n]$ with coefficient vectors $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^n$ we have

$$g_1(\vec{x}) \cdot g_2(\vec{x}) = \frac{1}{2} g_1(\vec{x}) \cdot g_2(\vec{x}) + \frac{1}{2} g_2(\vec{x}) \cdot g_1(\vec{x}) = \frac{1}{2} \vec{x}^\mathsf{T} \vec{v}_1 \vec{v}_2^\mathsf{T} \vec{x} + \frac{1}{2} \vec{x}^\mathsf{T} \vec{v}_2 \vec{v}_1^\mathsf{T} \vec{x} = \vec{x}^\mathsf{T} A \vec{x}, \qquad (3.1)$$

---

[10]Specifically, to deduce Theorem 3.3 from [14, Theorem 1.11], we take the even number ℓ in [14, Theorem 1.11] to be $\ell = t$ if t is even, and $\ell = t - 1$ if t is odd. In both cases, we can verify the inequality $2^{-\ell} \binom{\ell}{\ell/2} \leq t^{-1/2}$, and we obtain the desired bound by taking $\mathcal{A}_1, \ldots, \mathcal{A}_\ell$ in [14, Theorem 1.11] to be a partition of the columns of M into ℓ disjoint subsets each containing a basis of $\mathbb{R}^k$ (then $r_1 = \cdots = r_\ell = k$).

where $A \in \mathbb{R}^{n \times n}$ is the symmetric matrix given by $A = \frac{1}{2}(\vec{v}_1 \vec{v}_2^\mathsf{T} + \vec{v}_2 \vec{v}_1^\mathsf{T})$. More generally, every quadratic form (i.e., every quadratic polynomial with no linear and no constant term), can be written in the form $\vec{x}^\mathsf{T} A \vec{x}$ for a unique symmetric matrix $A \in \mathbb{R}^{n \times n}$.

Another important ingredient for our proof is the following *decoupling lemma*.

Lemma 3.6. *If an event $\mathcal{E}(X,Y)$ depends on independent random objects X, Y, and X' is an independent copy of X, then $\Pr[\mathcal{E}(X,Y)] \leq \Pr[\mathcal{E}(X,Y) \text{ and } \mathcal{E}(X',Y)]^{1/2}$.*

Lemma 3.6 is a slight variant of a lemma of Costello, Tao and Vu [6, Lemma 4.7], who popularised decoupling as a tool for polynomial anticoncentration. The particular statement of Lemma 3.6 appears (for example) as [5, Lemma 14]. For the reader's convenience, we include the proof (which is a simple application of the Cauchy–Schwarz inequality).

Proof of Lemma 3.6. By the Cauchy–Schwarz inequality, we have

$$\Pr[\mathcal{E}(X,Y) \text{ and } \mathcal{E}(X',Y)] = \mathbb{E}_Y\Big[\Pr[\mathcal{E}(X,Y) \text{ and } \mathcal{E}(X',Y) \,|\, Y]\Big] = \mathbb{E}_Y\Big[\Pr[\mathcal{E}(X,Y) \,|\, Y]^2\Big]$$

$$\geq \mathbb{E}_Y\Big[\Pr[\mathcal{E}(X,Y) \,|\, Y]\Big]^2 = \Pr[\mathcal{E}(X,Y)]^2.$$

Taking square roots on both sides gives the desired inequality. $\qquad\square$

We will also need a simple lemma usually attributed to Odlyzko [41].

Lemma 3.7. *Consider a matrix $M \in \mathbb{R}^{k \times n}$ with $\operatorname{rank} M = k$ and a vector $\vec{w} \in \mathbb{R}^k$. For a sequence of independent Rademacher random variables $\vec{\xi} = (\xi_1, \ldots, \xi_n) \in \{-1, 1\}^n$, we then have $\mathbb{P}[M\vec{\xi} = \vec{w}] \leq 2^{-k}$.*

Proof. We can interpret $M\vec{\xi} = \vec{w}$ as a system of linear equations (in the variables $\xi_1, \ldots, \xi_n$). Bringing this system into row echelon form, it has $n - k$ free variables (which determine the values of the k remaining variables). After exposing the $n - k$ entries of $\vec{\xi}$ corresponding to the free variables, there is at most one possibility for each of the remaining k entries of $\vec{\xi}$ satisfying this system of equations. Thus, the probability of having $M\vec{\xi} = \vec{w}$ is at most 2^{-k}. $\qquad\square$

Finally, we will need a simple numerical inequality.

Lemma 3.8. *For real numbers $a, b, c \geq 0$ with $a^2 \leq ab + c$, we have $a \leq b + \sqrt{c}$.*

Proof. Note that for all $x, y \geq 0$ we have the inequality $\sqrt{x + y} \leq \sqrt{x} + \sqrt{y}$. Now, by the quadratic formula, $a^2 \leq ab + c$ implies $a \leq (b + \sqrt{b^2 + 4c})/2 \leq (b + \sqrt{b^2} + \sqrt{4c})/2 = b + \sqrt{c}$. $\qquad\square$

4 Inductive decoupling from a geometric point of view

Note that for a quadratic polynomial Q and a sequence $\vec{\xi} = (\xi_1, \ldots, \xi_n) \in \{-1, 1\}^n$ of independent Rademacher random variables, the event $Q(\vec{\xi}) = 0$ is precisely the event that $\vec{\xi}$ falls in the vanishing locus of Q. Moreover, if the quadratic part of Q has rank less than r (i.e., if Q can be expressed as a linear combination of $r - 1$ squares of linear forms, plus an additional linear form, plus a constant term), then it is possible to interpret the event $Q(\vec{\xi}) = 0$ as the event that $\xi_1 \vec{a}_1 + \cdots + \xi_n \vec{a}_n \in \mathcal{Z}$, where $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{R}^r$ are vectors in r-dimensional space, and $\mathcal{Z}$ is the vanishing locus of some r-variable quadratic polynomial (indeed, the entries of each $\vec{a}_i$ correspond to the coefficients of ξ_i in each of the linear forms described above).

In this section, we obtain an essentially optimal bound (in Theorem 4.2 below) on probabilities of the form $\Pr[\xi_1 \vec{a}_1 + \cdots + \xi_n \vec{a}_n \in \mathcal{Z}]$, where $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{R}^r$ are vectors satisfying some robust nondegeneracy condition and $\mathcal{Z}$ is a quadric in $\mathbb{R}^r$ (or more generally, a quadric inside some affine-linear subspace of $\mathbb{R}^r$), under the
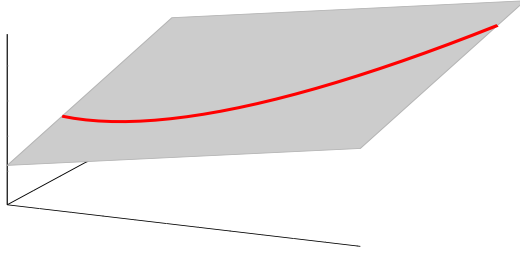
Figure 4.1. An example of a quadric on a 2-dimensional affine plane in $\mathbb{R}^3$.

assumption that the dimension r does not grow with n. This may be viewed as a warm-up to the full proof of Theorem 1.1: it is proved via the same inductive decoupling scheme, but has fewer technicalities.

Using the connection described in the first paragraph above (with an additional "dropping to a subspace" argument, of the type we will later see in Section 6), one can use Theorem 4.2 to prove Theorem 1.1 in the special case where the quadratic part of Q has bounded rank. Actually, Theorem 4.2 is also an ingredient in the full proof of our main theorem (Theorem 1.1). We also remark that the result of this section fits nicely in the "geometric Littlewood–Offord" framework of Fox, Kwan and Spink [17]. In particular, the case of Theorem 4.2 where $\mathcal{Z} \subseteq \mathbb{R}^4$ is a sphere in four dimensions was explicitly raised as the simplest open case of the main problem in [17].

Before stating the main result of this section, we record some relevant terminology.

Definition 4.1. *As usual, for a polynomial $P \in \mathbb{R}[x_1, \ldots, x_d]$ we write $\mathrm{V}(P) = \{\vec{x} \in \mathbb{R}^d : P(\vec{x}) = 0\}$ for the vanishing locus of P. A* quadric *$\mathcal{Z} \subsetneq \mathbb{R}^d$ is the vanishing locus $\mathcal{Z} = \mathrm{V}(P)$ of some nonzero quadratic polynomial $P \in \mathbb{R}[x_1, \ldots, x_d]$. We say that a quadric $\mathcal{Z} \subsetneq \mathbb{R}^d$ is* irreducible *if $\mathcal{Z} = \mathrm{V}(P)$ for an irreducible quadratic polynomial $P \in \mathbb{R}[x_1, \ldots, x_d]$.*

For a d-dimensional affine-linear subspace $\mathcal{W} \subseteq \mathbb{R}^r$, we say that $\mathcal{Z} \subsetneq \mathcal{W}$ is a quadric on $\mathcal{W}$ *if it is the image of some quadric $\mathrm{V}(P) \subsetneq \mathbb{R}^d$ under an affine-linear isomorphism $\phi : \mathbb{R}^d \to \mathcal{W}$. Equivalently, $\mathcal{Z} \subsetneq \mathcal{W}$ is a quadric on $\mathcal{W}$ if and only if $\mathcal{Z} = \mathcal{W} \cap \mathrm{V}(P)$ for some quadratic polynomial $P \in \mathbb{R}[x_1, \ldots, x_r]$ with $\mathcal{W} \not\subseteq \mathrm{V}(P)$. We say that a quadric $\mathcal{Z} \subsetneq \mathcal{W}$ is* irreducible *if it is the image of some irreducible quadric $\mathrm{V}(P) \subsetneq \mathbb{R}^d$ under an affine-linear isomorphism $\phi : \mathbb{R}^d \to \mathcal{W}$.*

As an example, see Figure 4.1 showing a quadric on a 2-dimensional affine plane in $\mathbb{R}^3$. Now we are ready to state the main result of this section.

Theorem 4.2. *Let $0 \leq d < r$ be integers. Let $\mathcal{Z} \subsetneq \mathcal{W}$ be a quadric on a $(d+1)$-dimensional affine-linear subspace $\mathcal{W} \subseteq \mathbb{R}^r$. Consider vectors $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{R}^r$ such that one can form t disjoint bases of $\mathbb{R}^r$ from the vectors $\vec{a}_1, \ldots, \vec{a}_n$, where t is a positive integer divisible by 2^d. Let $(\xi_1, \ldots, \xi_n) \in \{-1, 1\}^n$ be a sequence of independent Rademacher random variables. Then*

$$\Pr[\xi_1 \vec{a}_1 + \cdots + \xi_n \vec{a}_n \in \mathcal{Z}] \leq \frac{2^{dr+1}}{t^{(r-d)/2}}.$$

If $\mathcal{Z} \subsetneq \mathcal{W}$ is a quadric on a $(d+1)$-dimensional affine-linear subspace $\mathcal{W}$, then $\mathcal{Z}$ has dimension at most d (we do not formally define what "dimension" means in this context, as this is not needed for our arguments, but appeal to the reader's intuition). Note that the form of the bound in the above theorem (with $t^{(r-d)/2}$ in the denominator) is the same as in Corollary 3.4 for d-dimensional affine-linear subspaces.

For the proof of Theorem 4.2, we will rely on the following algebraic fact.

Lemma 4.3. *Let $d \geq 2$, and let $\mathcal{Z} \subsetneq \mathbb{R}^d$ be an irreducible quadric. Then, at least one of the following holds:*

13

(i) *There is a direction $\vec{v} \in \mathbb{R}^d \setminus \{\vec{0}\}$ such that $\mathcal{Z} + \mathbb{R}\vec{v} = \mathcal{Z}$ (i.e., $\mathcal{Z}$ is invariant under translation along the direction $\vec{v}$), or*

(ii) *for any vectors $\vec{x}, \vec{y} \in \mathbb{R}^d$ with $\vec{x} \neq \vec{y}$, the intersection $(\mathcal{Z} - \vec{x}) \cap (\mathcal{Z} - \vec{y})$ is a quadric on a $(d-1)$-dimensional affine-linear subspace $\mathcal{W}_{\vec{x},\vec{y}} \subsetneq \mathbb{R}^d$.*

Proof. Let $\mathcal{Z} = \mathrm{V}(P)$ for a nonzero quadratic polynomial $P \in \mathbb{R}[x_1, \ldots, x_d]$, and suppose that (i) does not hold. Note that then $\mathcal{Z} - \vec{x} = \{\vec{w} \in \mathbb{R}^d : P(\vec{w} + \vec{x}) = 0\}$ for any $\vec{x} \in \mathbb{R}^d$. For any $\vec{x}, \vec{y} \in \mathbb{R}^d$ with $\vec{x} \neq \vec{y}$, consider the linear polynomial $L_{\vec{x},\vec{y}} : \vec{w} \mapsto P(\vec{w} + \vec{y}) - P(\vec{w} + \vec{x})$ and write $\mathcal{W}_{\vec{x},\vec{y}} = \mathrm{V}(L_{\vec{x},\vec{y}})$, so $(\mathcal{Z} - \vec{x}) \cap (\mathcal{Z} - \vec{y}) = (\mathcal{Z} - \vec{x}) \cap \mathcal{W}_{\vec{x},\vec{y}}$. This already shows that $(\mathcal{Z} - \vec{x}) \cap (\mathcal{Z} - \vec{y})$ is the vanishing locus of a quadratic polynomial on $\mathcal{W}_{\vec{x},\vec{y}}$. To verify (ii), we will check that $L_{\vec{x},\vec{y}}$ is not the zero polynomial (which shows that $\dim \mathcal{W}_{\vec{x},\vec{y}} = d-1$), and that $\mathcal{W}_{\vec{x},\vec{y}} \not\subseteq \mathcal{Z} - \vec{x}$ (which implies $(\mathcal{Z} - \vec{x}) \cap (\mathcal{Z} - \vec{y}) = (\mathcal{Z} - \vec{x}) \cap \mathcal{W}_{\vec{x},\vec{y}} \subsetneq \mathcal{W}_{\vec{x},\vec{y}}$).

First, the reason $L_{\vec{x},\vec{y}}$ cannot be the zero polynomial is that (i) does not hold. Indeed, if $L_{\vec{x},\vec{y}}$ were the zero polynomial, then for all $\vec{w} \in \mathbb{R}^d$ and $\lambda \in \mathbb{Z}$ we would have $P(\vec{w}) = P(\vec{w} + \lambda(\vec{x} - \vec{y}))$. That is to say, for all $\vec{w} \in \mathbb{R}^d$ the quadratic polynomial $\lambda \mapsto P(\vec{w}) - P(\vec{w} + \lambda(\vec{x} - \vec{y}))$ would have infinitely many zeroes, so would be the zero polynomial, meaning that $P(\vec{w}) = P(\vec{w} + \lambda(\vec{x} - \vec{y}))$ for all $\lambda \in \mathbb{R}$ and $\vec{w} \in \mathbb{R}^d$. Hence we would have $\mathcal{Z} + \lambda(\vec{x} - \vec{y}) = \mathcal{Z}$ for all $\lambda \in \mathbb{R}$, so (i) would hold for $\vec{v} = \vec{x} - \vec{y}$.

Finally, it remains to show $\mathcal{W}_{\vec{x},\vec{y}} \not\subseteq \mathcal{Z} - \vec{x}$. Indeed, otherwise $\mathcal{W}_{\vec{x},\vec{y}}$ would be an irreducible component of $\mathcal{Z} - \vec{x}$, but by our assumptions $\mathcal{Z} - \vec{x}$ is irreducible (since $\mathcal{Z}$ is). We also have $\mathcal{Z} - \vec{x} \neq \mathcal{W}_{\vec{x},\vec{y}}$, since $\mathcal{Z} - \vec{x}$ not invariant under translation along the direction of any nonzero vector (since $\mathcal{Z}$ does not satisfy (i)). So we indeed have $\mathcal{W}_{\vec{x},\vec{y}} \not\subseteq \mathcal{Z} - \vec{x}$. $\qquad\square$

Now we prove Theorem 4.2.

Proof of Theorem 4.2. We proceed by induction on d. In the base case $d = 0$, our quadric $\mathcal{Z}$ consists of at most two points, so the theorem statement holds by Corollary 3.4 (taking $\mathcal{W}$ in Corollary 3.4 to be 0-dimensional, i.e., a single point). Assume now that $d \geq 1$ (and $r \geq d+1$), and that the theorem statement holds for smaller values of d. Let $\vec{X} = \xi_1 \vec{a}_1 + \cdots + \xi_n \vec{a}_n$.

Step 1: The reducible case. First, it is easy to handle the case where the quadric $\mathcal{Z} \subsetneq \mathcal{W}$ is reducible, i.e., where the irreducible components of $\mathcal{Z}$ are two affine-linear subspaces of dimension d. Indeed, suppose that $\mathcal{Z} = \mathcal{V}_1 \cup \mathcal{V}_2$ for two d-dimensional affine-linear subspaces $\mathcal{V}_1, \mathcal{V}_2 \subsetneq \mathcal{W}$. By Corollary 3.4 (i.e., by the version of Halász' theorem in [14]), we have

$$\Pr[\vec{X} \in \mathcal{Z}] \leq \Pr[\vec{X} \in \mathcal{V}_1] + \Pr[\vec{X} \in \mathcal{V}_2] \leq \frac{2}{t^{(r-d)/2}},$$

which implies the desired result.

Step 2: The translation-invariant case. It is also easy to handle the case where there is a direction $\vec{v} \neq \vec{0}$ such that $\mathcal{Z} + \mathbb{R}\vec{v} = \mathcal{Z}$, because then we can project our entire problem along the direction of $\vec{v}$ to obtain a lower-dimensional problem (noting that then we also have $\mathcal{W} + \mathbb{R}\vec{v} = \mathcal{W}$). Indeed, consider a linear map $\phi : \mathbb{R}^r \to \mathbb{R}^{r-1}$ with kernel $\mathrm{span}(\vec{v})$, and observe that $\phi(\mathcal{Z})$ is a quadric on the $(d-1)$-dimensional affine-linear subspace $\phi(\mathcal{W})$. Then, we have

$$\Pr[\vec{X} \in \mathcal{Z}] = \Pr[\phi(\vec{X}) \in \phi(\mathcal{Z})] \leq \frac{2^{(d-1)(r-1)+1}}{t^{(r-d)/2}},$$

by our induction hypothesis (noting that among $\phi(\vec{a}_1), \ldots, \phi(\vec{a}_n)$ one can still form t disjoint bases), and the desired result follows.

Step 3: Decoupling. Now, we can assume that $\mathcal{Z} \subseteq \mathcal{W}$ is an irreducible quadric and that there is no direction $\vec{v} \neq \vec{0}$ such that $\mathcal{Z} + \mathbb{R}\vec{v} = \mathcal{Z}$. Let $\tilde{\mathcal{W}} \subseteq \mathbb{R}^r$ be the $(d+1)$-dimensional linear subspace parallel to $\mathcal{W}$ (i.e., $\mathcal{W} = \tilde{\mathcal{W}} + \vec{w}$ for some $\vec{w} \in \mathbb{R}^r$). For any $\vec{x} \neq \vec{y}$ with $\vec{x} - \vec{y} \in \tilde{\mathcal{W}}$, by Lemma 4.3 (with an affine-linear isomorphism $\mathbb{R}^d \to \mathcal{W} - \vec{x}$), the intersection $(\mathcal{Z} - \vec{x}) \cap (\mathcal{Z} - \vec{y})$ is a quadric on a d-dimensional affine-linear subspace $\mathcal{W}_{\vec{x},\vec{y}} \subsetneq \tilde{\mathcal{W}} + \vec{w}_{\vec{x},\vec{y}}$ for some $\vec{w}_{\vec{x},\vec{y}} \in \mathbb{R}^r$. The next step is to split our random variable into two parts and use the decoupling lemma (Lemma 3.6) to relate our probability $\Pr[\vec{X} \in \mathcal{Z}]$ to probabilities that certain

14

random variables lie in quadrics of the form $(\mathcal{Z} - \vec{x}) \cap (\mathcal{Z} - \vec{y})$ (these probabilities can then be bounded via the induction hypothesis).

Let $[n] = I \cup J$ be a partition of the index set $[n]$ into two subsets I, J such that one can form $t/2$ disjoint bases from the vectors $\vec{a}_i$ for $i \in I$, and one can also form $t/2$ disjoint bases from the vectors $\vec{a}_j$ for $j \in J$. Let $\vec{X}_I = \sum_{i \in I} \xi_i \vec{a}_i$ and let $\vec{X}_J = \sum_{j \in J} \xi_j \vec{a}_j$ (so $\vec{X} = \vec{X}_I + \vec{X}_J$) and let $\vec{X}_I'$ be an independent copy of the random variable $\vec{X}_I$. By decoupling (Lemma 3.6) we have

$$\Pr[\vec{X} \in \mathcal{Z}]^2 \leq \Pr[\vec{X}_I + \vec{X}_J \in \mathcal{Z} \text{ and } \vec{X}_I' + \vec{X}_J \in \mathcal{Z}]$$
$$= \Pr[\vec{X}_J \in (\mathcal{Z} - \vec{X}_I) \cap (\mathcal{Z} - \vec{X}_I') \text{ and } \vec{X}_I \neq \vec{X}_I'] + \Pr[\vec{X}_I + \vec{X}_J \in \mathcal{Z} \text{ and } \vec{X}_I' = \vec{X}_I]. \qquad (4.1)$$

Step 4: Dealing with degenerate intersection. We now study the second term in (4.1) (which can be thought of as a lower-order "error term" corresponding to the possibility that decoupling does not actually reduce the dimension of our problem).

For any outcome of $\vec{X}_I$, we have $\Pr[\vec{X}_I' = \vec{X}_I \,|\, \vec{X}_I] \leq (t/2)^{-r/2}$ by Halász' theorem (Corollary 3.4, taking $\mathcal{W}$ to be a single point). So,

$$\Pr[\vec{X}_I + \vec{X}_J \in \mathcal{Z} \text{ and } \vec{X}_I' = \vec{X}_I] \leq \Pr[\vec{X}_I + \vec{X}_J \in \mathcal{Z}] \cdot (t/2)^{-r/2} = \Pr[\vec{X} \in \mathcal{Z}] \cdot (t/2)^{-r/2}. \qquad (4.2)$$

Step 5: Inductively bounding the main term. Now we deal with the first term in (4.1). Recalling that $\tilde{\mathcal{W}} \subseteq \mathbb{R}^r$ is the $(d+1)$-dimensional linear subspace parallel to $\mathcal{W}$, note that it is impossible to have $\vec{X}_J \in (\mathcal{Z} - \vec{X}_I) \cap (\mathcal{Z} - \vec{X}_I')$ if $\vec{X}_I - \vec{X}_I' \notin \tilde{\mathcal{W}}$ (since then $(\mathcal{Z} - \vec{X}_I) \cap (\mathcal{Z} - \vec{X}_I') \subseteq (\mathcal{W} - \vec{X}_I) \cap (\mathcal{W} - \vec{X}_I') = \emptyset$). So, we combine our induction hypothesis with a bound on the probability that $\vec{X}_I - \vec{X}_I' \in \tilde{\mathcal{W}}$.

For all outcomes of $\vec{X}_I'$, by Halász' theorem (Corollary 3.4) we have

$$\Pr[\vec{X}_I - \vec{X}_I' \in \tilde{\mathcal{W}} \,|\, \vec{X}_I'] = \Pr[\vec{X}_I \in \tilde{\mathcal{W}} + \vec{X}_I' \,|\, \vec{X}_I'] \leq (t/2)^{-(r-d-1)/2}.$$

Also, for any outcomes of $\vec{X}_I, \vec{X}_I'$ such that $\vec{X}_I \neq \vec{X}_I'$ and $\vec{X}_I - \vec{X}_I' \in \tilde{\mathcal{W}}$, by the discussion in Step 3, the intersection $(\mathcal{Z} - \vec{X}_I) \cap (\mathcal{Z} - \vec{X}_I')$ is a quadric on a d-dimensional affine-linear subspace $\mathcal{W}_{\vec{X}_I, \vec{X}_I'} \subseteq \mathbb{R}^r$. So by our induction hypothesis we have

$$\Pr[\vec{X}_J \in (\mathcal{Z} - \vec{X}_I) \cap (\mathcal{Z} - \vec{X}_I') \,|\, \vec{X}_I, \vec{X}_I'] \leq \frac{2^{(d-1)r+1}}{(t/2)^{(r-d+1)/2}}.$$

Thus, we obtain

$$\Pr[\vec{X}_J \in (\mathcal{Z} - \vec{X}_I) \cap (\mathcal{Z} - \vec{X}_I') \text{ and } \vec{X}_I \neq \vec{X}_I'] \leq (t/2)^{-(r-d-1)/2} \cdot \frac{2^{(d-1)r+1}}{(t/2)^{(r-d+1)/2}} = \frac{2^{(d-1)r+1}}{(t/2)^{r-d}} \leq \frac{2^{dr+1}}{t^{r-d}}. \qquad (4.3)$$

Step 6: Concluding. We can now deduce from (4.1), (4.2) and (4.3) that

$$\Pr[X \in \mathcal{Z}]^2 \leq \Pr[X \in \mathcal{Z}] \cdot (t/2)^{-r/2} + \frac{2^{dr+1}}{t^{r-d}} = \Pr[X \in \mathcal{Z}] \cdot \frac{2^{r/2}}{t^{r/2}} + \frac{2^{dr+1}}{t^{r-d}}.$$

So, by Lemma 3.8 we have (also using that $d \geq 1$ and $r \geq d + 1 \geq 2$)

$$\Pr[X \in \mathcal{Z}] \leq \frac{2^{r/2}}{t^{r/2}} + \frac{2^{(dr+1)/2}}{t^{(r-d)/2}} \leq \frac{2^{dr+1}}{t^{(r-d)/2}},$$

as desired. $\qquad \square$

5 Proof strategy for the general case

Now we turn to the general (not necessarily low-rank) case of Theorem 1.1. Actually, we consider the following variation on Theorem 1.1, with a slightly more technical assumption on Q (namely, we need to assume that the *quadratic part* of Q "robustly depends on many different variables"). This assumption is very similar to assumptions in some previous Littlewood–Offord-type theorems [37, 43].

Theorem 5.1. *Let $Q \in \mathbb{R}[x_1, \ldots, x_n]$ be a multivariate quadratic polynomial with quadratic part $\vec{x}^{\mathsf{T}} A \vec{x}$ for some symmetric matrix $A \in \mathbb{R}^{n \times n}$. Let $s \geq 1$ be an integer, and assume that for every subset $S \subseteq [n]$ with $|S| \geq n - s$, the submatrix $A[S \times S]$ has at least one nonzero entry outside its diagonal. Then for a sequence $\vec{\xi} \in \{-1, 1\}^n$ of independent Rademacher random variables, we have*

$$\Pr[Q(\vec{\xi}) = 0] \leq \frac{C'}{\sqrt{s}},$$

for some absolute constant C'.

The reason that the assumption in Theorem 5.1 only pays attention to the nondiagonal entries of A is that the diagonal entries correspond to square terms of the form x_i^2. If $x_i \in \{-1, 1\}$ then x_i^2 is always equal to 1, so such square terms can be treated as constants (and therefore they "do not really contribute" to the quadratic part of Q).

In Section 11 we will show how to deduce Theorem 1.1 (and Theorem 1.2 for general distributions) from Theorem 5.1. For most of the rest of the paper, we will focus on proving Theorem 5.1.

At a high level, the strategy to prove Theorem 5.1 is similar to the proof of Theorem 4.2: in order to estimate the probability $\Pr[Q(\vec{\xi}) = 0]$ for a given quadratic polynomial $Q \in \mathbb{R}[x_1, \ldots, x_n]$, we inductively estimate probabilities of the form $\Pr[Q(\vec{\xi}) = 0$ and $M\vec{\xi} = \vec{w}]$ for a given matrix $M \in \mathbb{R}^{k \times n}$ and a given vector $\vec{w} \in \mathbb{R}^k$ (i.e., we estimate the probability that $\vec{\xi}$ lies in a given quadric on a given affine-linear subspace). However, there are additional difficulties in comparison to the proof of Theorem 4.2 in the previous section.

Most importantly, recall that in Theorem 4.2 we worked with a random vector $\xi_1 \vec{a}_1 + \ldots \xi_n \vec{a}_n$ which has "a lot of anticoncentration in each direction" (we assumed that there are many disjoint bases among the vectors $\vec{a}_1, \ldots, \vec{a}_n$). This was only possible since the dimension of our space was much less than n: to be precise, that proof approach can only obtain bounds of the form $O(1/\sqrt{n})$ when we have $\Omega(n)$ disjoint bases among the vectors $\vec{a}_1, \ldots, \vec{a}_n$, which is only possible when we are working in $O(1)$-dimensional space.

In the proof of Theorem 5.1 we work directly with the random vector $\vec{\xi} \in \{-1, 1\}^n$ in n-dimensional space. We cannot ensure good anticoncentration in all directions, so we need to restrict the affine-linear subspaces we can consider. Specifically, we prove bounds on $\Pr[Q(\vec{\xi}) = 0$ and $M\vec{\xi} = \vec{w}]$ only when the matrix M satisfies a Halász-type robust rank condition (as in Definition 3.1). This means that we now need to maintain this robust rank condition as M changes over the course of the induction.

Also, we encounter much more delicate nondegeneracy issues than in the proof of Theorem 4.2. In particular, even if Q satisfies the nondegeneracy condition in Theorem 5.1 (robustly depending on many variables), it may become very degenerate when restricted to a subspace of the form $\{\vec{x} : M\vec{x} = \vec{w}\}$ for a matrix $M \in \mathbb{R}^{k \times n}$ and a vector $\vec{w} \in \mathbb{R}^k$. For example, consider the case where n is divisible by 2 and $Q \in \mathbb{R}[x_1, \ldots, x_n]$, $M \in \mathbb{R}^{1 \times n}$ and $\vec{w} \in \mathbb{R}^1$ are defined by

$$Q(\vec{x}) = (x_1 + \cdots + x_{n/2})(x_{n/2+1} + \cdots + x_n), \qquad M = (1, \ldots, 1, 0, \ldots, 0), \qquad \vec{w} = 0 \qquad (5.1)$$

(where the first $n/2$ entries of $M \in \mathbb{R}^{1 \times n}$ are "1" and the last $n/2$ entries are "0"). In this case Q is always zero on the subspace $\{\vec{x} \in \mathbb{R}^n : M\vec{x} = \vec{w}\}$. So, in order to be able to obtain a sensible bound, we need to ensure that Q satisfies a nondegeneracy condition *with respect to M*.

In the rest of this section we describe the strategy of the proof of Theorem 5.1 in a bit more detail, stating several key lemmas and definitions along the way. First, we elaborate on the "nondegeneracy with respect to M" condition mentioned above. To formulate this condition (as well as other similar conditions appearing later in the proof), we define the notion of a (T, U)-*perturbation*, as follows.

Definition 5.2. *For matrices $A \in \mathbb{R}^{n \times m}$ and $T \in \mathbb{R}^{k \times m}$ and $U \in \mathbb{R}^{k \times n}$, a (T, U)-perturbation of A is a matrix $A' \in \mathbb{R}^{n \times m}$ of the form $A' = A + LT + U^{\mathsf{T}}R$ for some matrices $L \in \mathbb{R}^{n \times k}$ and $R \in \mathbb{R}^{k \times m}$, i.e., some matrix that can be obtained from A by adding linear combinations of rows of T to its rows, and adding linear combinations of rows of U to its columns.*

As the degenerate $k = 0$ case of the above definition, note that if $T \in \mathbb{R}^{0 \times m}$ and $U \in \mathbb{R}^{0 \times n}$ are "empty matrices", then $A' \in \mathbb{R}^{n \times m}$ is a (T, U)-perturbation of A if and only if $A' = A$.

For $Q, M, \vec{w}$ as defined in the example in (5.1), we can write $Q(\vec{x}) = \vec{x}^{\mathsf{T}}A\vec{x}$ for a symmetric matrix $A \in \mathbb{R}^{n \times n}$ (with a block structure where the bottom left block and the top right block, each of size $(n/2) \times (n/2)$, have all entries being $1/2$, and all entries outside these blocks are 0). It turns out that this matrix A is an (M, M)-perturbation of the zero matrix in $\mathbb{R}^{n \times n}$. Roughly speaking, this is the reason for the degenerate behaviour of Q on the subspace $\{\vec{x} : M\vec{x} = \vec{w}\}$. In general, the notion of an (M, M)-perturbation gives a condition under which two $n \times n$ matrices give rise to quadratic polynomials which are "essentially the same" on an affine-linear subspace of the form $\{\vec{x} : M\vec{x} = \vec{w}\}$, as follows.

Lemma 5.3. *Fix a matrix $M \in \mathbb{R}^{k \times n}$, a vector $\vec{w} \in \mathbb{R}^k$, and a quadratic polynomial $Q \in \mathbb{R}[x_1, \ldots, x_n]$ with quadratic part $\vec{x}^{\mathsf{T}}A\vec{x}$ (where $A \in \mathbb{R}^{n \times n}$).*

Consider an (M, M)-perturbation A' of A, and let $A^* \in \mathbb{R}^{n \times n}$ be a matrix which agrees with A' on its off-diagonal entries. Then, there is a quadratic polynomial $Q^* \in [x_1, \ldots, x_n]$ with quadratic part $\vec{x}^{\mathsf{T}}A^*\vec{x}$ such that $Q^*(\vec{\xi}) = Q(\vec{\xi})$ for all $\vec{\xi} \in \{-1, 1\}^n$ with $M\vec{\xi} = \vec{w}$. In particular, for a sequence $\vec{\xi} \in \{-1, 1\}^n$ of independent Rademacher random variables, we have

$$\Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}] = \Pr[Q^*(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}].$$

Proof. Let $A^* = A' + D = A + LM + M^{\mathsf{T}}R + D$ for some matrices $L \in \mathbb{R}^{n \times k}$ and $R \in \mathbb{R}^{k \times n}$ and a diagonal matrix $D \in \mathbb{R}^{n \times n}$, and let $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ be the diagonal entries of D. Then

$$\vec{x}^{\mathsf{T}}A^*\vec{x} = \vec{x}^{\mathsf{T}}A\vec{x} + \vec{x}^{\mathsf{T}}LM\vec{x} + \vec{x}^{\mathsf{T}}M^{\mathsf{T}}R\vec{x} + \vec{x}^{\mathsf{T}}D\vec{x} = \vec{x}^{\mathsf{T}}A\vec{x} + (M\vec{x})^{\mathsf{T}}L^{\mathsf{T}}\vec{x} + (M\vec{x})^{\mathsf{T}}R\vec{x} + (\lambda_1 x_1^2 + \cdots + \lambda_n x_n^2).$$

Thus, for every $\vec{\xi} \in \{-1, 1\}^n$ with $M\vec{\xi} = \vec{w}$, we have

$$\vec{\xi}^{\mathsf{T}}A^*\vec{\xi} = \vec{\xi}^{\mathsf{T}}A\vec{\xi} + \vec{w}^{\mathsf{T}}L^{\mathsf{T}}\vec{\xi} + \vec{w}^{\mathsf{T}}R\vec{x} + (\lambda_1 \xi_1^2 + \cdots + \lambda_n \xi_n^2) = \vec{\xi}^{\mathsf{T}}A\vec{\xi} + \vec{w}^{\mathsf{T}}(L^{\mathsf{T}} + R)\vec{\xi} + (\lambda_1 + \cdots + \lambda_n).$$

Hence, we can define the desired quadratic polynomial $Q^* \in [x_1, \ldots, x_n]$ with quadratic part $\vec{x}^{\mathsf{T}}A^*\vec{x}$ by

$$Q^*(\vec{x}) = Q(\vec{x}) + \vec{x}^{\mathsf{T}}A^*\vec{x} - \vec{x}^{\mathsf{T}}A\vec{x} - \vec{w}^{\mathsf{T}}(L^{\mathsf{T}} + R)\vec{x} - (\lambda_1 + \cdots + \lambda_n),$$

and we indeed have $Q^*(\vec{\xi}) = Q(\vec{\xi})$ for all $\vec{\xi} \in \{-1, 1\}^n$ with $M\vec{\xi} = \vec{w}$. $\qquad\square$

Now, recall that our strategy to prove Theorem 5.1 is to inductively upper-bound probabilities of the form $\Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}]$, assuming that M satisfies a robust rank condition, and assuming a nondegeneracy condition on Q with respect to M. It will be convenient to introduce some notation for the maximum possible such probability, as follows.

Definition 5.4. *For an integer $k \geq 0$ and real number $s \geq 0$, let us define*

$$f(k, s) = \sup_{(n, Q, M, \vec{w})} \Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}],$$

where the supremum is taken over all quadruples $(n, Q, M, \vec{w})$, where n is a positive integer, $Q \in \mathbb{R}[x_1, \ldots, x_n]$ is a quadratic polynomial, $M \in \mathcal{H}^{k \times n}(s)$ and $\vec{w} \in \mathbb{R}^k$, such that the following condition holds:

(∗) *If we write the quadratic part of $Q(\vec{x})$ as $\vec{x}^{\mathsf{T}}A\vec{x}$ for a symmetric matrix $A \in \mathbb{R}^{n \times n}$, then for every subset $S \subseteq [n]$ with $|S| \geq n - s$, and every (M, M)-perturbation A' of A, the submatrix $A'[S \times S]$ has at least one nonzero entry outside the diagonal.*

For each such quadruple $(n, Q, M, \vec{w})$, the probability above is taken with respect to a sequence of independent Rademacher random variables $\vec{\xi} \in \{-1, 1\}^n$.

Note that we always have $f(k, s) \leq 1$, since $f(k, s)$ is defined as a supremum of certain probabilities (which are all at most 1). Also note that for $0 \leq s' \leq s$, we always have $f(k, s) \leq f(k, s')$ (since the supremum in the definition of $f(k, s')$ is taken over a wider range of quadruples $(n, Q, M, \vec{w})$ than for $f(k, s)$).

Note that for a polynomial $Q \in \mathbb{R}[x_1, \ldots, x_n]$ and an integer $s \geq 1$ as in Theorem 5.1, condition $(*)$ is satisfied for $k = 0$, the empty matrix $M \in \mathbb{R}^{0 \times n}$ and the empty vector $\vec{w} \in \mathbb{R}^0$. Indeed, writing the quadratic part of Q as $\vec{x}^\mathsf{T} A \vec{x}$ for a symmetric matrix $A \in \mathbb{R}^{n \times n}$, the only (M, M) perturbation A' of A is $A' = A$, and by the assumption in Theorem 5.1 the matrix $A'[S \times S] = A[S \times S]$ has at least one nonzero entry outside the diagonal for every subset $S \subseteq [n]$ with $|S| \geq n - s$. Therefore we have (noting that the condition $M\vec{\xi} = \vec{w} \in \mathbb{R}^0$ is vacuous)

$$\Pr[Q(\vec{\xi}) = 0] = \Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}] \leq f(0, s).$$

Thus, proving Theorem 5.1 amounts to showing that $f(0, s) \leq C'/\sqrt{s}$ for some absolute constant C'.

Now, the following recursive upper bound on $f(k, s)$ is the main ingredient in our proof of Theorem 5.1.

Theorem 5.5. *For any integer $k \geq 0$ and any real number $s > 0$, we have*

$$f(k, s) \leq \max\left\{ s_*^{-(k+1)/2}, \quad s_*^{-(k+2)/2} + s_*^{-k/4} \cdot f(k+1, s_*)^{1/2} \right\},$$

where $s_* = s/(k+2)^{500}$.

Using this recursive bound, we can obtain an upper bound for $f(k, s)$ inductively (the formula for this upper bound is rather complicated, so we postpone this calculation to Section 10). With another straightforward (though somewhat technical) calculation, one can then show that in the $k = 0$ case this upper bound implies Theorem 5.1. This latter calculation can also be found in Section 10.

To prove Theorem 5.5, we need to upper-bound $\Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}]$ for $Q, M, \vec{w}$ satisfying the conditions in Definition 5.4, in terms of a probability of the same form but with slightly different parameters (most importantly, with "$k + 1$" in place of "k"). To do so, we use our inductive decoupling scheme outlined in Section 2.3: we consider a partition $[n] = I \cup J$ of the index set into two parts I and J, and, using the decoupling inequality in Lemma 3.6, we will obtain an upper bound on $\Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}]$ involving conditional probabilities of the form

$$\Pr\left[Q_{\vec{\xi}[I]}(\vec{\xi}[J]) = 0 \text{ and } M_{\vec{\xi}[I], \vec{\xi}'[I]} \vec{\xi}[J] = \vec{w}_{\vec{\xi}[I], \vec{\xi}'[I]} \,\Big|\, \vec{\xi}[I], \vec{\xi}'[I] \right], \tag{5.2}$$

where $M_{\vec{\xi}[I], \vec{\xi}'[I]} \in \mathbb{R}^{(k+1) \times J}$ is a $(k+1) \times |J|$ matrix depending on $\vec{\xi}[I]$ and $\vec{\xi}'[I]$, and $Q_{\vec{\xi}[I]} \in \mathbb{R}[x_j : j \in J]$ is a quadratic polynomial whose coefficients depend on $\vec{\xi}[I]$, and $\vec{w}_{\vec{\xi}[I], \vec{\xi}'[I]} \in \mathbb{R}^{k+1}$ is a vector depending on $\vec{\xi}[I]$ and $\vec{\xi}'[I]$.

In our proof of Theorem 5.5, we wish to upper-bound conditional probabilities of the form in (5.2) by $f(k+1, s_*)$. Recalling Definition 5.4, this requires that $M_{\vec{\xi}[I], \vec{\xi}'[I]} \in \mathcal{H}^{(k+1) \times n}(s_*)$ (i.e, that $M_{\vec{\xi}[I], \vec{\xi}'[I]}$ robustly has rank at least $k + 1$). So, as previously outlined, an important ingredient in the proof is a "robust rank inheritance" lemma, which implies that this robust rank condition for $M_{\vec{\xi}[I], \vec{\xi}'[I]}$ holds with sufficiently high probability.

Now, the way in which $M_{\vec{\xi}[I], \vec{\xi}'[I]}$ is derived from M depends on Q: specifically, one can check that $M_{\vec{\xi}[I], \vec{\xi}'[I]}$ is obtained from $M[[k] \times J]$ by adding the vector $2A[J \times I](\vec{\xi}[I] - \vec{\xi}'[I])$ as an additional row, where $\vec{x}^\mathsf{T} A \vec{x}$ is the quadratic part of $Q(\vec{x})$. The statement of our robust rank inheritance lemma requires an assumption on Q; to specify this assumption we need another definition.

Definition 5.6. *For integers $r \geq 1$ and $0 \leq k \leq m \leq n$ and $s \geq 0$, let $\mathcal{M}_r^{k,m,n}(s) \subseteq \mathbb{R}^{k \times m} \times \mathbb{R}^{k \times n} \times \mathbb{R}^{n \times m}$ be the set of triples of matrices $(T, U, A) \in \mathbb{R}^{k \times m} \times \mathbb{R}^{k \times n} \times \mathbb{R}^{n \times m}$ for which there exist disjoint subsets $I_1, \ldots, I_s \subseteq [m]$ and disjoint subsets $J_1, \ldots, J_s \subseteq [n]$ of size $|I_1| = \cdots = |I_s| = |J_1| = \cdots = |J_s| = k + r$ such that*

(a) For $t = 1, \ldots, s$, the submatrix $T[[k] \times I_t]$ has rank k.
(b) For $t = 1, \ldots, s$, the submatrix $U[[k] \times J_t]$ has rank k.
(c) For $t = 1, \ldots, s$, every $(T[[k] \times I_t], U[[k] \times J_t])$-perturbation of the matrix $A[J_t \times I_t]$ has rank at least r.

Remark 5.1. If $(T, U, A) \in \mathcal{M}_r^{k,m,n}(s)$ for any $r \geq 1$, then we automatically have $T \in \mathcal{H}^{k \times m}(s)$ and $U \in \mathcal{H}^{k \times n}(s)$. Also note that the property $(T, U, A) \in \mathcal{M}_r^{k,m,n}(s)$ is invariant under rescaling any of the matrices T, U and A.

In our proof of Theorem 5.5, we consider a partition $[n] = I \cup J$ as outlined above, and take the matrix "A" in Definition 5.6 to be the matrix $A[J \times I]$ together with $T = M[[k] \times I]$ and $U = M[[k] \times J]$. In this case, $(M[[k] \times I], M[[k] \times J], A[J \times I]) \in \mathcal{M}_r^{k,I,J}(s)$ says (roughly speaking) that $A'[J \times I]$ robustly has rank at least r for any (M, M)-perturbation A' of A.

Now, our robust rank inheritance lemma is as follows.

Lemma 5.7. *Let $s \geq 1$ and $0 \leq k \leq m \leq n$ be integers. Consider $(T, U, A) \in \mathcal{M}_2^{k,m,n}(s)$, and vectors $\vec{y} \in \mathbb{R}^k$ and $\vec{b} \in \mathbb{R}^n$. Let $\vec{\xi} \in \{-1, 1\}^m$ be a sequence of independent Rademacher random variables, and write $U_{\vec{\xi}}' \in \mathbb{R}^{(k+1) \times n}$ for the (random) matrix obtained by appending the vector $A\vec{\xi} - \vec{b}$ as an additional row to U. Then we have*

$$\Pr[T\vec{\xi} = \vec{y} \text{ and } U_{\vec{\xi}}' \notin \mathcal{H}^{(k+1) \times n}(s/6)] \leq \left(\frac{s}{10^{61}(k+2)^{20}} \right)^{-(k+2)/2}.$$

We will prove Lemma 5.7 in Section 7. We remark that on the right-hand side of the inequality, both the "$+2$" in the exponent $(k+2)/2$ and the "$+2$" in the denominator are due to the "2" in the assumption $(T, U, A) \in \mathcal{M}_2^{k,m,n}(s)$; in general we would get a "$+r$" in both of these places if we assumed that $(T, U, A) \in \mathcal{M}_2^{k,m,n}(s)$. The exponent $(k+2)/2$ here leads to the exponent $(k+2)/2$ appearing in Theorem 5.5, and the "$+2$" there is crucial in order to deduce Theorem 5.1 from Theorem 5.5 (just having "$+1$" there would not suffice).

In order to actually apply Lemma 5.7 in the proof of Theorem 5.5, we would like to show that there is a partition $[n] = I \cup J$ such that $(M[[k] \times I], M[[k] \times J], A[J \times I]) \in \mathcal{M}_2^{k,I,J}(s)$ for a suitable value of s. Such a partition might not in general exist, because we are making no assumption that A itself even has rank at least 2. However, if every (M, M)-perturbation of A robustly has rank at least 2 even after changing the diagonal entries, we can find a suitable partition $[n] = I \cup J$ using the following lemma.

Lemma 5.8. *Let $0 \leq k \leq n$ be integers and let $s \geq 4k + 8$ be a real number. Let $M \in \mathcal{H}^{k \times n}(s)$ and let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix such that $\operatorname{rank} A^*[S \times S] \geq 2$ for any subset $S \subseteq [n]$ of size $|S| \geq n - s$ and any matrix $A^* \in \mathbb{R}^{n \times n}$ that agrees with some (M, M)-perturbation of A in all off-diagonal entries. Then there is a partition $[n] = I \cup J$, with $|I| \leq s$, such that*

$$(M[[k] \times I], M[[k] \times J], A[J \times I]) \in \mathcal{M}_2^{k,I,J}(\lfloor s/(4k + 8) \rfloor).$$

We prove Lemma 5.8 in Section 8. Roughly speaking, we are able to *greedily* find the desired subsets $I_1, \ldots, I_s, J_1, \ldots, J_s$ in the definition of $\mathcal{M}_2^{k,m,n}(s)$ in Definition 5.6.

To summarise, the proof of Theorem 5.5 combines a number of ingredients. If A does not satisfy the assumption in Lemma 5.8 (i.e., if some (M, M)-perturbation of A does not robustly have rank at least 2 after changing the diagonal entries), then we complete the proof using Theorem 4.2 (in the low-rank case, there are various types of degeneracies to consider, which can be gracefully handled with a geometric point of view). Otherwise, we apply Lemma 5.8 to find a suitable partition $I \cup J$, and we apply the decoupling inequality in Lemma 3.6 to $(\vec{\xi}[I], \vec{\xi}[J])$. We manipulate the resulting expression in roughly the same way as in the proof of Theorem 4.2, and then bound relevant quantities using Lemma 5.7 and Corollary 3.5. After proving Theorem 5.5, we can obtain Theorem 5.1 with somewhat technical calculations (presented in Section 10).

In the next section we state and prove a consequence of Theorem 4.2 which will be suitable to handle the low-rank case of Theorem 5.5. In Section 7 we prove Lemma 5.7 and in Section 8 we prove Lemma 5.8. The proof of Theorem 5.5 then appears in Section 9.

6   The low-rank case

In this section, we show how to use Theorem 4.2 to deduce the following proposition, which handles the low-rank case of Theorem 5.5.

Proposition 6.1. *Consider integers $0 \leq k \leq n$ and $s \geq 0$ and $r \geq 1$. Let $M \in \mathcal{H}^{k \times n}(s)$, let $\vec{w} \in \mathbb{R}^k$ and let $Q \in \mathbb{R}[x_1, \ldots, x_n]$ be a quadratic polynomial. Writing the quadratic part of $Q(\vec{x})$ as $\vec{x}^\mathsf{T} A \vec{x}$ for a symmetric matrix A, assume that $\operatorname{rank} A \leq r - 1$ and that there is no subset $I \subseteq [n]$ of size $|I| \geq n - s$ such that the matrix $A[I \times I]$ is an $(M[[k] \times I], M[[k] \times I])$-perturbation of the zero matrix in $\mathbb{R}^{I \times I}$.*

Then for a sequence $\vec{\xi} \in \{-1, 1\}^n$ of independent Rademacher random variables we have

$$\Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}] \leq \left(\frac{s}{2^{3r^2}(k+r)^2} \right)^{-(k+1)/2}. \tag{6.1}$$

In our proof of Theorem 5.5, we will use Proposition 6.1 only for $r = 5$, but we still state it here for general r (as the proof works for any r).

Deducing Proposition 6.1 from Theorem 4.2 mostly consists of translating from "geometric" to "algebraic" language. To give a some idea of how this works: note that under the assumptions of Proposition 6.1, we can write

$$Q(\vec{x}) = \lambda_1 \cdot (g^{(1)}(\vec{x}))^2 + \cdots + \lambda_{r-1} \cdot (g^{(r-1)}(\vec{x}))^2 + g^{(r)}(\vec{x}) + c$$

for some $\lambda_1, \ldots, \lambda_{r-1}, c \in \mathbb{R}$ and some linear forms $g^{(1)}, \ldots, g^{(r-1)}, g^{(r)} \in \mathbb{R}[x_1, \ldots, x_n]$. That is to say, we can write

$$Q(\vec{x}) = P(g^{(1)}(\vec{x}), \ldots, g^{(r)}(\vec{x})) = P(x_1 \vec{a}_1 + \cdots + x_n \vec{a}_n)$$

for the quadratic polynomial $P(\vec{y}) = \lambda_1 y_1^2 + \cdots + \lambda_{r-1} y_{r-1}^2 + y_r + c$, where for $j = 1, \ldots, n$, the vector $\vec{a}_j \in \mathbb{R}^r$ records the coefficients of x_j in the linear forms $g^{(1)}, \ldots, g^{(r)}$. So, the event $\{Q(\vec{\xi}) = 0\}$ can be interpreted as the event that $\xi_1 \vec{a}_1 + \cdots + \xi_n \vec{a}_n$ falls in the vanishing locus of the polynomial P. The joint event $\{Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}\}$ can be given a similar geometric interpretation: roughly speaking, we consider k additional linear forms $g^{(r+1)}, \ldots, g^{(r+k)} \in \mathbb{R}[x_1, \ldots, x_n]$ corresponding to the k rows of M, augmenting $\vec{a}_1, \ldots, \vec{a}_n$ accordingly with k additional entries each, and then we consider the event that the $(r + k)$-dimensional random vector $\xi_1 \vec{a}_1 + \cdots + \xi_n \vec{a}_n$ falls into a certain quadric $\mathcal{Z}$ (still described by the same polynomial P) on the codimension-k affine-linear subspace $\mathcal{W} \subseteq \mathbb{R}^{r+k}$ given by $\mathcal{W} = \{\vec{y} \in \mathbb{R}^{r+k} : \vec{y}[\{r+1, \ldots, r+k\} = \vec{w}\}$ (this subspace $\mathcal{W} \subseteq \mathbb{R}^{r+k}$ corresponds to the system of equations $M\vec{\xi} = \vec{w}$).

This more-or-less explains how to reduce Proposition 6.1 to the setting of Theorem 4.2. The only complication is that it may not be possible to find many disjoint bases of $\mathbb{R}^{r+k}$ among the vectors $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{R}^{r+k}$, as is necessary to apply Theorem 4.2 (in fact, the vectors $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{R}^{r+k}$ might not span $\mathbb{R}^{r+k}$ at all). In this case, we need to "drop to a subspace" that is robustly spanned by most of the $\vec{a}_1, \ldots, \vec{a}_n$ (i.e., we need to identify some large subset $I \subseteq [n]$, such that we can find many disjoint bases of $\operatorname{span}(\vec{a}_i : i \in I)$ among the vectors $\vec{a}_i$ for $i \in I$). This, roughly speaking, corresponds to considering only some subset of the linear forms $g^{(1)}, \ldots, g^{(r+k)}$, such that the remaining linear forms are "close to" being linear combinations of these linear forms (i.e., each of the remaining linear forms agrees with such a linear combination in its coefficients of the variables x_i for $i \in I$). The quadratic polynomial $P(\vec{y})$ considered above then needs to be changed appropriately in order to reflect these relations between the linear forms $g^{(1)}, \ldots, g^{(r+k)}$.

The following lemma encapsulates this translation discussed above, going from the setting of Proposition 6.1 to the setting of Theorem 4.2.

Lemma 6.2. *Consider integers $0 \le k \le n$ and $s \ge 0$ and $r \ge 1$. Let $M \in \mathcal{H}^{k \times n}(s)$ and let $Q \in \mathbb{R}[x_1 \ldots, x_n]$ be a quadratic polynomial. Writing the quadratic part of $Q(\vec{x})$ as $\vec{x}^\mathsf{T} A \vec{x}$ for a symmetric matrix $A \in \mathbb{R}^{n \times n}$, assume that rank $A \le r - 1$ and that there is no subset $I \subseteq [n]$ of size $|I| \ge n - s$ such that the matrix $A[I \times I]$ is an $(M[[k] \times I], M[[k] \times I])$-perturbation of the zero matrix in $\mathbb{R}^{I \times I}$.*

Then, we can find a positive integer $\ell \le r$, a partition $[n] = I \cup S$ with $|S| \le s$, linear forms $g^{(1)}, \ldots, g^{(\ell+k)} \in \mathbb{R}[x_1, \ldots, x_n]$, and a quadratic polynomial $P \in \mathbb{R}[y_1, \ldots, y_{\ell+k}, (x_i)_{i \in S}]$ such that the following conditions hold.

(i) $Q(x_1, \ldots, x_n) = P(g^{(1)}(\vec{x}), \ldots, g^{(\ell+k)}(\vec{x}), (x_i)_{i \in S})$.
(ii) For each $j = 1, \ldots, k$, the coefficient vector of $g^{(\ell+j)}$ is precisely the j-th row of M.
(iii) Writing $M' \in \mathbb{R}^{(\ell+k) \times n}$ for the $(\ell + k) \times n$ matrix whose j-th row is the coefficient vector of $g^{(j)}$ for $j = 1, \ldots, \ell + k$, we have $M'[[\ell + k] \times I] \in \mathcal{H}^{(\ell+k) \times I}(s/r)$.
(iv) There exist $j, j' \in [\ell]$ such that the coefficient of $y_j y_{j'}$ in P is nonzero.

We remark that the vectors $\vec{a}_i$ for $i \in I$ in the informal explanation above correspond to the columns of the matrix $M'[[\ell + k] \times I]$ in Lemma 6.2. Condition (iii) ensures that we can find many disjoint bases of $\mathbb{R}^{\ell+k}$ among these vectors $\vec{a}_i$ for $i \in I$ (recall that this is required to apply Theorem 4.2 to these vectors). Condition (iv) ensures that the polynomial P does not vanish on our entire subspace $\mathcal{W}$ (corresponding to a system of equations of the form $M\vec{\xi} = \vec{w}$).

Proof. Let us consider the minimum number $\ell \in \{0, 1, \ldots, r\}$ such that there exists a partition $[n] = I \cup S$ with $|S| \le s \cdot (r - \ell)/r$, linear forms $g^{(1)}, \ldots, g^{(\ell+k)} \in \mathbb{R}[x_1, \ldots, x_n]$, and a quadratic polynomial $P \in \mathbb{R}[y_1, \ldots, y_{\ell+k}, (x_i)_{i \in S}]$, such that conditions (i) and (ii) hold (i.e., we have $Q(x_1, \ldots, x_n) = P(g^{(1)}(\vec{x}), \ldots, g^{(\ell+k)}(\vec{x}), (x_i)_{i \in S})$ and for each $j = 1, \ldots, k$, the coefficient vector of $g^{(\ell+j)}$ is precisely the j-th row of M).

First, in order to see that this minimum number ℓ is well-defined, let us check that $\ell = r$ satisfies the conditions. Indeed, given that rank $A \le r - 1$, we can write

$$Q(\vec{x}) = \lambda_1 \cdot (g^{(1)}(\vec{x}))^2 + \cdots + \lambda_{r-1} \cdot (g^{(r-1)}(\vec{x}))^2 + g^{(r)}(\vec{x}) + c$$

for some $\lambda_1, \ldots, \lambda_{r-1}, c \in \mathbb{R}$ and some linear forms $g^{(1)}, \ldots, g^{(r-1)}, g^{(r)} \in \mathbb{R}[x_1, \ldots, x_n]$. That is to say, defining $g^{(r+1)}, \ldots, g^{(r+k)} \in \mathbb{R}[x_1, \ldots, x_n]$ to be the linear forms whose coefficient vectors are given by the rows of M as in condition (ii), we can write

$$Q(\vec{x}) = P(g^{(1)}(\vec{x}), \ldots, g^{(r+k)}(\vec{x})),$$

where $P(y_1, \ldots, y_{r+k}) = \lambda_1 y_1^2 + \cdots + \lambda_{r-1} y_{r-1}^2 + y_r + c$ (formally this is a polynomial in $r + k$ variables, despite the fact that $y_{r+1}, \ldots, y_{r+k}$ do not actually appear in it). Note that (ii) holds by definition, and taking $I = [n]$ and $S = \emptyset$, condition (i) holds as well and we have $|S| = 0 \le s \cdot (r - r)/r$. Thus, $\ell = r$ has the required properties, and the minimum number ℓ above is well-defined.

Now let ℓ be this minimum number, and choose the partition $[n] = I \cup S$, the linear forms $g^{(1)}, \ldots, g^{(\ell+k)} \in \mathbb{R}[x_1, \ldots, x_n]$, and the quadratic polynomial P accordingly with the properties above. Note that $|S| \le s \cdot (r - \ell)/r \le s$, and consequently $|I| \ge n - s$. It remains to show that $\ell \ge 1$ and that conditions (iii) and (iv) hold.

Let us first show (iv). Suppose for contradiction that in the quadratic polynomial P, the coefficient of $y_j y_{j'}$ is zero for all $j, j' \in [\ell]$. Then the quadratic part of $Q(x_1, \ldots, x_n) = P(g^{(1)}(\vec{x}), \ldots, g^{(\ell+k)}(\vec{x}), (x_i)_{i \in S})$ can be written as a linear combination of terms of the form $g^{(j)}(\vec{x}) x_h$ with $j \in \{\ell + 1, \ldots, \ell + k\}$ and $h \in [n]$, and terms of the form $x_h x_i$ with $h \in [n]$ and $i \in S$. By (3.1), this leads to a representation of the symmetric matrix A as a linear combination of matrices of the form $\vec{v}_j \vec{e}_h^\mathsf{T}$, $\vec{e}_h \vec{v}_j^\mathsf{T}$, $\vec{e}_i \vec{e}_h^\mathsf{T}$ and $\vec{e}_h \vec{e}_i^\mathsf{T}$ for $j \in \{\ell + 1, \ldots, \ell + k\}$, $h \in [n]$ and $i \in S$, where $\vec{e}_1, \ldots, \vec{e}_n$ are the standard basis vectors in $\mathbb{R}^n$, and $\vec{v}_{\ell+1}, \ldots, \vec{v}_{\ell+k} \in \mathbb{R}^n$ denote the coefficient vectors of $g^{(\ell+1)}, \ldots, g^{(\ell+k)}$ (i.e., the row vectors of M, by condition (ii)). Hence the matrix $A[I \times I]$ is a linear combination of matrices of the form $\vec{v}_j[I] \vec{e}_h^\mathsf{T}[I]$ and $\vec{e}_h[I] \vec{v}_j^\mathsf{T}[I]$ for $j \in \{\ell + 1, \ldots, \ell + k\}$ and $h \in I$. Recalling that $\vec{v}_{\ell+1}, \ldots, \vec{v}_{\ell+k}$ are the row vectors of the matrix M, this means that $A[I \times I]$ is a $(M[[k] \times I], M[[k] \times I])$-perturbation of the zero matrix, contradicting our assumption. So indeed, for some

21

monomial $y_j y_{j'}$ with $j, j' \in [\ell]$, the coefficient of $y_j y_{j'}$ in P must be nonzero. This also automatically implies that $\ell \geq 1$.

It now remains to show (iii), i.e., to show that for the matrix $M' \in \mathbb{R}^{(\ell+k) \times n}$ whose rows are given by the coefficient vectors of $g^{(1)}, \ldots, g^{(\ell+k)}$ we have $M'[[\ell + k] \times I] \in \mathcal{H}^{(\ell+k) \times I}(s/r)$. Assume the contrary; then there exists a nontrivial linear combination of the rows of $M'[[\ell + k] \times I]$ yielding a vector with at most s/r nonzero entries. By (ii), the rows of the matrix $M'[\{\ell + 1, \ldots, \ell + k\} \times [n]]$ agree with the rows of the matrix M, and we therefore have $M'[\{\ell+1, \ldots, \ell+k\} \times [n]] \in \mathcal{H}^{k \times n}(s)$ (recalling that $M \in \mathcal{H}^{k \times n}(s)$). Using $|S| \leq s \cdot (r - \ell)/r \leq s \cdot (r-1)/r$, we can conclude that $M'[\{\ell+1, \ldots, \ell+k\} \times I] \in \mathcal{H}^{k \times I}(s - |S|) \subseteq \mathcal{H}^{k \times I}(s/r)$, so our linear combination cannot only involve rows with indices in $\{\ell + 1, \ldots, \ell + k\}$. Thus, we may assume without loss of generality that it involves the first row of $M'[[\ell + k] \times I]$, meaning that this first row differs in at most s/r entries from some linear combination of the other rows of $M'[[\ell + k] \times I]$. In other words, we can write $g^{(1)}$ as a linear combination of $g^{(2)}, \ldots, g^{(\ell+k)}$ and x_i for $i \in S \cup S'$, for some subset $S' \subseteq [n]$ of size $|S'| \leq s/r$. But this means that $Q(x_1, \ldots, x_n) = P(g^{(1)}(\vec{x}), \ldots, g^{(\ell+k)}(\vec{x}), (x_i)_{i \in S})$ can be written as $P^*(g^{(2)}(\vec{x}), \ldots, g^{(\ell+k)}(\vec{x}), (x_i)_{i \in S \cup S'})$ for some quadratic polynomial $P^* \in \mathbb{R}[y_2, \ldots, y_{\ell+k}, (x_i)_{i \in S \cup S'}]$. Since $|S \cup S'| \leq |S| + |S'| \leq s \cdot (r - \ell)/r + s/r = s \cdot (r - \ell + 1)/r$, this contradicts the minimality of ℓ (we now have a suitable representation of Q in terms of the $(\ell - 1) + k$ linear forms $g^{(2)}, \ldots, g^{(\ell+k)}$). So we must have $M'[[\ell + k] \times I] \in \mathcal{H}^{(\ell+k) \times I}(s/r)$, as desired. $\qquad\square$

Using Lemma 6.2 and Theorem 4.2, we now prove Proposition 6.1.

Proof of Proposition 6.1. First, note that we may assume that $s \geq 2^{3r^2}(k + r)^2$, since otherwise (6.1) is trivially true. Now, apply Lemma 6.2 to obtain a positive integer $\ell \leq r$, a partition $[n] = I \cup S$, linear forms $g^{(1)}, \ldots, g^{(\ell+k)} \in \mathbb{R}[x_1, \ldots, x_n]$ and a quadratic polynomial $P \in \mathbb{R}[y_1, \ldots, y_{\ell+k}, (x_i)_{i \in S}]$ satisfying conditions (i) to (iv) in the statement of the lemma.

Our plan is to show that the bound in (6.1) holds even if we condition on an arbitrary outcome of $\vec{\xi}[S]$ (leaving only the randomness in $\vec{\xi}[I]$). For any outcome of $\vec{\xi}[S]$, when plugging in $x_i = \vec{\xi}[i]$ for $i \in S$ into the polynomial $Q(\vec{x})$ and the linear forms $g^{(1)}(\vec{x}), \ldots, g^{(\ell+k)}(\vec{x})$, we obtain a polynomial $Q_{\vec{\xi}[S]}(\vec{x}[I])$ and linear functions $g_*^{(1)}(\vec{x}[I]) + c_{\vec{\xi}[S]}^{(1)}, \ldots, g_*^{(\ell+k)}(\vec{x}[I]) + c_{\vec{\xi}[S]}^{(\ell+k)}$ (where $g_*^{(1)}(\vec{x}[I]), \ldots, g_*^{(\ell+k)}(\vec{x}[I])$ are the linear forms obtained from $g^{(1)}(\vec{x}), \ldots, g^{(\ell+k)}(\vec{x})$ by omitting all terms with variables x_i for $i \in S$, and where $c_{\vec{\xi}[S]}^{(1)}, \ldots, c_{\vec{\xi}[S]}^{(\ell+k)} \in \mathbb{R}$ are real numbers that may depend on $\vec{\xi}[S]$) and by (i) we obtain that $Q_{\vec{\xi}[S]}(\vec{x}[I]) = P(g_*^{(1)}(\vec{x}[I]) + c_{\vec{\xi}[S]}^{(1)}, \ldots, g_*^{(\ell+k)}(\vec{x}[I]) + c_{\vec{\xi}[S]}^{(\ell+k)}, (\xi[i])_{i \in S})$. For any outcome of $\vec{\xi}[S]$, we can furthermore write $P(y_1 + c_{\vec{\xi}[S]}^{(1)}, \ldots, y_{\ell+k} + c_{\vec{\xi}[S]}^{(\ell+k)}, (\xi[i])_{i \in S}) = P_{\vec{\xi}[S]}(y_1, \ldots, y_{\ell+k})$ for a polynomial $P_{\vec{\xi}[S]} \in \mathbb{R}[y_1, \ldots, y_{\ell+k}]$ whose coefficients may depend on $\vec{\xi}[S]$. Then we always have $Q_{\vec{\xi}[S]}(\vec{x}[I]) = P_{\vec{\xi}[S]}(g_{\vec{\xi}[S]}^{(1)}(\vec{x}[I]]), \ldots, g_{\vec{\xi}[S]}^{(\ell+k)}(\vec{x}[I]))$. Recall that the coefficient vectors of the linear forms $g_*^{(1)}(\vec{x}[I]), \ldots, g_*^{(\ell+k)}(\vec{x}[I])$ are obtained by restricting the coefficient vectors of $g^{(1)}(\vec{x}), \ldots, g^{(\ell+k)}(\vec{x})$ to the index set I. In particular, by condition (ii), for $j = 1, \ldots, k$ the coefficient vector of $g_*^{(\ell+j)}(\vec{x}[I])$ is the restriction $M[[j] \times I]$ of the j-th row of M to I. We then have

$$\Pr\left[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w} \,\middle|\, \vec{\xi}[S]\right]$$
$$= \Pr\left[Q_{\vec{\xi}[S]}(\vec{\xi}[I]) = 0 \text{ and } M[[k] \times I]\vec{\xi}[I] = \vec{w}_{\vec{\xi}[S]} \,\middle|\, \vec{\xi}[S]\right]$$
$$= \Pr\left[P_{\vec{\xi}[S]}(g_*^{(1)}(\vec{\xi}[I]), \ldots, g_*^{(\ell+k)}(\vec{\xi}[I])) = 0 \text{ and } g_*^{(\ell+j)}(\vec{\xi}[I]) = \vec{w}_{\vec{\xi}[S]}[j] \text{ for } j = 1, \ldots, k \,\middle|\, \vec{\xi}[S]\right]$$

for any outcome of $\vec{\xi}[S]$, where $\vec{w}_{\vec{\xi}[S]} = \vec{w} - M[[k] \times S]\vec{\xi}[S] \in \mathbb{R}^k$. We wish to show that for all outcomes of $\vec{\xi}[S]$, the above conditional probability is bounded by the right-hand side of (6.1).

Recall that $P \in \mathbb{R}[y_1, \ldots, y_{\ell+k}, (x_i)_{i \in S}]$ is a quadratic polynomial, and by condition (iv), for some $j, j' \in [\ell]$ the coefficient of $y_j y_{j'}$ in P is nonzero. This means that the coefficient of $y_j y_{j'}$ in $P_{\vec{\xi}[S]}(y_1, \ldots, y_{\ell+k}) \in \mathbb{R}[y_1, \ldots, y_{\ell+k}]$ is still nonzero, for any outcome of $\vec{\xi}[S]$.

Since the coefficient vectors of the linear forms $g^{(1)}(\vec{x}), \ldots, g^{(\ell+k)}(\vec{x}) \in \mathbb{R}[x_1, \ldots, x_n]$ are the rows of the matrix M' in condition (iii), the coefficient vectors of $g_*^{(1)}(\vec{x}[I]), \ldots, g_*^{(\ell+k)}(\vec{x}[I]) \in \mathbb{R}[x_i : i \in I]$ are the rows of the matrix $M'[[\ell+k] \times I]$. Now, define the vectors $\vec{a}_i \in \mathbb{R}^{\ell+k}$, for $i \in I$, to be the columns of the matrix $M'[[\ell+k] \times I]$. Then for any outcome of $\vec{\xi}[I]$, the vector $(g_*^{(1)}(\vec{\xi}[I]), \ldots, g_*^{(\ell+k)}(\vec{\xi}[I]))$ agrees with $\sum_{i \in I} \vec{\xi}[i]\vec{a}_i$.

Furthermore, as $M'[[\ell+k] \times I] \in \mathcal{H}^{(\ell+k) \times I}(s/r)$, by Lemma 3.2(ii) the matrix $M'[[\ell+k] \times I]$ must contain $\lceil s/(r(\ell+k)) \rceil \geq \lceil s/(k+r)^2 \rceil$ disjoint nonsingular $(\ell+k) \times (\ell+k)$ submatrices, so among the vectors $\vec{a}_i$ for $i \in I$ we can form $\lceil s/(k+r)^2 \rceil$ disjoint bases of $\mathbb{R}^{\ell+k}$. Consider the largest integer $t \leq \lceil s/(k+r)^2 \rceil$ which is divisible by $2^{\ell-1}$, and note that $t \geq \lceil s/(k+r)^2 \rceil/2 \geq s/(2(k+r)^2)$, since $\lceil s/(k+r)^2 \rceil \geq 2^{3r^2} \geq 2^{\ell-1}$.

Now, for any outcome of $\vec{\xi}[S]$, we obtain

$$
\Pr\left[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w} \,\middle|\, \vec{\xi}[S]\right]
$$
$$
= \Pr\left[P_{\vec{\xi}[S]}\left(\sum_{i \in I} \vec{\xi}[i]\vec{a}_i\right) = 0 \text{ and } \sum_{i \in I} \vec{\xi}[i]\vec{a}_i[\ell+j] = \vec{w}_{\vec{\xi}[S]}[j] \text{ for } j = 1, \ldots, k \,\middle|\, \vec{\xi}[S]\right]
$$
$$
= \Pr\left[P_{\vec{\xi}[S]}\left(\sum_{i \in I} \vec{\xi}[i]\vec{a}_i\right) = 0 \text{ and } \sum_{i \in I} \vec{\xi}[i]\vec{a}_i \in \mathcal{W}_{\vec{\xi}[S]} \,\middle|\, \vec{\xi}[S]\right] = \Pr\left[\sum_{i \in I} \vec{\xi}[i]\vec{a}_i \in \mathcal{Z}_{\vec{\xi}[S]} \,\middle|\, \vec{\xi}[S]\right],
$$

where $\mathcal{W}_{\vec{\xi}[S]} \subseteq \mathbb{R}^{\ell+k}$ is the ℓ-dimensional affine-linear subspace consisting of the points $\vec{y} \in \mathbb{R}^{\ell+k}$ with $\vec{y}[\ell+j] = \vec{w}_{\vec{\xi}[S]}[j]$ for $j = 1, \ldots, k$, and $\mathcal{Z}_{\vec{\xi}[S]} \subseteq \mathcal{W}_{\vec{\xi}[S]}$ is the subset of $\mathcal{W}_{\vec{\xi}[S]}$ given by $\mathcal{Z}_{\vec{\xi}[S]} = \{\vec{y} \in \mathcal{W}_{\vec{\xi}[S]} : P_{\vec{\xi}[S]}(\vec{y}) = 0\}$.

We claim that for any outcome of $\vec{\xi}[S]$, we have $\mathcal{Z}_{\vec{\xi}[S]} \subsetneq \mathcal{W}_{\vec{\xi}[S]}$, i.e., $\mathcal{Z}_{\vec{\xi}[S]}$ is a quadric on $\mathcal{W}_{\vec{\xi}[S]}$. Indeed, if we had $\mathcal{Z}_{\vec{\xi}[S]} = \mathcal{W}_{\vec{\xi}[S]}$, then the polynomial $P_{\vec{\xi}[S]}$ would be identically zero on the entire subspace $\mathcal{W}_{\vec{\xi}[S]} \subseteq \mathbb{R}^{\ell+k}$. Note that on the space $\mathcal{W}_{\vec{\xi}[S]}$, we can identify $P_{\vec{\xi}[S]}$ with some quadratic polynomial in the variables $y_1, \ldots, y_\ell$ (obtained by substituting $y_{\ell+j} = \vec{w}_{\vec{\xi}[S]}[j]$ for $j = 1, \ldots, k$). This polynomial has a nonzero coefficient for some monomial of the form $y_j y_{j'}$ with $j, j' \in [\ell]$, since $P_{\vec{\xi}[S]}$ also has a nonzero coefficient for such a monomial. Hence the polynomial $P_{\vec{\xi}[S]}$ cannot vanish on the entire subspace $\mathcal{W}_{\vec{\xi}[S]}$, so indeed $\mathcal{Z}_{\vec{\xi}[S]} \subsetneq \mathcal{W}_{\vec{\xi}[S]}$.

Recalling that among the vectors $\vec{a}_i$ for $i \in I$ we can form t disjoint bases of $\mathbb{R}^{\ell+k}$, we can now apply Theorem 4.2 (after conditioning on any outcome of $\vec{\xi}[S]$) and obtain the bound

$$
\Pr\left[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w} \,\middle|\, \vec{\xi}[S]\right] = \Pr\left[\sum_{i \in I} \vec{\xi}[i]\vec{a}_i \in \mathcal{Z}_{\vec{\xi}[S]} \,\middle|\, \vec{\xi}[S]\right]
$$
$$
\leq \frac{2^{(\ell-1)(\ell+k)+1}}{t^{((\ell+k)-(\ell-1))/2}} \leq \frac{2^{(\ell-1)\ell(k+1)+1}}{t^{(k+1)/2}} \leq \frac{2^{\ell^2(k+1)}}{(s/(2(k+r)^2))^{(k+1)/2}}
$$
$$
\leq \left(\frac{2^{3\ell^2}(k+r)^2}{s}\right)^{(k+1)/2} = \left(\frac{s}{2^{3r^2}(k+r)^2}\right)^{-(k+1)/2},
$$

as desired. $\qquad\square$

7   Robust rank inheritance

In this section, we prove Lemma 5.7 (the robust rank inheritance lemma). Actually, we will deduce Lemma 5.7 from the following somewhat simpler statement with just two matrices T, A instead of three matrices T, U, A (to deduce Lemma 5.7, we will take $r = 2$ in the statement below).

Lemma 7.1. *Let $r \geq 1$ and $s \geq 1$ and $0 \leq k \leq m \leq n$ be integers. Consider matrices $T \in \mathbb{R}^{k \times m}$ and $A \in \mathbb{R}^{n \times m}$, and vectors $\vec{w} \in \mathbb{R}^k$ and $\vec{v} \in \mathbb{R}^n$. Assume that there exist disjoint subsets $I_1, \ldots, I_s \subseteq [m]$ and disjoint subsets $J_1, \ldots, J_s \subseteq [n]$ of size $|I_1| = \cdots = |I_s| = |J_1| = \cdots = |J_s| = k + r$ such that:*

(i) *For* $t = 1, \ldots, s$, *the submatrix* $T[[k] \times I_t]$ *has rank* k.

(ii) *For* $t = 1, \ldots, s$, *every* $(T[[k] \times I_t], 0)$-*perturbation of the matrix* $A[J_t \times I_t]$ *has rank at least* r.

Then, for a sequence of independent Rademacher random variables $\vec{\xi} = (\xi_1, \ldots, \xi_m) \in \{-1, 1\}^m$, *we have*

$$\Pr[T\vec{\xi} = \vec{w} \text{ and } A\vec{\xi} - \vec{v} \text{ has at most } s/6 \text{ nonzero coordinates}] \leq \left(\frac{s}{10^{60}(k+r)^{20}} \right)^{-(k+r)/2}. \tag{7.1}$$

Note that if $k = 0$, then condition (i) is vacuous, and in condition (ii), there are no nontrivial $(T[[k] \times I_t], 0)$-perturbations. This case of Lemma 7.1 implies Lemma 2.1 (the "Hamming norm" anticoncentration inequality mentioned in Section 2); the formal deduction can be found at the end of this section.

Also, note that assumptions (i) and (ii) in particular imply that for all $t = 1, \ldots, s$ we have

$$\mathrm{rank}\left(\frac{A[J_t \times I_t]}{T[[k] \times I_t]} \right) \geq k + r. \tag{7.2}$$

Indeed, if this rank were at most $k + r - 1$, then one would be able to form a basis of the row span of the above matrix using the k rows of $T[[k] \times I_t]$ and up to $r - 1$ rows of $A[J_t \times I_t]$ (recall that by assumption (i) the rows of the matrix $T[[k] \times I_t]$ are linearly independent). But then it would be possible to add linear combinations of the rows of $T[[k] \times I_t]$ to the rows of $A[J_t \times I_t]$ to obtain a matrix of rank at most $r - 1$, contradicting assumption (ii).

In order to prove Lemma 7.1 we will employ a *witness-counting* approach (as outlined in Section 2.4). We need to bound the probability that $T\vec{\xi} = \vec{w}$ and $A\vec{\xi} - \vec{v}$ has more than $n - s/6$ zero coordinates. Note that we can use Halász' inequality (Theorem 3.3) to upper-bound the probability that $T\vec{\xi} = \vec{w}$ and a particular set of coordinates of $A\vec{\xi} - \vec{v}$ are zero. However, it is far too wasteful to simply take a union bound over all subsets of $n - s/6$ coordinates. Instead, we consider certain types of "witness" sequences $(h_1, \ldots, h_z) \in [n]^z$ of indices where the coordinates of $A\vec{\xi} - \vec{v}$ are zero. If $A\vec{\xi} - \vec{v}$ has many zero coordinates, there must be many such "witness" sequences, but on the other hand we can bound the expected number of such "witness" sequences by considering the probability that $(A\vec{\xi} - \vec{v})[\{h_1, \ldots, h_z\}] = \vec{0}$ (and $T\vec{\xi} = \vec{w}$) for a given such sequence $(h_1, \ldots, h_z)$.

The following lemma states that under the assumptions of Lemma 7.1, one can find a large submatrix of A which "has its rank robustly" (the assumptions of the assumptions of Lemma 7.1 guarantee that A robustly has rank *at least* r, but the rank of A may actually be much larger than r, and that larger rank may be "fragile"). More precisely, we can find such a submatrix even within any large specified set H of rows.

Lemma 7.2. *Let* $r, k, s, m, n \in \mathbb{Z}$, *the matrices* $T \in \mathbb{R}^{k \times m}$ *and* $A \in \mathbb{R}^{n \times m}$, *and the sets* $J_1, \ldots, J_s \subseteq [n]$ *be as in Lemma 7.1. Then for any subset* $H \subseteq J_1 \cup \cdots \cup J_s$ *of size* $|H| \geq (k + r)s - (2/3)s$, *there are subsets* $J \subseteq H$ *and* $I \subseteq [m]$ *with* $|J| \geq |H| - s/6$ *and* $|I| \geq m - s/6$ *and an integer* $z \geq r$ *such that*

$$\mathrm{rank}\left(\frac{A[J \times I]}{T[[k] \times I]} \right) = \mathrm{rank}\left(\frac{A[J' \times I']}{T[[k] \times I']} \right) = k + z \tag{7.3}$$

for all subsets $J' \subseteq J$ *and* $I' \subseteq I$ *of sizes* $|J'| \geq |J| - s/(12z^2)$ *and* $|I'| \geq |I| - s/(12z^2)$.

Proof. For every integer $z \geq 0$, define $f(z) = \sum_{y=z+1}^{\infty} s/(12y^2)$ and note that $f(z) \leq \sum_{y=1}^{\infty} s/(12y^2) < s/6$ for all $z \geq 0$. Now consider the minimum integer $z \geq 0$ such that there exist subsets $J \subseteq H$ and $I \subseteq [m]$ with $|J| \geq |H| - f(z)$ and $|I| \geq m - f(z)$ and

$$\mathrm{rank}\left(\frac{A[J \times I]}{T[[k] \times I]} \right) \leq k + z$$

(such an integer z exists, for example $z = m$ satisfies the condition when taking $J = H$ and $I = [m]$). We claim that $z \geq r$. Indeed, note that $|J| \geq |H| - f(z) > (k+r)s - (2/3)s - s/6 = |J_1 \cup \cdots \cup J_s| - (5s/6)$ and

24

$|I| \geq m - f(z) > m - s/6$, so there are strictly fewer than s indices $t \in \{1, \dots, s\}$ such that $I_t \setminus I \neq \emptyset$ or $J_t \setminus J \neq \emptyset$. Hence there must be some index $t \in \{1, \dots, s\}$ with $I_t \subseteq I$ and $J_t \subseteq J$, and we have

$$k + z \geq \mathrm{rank}\begin{pmatrix} A[J \times I] \\ T[[k] \times I] \end{pmatrix} \geq \mathrm{rank}\begin{pmatrix} A[J_t \times I_t] \\ T[[k] \times I_t] \end{pmatrix} \geq k + r,$$

where the last step follows from the assumptions in Lemma 7.1, see (7.2). So we indeed have $z \geq r \geq 1$.

Finally, for any subsets $I' \subseteq I$ and $J' \subseteq J$ of sizes $|I'| \geq |I| - s/(12z^2)$ and $|J'| \geq |J| - s/(12z^2)$, we have $|I'| \geq m - f(z) - s/(12z^2) = m - f(z-1)$ and $|J'| \geq |H| - f(z) - s/(12z^2) = |H| - f(z-1)$, so by the choice of z we must have

$$k + z \geq \mathrm{rank}\begin{pmatrix} A[J \times I] \\ T[[k] \times I] \end{pmatrix} \geq \mathrm{rank}\begin{pmatrix} A[J' \times I'] \\ T[[k] \times I'] \end{pmatrix} \geq k + (z-1) + 1 = k + z,$$

implying (7.3). □

We can use the robust-rank submatrix guaranteed by Corollary 7.3 to find many sequences $(h_1, \dots, h_z)$ to be used as "witnesses", as follows.

Corollary 7.3. *Let $r, k, s, m, n \in \mathbb{Z}$, the matrices $T \in \mathbb{R}^{k \times m}$ and $A \in \mathbb{R}^{n \times m}$, and the sets $J_1, \dots, J_s \subseteq [n]$ be as in Lemma 7.1. Then for any subset $H \subseteq J_1 \cup \dots \cup J_s$ of size $|H| \geq (k+r)s - (2/3)s$, there is some integer $z \geq r$ such that there are at least $s^z/(12z^2)^z$ sequences $(h_1, \dots, h_z) \in H^z$ with*

$$\begin{pmatrix} A[\{h_1, \dots, h_z\} \times [m]] \\ T \end{pmatrix} \in \mathcal{H}^{(k+z) \times m}(s/(12z^2)). \tag{7.4}$$

Proof. Let $J \subseteq H$ and $I \subseteq [m]$ and $z \geq r$ be as in Lemma 7.2. We first claim that there are at least $s^z/(12z^2)^z$ sequences $(h_1, \dots, h_z) \in J^z \subseteq H^z$ such that

$$\mathrm{row\text{-}span}\begin{pmatrix} A[\{h_1, \dots, h_z\} \times I] \\ T[[k] \times I] \end{pmatrix} = \mathrm{row\text{-}span}\begin{pmatrix} A[J \times I] \\ T[[k] \times I] \end{pmatrix} \tag{7.5}$$

Indeed, the matrix on the right-hand side has rank $k + z$ by (7.3), and the rows of $T[[k] \times I]$ are linearly independent by assumption (i) in Lemma 7.1 (since $|I| \geq m - s/6$, there needs to be at least one index $t \in \{1, \dots, s\}$ with $I_t \subseteq I$, meaning that $\mathrm{rank}\, T[[k] \times I] = T[[k] \times I_t] = k$). So we can form a basis of the row span of the matrix on the right-hand side by taking the rows of $T[[k] \times I]$ and adding, one by one, z different rows of $A[J \times I]$. By (7.3), at every step we have at least $s/(12z^2)$ choices for a new row to add to our basis (indeed, the index set $J' \subseteq J$ of all rows in the span of the already selected basis elements must have size $|J'| < |J| - s/(12z^2)$, since this span has dimension less than $k + z$ and so otherwise we would have a contradiction to (7.3) with $I' = I$). Thus, there are indeed at least $s^z/(12z^2)^z$ choices for the sequence $(h_1, \dots, h_z) \in J^z$ of indices of the rows of $A[J \times I]$ selected in this process. For each such sequence (7.5) holds.

It remains to show that every sequence $(h_1, \dots, h_z) \in J^z$ satisfying (7.5) also satisfies (7.4). So, consider any $(h_1, \dots, h_z) \in J^z$ satisfying (7.5). To show (7.4), it suffices to show that

$$\begin{pmatrix} A[\{h_1, \dots, h_z\} \times I] \\ T[[k] \times I] \end{pmatrix} \in \mathcal{H}^{(k+z) \times I}(s/(12z^2)). \tag{7.6}$$

Suppose the contrary; then there exists a subset $I' \subseteq I$ of size $|I'| \geq |I| - s/(12z^2)$ such that

$$\mathrm{rank}\begin{pmatrix} A[\{h_1, \dots, h_z\} \times I'] \\ T[[k] \times I'] \end{pmatrix} < k + z.$$

But now by (7.5) we have

$$\mathrm{row\text{-}span}\begin{pmatrix} A[J \times I'] \\ T[[k] \times I'] \end{pmatrix} = \mathrm{row\text{-}span}\begin{pmatrix} A[\{h_1, \dots, h_z\} \times I'] \\ T[[k] \times I'] \end{pmatrix}$$

25

and hence

$$\mathrm{rank}\begin{pmatrix} A[J \times I'] \\ T[[k] \times I'] \end{pmatrix} = \mathrm{rank}\begin{pmatrix} A[\{h_1, \ldots, h_z\} \times I'] \\ T[[k] \times I'] \end{pmatrix} < k + z,$$

contradicting (7.3). So we indeed have (7.6), as desired. □

Now we prove Lemma 7.1. We need two slightly different implementations of our witness-counting strategy described above, with different conditions for the "witness" sequences $(h_1, \ldots, h_z)$ we are counting. In the case where A "robustly has very high rank", we can get away with rather crude arguments using Odlyzko's lemma. In the complementary case, we need to use Corollary 7.3 and Halász' inequality (we cannot simply use Corollary 7.3 and Halász' inequality in both cases, because the conclusions of Lemma 7.2 and Corollary 7.3 are very weak when z is very large).

Proof of Lemma 7.1. Let[11] $L = \lceil 2(k+r)\log s \rceil$, let $J_* = J_1 \cup \cdots \cup J_s \subseteq [n]$ and note that $|J_*| = (k+r)s$. Furthermore note that (7.1) trivially holds if $s \le 10^{60}(k+r)^{20}$, so we may assume that $s > 10^{60}(k+r)^{20} \ge 10^{60}$ and hence $\log s \le s^{1/20}$. Then we in particular have

$$L = \lceil 2(k+r)\log s \rceil \le 4(k+r) \cdot \log s \le \frac{s^{1/20}}{250} \cdot s^{1/20} = \frac{s^{1/10}}{250}. \tag{7.7}$$

Step 1: The robust-high-rank case. First, suppose that for every subset $J' \subseteq J_*$ with $|J'| \ge |J_*| - s/2$, we have $\mathrm{rank}\, A[J' \times [m]] \ge L$.

Then, for every subset $S \subseteq J_*$ of size $|S| < L$, the row span of $A[S \times [m]]$ has dimension less than L and so it can contain at most $|J_*| - s/2$ rows of the matrix $A[J_* \times [m]]$. This means that there are at least $s/2$ indices $h \in J_*$ such that $\mathrm{rank}\, A[(S \cup \{h\}) \times [m]] = \mathrm{rank}\, A[S \times [m]] + 1$. Let $H_S \subseteq J_*$ denote the set of the $\lceil s/2 \rceil$ smallest such indices $h \in J_*$.

For our double-counting argument, we will consider sequences $(h_1, \ldots, h_L) \in J_*^L$ with $h_t \in H_{\{h_1, \ldots, h_{t-1}\}}$ for $t = 1, \ldots, L$. For every such sequence, we have $\mathrm{rank}\, A[\{h_1, \ldots, h_t\} \times [m]] = \mathrm{rank}\, A[\{h_1, \ldots, h_{t-1}\} \times [m]] + 1$ for $t = 1, \ldots, L$ and hence $\mathrm{rank}\, A[\{h_1, \ldots, h_L\} \times [m]] = L$. Furthermore, note that the total number of such sequences is exactly $\lceil s/2 \rceil^L$ (since after choosing $h_1, \ldots, h_{t-1}$ we have exactly $|H_{\{h_1, \ldots, h_{t-1}\}}| = \lceil s/2 \rceil$ choices for h_t).

We claim that for every outcome of $\vec{\xi} \in \{-1, 1\}^m$ such that $A\vec{\xi} - \vec{v}$ has at most $s/6$ nonzero coordinates, there are at least $(s/3)^L$ sequences $(h_1, \ldots, h_L) \in J_*^L$ with $h_t \in H_{\{h_1, \ldots, h_{t-1}\}}$ for $t = 1, \ldots, L$, such that $(A\vec{\xi} - \vec{v})[\{h_1, \ldots, h_z\}] = \vec{0}$. Indeed, choosing $h_1, \ldots, h_L$ one at a time, at every step we need to choose $h_t \in H_{\{h_1, \ldots, h_{t-1}\}}$ with $(A\vec{\xi} - \vec{v})[h_t] = 0$. At most $s/6$ of the $\lceil s/2 \rceil$ elements of $H_{\{h_1, \ldots, h_{t-1}\}}$ fail this condition, so we indeed have at least $\lceil s/2 \rceil - s/6 \ge s/3$ choices at every step. Hence are indeed at least $(s/3)^L$ such sequences $(h_1, \ldots, h_L) \in J_*^L$.

Thus, the expected number of sequences $(h_1, \ldots, h_L) \in J_*^L$ with $h_t \in H_{\{h_1, \ldots, h_{t-1}\}}$ for $t = 1, \ldots, L$ such that $(A\vec{\xi} - \vec{v})[\{h_1, \ldots, h_z\}] = \vec{0}$ is at least $\Pr[A\vec{\xi} - \vec{v}$ has at most $s/6$ nonzero coordinates$] \cdot (s/3)^L$. On the other hand, for every given sequence $(h_1, \ldots, h_L) \in J_*^L$ with $h_t \in H_{\{h_1, \ldots, h_{t-1}\}}$ for $t = 1, \ldots, L$, having $(A\vec{\xi} - \vec{v})[\{h_1, \ldots, h_z\}] = \vec{0}$ is equivalent to $A[\{h_1, \ldots, h_L\} \times [m]]\vec{\xi} = \vec{v}[\{h_1, \ldots, h_L\}]$ and by Odlyzko's lemma (Lemma 3.7), this happens with probability at most 2^{-L} (recalling that $\mathrm{rank}\, A[\{h_1, \ldots, h_L\} \times [m]] = L$). Hence the expected number of such sequences $(h_1, \ldots, h_L)$ is at most $\lceil s/2 \rceil^L \cdot 2^{-L}$. Thus, we obtain

$$\Pr[A\vec{\xi} - \vec{v} \text{ has at most } s/6 \text{ nonzero coordinates}] \le \frac{\lceil s/2 \rceil^L \cdot 2^{-L}}{(s/3)^L} \le \frac{(5s/9)^L \cdot 2^{-L}}{(s/3)^L} = (5/6)^L \le s^{-(k+r)/2},$$

recalling that $L = \lceil 2(k+r)\log s \rceil$ and using that $(5/6)^4 < 1/2$. This in particular implies (7.1).

Step 2: Covering events for the low-rank case. Now, we may assume that for some subset $J' \subseteq J_*$ with $|J'| \ge |J_*| - s/2$, we have $\mathrm{rank}\, A[J' \times [m]] \le L$.

---

[11]Recall that $\log s$ denotes the base-2 logarithm of s.

For every outcome of $\vec{\xi} \in \{-1, 1\}^m$, let $H_{\vec{\xi}} \subseteq J'$ be the set of indices $h \in J'$ with $(A\vec{\xi} - \vec{v})[h] = 0$. Note that whenever $A\vec{\xi} - \vec{v}$ has at most $s/6$ nonzero coordinates, we have $|H_{\vec{\xi}}| \geq |J'| - s/6 \geq |J_*| - s/2 - s/6 = (k+r)s - (2/3)s$, and so by Corollary 7.3 there is some integer $z \geq r$ such that there are at least $s^z/(12z^2)^z$ sequences $(h_1, \ldots, h_z) \in H_{\vec{\xi}}^z$ satisfying (7.4). Note that for each such sequence, by (7.4) we must in particular have

$$k + z = \operatorname{rank}\left(\begin{matrix} A[\{h_1, \ldots, h_z\} \times [m]] \\ T \end{matrix}\right) \leq k + \operatorname{rank} A[\{h_1, \ldots, h_z\} \times [m]] \leq k + \operatorname{rank} A[J' \times [m]] \leq k + L,$$

and consequently $z \leq L$.

Thus, for every outcome of $\vec{\xi} \in \{-1, 1\}^m$ such that $A\vec{\xi} - \vec{v}$ has at most $s/6$ nonzero coordinates, there is some integer $z \in \{r, r+1, \ldots, L\}$ such that there are at least $s^z/(12z^2)^z$ sequences $(h_1, \ldots, h_z) \in J_*^z$ satisfying (7.4) and $(A\vec{\xi} - \vec{v})[\{h_1, \ldots, h_z\}] = \vec{0}$. For every $z \in \{r, r+1, \ldots, L\}$, let $\mathcal{E}_z$ be the event that $T\vec{\xi} = \vec{w}$ and that there are at least $s^z/(12z^2)^z$ sequences $(h_1, \ldots, h_z) \in J_*^z$ satisfying (7.4) and $(A\vec{\xi} - \vec{v})[\{h_1, \ldots, h_z\}] = \vec{0}$. Then

$$\Pr[T\vec{\xi} = \vec{w} \text{ and } A\vec{\xi} - \vec{v} \text{ has at most } s/6 \text{ nonzero coordinates}] \leq \sum_{z=r}^{L} \Pr[\mathcal{E}_z] \qquad (7.8)$$

and it remains to bound the probability of the events $\mathcal{E}_z$ for $z = r, \ldots, L$.

Step 3: Double-counting. Let $z \in \{r, r+1, \ldots, L\}$ (and note that then in particular $z \geq r \geq 1$).

For every sequence $(h_1, \ldots, h_z) \in J_*^z$ satisfying (7.4), we have the equivalence

$$T\vec{\xi} = \vec{w} \text{ and } (A\vec{\xi} - \vec{v})[\{h_1, \ldots, h_z\}] = \vec{0} \quad \Longleftrightarrow \quad \left(\begin{matrix} A[\{h_1, \ldots, h_z\} \times [m]] \\ T \end{matrix}\right)\vec{\xi} = \left(\begin{matrix} \vec{v}[\{h_1, \ldots, h_z\}] \\ \vec{w} \end{matrix}\right).$$

So, by the version of Halász' inequality in Corollary 3.5, we have

$$\Pr\left[T\vec{\xi} = \vec{w} \text{ and } (A\vec{\xi} - \vec{v})[\{h_1, \ldots, h_z\}] = \vec{0}\right] \leq \left(\frac{s}{12z^2(z+k)}\right)^{-(k+z)/2} \leq (4(k+z)^2)^{k+z} \cdot s^{-(k+z)/2}$$

for every sequence $(h_1, \ldots, h_z) \in J_*^z$ satisfying (7.4).

Hence, the expected number of sequences $(h_1, \ldots, h_z) \in J_*^z$ satisfying (7.4) such that $(A\vec{\xi} - \vec{v})[\{h_1, \ldots, h_z\}] = \vec{0}$ and $T\vec{\xi} = \vec{w}$ hold, is at most $|J_*|^z \cdot (4(k+z)^2)^{k+z} \cdot s^{-(k+z)/2}$. On the other hand, whenever the event $\mathcal{E}_z$ occurs, there are at least $s^z/(12z^2)^z$ such sequences. We deduce that

$$\Pr[\mathcal{E}_z] \leq \frac{|J_*|^z \cdot (4(k+z)^2)^{k+z} \cdot s^{-(k+z)/2}}{s^z/(12z^2)^z} \leq \frac{((k+r)s)^z(48(k+z)^4)^{k+z}}{s^z \cdot s^{(k+z)/2}} \leq \frac{(48(k+z)^5)^{k+z}}{s^{(k+z)/2}}$$

for every $z \in \{r, \ldots, L\}$.

Step 4: Summing up. We now obtain

$$\sum_{z=r}^{L} \Pr[\mathcal{E}_z] \leq \sum_{z=r}^{L} \frac{(48(k+z)^5)^{k+z}}{s^{(k+z)/2}} = \sum_{z=r}^{L} a(z), \quad \text{where } a(z) = \frac{(48(k+z)^5)^{k+z}}{s^{(k+z)/2}} \text{ for } z = r, \ldots, L.$$

For $z = r, \ldots, L-1$ we compute (recalling that $k \leq \lceil 2(k+r)\log s \rceil = L \leq s^{1/10}/250$ by (7.7) and hence $k + z + 1 \leq k + L \leq 2L \leq s^{1/10}/100$)

$$\frac{a(z+1)}{a(z)} = \frac{48(k+z+1)^5}{s^{1/2}} \cdot \left(\frac{k+z+1}{k+z}\right)^{5(k+z)} \leq \frac{48(s^{1/10}/100)^5}{s^{1/2}} \cdot e^5 \leq \frac{48 \cdot 3^5}{100^5} \leq \frac{1}{2}.$$

So, we have $a(z) \leq 2^{-(z-r)}a(r)$ for $z = r, \ldots, L$ (so $a(z)$ is bounded by a geometric series with common ratio $1/2$), and consequently

$$\sum_{z=r}^{L} \Pr[\mathcal{E}_z] \leq \sum_{z=r}^{L} a(z) \leq \sum_{z=r}^{L} 2^{-(z-r)}a(r) \leq 2 \cdot a(r) = \frac{2 \cdot (48(k+r)^5)^{k+r}}{s^{(k+r)/2}} \leq \left(\frac{s}{10^{60}(k+r)^{20}}\right)^{-(k+r)/2}.$$

Together with (7.8), this implies (7.1). $\qquad \square$

Next, we deduce Lemma 5.7 from Lemma 7.1.

Proof of Lemma 5.7. Recalling that $(T, U, A) \in \mathcal{M}_2^{k,m,n}(s)$, let $I_1, \ldots, I_s, J_1, \ldots, J_s$ be index sets as in the definition of $\mathcal{M}_2^{k,m,n}(s)$ in Definition 5.6 (recall in particular that $I_1, \ldots, I_s \subseteq [m]$ are disjoint sets and $J_1, \ldots, J_s \subseteq [n]$ are disjoint sets).

Note that $U \in \mathcal{H}^{k \times n}(s)$, due to condition (b) in Definition 5.6. Write $\mathcal{W}_U \subseteq \mathbb{R}^n$ for the row span of U, and note that we have $U'_{\vec{\xi}} \notin \mathcal{H}^{(k+1) \times n}(s/6)$ if and only if $A\vec{\xi} - \vec{b}$ agrees in at least $n - s/6$ coordinates with a vector $\vec{v} \in \mathcal{W}_U$. Hence

$$\Pr\left[T\vec{\xi} = \vec{y} \text{ and } U'_{\vec{\xi}} \notin \mathcal{H}^{(k+1) \times n}(s/6)\right]$$

$$= \Pr\left[T\vec{\xi} = \vec{y} \text{ and there is } S \subseteq [n] \text{ with } |S| \geq n - s/6 \text{ and } (A\vec{\xi} - \vec{b})[S] = \vec{v}[S] \text{ for some } \vec{v} \in \mathcal{W}_U\right]$$

$$\leq \frac{2}{s} \sum_{t=1}^{s} \Pr\left[T\vec{\xi} = \vec{y} \text{ and there is } J_t \subseteq S \subseteq [n] \text{ with } |S| \geq n - s/6 \text{ and } (A\vec{\xi} - \vec{b})[S] = \vec{v}[S] \text{ for some } \vec{v} \in \mathcal{W}_U\right],$$

where the inequality follows by observing that every set $S \subseteq [n]$ of size $|S| \geq n - s/6$ satisfies $J_t \subseteq S$ for at least $s - s/6 \geq s/2$ indices $t \in \{1, \ldots, s\}$. So it suffices to show that for every fixed $t = 1, \ldots, s$ we have

$$\Pr\left[T\vec{\xi} = \vec{y} \text{ and there is } J_t \subseteq S \subseteq [n] \text{ with } |S| \geq n - s/6 \text{ and } (A\vec{\xi} - \vec{b})[S] = \vec{v}[S] \text{ for some } \vec{v} \in \mathcal{W}_U\right]$$

$$\leq \left(\frac{s}{10^{60}(k+2)^{20}}\right)^{-(k+2)/2}. \tag{7.9}$$

By condition (b) in Definition 5.6 we can choose a subset $J'_t \subseteq J_t$ of size $|J'_t| = k$ such that $U[[k] \times J'_t]$ is nonsingular . Without loss of generality we may assume that $A[J'_t \times [m]]$ is the all-zero matrix, since the event on the left-hand side of (7.9) does not change if we add linear combinations of the rows of U to the columns of A. But this assumption means that $(A\vec{\xi} - \vec{b})[J'_t] = -\vec{b}[J'_t]$, and recalling that $U[[k] \times J'_t]$ is nonsingular, there is only exactly one vector $\vec{v}_t \in \mathcal{W}_U$ with $\vec{v}_t[J'_t] = -\vec{b}[J'_t]$. Hence the event on the left-hand side of (7.9) can only happen with $\vec{v} = \vec{v}_t$, and we obtain

$$\Pr\left[T\vec{\xi} = \vec{y} \text{ and there is } J_t \subseteq S \subseteq [n] \text{ with } |S| \geq n - s/6 \text{ and } (A\vec{\xi} - \vec{b})[S] = \vec{v}[S] \text{ for some } \vec{v} \in \mathcal{W}_U\right]$$

$$= \Pr\left[T\vec{\xi} = \vec{y} \text{ and there is } J_t \subseteq S \subseteq [n] \text{ with } |S| \geq n - s/6 \text{ such that } (A\vec{\xi} - \vec{b})[S] = \vec{v}_t[S]\right]$$

$$\leq \Pr[T\vec{\xi} = \vec{y} \text{ and } A\vec{\xi} - \vec{b} - \vec{v}_t \text{ has at least } n - s/6 \text{ zero coordinates}]$$

$$= \Pr[T\vec{\xi} = \vec{y} \text{ and } A\vec{\xi} - \vec{b} - \vec{v}_t \text{ has at most } s/6 \text{ nonzero coordinates}].$$

The desired bound (7.9) now follows from Lemma 7.1 with $r = 2$ (note that conditions (i) and (ii) in Lemma 7.1 hold by conditions (a) and (c) in Definition 5.6, respectively). $\qquad\square$

We end this section by deducing Lemma 2.1 (the "Hamming norm" anticoncentration inequality stated in the outline in Section 2) from Lemma 7.1.

Proof of Lemma 2.1. Let $C_r = (10r)^{30r}$ and $c_r = 1/(6r)$. As in the statement of the lemma, let $A \in \mathbb{R}^{m \times n}$ be a matrix which has rank at least r after deletion of any t rows and t columns.

We claim that for $s = \lceil t/r \rceil$ we can find disjoint subsets $I_1, \ldots, I_s \subseteq [n]$ and disjoint subsets $J_1, \ldots, J_s \subseteq [m]$ of size $|I_1| = \cdots = |I_s| = |J_1| = \cdots = |J_s| = r$ such that rank $A[J_\ell \times I_\ell] = r$ for $\ell = 1, \ldots, s$. Indeed, we can find such subsets greedily: after having found $I_1, \ldots, I_\ell \subseteq [n]$ and $J_1, \ldots, J_\ell \subseteq [m]$ for some $\ell < t/r$, we can delete all columns with indices in $I_1 \cup \cdots \cup I_\ell$ and all rows with indices in $J_1 \cup \cdots \cup J_\ell$ from A and the resulting matrix must have rank at least r, so there must be subsets $I_{\ell+1} \subseteq [n] \setminus (I_1 \cup \cdots \cup I_\ell)$ and $J_{\ell+1} \subseteq [m] \setminus (J_1 \cup \cdots \cup J_\ell)$ with rank $A[J_\ell \times I_\ell] = r$ and $|J_{\ell+1}| = |I_{\ell+1}| = r$.

Having found such subsets $I_1, \ldots, I_s \subseteq [n]$ and $J_1, \ldots, J_s \subseteq [m]$, we can now apply Lemma 7.1 with $k = 0$ and $T \in \mathbb{R}^{0 \times n}$ being the empty matrix and $\vec{w} \in \mathbb{R}^0$ being the empty vector. Note that condition (i) is the vacuously true, and condition (ii) is true because for every $\ell = 1, \ldots, s$ the only $(T[[k] \times I_t], 0)$-perturbation of the matrix $A[J_t \times I_t]$ is the matrix $A[J_t \times I_t]$ itself, which has rank r. Thus, for a sequence $\vec{\xi} = (\xi_1, \ldots, \xi_n) \in \{-1, 1\}^n$ of independent Rademacher random variables, and for any vector $\vec{v} \in \mathbb{R}^m$, we have

$$\Pr[A\vec{\xi} \text{ differs from } \vec{v} \text{ in fewer than } t/(6r) \text{ coordinates}] \leq \Pr[A\vec{\xi} - \vec{v} \text{ has at most } s/6 \text{ nonzero coordinates}]$$

$$\leq \left(\frac{s}{10^{60} r^{20}} \right)^{-r/2} \leq \left(\frac{t}{10^{60} r^{21}} \right)^{-r/2} \leq (10r)^{30r} \cdot t^{-r/2}$$

(noting that $T\vec{\xi} = \vec{w}$ holds vacuously for all $\vec{\xi} \in \{-1, 1\}^n$). $\qquad\qquad\square$

8 Splitting the index set

In this section, we prove Lemma 5.8, splitting our index set $[n]$ into disjoint subsets I and J such that the conditions in Definition 5.6 are satisfied. This amounts to finding disjoint subsets $I_1, \ldots, I_s, J_1, \ldots, J_s \subseteq [n]$ satisfying certain robust-rank-2 conditions. Most of the work is in finding a single pair of subsets I_1, J_1 satisfying the desired property; we will then be able to find our subsets in a greedy fashion, in a similar way to the proof of Lemma 3.2(ii).

Proof of Lemma 5.8. Let $0 \leq k \leq n$, $s \geq 4k + 8$, $M \in \mathcal{H}^{k \times n}(s)$ and $A \in \mathbb{R}^{n \times n}$ be as in the statement of the lemma. Recall that A is a symmetric matrix and that we are assuming rank $A^*[S \times S] \geq 2$ for any subset $S \subseteq [n]$ of size $|S| \geq n - s$ and any matrix $A^* \in \mathbb{R}^{n \times n}$ that agrees with some (M, M)-perturbation of A in all off-diagonal entries.

Step 1: Setup. For $\ell = \lfloor s/(4k + 8) \rfloor$, we wish to find disjoint subsets $I_1, \ldots, I_\ell, J_1, \ldots, J_\ell \subseteq [n]$ of size $|I_1| = \cdots = |I_\ell| = |J_1| = \cdots = |J_\ell| = k + 2$ such that the following three conditions hold for $t = 1, \ldots, \ell$:

(a) The submatrix $M[[k] \times I_t]$ has rank k.
(b) The submatrix $M[[k] \times J_t]$ has rank k.
(c) Every $(M[[k] \times I_t], M[[k] \times J_t])$-perturbation of the matrix $A[J_t \times I_t]$ has rank at least 2.

This suffices, since we can then take $I = I_1 \cup \cdots \cup I_\ell$ and $J = [n] \setminus I$ to obtain a partition $[n] = I \cup J$ satisfying the conclusion of the lemma (the conditions (a)–(c) in Definition 5.6 precisely correspond to conditions (a) to (c) above, and we have $k \leq |I| \leq |J|$ since $|I| = \ell(k + 2)$ and $|J| \geq |J_1 \cup \cdots \cup J_\ell| = \ell(k + 2)$).

To find such disjoint subsets $I_1, \ldots, I_\ell, J_1, \ldots, J_\ell \subseteq [n]$, it suffices to show that for any subset $R \subseteq [n]$ with $|R| > n - s/2$, we can find disjoint subsets $I_1, J_1 \subseteq R$ of size $|I_1| = |J_1| = k + 2$ such that conditions (a)–(c) above hold for $t = 1$. Indeed, then we can then construct the desired subsets $I_1, \ldots, I_\ell, J_1, \ldots, J_\ell \subseteq [n]$ greedily (at every step choosing $I_t, J_t \subseteq [n] \setminus (I_1 \cup \cdots \cup I_{t-1} \cup J_1 \cup \cdots \cup J_{t-1})$). Showing this will be our objective for the rest of the proof, so let us fix a subset $R \subseteq [n]$ with $|R| > n - s/2$.

Since $M \in \mathcal{H}^{k \times n}(s)$, we can find disjoint subsets $I_1', J_1' \subseteq R$ of size $|I_1'| = |J_1'| = k$ with rank $M[[k] \times I_1'] = $ rank $M[[k] \times J_1'] = k$. We will obtain I_1, J_1 by augmenting I_1' and J_1' with two additional indices each (we will find suitable indices for this using our assumption on A).

Step 2: Augmenting the index sets. Since the $k \times k$ matrices $M[[k] \times I_1']$ and $M[[k] \times J_1']$ are both nonsingular, we can add linear combinations of the rows of M to the rows and columns of A to obtain an (M, M)-perturbation A' of A such that all entries of $A'[J_1' \times [n]]$ and all entries of $A'[[n] \times I_1']$ are zero. Define $R' = R \setminus (I_1' \cup J_1')$ and note that $|R'| \geq n - s/2 - 2k \geq n - s + 4$. By our assumption on A, the submatrix $A'[R' \times R']$ must have some nonzero off-diagonal entry $a'_{j,i}$ with distinct $i, j \in R'$ (if all off-diagonal entries were zero, by modifying the diagonal entries of A' we would be able to find a matrix $A^* \in \mathbb{R}^{n \times n}$ agreeing with the (M, M)-perturbation A' of A in all off-diagonal entries such that rank $A^*[R' \times R'] = 0$, contradicting our assumption on A).

Now, let A^* be the matrix obtained from A' by adjusting the diagonal entries $a'_{h,h}$ for $h \in R' \setminus \{i,j\}$ in such a way that each 2×2 submatrix of the form $A^*[\{j,h\} \times \{i,h\}]$ (for any $h \in R' \setminus \{i,j\}$) is singular. Since A^* agrees with the (M,M)-perturbation A' of A in all off-diagonal entries, we have rank $A^*[S \times S] \geq 2$ for every subset $S \subseteq [n]$ of size $|S| \geq n - s$, and in particular for the subset $S = R' \setminus \{i,j\}$. Hence we obtain rank $A^*[(R' \setminus \{i\}) \times (R' \setminus \{j\})] \geq 2$, and also note that the (j,i)-entry of A^* is $a^*_{j,i} = a'_{j,i} \neq 0$.

Now, if a matrix has rank at least 2, then for any nonzero row (respectively, column), we can find a second linearly independent row (respectively, column). Applying this fact to the matrix $A^*[(R' \setminus \{i\}) \times (R' \setminus \{j\})]$ and the row with index j, we can find an index $j' \in R' \setminus \{i,j\}$ such that rank $A^*[\{j,j'\} \times (R' \setminus \{j\})] = 2$ (i.e., such that the row with index j' is linearly independent from the row with index j). Applying the fact again to the matrix $A^*[\{j,j'\} \times (R' \setminus \{j\})]$ and the column with index i, we can find an index $i' \in R' \setminus \{j,i\}$ such that rank $A^*[\{j,j'\} \times \{i,i'\}] = 2$. Now, we must have $j' \neq i'$ (since rank $A^*[\{j,h\} \times \{i,h\}] \leq 1$ for all $h \in R' \setminus \{j,i\}$). So, $\{j,j'\}$ and $\{i,i'\}$ are disjoint subsets of $R' = R \setminus (I'_1 \cup J'_1)$, and we have rank $A'[\{j,j'\} \times \{i,i'\}] = $ rank $A^*[\{j,j'\} \times \{i,i'\}] = 2$. Defining $I_1 = I'_1 \cup \{i,i'\}$ and $J_1 = J'_1 \cup \{i,i'\}$, we obtain disjoint subsets of R of size $k + 2$.

Step 3: Proving the full-rank condition. We now need to show that the subsets $I_1, J_1 \subseteq R \subseteq [n]$ satisfy conditions (a)–(c) above for $t = 1$. This is equivalent to showing that the $(2k + 2) \times (2k + 2)$ matrix

$$\begin{pmatrix} A[J_1 \times I_1] & M[[k] \times J_1]^{\mathsf{T}} \\ M[[k] \times I_1] & 0 \end{pmatrix}$$

has full rank $2k + 2$. But note that the above matrix can be reduced to

$$\begin{pmatrix} 0 & 0 & M[[k] \times J'_1]^{\mathsf{T}} \\ 0 & A'[\{j,j'\} \times \{i,i'\}] & M[[k] \times \{j,j'\}]^{\mathsf{T}} \\ M[[k] \times I'_1] & M[[k] \times \{i,i'\}] & 0 \end{pmatrix}$$

by elementary row and column operations (precisely the row and column operations that were used to obtain A' from A in the previous step of the proof). Recalling that the matrices $M[[k] \times J'_1]$, $M[[k] \times I'_1]$ and $A'[\{j,j'\} \times \{i,i'\}]$ are nonsingular, the desired result follows. $\qquad\square$

9 Proof of the recursive bound in Theorem 5.5

In this section, we finally prove Theorem 5.5, using the results established in the previous three sections.

Proof of Theorem 5.5. As in the statement of the theorem, let $k \geq 0$ be an integer, $s > 0$ be a real number and define $s_* = s/(k+2)^{500}$. To show the desired bound on $f(k,s)$, we need to show that for any quadruple $(n, Q, M, \vec{w})$ as in Definition 5.4, we have

$$\Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}] \leq \max\left\{ s_*^{-(k+1)/2}, \quad s_*^{-(k+2)/2} + s_*^{-k/4} \cdot f(k+1, s_*)^{1/2} \right\}, \tag{9.1}$$

where the probability is taken with respect to a sequence of independent Rademacher random variables $\vec{\xi} \in \{-1,1\}^n$. This is clearly true if $s_* \leq 1$, so we may assume without loss of generality that $s_* > 1$ and hence $s > (k+2)^{500} \geq 2^{500}$.

So let n be let a positive integer, $Q \in \mathbb{R}[x_1, \ldots, x_n]$ be a quadratic polynomial, let $M \in \mathcal{H}^{k \times n}(s)$ and $\vec{w} \in \mathbb{R}^k$. Let us write the quadratic part of $Q(\vec{x})$ as $\vec{x}^{\mathsf{T}} A \vec{x}$ for a symmetric matrix $A \in \mathbb{R}^{n \times n}$, and assume that for every subset $S \subseteq [n]$ with $|S| \geq n - s$, and every (M,M)-perturbation A' of A, the submatrix $A'[S \times S]$ has at least one nonzero entry outside the diagonal (this is condition $(*)$ in Definition 5.4).

As the first step of the proof, we treat the case where the matrix A does not robustly have rank at least 2. In the remaining steps of the proof we can then assume that the matrix A does have rank at least 2 robustly.

Step 1: The low-rank case. Suppose that there is a matrix $B \in \mathbb{R}^{n \times n}$ which can be obtained from an (M,M)-perturbation of A by changing its diagonal entries, and a set S of size $|S| \geq n - 2s/3$, such

that rank $B[S \times S] \leq 2$. Note that then B^T can also be obtained from an (M, M)-perturbation of $A^\mathsf{T} = A$ by changing its diagonal entries, and satisfies rank $B^\mathsf{T}[S \times S] \leq 2$. Consider the symmetric matrix $A^* = \frac{1}{2}(B + B^\mathsf{T})$; note that A^* can be obtained from an (M, M)-perturbation of $\frac{1}{2}(A + A) = A$ by changing its diagonal entries, and we have rank $A^*[S \times S] \leq 4$.

Let $Q^* \in \mathbb{R}[x_1, \ldots, x_n]$ be a quadratic polynomial with quadratic part $\vec{x}^\mathsf{T} A^* \vec{x}$ such that $Q(\vec{\xi}) = Q^*(\vec{\xi})$ for all $\vec{\xi} \in \{-1, 1\}^n$ with $M\vec{\xi} = \vec{w}$ (such a polynomial Q^* exists by Lemma 5.3).

Now, let $T = [n] \setminus S$, and note that it suffices to prove the probability bound in (9.1) conditioned on every possible outcome for $\vec{\xi}[T]$. For any given outcome of $\vec{\xi}[T]$, we can write $Q^*(\vec{\xi})$ as $Q^*_{\vec{\xi}[T]}(\vec{\xi}[S])$, for some quadratic polynomial $Q^*_{\vec{\xi}[T]}$ in the variables $\vec{\xi}[i]$ for $i \in S$, with quadratic part $\vec{x}[S]^\mathsf{T} A^*[S \times S] \vec{x}[S]$ (and whose linear and constant coefficients depend on $\vec{\xi}[T]$). Simply put, this polynomial is obtained from Q^* by plugging in the given values of $\vec{\xi}[i]$ for $i \in T$. Recall that we always have $Q(\vec{\xi}) = Q^*(\vec{\xi}) = Q^*_{\vec{\xi}[T]}(\vec{\xi}[S])$ (for any $\vec{\xi} \in \{-1, 1\}^n$ with $M\vec{\xi} = \vec{w}$).

Now, we claim that for any subset $S' \subseteq S$ of size $|S'| \geq |S| - s/3 \geq n - s$, the matrix $A^*[S' \times S']$ cannot be a $(M[[k] \times S'], M[[k] \times S'])$-perturbation of the zero matrix in $\mathbb{R}^{S' \times S'}$. Indeed, if $A^*[S' \times S']$ was a $(M[[k] \times S'], M[[k] \times S'])$-perturbation of the zero matrix, then from the matrix $A[S' \times S']$ one could obtain the zero matrix by taking an $(M[[k] \times S'], M[[k] \times S'])$-perturbation and changing its diagonal entries. But this means that there is some (M, M)-perturbation A' of A such $A'[S' \times S']$ agrees with the zero matrix in $\mathbb{R}^{S' \times S'}$ in all off-diagonal entries. This means that $A'[S' \times S']$ does not have any nonzero entries outside its diagonal, contradicting our assumption on A made above (coming from condition $(*)$ Definition 5.4). This means that for any subset $S' \subseteq S$ of size $|S'| \geq |S| - s/3$, the matrix $A^*[S' \times S']$ is indeed not a $(M[[k] \times S'], M[[k] \times S'])$-perturbation of the zero matrix.

Furthermore $M[[k] \times S] \in \mathcal{H}^{k \times S}(s/3)$ and rank $A^*[S \times S] \leq 4$, so we can apply Proposition 6.1 with $r = 5$ to the matrix $M[[k] \times S]$, the vector $\vec{w} - M[[k] \times T]\vec{\xi}[T] \in \mathbb{R}^k$, and the quadratic polynomial $Q^*_{\vec{\xi}[T]}$ in the variables $\vec{\xi}[i]$ for $i \in S$. The conclusion of the proposition then gives

$$\Pr\Big[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w} \,\Big|\, \vec{\xi}[T]\Big] = \Pr\Big[Q^*_{\vec{\xi}[T]}(\vec{\xi}[S]) = 0 \text{ and } M[[k] \times S]\vec{\xi}[S] = \vec{w} - M[[k] \times T]\vec{\xi}[T] \,\Big|\, \vec{\xi}[T]\Big]$$

$$\leq \left(\frac{\lfloor s/3 \rfloor}{2^{75}(k+5)^2}\right)^{-(k+1)/2} \leq s_*^{-(k+1)/2}.$$

for every possible outcome of $\vec{\xi}[T]$. Hence $\Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}] \leq s_*^{-(k+1)/2}$, which in particular proves (9.1).

Step 2: Decoupling. From now on, we may assume that rank $B[S \times S] \geq 3$ for any matrix $B \in \mathbb{R}^{n \times n}$ which can be obtained from an (M, M)-perturbation of A by changing its diagonal entries, and any set S of size $|S| \geq n - 2s/3$. This means in particular that the assumption in Lemma 5.8 is satisfied for $s/2 \geq 4k + 8$, and therefore Lemma 5.8 gives us a partition $[n] = I \cup J$ with $|I| \leq s/2$, such that

$$\Big(M[[k] \times I], M[[k] \times J], A[J \times I]\Big) \in \mathcal{M}_2^{k,I,J}\Big(\lfloor s/(8k+16) \rfloor\Big).$$

Recalling the second part of Remark 5.1, this implies

$$\Big(M[[k] \times I], M[[k] \times J], -2A[J \times I]\Big) \in \mathcal{M}_2^{k,I,J}\Big(\lfloor s/(8k+16) \rfloor\Big) \subseteq \mathcal{M}_2^{k,I,J}\Big(6 \cdot \lfloor s/(48k+96) \rfloor\Big) \qquad (9.2)$$

(which in particular means that $6 \cdot \lfloor s/(48k+96) \rfloor \leq |I| \leq |J|$).

Let $\vec{\xi}'[I]$ be an independent copy of $\vec{\xi}[I]$ (i.e., consider independent random variables $\vec{\xi}'[i] \in \{-1, 1\}$ for $i \in I$, independent from all entries of $\vec{\xi}$). Now, let us extend $\vec{\xi}'[I]$ to a vector $\vec{\xi}' \in \mathbb{R}^n$ by defining $\vec{\xi}'[j] = \vec{\xi}[j]$ for all $j \in J$. Note that then $\vec{\xi}'$ consists of the data $(\vec{\xi}'[I], \vec{\xi}[J])$, while $\vec{\xi}$ consists of the data $(\vec{\xi}[I], \vec{\xi}[J])$. By decoupling (Lemma 3.6) we have

$$\Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}]^2$$

$$\leq \Pr[Q(\vec{\xi}) = Q(\vec{\xi}') = 0 \text{ and } M\vec{\xi} = M\vec{\xi}' = \vec{w}]$$
$$= \Pr[Q(\vec{\xi}) = 0 \text{ and } Q(\vec{\xi}) - Q(\vec{\xi}') = 0 \text{ and } M\vec{\xi} = \vec{w} \text{ and } M(\vec{\xi} - \vec{\xi}') = \vec{0}]$$
$$= \Pr\Big[Q(\vec{\xi}) = 0 \text{ and } Q(\vec{\xi}) - Q(\vec{\xi}') = 0 \text{ and } M\vec{\xi} = \vec{w} \text{ and } M[[k] \times I](\vec{\xi}[I] - \vec{\xi}'[I]) = \vec{0}\Big]. \tag{9.3}$$

Note that the event $M[[k] \times I](\vec{\xi}[I] - \vec{\xi}'[I]) = \vec{0}$ depends only on the outcomes of $\vec{\xi}[I]$ and $\vec{\xi}'[I]$ (and not on the outcome of $\vec{\xi}[J]$). We will later bound the probability of this event using Corollary 3.5. For fixed outcomes of $\vec{\xi}[I]$ and $\vec{\xi}'[I]$, we can express the conditions $Q(\vec{\xi}) = 0$ and $Q(\vec{\xi}) - Q(\vec{\xi}') = 0$ and $M\vec{\xi} = \vec{w}$ in terms of $\vec{\xi}[J]$; namely we can write them in the form $Q_{\vec{\xi}[I]}(\vec{\xi}[J]) = 0$ and $M_{\vec{\xi}[I],\vec{\xi}'[I]}\vec{\xi}[J] = \vec{w}_{\vec{\xi}[I],\vec{\xi}'[I]}$ for some quadratic polynomial $Q_{\vec{\xi}[I]}$ and some matrix $M_{\vec{\xi}[I],\vec{\xi}'[I]}$ (which depend on the outcomes of $\vec{\xi}[I]$ and $\vec{\xi}'[I]$).

More specifically, for a given outcome of $\vec{\xi}[I]$, the event

$$\{Q(\vec{\xi}) = 0\} \quad \text{can be written as} \quad \{Q_{\vec{\xi}[I]}(\vec{\xi}[J]) = 0\}, \tag{9.4}$$

where $Q_{\vec{\xi}[I]}$ is the quadratic polynomial in the variables x_j for $j \in J$ obtained from Q by replacing each variable x_i for $i \in I = [n] \setminus J$ by the given value of $\vec{\xi}[i]$. Then, by definition, we have $Q_{\vec{\xi}[I]}(\vec{\xi}[J]) = Q(\vec{\xi})$ for any outcome of $\vec{\xi}[J]$. Also note that the quadratic part of the polynomial $Q_{\vec{\xi}[I]}$ is given by $\vec{x}[J]^\mathsf{T} A[J \times J]\vec{x}[J]$.

Furthermore, for given outcomes of $\vec{\xi}[I]$ and $\vec{\xi}'[I]$, the event

$$\{M\vec{\xi} = \vec{w} \text{ and } Q(\vec{\xi}) - Q(\vec{\xi}') = 0\} \quad \text{can be written as} \quad \{M_{\vec{\xi}[I],\vec{\xi}'[I]}\vec{\xi}[J] = \vec{w}_{\vec{\xi}[I],\vec{\xi}'[I]}\} \tag{9.5}$$

where the matrix $M_{\vec{\xi}[I],\vec{\xi}'[I]} \in \mathbb{R}^{(k+1) \times J}$ and the vector $\vec{w}_{\vec{\xi}[I],\vec{\xi}'[I]} \in \mathbb{R}^{k+1}$ are defined as follows: When plugging in the given values for $\vec{\xi}[i]$ for $i \in I$, we can interpret $M\vec{\xi} = \vec{w}$ as a system of linear equations in the variables $\vec{\xi}[j]$ for $j \in J = [n] \setminus I$. This system of linear equations has the form $M[[k] \times J]\vec{\xi}[J] = \vec{w} - M[[k] \times I](\vec{\xi}[I] - \vec{\xi}'[I])$. Furthermore, when plugging the given values for $\vec{\xi}[i]$ and $\vec{\xi}'[i]$ for $i \in I$ into $Q(\vec{\xi}) - Q(\vec{\xi}') = 0$, we obtain another linear equation in the variables $\vec{\xi}[j]$ for $j \in J$ (indeed, the quadratic terms in the variables $\vec{\xi}[j]$ for $j \in J$ cancel out in the difference $Q(\vec{\xi}) - Q(\vec{\xi}')$). The coefficient vector of this linear equation is precisely $2A[J \times I](\vec{\xi}[I] - \vec{\xi}'[I])$. Appending this additional linear equation to our previous system of linear equations in the variables $\vec{\xi}[j]$ for $j \in J$, we obtain a system of $k + 1$ equations that we can express in the form $M_{\vec{\xi}[I],\vec{\xi}'[I]}\vec{\xi}[J] = \vec{w}_{\vec{\xi}[I],\vec{\xi}'[I]}$ for a matrix $M_{\vec{\xi}[I],\vec{\xi}'[I]} \in \mathbb{R}^{(k+1) \times J}$ and a vector $\vec{w}_{\vec{\xi}[I],\vec{\xi}'[I]} \in \mathbb{R}^{k+1}$. Note that the first k rows of the matrix $M_{\vec{\xi}[I],\vec{\xi}'[I]} \in \mathbb{R}^{(k+1) \times J}$ are given by $M[[k] \times J]$, and the last row is given by (the transpose of) the vector $2A[J \times I](\vec{\xi}[I] - \vec{\xi}'[I])$.

For given outcomes of $\vec{\xi}[I]$ and $\vec{\xi}'[I]$, we would like to bound the probability of having $Q_{\vec{\xi}[I]}(\vec{\xi}[J]) = 0$ and $M_{\vec{\xi}[I],\vec{\xi}'[I]}\vec{\xi}[J] = \vec{w}_{\vec{\xi}[I],\vec{\xi}'[I]}$ (subject to the randomness of $\vec{\xi}[J]$). If $M_{\vec{\xi}[I],\vec{\xi}'[I]} \in \mathcal{H}^{(k+1) \times J}(s')$ for some suitable s', then we will be able to bound this probability by $f(k + 1, s')$; see Definition 5.4. In the next step, we first handle the case where $M_{\vec{\xi}[I],\vec{\xi}'[I]} \notin \mathcal{H}^{(k+1) \times J}(s')$.

Step 3: Failure of the Halász condition. We define $s' = \lfloor s/(48k + 96) \rfloor$, then $k \leq 6s' \leq |I|$ (recalling (9.2) and $s > (k + 2)^{500}$). Our goal in this step is to bound the probability that the outcomes of $\vec{\xi}[I]$ and $\vec{\xi}'[I]$ are such that $M[[k] \times I](\vec{\xi}[I] - \vec{\xi}'[I]) = \vec{0}$ and $M_{\vec{\xi}[I],\vec{\xi}'[I]} \notin \mathcal{H}^{(k+1) \times J}(s')$. To do so, let us condition on an arbitrary outcome of $\vec{\xi}[I]$. Let us apply Lemma 5.7 with $0 \leq k \leq 6s' \leq |I| \leq |J|$, the random vector $\vec{\xi}'[I]$, the matrices $T = M[[k] \times I]$, $U = M[[k] \times J]$ and $-2A[J \times I]$ (noting that then $(T, U, -2A[J \times I]) \in \mathcal{M}_2^{k,I,J}(6s')$ by (9.2)), as well as the vectors $\vec{y} = M[[k] \times I]\vec{\xi}[I]$ and $\vec{b} = -2A[J \times I]\vec{\xi}[I]$. Note that then the random matrix $U'_{\vec{\xi}}$ in Lemma 5.7 is precisely the matrix $M_{\vec{\xi}[I],\vec{\xi}'[I]}$ (namely, the matrix obtained from $U = M[[k] \times J]$ by appending the vector $-2A[J \times I]\vec{\xi}'[I] - \vec{b} = 2A[J \times I](\vec{\xi}[I] - \vec{\xi}'[I])$ as an additional row). So, we obtain

$$\Pr\Big[M[[k] \times I](\vec{\xi}[I] - \vec{\xi}'[I]) = \vec{0} \text{ and } M_{\vec{\xi}[I],\vec{\xi}'[I]} \notin \mathcal{H}^{(k+1) \times J}(s') \,\Big|\, \vec{\xi}[I]\Big] \leq \left(\frac{6s'}{10^{61}(k + 2)^{20}}\right)^{-(k+2)/2} \leq s_*^{-(k+2)/2} \tag{9.6}$$

32

for any outcome of $\vec{\xi}[I]$ (here, the probability is with respect to the randomness of $\vec{\xi}'[I]$). In the last step, we used that $s' = \lfloor s/(48k + 96)\rfloor$ and therefore $s'/(10^{61}(k + 2)^{20}) \geq s/(10^{63}(k + 2)^{21}) \geq s/(k + 2)^{273} > s_*$ (since $k + 2 \geq 2$).

Step 4: Bounding the main term. Recall from Step 2 that $\operatorname{rank} B[S \times S] \geq 3$ for any matrix $B \in \mathbb{R}^{n \times n}$ which can be obtained from an (M, M)-perturbation of A by changing its diagonal entries, and any subset $S \subseteq J$ of size $|S| \geq |J| - s'$ (here we are using that $|J| - s' \geq n - 2s/3$ as $s' \leq s/6$ and $|J| = n - |I| \geq n - s/2$). Recalling that the matrix $M_{\vec{\xi}[I],\vec{\xi}'[I]}$ is obtained from $M[[k] \times J]$ by adding one row, this implies that $\operatorname{rank} B[S \times S] \geq 1$ for any matrix $B \in \mathbb{R}^{J \times J}$ obtained from an $(M_{\vec{\xi}[I],\vec{\xi}'[I]}, M_{\vec{\xi}[I],\vec{\xi}'[I]})$-perturbation of $A[J \times J]$ by changing its diagonal entries, and any subset $S \subseteq J$ of size $|S| \geq |J| - s'$. Thus, $A'[S \times S]$ must have at least one nonzero entry outside the diagonal for every subset $S \subseteq J$ of size $|S| \geq |J| - s'$ and every $(M_{\vec{\xi}[I],\vec{\xi}'[I]}, M_{\vec{\xi}[I],\vec{\xi}'[I]})$-perturbation A' of $A[J \times J]$.

This means that whenever we have $M_{\vec{\xi}[I],\vec{\xi}'[I]} \in \mathcal{H}^{(k+1) \times J}(s')$, the quadruple $(|J|, Q_{\vec{\xi}[I]}, M_{\vec{\xi}[I],\vec{\xi}'[I]}, \vec{w}_{\vec{\xi}[I],\vec{\xi}'[I]})$ satisfies the conditions in Definition 5.4, with parameters $k + 1$ and s' (recall that $Q_{\vec{\xi}[I]}$ has quadratic part $\vec{x}[J]^\mathsf{T} A[J \times J]\vec{x}[J]$). Hence, for any outcomes of $\vec{\xi}[I]$ and $\vec{\xi}'[I]$ such that $M_{\vec{\xi}[I],\vec{\xi}'[I]} \in \mathcal{H}^{(k+1) \times J}(s')$, we have the conditional probability bound

$$\Pr\left[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w} \text{ and } Q(\vec{\xi}) - Q(\vec{\xi}') = 0 \,\middle|\, \vec{\xi}[I], \vec{\xi}'[I]\right]$$

$$= \Pr\left[Q_{\vec{\xi}[I]}(\vec{\xi}[J]) = 0 \text{ and } M_{\vec{\xi}[I],\vec{\xi}'[I]}\vec{\xi}[J] = \vec{w}_{\vec{\xi}[I],\vec{\xi}'[I]} \,\middle|\, \vec{\xi}[I], \vec{\xi}'[I]\right] \leq f(k + 1, s') \leq f(k + 1, s_*) \quad (9.7)$$

(for the first step, recall (9.4) and (9.5), and for the last step note that $s' = \lfloor s/(48k + 96)\rfloor \geq s/(k + 2)^{500} = s_*$).

Step 5: Concluding. For ease of notation, let us abbreviate $\mathcal{H}^{(k+1) \times J}(s')$ by $\mathcal{H}$. By (9.3) we have

$$\Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}]^2$$

$$\leq \Pr\left[Q(\vec{\xi}) = 0 \text{ and } Q(\vec{\xi}) - Q(\vec{\xi}') = 0 \text{ and } M\vec{\xi} = \vec{w} \text{ and } M[[k] \times I](\vec{\xi}[I] - \vec{\xi}'[I]) = \vec{0}\right]$$

$$\leq \Pr\left[Q(\vec{\xi}) = 0 \text{ and } Q(\vec{\xi}) - Q(\vec{\xi}') = 0 \text{ and } M\vec{\xi} = \vec{w} \text{ and } M[[k] \times I](\vec{\xi}[I] - \vec{\xi}'[I]) = \vec{0} \text{ and } M_{\vec{\xi}[I],\vec{\xi}'[I]} \in \mathcal{H}\right]$$

$$+ [Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w} \text{ and } M[[k] \times I](\vec{\xi}[I] - \vec{\xi}'[I]) = \vec{0} \text{ and } M_{\vec{\xi}[I],\vec{\xi}'[I]} \notin \mathcal{H}]$$

$$\leq \Pr\left[M[[k] \times I](\vec{\xi}[I] - \vec{\xi}'[I]) = \vec{0}\right] \sup_{\substack{\vec{\xi}[I],\vec{\xi}'[I] \\ M_{\vec{\xi}[I],\vec{\xi}'[I]} \in \mathcal{H}}} \Pr\left[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w} \text{ and } Q(\vec{\xi}) - Q(\vec{\xi}') = 0 \,\middle|\, \vec{\xi}[I], \vec{\xi}'[I]\right]$$

$$+ \Pr\left[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}\right] \sup_{\vec{\xi}[I]} \Pr\left[M[[k] \times I](\vec{\xi}[I] - \vec{\xi}'[I]) = \vec{0} \text{ and } M_{\vec{\xi}[I],\vec{\xi}'[I]} \notin \mathcal{H} \,\middle|\, \vec{\xi}[I]\right].$$

Note that $M[[k] \times I] \in \mathcal{H}^{k \times I}(6s')$ by (9.2) and Remark 5.1. Therefore, by Corollary 3.5 (applied with $\vec{w} = M[[k] \times I]\vec{\xi}'[I]$ for any fixed outcome of $\vec{\xi}'[I]$) we have $\Pr\left[M[[k] \times I](\vec{\xi}[I] - \vec{\xi}'[I]) = \vec{0}\right] \leq (6s'/k)^{-k/2} \leq s_*^{-k/2}$ if $k \geq 1$ (recalling that $s' = \lfloor s/(48k + 96)\rfloor$ and $s_* = s/(k + 2)^{500}$). If $k = 0$, then we trivially have $\Pr\left[M[[k] \times I](\vec{\xi}[I] - \vec{\xi}'[I]) = \vec{0}\right] \leq s_*^{-k/2}$, so this inequality holds in either case. Plugging this observation, as well as (9.7) and (9.6) into the above chain of inequalities, we obtain

$$\Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}]^2 \leq s_*^{-k/2} \cdot f(k + 1, s_*) + \Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}] \cdot s_*^{-(k+2)/2}$$

Applying Lemma 3.8, this yields

$$\Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}] \leq s_*^{-(k+2)/2} + s_*^{-k/4} \cdot f(k + 1, s_*)^{1/2},$$

implying the desired bound in (9.1). $\qquad\square$

10 Deducing a non-recursive bound

In this section, we prove Theorem 5.1. First, from the recursive bound for $f(k, s)$ in Theorem 5.5, one can deduce the following non-recursive bound.

Corollary 10.1. *For any integers $0 \leq k \leq \ell$ and any real $s > 0$ we have*

$$f(k, s) \leq (s_{k,\ell})^{-\ell/2^{\ell-k+1}} \prod_{j=k}^{\ell-1} (s_{k,j})^{-j/2^{j-k+2}} + \sum_{i=k}^{\ell-1} (s_{k,i})^{-(i+2)/2^{i-k+1}} \prod_{j=k}^{i-1} (s_{k,j})^{-j/2^{j-k+2}}, \tag{10.1}$$

where for $i = k, \ldots, \ell$ we define

$$s_{k,i} = \frac{s}{(i+2)^{500(i-k+1)}}.$$

In order to analyse the bound on the left-hand side of (10.1), we make the following simple observation.

Lemma 10.2. *For any integers $0 \leq k \leq i$, we have*

$$\sum_{j=k}^{i-1} \frac{j}{2^{j-k+2}} = \frac{k+1}{2} - \frac{i+1}{2^{i-k+1}}.$$

Proof. The identity can easily be shown by induction on $i - k$. The base case $i - k = 0$ (i.e., $i = k$) is trivial, as both sides are zero. For the induction step, with $i - k \geq 1$, we compute

$$\sum_{j=k}^{i-1} \frac{j}{2^{j-k+2}} = \sum_{j=k}^{i-2} \frac{j}{2^{j-k+2}} + \frac{i-1}{2^{i-k+1}} = \frac{k+1}{2} - \frac{i}{2^{i-k}} + \frac{i-1}{2^{i-k+1}} = \frac{k+1}{2} - \frac{i+1}{2^{i-k+1}}. \qquad \square$$

Using the above observation, let us now deduce Corollary 10.1 from Theorem 5.5 by induction on $\ell - k$ and some tedious but straightforward calculations.

Proof of Corollary 10.1. We prove the desired bound by induction on $\ell - k$. If $\ell - k = 0$ (i.e., if $\ell = k$), we need to show that $f(k, s) \leq (s_{k,k})^{-k/2}$. If $k = 0$, this is trivially true as $f(k, s) \leq 1 = (s_{0,0})^{-0/2}$. To check $f(k, s) \leq (s_{k,k})^{-k/2}$ for $k \geq 1$, recall that for every quadruple $(n, Q, M, \vec{w})$ in the supremum in the definition of $f(k, s)$ in Definition 5.4, we have $M \in \mathcal{H}^{k \times n}(s)$ and therefore by Corollary 3.5

$$\Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}] \leq \Pr[M\vec{\xi} = \vec{w}] \leq (s/k)^{-k/2} \leq \left(\frac{s}{(k+2)^{500}} \right)^{-k/2} = (s_{k,k})^{-k/2}.$$

This shows $f(k, s) \leq (s_{k,k})^{-k/2}$, as desired.

Let us now assume that $\ell - k \geq 1$ and that we already proved the desired bound for all smaller values of $\ell - k$. Note that now by Theorem 5.5 we have

$$f(k, s) \leq \max\left\{ s_*^{-(k+1)/2}, \quad s_*^{-(k+2)/2} + s_*^{-k/4} \cdot f(k+1, s_*)^{1/2} \right\},$$

with $s_* = s/(k+2)^{500}$. So it suffices to show that both terms in this maximum are bounded by the left-hand side of (10.1).

First, note that $s_* = s/(k+2)^{500} \geq s_{k,i}$ for all $i = k, \ldots, \ell$. This means that in the case $s_* \leq 1$, we also have $s_{k,i} \leq 1$ for all $i = k, \ldots, \ell$ and hence (10.1) follows trivially from $f(k, s) \leq 1$. We may therefore assume that $s_* > 1$.

For the first term in the maximum above, note that by Lemma 10.2 applied to $i = \ell$, we have

$$\frac{\ell}{2^{\ell-k+1}} + \sum_{j=k}^{\ell-1} \frac{j}{2^{j-k+2}} = \frac{\ell}{2^{\ell-k+1}} + \frac{k+1}{2} - \frac{\ell+1}{2^{\ell-k+1}} = \frac{k+1}{2} - \frac{1}{2^{\ell-k+1}} < \frac{k+1}{2}$$

Hence

$$s_*^{-(k+1)/2} \leq (s_*)^{-\ell/2^{\ell-k+1}} \prod_{j=k}^{\ell-1} (s_*)^{-j/2^{j-k+2}} \leq (s_{k,\ell})^{-\ell/2^{\ell-k+1}} \prod_{j=k}^{\ell-1} (s_{k,j})^{-j/2^{j-k+2}},$$

34

which in particular shows that $s_*^{-(k+1)/2}$ is bounded by the left-hand side of (10.1).

To bound the second term, first note that by the inductive assumption we have

$$f(k+1, s_*) \le (s'_{k+1,\ell})^{-\ell/2^{\ell-k}} \prod_{j=k+1}^{\ell-1} (s'_{k+1,j})^{-j/2^{j-k+1}} + \sum_{i=k+1}^{\ell-1} (s'_{k+1,i})^{-(i+2)/2^{i-k}} \prod_{j=k+1}^{i-1} (s'_{k+1,j})^{-j/2^{j-k+1}},$$

defining

$$s'_{k+1,i} = \frac{s_*}{(i+2)^{500(i-k)}} = \frac{s}{(i+2)^{500(i-k)} \cdot (k+2)^{500}}.$$

for $i = k+1, \ldots, \ell$. Note that we have $s'_{k+1,i} \ge s_{k,i}$ for all $i = k+1, \ldots, \ell$, implying

$$f(k+1, s_*) \le (s_{k,\ell})^{-\ell/2^{\ell-k}} \prod_{j=k+1}^{\ell-1} (s_{k,j})^{-j/2^{j-k+1}} + \sum_{i=k+1}^{\ell-1} (s_{k,i})^{-(i+2)/2^{i-k}} \prod_{j=k+1}^{i-1} (s_{k,j})^{-j/2^{j-k+1}}.$$

As $\sqrt{x+y} \le \sqrt{x} + \sqrt{y}$ for all $x, y \ge 0$, this implies

$$f(k+1, s_*)^{1/2} \le (s_{k,\ell})^{-\ell/2^{\ell-k+1}} \prod_{j=k+1}^{\ell-1} (s_{k,j})^{-j/2^{j-k+2}} + \sum_{i=k+1}^{\ell-1} (s_{k,i})^{-(i+2)/2^{i-k+1}} \prod_{j=k+1}^{i-1} (s_{k,j})^{-j/2^{j-k+2}}.$$

Also using that $s_{k,k} = s_*$, we now obtain

$$s_*^{-(k+2)/2} + s_*^{-k/4} \cdot f(k+1, s_*)^{1/2}$$

$$\le (s_{k,k})^{-(k+2)/2} + (s_{k,\ell})^{-\ell/2^{\ell-k+1}} (s_{k,k})^{-k/4} \prod_{j=k+1}^{\ell-1} (s_{k,j})^{-j/2^{j-k+2}}$$

$$+ \sum_{i=k+1}^{\ell-1} (s_{k,i})^{-(i+2)/2^{i-k+1}} (s_{k,k})^{-k/4} \prod_{j=k+1}^{i-1} (s_{k,j})^{-j/2^{j-k+2}}$$

$$= (s_{k,k})^{-(k+2)/2} + (s_{k,\ell})^{-\ell/2^{\ell-k+1}} \prod_{j=k}^{\ell-1} (s_{k,j})^{-j/2^{j-k+2}} + \sum_{i=k+1}^{\ell-1} (s_{k,i})^{-(i+2)/2^{i-k+1}} \prod_{j=k}^{i-1} (s_{k,j})^{-j/2^{j-k+2}}$$

$$= (s_{k,\ell})^{-\ell/2^{\ell-k+1}} \prod_{j=k}^{\ell-1} (s_{k,j})^{-j/2^{j-k+2}} + \sum_{i=k}^{\ell-1} (s_{k,i})^{-(i+2)/2^{i-k+1}} \prod_{j=k}^{i-1} (s_{k,j})^{-j/2^{j-k+2}}.$$

Hence the second term in the maximum above also satisfies the desired bound, completing the proof. $\qquad\square$

Finally, we show how to deduce Theorem 5.1 from the $k = 0$ case of Corollary 10.1.

Proof of Theorem 5.1. Let $Q \in \mathbb{R}[x_1, \ldots, x_n]$, $A \in \mathbb{R}^{n \times n}$ and s be as in Theorem 5.1. We may assume that $s \ge 4$ (otherwise the bound in Theorem 5.1 holds trivially as long as $C' \ge 2$). Let $k = 0$, let $M \in \mathcal{H}^{0 \times n}(s)$ be the empty $0 \times n$ matrix, and let $\vec{w} \in \mathbb{R}^0$ be the empty vector. Note that the quadruple $(n, Q, M, \vec{w})$ satisfies condition $(*)$ in Definition 5.4, because the only (M, M)-perturbation of A is the matrix $A' = A$, and by the assumption in Theorem 5.1, for every subset $S \subseteq [n]$ of size $|S| \ge n - s$, the submatrix $A[S \times S]$ has at least one nonzero entry outside its diagonal. Thus, for a sequence $\vec{\xi} \in \{-1, 1\}^n$ of independent Rademacher random variables, we have

$$\Pr[Q(\vec{\xi}) = 0] = \Pr[Q(\vec{\xi}) = 0 \text{ and } M\vec{\xi} = \vec{w}] \le f(0, s)$$

and hence by Corollary 10.1

$$\Pr[Q(\vec{\xi}) = 0] \le (s_{0,\ell})^{-\ell/2^{\ell+1}} \prod_{j=0}^{\ell-1} (s_{0,j})^{-j/2^{j+2}} + \sum_{i=0}^{\ell-1} (s_{0,i})^{-(i+2)/2^{i+1}} \prod_{j=0}^{i-1} (s_{0,j})^{-j/2^{j+2}}$$

for every integer $\ell \geq 0$ (with $s_{0,i} = s/(i+2)^{500(i+1)}$ for $i = 0, \ldots, \ell$ as defined in Corollary 10.1). Then for every integer $\ell \geq 0$ we obtain

$$\Pr[Q(\vec{\xi}) = 0] \leq \prod_{i=0}^{\ell} (i+2)^{500(i+1)(i+2)/2^{i+1}} \cdot \left(s^{-\ell/2^{\ell+1}} \prod_{j=0}^{\ell-1} s^{-j/2^{j+2}} + \sum_{i=0}^{\ell-1} s^{-(i+2)/2^{i+1}} \prod_{j=0}^{i-1} s^{-j/2^{j+2}} \right)$$

$$= \exp\left(500 \sum_{i=0}^{\ell} \frac{(i+2)^2 \ln(i+2)}{2^{i+1}} \right) \cdot \left(s^{-1/2 + 1/2^{\ell+1}} + \sum_{i=0}^{\ell-1} s^{-1/2 - 1/2^{i+1}} \right)$$

using Lemma 10.2 for $k = 0$. Noting that the series $\sum_{i=0}^{\infty} (i+2)^2 \ln(i+2)/2^{i+1} = \sum_{i=2}^{\infty} i^2 \ln i/2^{i-1}$ converges, for all integers $\ell \geq 0$ we obtain

$$\Pr[Q(\vec{\xi}) = 0] \leq C_1 \cdot \left(s^{-1/2 + 1/2^{\ell+1}} + \sum_{i=0}^{\ell-1} s^{-1/2 - 1/2^{i+1}} \right) = C_1 \cdot s^{-1/2} \cdot \left(s^{1/2^{\ell+1}} + \sum_{i=0}^{\ell-1} s^{-2^{\ell-i}/2^{\ell+1}} \right)$$

for some absolute constant $C_1 \geq 1$. Plugging in $\ell = \lfloor \log \log s \rfloor - 1$, we have

$$2 \leq s^{1/2^{\ell+1}} \leq 4$$

and therefore

$$\Pr[Q(\vec{\xi}) = 0] \leq C_1 \cdot s^{-1/2} \cdot \left(4 + \sum_{i=0}^{\ell-1} (1/2)^{2^{\ell-i}} \right) \leq \frac{C_1}{\sqrt{s}} \cdot \left(4 + \sum_{i=0}^{\ell-1} (1/2)^{\ell-i} \right) \leq \frac{5C_1}{\sqrt{s}}.$$

Setting $C' = 5C_1$, this gives the desired bound. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

11   Deduction of main results

In this section, we deduce our main theorem (Theorem 1.1) and its generalisation to arbitrary distributions (Theorem 1.2) from the slightly more technical statement of Theorem 5.1. Note that Theorem 1.2 directly implies Theorem 1.1, taking $\xi_1, \ldots, \xi_n$ to be independent Rademacher random variables, and taking $\delta = 1/2$. So, we will just prove Theorem 1.2 (however, we remark that the proof gets easier and certain steps can be skipped if one is only interested in Theorem 1.1).

First, to apply Theorem 5.1 in the general setting of Theorem 1.2, we use that any discrete random variable can be expressed in a way such that after some further conditioning one basically obtains a Rademacher random variable. The following lemma strengthens an observation made by Meka, Nguyen and Vu [37] for a similar purpose.

Lemma 11.1. *For every discrete random variable $\zeta \in \mathbb{R}$, we can find a representation of the form*

$$\zeta = \alpha + \xi\beta$$

for a Rademacher random variable $\xi \in \{-1, 1\}$ and a discrete random vector $(\alpha, \beta) \in \mathbb{R}^2$ which is independent of ξ. Moreover, this representation can be chosen such that the distribution of (α, β) satisfies the following two conditions:

(a) There is at most one real number $a \in \mathbb{R}$ with $\Pr[(\alpha, \beta) = (a, 0)] > 0$
(b) If there is some value $z \in \mathbb{R}$ with $\Pr[\zeta = z] > 1/2$, then we always have $\alpha + \beta = z$ (for any outcome of the random vector $(\alpha, \beta) \in \mathbb{R}^2$).

Proof. If the random variable ζ is constant, i.e., if $\Pr[\zeta = z] = 1$ for some $z \in \mathbb{R}$, we can define the random vector $(\alpha, \beta) \in \mathbb{R}^2$ to always take the constant value $(z, 0)$. So let us from now on assume that ζ is not constant.

Case 1: There is a majority outcome. First, we consider the case where $\Pr[\zeta = z] \geq 1/2$ for some $z \in \mathbb{R}$. Let $\rho = \Pr[\zeta \neq z] \leq 1/2$. Then, we can take the random vector $(\alpha, \beta) \in \mathbb{R}^2$ to be equal to $(z, 0)$ with probability $1 - 2\rho$ and equal to $((z + Y)/2, (z - Y)/2)$ with probability 2ρ, where Y is a sample from the distribution of ζ conditioned on the event $\zeta \neq z$. Note that we then always have $\alpha + \beta = z$ (as $\alpha + \beta = z + 0 = z$ or $\alpha + \beta = (z + Y)/2 + (z - Y)/2 = z$). Also note for all $a \in \mathbb{R}$ with $a \neq z$ we have $\Pr[(\alpha, \beta) = (a, 0)] = 0$, since Y never takes the value z and so we can only have $\beta = 0$ when $\alpha = z$. Now, taking $\xi \in \{-1, 1\}$ to be a Rademacher random variable that is independent of the random vector (α, β), the expression $\alpha + \xi\beta$ evaluates to $z \pm 0 = z$ with probability $1 - 2\rho$, to $(z + Y)/2 + (z - Y)/2 = z$ with probability ρ and to $(z + Y)/2 - (z - Y)/2 = Y$ with probability ρ. Hence the distribution of $\alpha + \xi\beta$ coincides with the distribution of ζ. We can therefore find a coupling of ζ with (α, β) and ξ such that $(\alpha, \beta) \in \mathbb{R}^2$ and $\xi \in \{-1, 1\}$ are independent and $\zeta = \alpha + \xi\beta$.

Case 2: There is no majority outcome. Now, let us consider the case where $\Pr[\zeta = z] < 1/2$ for all $z \in \mathbb{R}$. Let x be a median of ζ (meaning that $\Pr[\zeta \geq x] \geq 1/2$ and $\Pr[\zeta \leq x] \geq 1/2$; for example we can take $x = \sup\{x \in \mathbb{R} : \Pr[\zeta \geq x] \geq 1/2\}$). As $\Pr[\zeta = x] < 1/2$, we can see that $0 < \Pr[\zeta < x] \leq 1/2$ and $0 < \Pr[\zeta > x] \leq 1/2$.

Now, let $\rho_1 = \Pr[\zeta < x]$ and $\rho_2 = \Pr[\zeta > x]$. Let us assume without loss of generality that $\rho_1 \geq \rho_2$ (the case $\rho_2 \geq \rho_1$ is analogous). Let Y_1 (respectively, Y_2) be a sample from the distribution of ζ conditioned on the event $\zeta < x$ (respectively, the event $\zeta > x$). We can now take the random vector $(\alpha, \beta) \in \mathbb{R}^2$ to be equal to $(x, 0)$ with probability $1 - 2\rho_1$, equal to $((x + Y_1)/2, (x - Y_1)/2)$ with probability $2\rho_1 - 2\rho_2$, and equal to $((Y_2 + Y_1)/2, (Y_2 - Y_1)/2)$ with probability $2\rho_2$. Note that we can only have $\beta = 0$ when $\alpha = x$, since we always have $Y_1 < x$ and $Y_2 > x$. Now, taking $\xi \in \{-1, 1\}$ to be a Rademacher random variable that is independent of the random vector $(\alpha, \beta) \in \mathbb{R}^2$, the expression $\alpha + \xi\beta$ evaluates to $x \pm 0 = x$ with probability $1 - 2\rho_1$, to $(x + Y_1)/2 + (x - Y_1)/2 = x$ with probability $\rho_1 - \rho_2$, to $(x + Y_1)/2 - (x - Y_1)/2 = Y_1$ with probability $\rho_1 - \rho_2$, to $(Y_2 + Y_1)/2 - (Y_2 - Y_1)/2 = Y_1$ with probability ρ_2, and to $(Y_2 + Y_1)/2 + (Y_2 - Y_1)/2 = Y_2$ with probability ρ_2. So all in all, $\alpha + \xi\beta$ evaluates to x with probability $1 - \rho_1 - \rho_2 = \Pr[\zeta = x]$, to Y_1 with probability ρ_1 and to Y_2 with probability ρ_2. Hence the distribution of $\alpha + \xi\beta$ agrees with the distribution of ζ, and we can again find the desired coupling such that $\zeta = \alpha + \xi\beta$. $\qquad\square$

In the setting of Lemma 11.1, note that if there is a real number $a \in \mathbb{R}$ with $\Pr[(\alpha, \beta) = (a, 0)] > 0$, then we have $\Pr[\beta = 0] = \Pr[(\alpha, \beta) = (a, 0)] \leq \Pr[\zeta = a]$ (here we used (a)). On the other hand, if there is no such $a \in \mathbb{R}$, we clearly have $\Pr[\beta = 0] = 0$. Thus, in any case we can conclude that

$$\Pr[\beta = 0] \leq \sup_{z \in \mathbb{R}} \Pr[\zeta = z]. \tag{11.1}$$

We will also need a generalisation of the Erdős–Littlewood–Offord theorem to arbitrary discrete random variables (i.e., not just for Rademacher random variables). The following theorem follows directly from the result of [30], and can also be deduced from the ordinary Erdős–Littlewood–Offord theorem.

Theorem 11.2. *Fix $\delta > 0$, and let $X_1, \ldots, X_t \in \mathbb{R}$ be independent discrete random variables satisfying $\sup_{z \in \mathbb{R}} \Pr[X_i = z] \leq 1 - \delta$ for all $i = 1, \ldots, t$. Then we have*

$$\sup_{z \in \mathbb{R}} \Pr[X_1 + \cdots + X_t = z] \leq \frac{C'_\delta}{\sqrt{t}},$$

for some constant C'_δ only depending on δ.

Let us first prove Theorem 1.2 under the additional assumption that all point probabilities of the variables $\zeta_1, \ldots, \zeta_n$ are at most $1 - \delta$. This does not require the full strength of Lemma 11.1 (we will not need (b) in Lemma 11.1).

Proposition 11.3. *The statement of Theorem 1.2 holds under the additional assumption that we have $\sup_{z \in \mathbb{R}} \Pr[\zeta_i = z] \leq 1 - \delta$ for all $i = 1, \ldots, n$.*

Proof. For ease of notation, we write $\vec{\zeta} = (\zeta_1, \ldots, \zeta_n)$. Let $C' \geq 1$ be an absolute constant such that Theorem 5.1 holds, and let $C_\delta = \max(8C'/\delta, 2C'_\delta)$ for the constant C'_δ in Theorem 11.2. Note that the claimed probability bound is trivially true if $\sqrt{m} \leq C_\delta$, so we may assume that $m \geq (C_\delta)^2 \geq 64/\delta^2$.

Recall that $Q \in \mathbb{R}[x_1, \ldots, x_n]$ is a quadratic polynomial, and that we are assuming that for any fixing box $R_1 \times \cdots \times R_n$ of Q (where $R_1, \ldots, R_n$ are nonempty subsets of the supports of $\zeta_1, \ldots, \zeta_n$, respectively) there are at least m indices $i \in \{1, \ldots, n\}$ with $\Pr[\zeta_i \in R_i] \leq 1 - \delta$. Note that this assumption is not affected by changing the constant term of Q. It therefore suffices to prove

$$\Pr[Q(\vec{\zeta}) = 0] \leq \frac{C_\delta}{\sqrt{m}}. \tag{11.2}$$

Indeed, for any $z \in \mathbb{R}$, applying this inequality to the polynomial $Q - z$ gives $\Pr[Q(\vec{\zeta}) = z] \leq C_\delta/\sqrt{m}$, as claimed. To prove (11.2), we distinguish two cases.

Case 1: Quadratic anticoncentration. First, consider the case that for some $\ell \geq m/4$ there exist distinct indices $i_1, \ldots, i_\ell, j_1, \ldots, j_\ell \in [n]$ such that for each $h = 1, \ldots, \ell$ the coefficient of $x_{i_h} x_{j_h}$ in the quadratic polynomial Q is nonzero (and let $c_h \neq 0$ denote this coefficient). In this case, our goal is to apply Theorem 5.1.

For each $i = 1, \ldots, n$, let us represent the random variable ζ_i as $\zeta_i = \alpha_i + \xi_i \beta_i$, for a Rademacher random variable $\xi_i \in \{-1, 1\}$ and a random vector $(\alpha_i, \beta_i) \in \mathbb{R}^2$, as in Lemma 11.1, in such a way that $\xi_1, \ldots, \xi_n$ and $(\alpha_1, \beta_1), \ldots, (\alpha_n, \beta_n)$ are all mutually independent. Note that by (11.1) and our assumption in the proposition, we have

$$\Pr[\beta_i = 0] \leq \sup_{z \in \mathbb{R}} \Pr[\zeta_i = z] \leq 1 - \delta$$

for $i = 1, \ldots, n$. Thus, for each $h = 1, \ldots, \ell$, we obtain

$$\Pr[\beta_{i_h} \beta_{j_h} \neq 0] = \Pr[\beta_{i_h} \neq 0] \cdot \Pr[\beta_{i_h} \neq 0] \geq \delta^2,$$

and furthermore the events $\beta_{i_h} \beta_{j_h} \neq 0$ are independent for all $h = 1, \ldots, \ell$. Thus, by a Chernoff bound (see for example [24, Theorem 2.1]), we have

$$\Pr\left[\text{fewer than } \delta^2 \ell/2 \text{ indices } h \in \{1, \ldots, \ell\} \text{ satisfy } \beta_{i_h} \beta_{j_h} \neq 0\right] \leq e^{-\delta^2 \ell/8} \leq e^{-\delta^2 m/32} \leq \frac{32}{\delta^2 m} \leq \frac{4/\delta}{\sqrt{m}}, \tag{11.3}$$

recalling $\ell \geq m/4$ and $m \geq 64/\delta^2$ (i.e., $\delta\sqrt{m} \geq 8$). Conditioning on any outcomes of $(\alpha_1, \beta_1), \ldots, (\alpha_n, \beta_n)$, such that $\beta_{i_h} \beta_{j_h} \neq 0$ for at least $\delta^2 \ell/2$ indices $h \in \{1, \ldots, \ell\}$, we can interpret $Q(\zeta_1, \ldots, \zeta_n)$ as a quadratic polynomial in the independent Rademacher random variables $\xi_1, \ldots, \xi_n$ by plugging in $\zeta_i = \alpha_i + \xi_i \beta_i$ for $i = 1, \ldots, n$. Note that for $h = 1, \ldots, \ell$ the coefficient of $\xi_{i_h} \xi_{j_h}$ in this quadratic polynomial is $c_h \beta_{i_h} \beta_{j_h}$. Hence there are at least $\delta^2 \ell/2 \geq \delta^2 m/8$ indices $h \in \{1, \ldots, \ell\}$ such that the coefficient of $\xi_{i_h} \xi_{j_h}$ in this quadratic polynomial is nonzero, and so this quadratic polynomial satisfies the condition in Theorem 5.1 for any positive integer $s < \delta^2 m/8$ (indeed, for at least $\delta^2 m/8$ indices $h \in \{1, \ldots, \ell\}$ the (i_h, j_h)-entry of the matrix A appearing in this condition is nonzero, and for every subset $S \subseteq [n]$ of size $|S| \geq n - s > n - \delta^2 m/8$, the submatrix $A[S \times S]$ must contain one of these nonzero entries). Taking $s = \lceil \delta^2 m/8 \rceil - 1 \geq \delta^2 m/8 - 1 \geq \delta^2 m/16$ (recalling that $m \geq 64/\delta^2$), by Theorem 5.1 we obtain

$$\Pr\left[Q(\vec{\zeta}) = 0 \,\big|\, (\alpha_1, \beta_1), \ldots, (\alpha_n, \beta_n)\right] \leq \frac{C'}{\sqrt{\delta^2 m/16}} = \frac{4C'/\delta}{\sqrt{m}}$$

when conditioning on any outcomes of $(\alpha_1, \beta_1), \ldots, (\alpha_n, \beta_n)$ such that $\beta_{i_h} \beta_{j_h} \neq 0$ for at least $\delta^2 \ell/2$ indices $h \in \{1, \ldots, \ell\}$. All in all, together with (11.3), this yields

$$\Pr[Q(\vec{\zeta}) = 0] \leq \frac{4C'/\delta}{\sqrt{m}} + \frac{4/\delta}{\sqrt{m}} \leq \frac{8C'/\delta}{\sqrt{m}} \leq \frac{C_\delta}{\sqrt{m}},$$

showing the desired bound (11.2).

Case 2: Linear anticoncentration. From now on we may assume that the condition in Case 1 does not hold. For the maximum possible ℓ, consider distinct indices $i_1, \ldots, i_\ell, j_1, \ldots, j_\ell \in [n]$ such that for $h = 1, \ldots, \ell$ the coefficient of $x_{i_h} x_{j_h}$ in Q is nonzero. By our assumption for this case, we have $\ell < m/4$. Let $J = \{i_1, \ldots, i_\ell, j_1, \ldots, j_\ell\}$ and $I = [n] \setminus J$. Note that then, by the maximality of ℓ, for any distinct $i, i' \in I$, the coefficient of $x_i x_{i'}$ in Q is zero.

Our plan is to condition on an arbitrary outcome of $\vec{\zeta}[J]$, and to apply Theorem 11.2 in the resulting conditional probability space (only using the randomness of ζ_i for $i \in I$).

For any outcome of $\vec{\zeta}[J]$, we can interpret $Q(\vec{\zeta})$ as a polynomial in the remaining variables ζ_i for $i \in I$ (with coefficients depending on $\vec{\zeta}[J]$). Writing $Q_{\vec{\zeta}[J]}$ for this polynomial, we always have $Q(\vec{\zeta}) = Q_{\vec{\zeta}[J]}(\vec{\zeta}[I])$. For any distinct $i, i' \in I$, the coefficient of $\zeta_i \zeta_{i'}$ in $Q_{\vec{\zeta}[J]}$ is zero, so $Q_{\vec{\zeta}[J]}$ can be written as a sum $\sum_{i \in I} P^{(i)}_{\vec{\zeta}[J]}(\zeta_i)$, where for each $i \in I$, the summand $P^{(i)}_{\vec{\zeta}[J]}(\zeta_i)$ is a quadratic polynomial in the single variable ζ_i (with coefficients depending on $\vec{\zeta}[J]$). We now have

$$\Pr\left[Q(\vec{\zeta}) = 0 \,\middle|\, \vec{\zeta}[J]\right] = \Pr\left[Q_{\vec{\zeta}[J]}(\vec{\zeta}[I]) = 0 \,\middle|\, \vec{\zeta}[J]\right] = \Pr\left[\sum_{i \in I} P^{(i)}_{\vec{\zeta}[J]}(\zeta_i) = 0 \,\middle|\, \vec{\zeta}[J]\right]$$

for every outcome of $\vec{\zeta}[J]$.

For any outcome of $\vec{\zeta}[J]$, let $T_{\vec{\zeta}[J]} \subseteq I = [n] \setminus J$ be the set of indices $i \in I$ with $\sup_{z \in \mathbb{R}} \Pr\left[P^{(i)}_{\vec{\zeta}[J]}(\zeta_i) = z \,\middle|\, \vec{\zeta}[J]\right] \leq 1 - \delta$. We claim that we must always have $|T_{\vec{\zeta}[J]}| \geq m/2$, due to the assumption in Theorem 1.2 concerning fixing boxes. Indeed, suppose for the purpose of contradiction that there is an outcome $\vec{w} = (w_j)_{j \in J} \in \mathbb{R}^J$ of $\vec{\zeta}[J]$ such that $|T_{\vec{w}}| < m/2$. Then, we have $|I \setminus T_{\vec{w}}| > n - m$ (recalling that $|I| = n - |J| = n - 2\ell > n - m/2$). We can construct a fixing box $R_1 \times \cdots \times R_n$ for the polynomial Q as follows:

- For $j \in J$, take $R_j = \{w_j\}$;
- For $t \in T_{\vec{w}}$, take $R_t = \{y_t\}$ for some arbitrary element y_t of the support of ζ_t;
- For $i \in I \setminus T_{\vec{w}}$, take $z_i \in \mathbb{R}$ such that $\Pr\left[P^{(i)}_{\vec{w}}(\zeta_i) = z_i\right] > 1 - \delta$ (such a value z_i exists by the definition of $T_{\vec{w}}$), and let R_i be the set of all y in the support of ζ_i such that $P^{(i)}_{\vec{w}}(y) = z_i$ (i.e., $R_i = (P^{(i)}_{\vec{w}})^{-1}(z_i) \cap \mathrm{supp}(\zeta_i)$).

Note that Q is constant on $R_1 \times \cdots \times R_n$: indeed, for any $(\zeta_1, \ldots, \zeta_n) \in R_1 \times \cdots \times R_n$ we have

$$Q(\zeta_1, \ldots, \zeta_n) = Q_{\vec{w}}(\vec{\zeta}[I]) = \sum_{i \in I} P^{(i)}_{\vec{w}}(\zeta_i) = \sum_{t \in T_{\vec{w}}} P^{(t)}_{\vec{w}}(y_t) + \sum_{i \in I \setminus T_{\vec{w}}} z_i.$$

So, $R_1 \times \cdots \times R_n$ is indeed a fixing box of Q. On the other hand we have $\Pr[\zeta_i \in R_i] = \Pr\left[P^{(i)}_{\vec{\zeta}[J]}(\zeta_i) = z_i\right] > 1 - \delta$ for all $i \in I \setminus T_{\vec{w}}$. As $|I \setminus T_{\vec{w}}| > n - m$, this means that there are strictly fewer than m indices $i \in \{1, \ldots, n\}$ with $\Pr[\zeta_i \in R_i] \leq 1 - \delta$, contradicting our assumption.

We have established that $|T_{\vec{\zeta}[J]}| \geq m/2$ for any outcome of $\vec{\zeta}[J]$. For any outcomes of $\vec{\zeta}[J]$ and $\vec{\zeta}[I \setminus T_{\vec{\zeta}[J]}]$, we now have

$$\Pr\left[\sum_{i \in I} P^{(i)}_{\vec{\zeta}[J]}(\zeta_i) = 0 \,\middle|\, \vec{\zeta}[J], \vec{\zeta}[I \setminus T_{\vec{\zeta}[J]}]\right] = \Pr\left[\sum_{i \in T_{\vec{\zeta}[J]}} P^{(i)}_{\vec{\zeta}[J]}(\zeta_i) = -\sum_{i \in I \setminus T_{\vec{\zeta}[J]}} P^{(i)}_{\vec{\zeta}[J]}(\zeta_i) \,\middle|\, \vec{\zeta}[J], \vec{\zeta}[I \setminus T_{\vec{\zeta}[J]}]\right]$$

$$\leq \frac{C'_\delta}{\sqrt{|T_{\vec{\zeta}[J]}|}} \leq \frac{C'_\delta}{\sqrt{m/2}} \leq \frac{2C'_\delta}{\sqrt{m}}$$

by Theorem 11.2. To be precise, for any possible outcomes of $\vec{\zeta}[J]$ and $\vec{\zeta}[I \setminus T_{\vec{\zeta}[J]}]$, we apply Theorem 11.2 in the conditional probability space given these outcomes, with the random variables $X_i = P^{(i)}_{\vec{\zeta}[J]}(\zeta_i)$ for $i \in T_{\vec{\zeta}[J]}$,

noting that by the definition of $T_{\vec{\zeta}[J]}$ we then have

$$\sup_{z \in \mathbb{R}} \Pr\left[X_i = z \,\Big|\, \vec{\zeta}[J], \vec{\zeta}[I \setminus T_{\vec{\zeta}[J]}]\right] \leq 1 - \delta$$

for each $i \in T_{\vec{\zeta}[J]}$. So overall we obtain

$$\Pr\left[Q(\vec{\zeta}) = 0 \,\Big|\, \vec{\zeta}[J]\right] = \Pr\left[\sum_{i \in I} P_{\vec{\zeta}[J]}^{(i)}(\zeta_i) = 0 \,\Big|\, \vec{\zeta}[J]\right] \leq \frac{2C'_\delta}{\sqrt{m}} \leq \frac{C_\delta}{\sqrt{m}}$$

for any outcome of $\vec{\xi}[J]$. This implies the desired bound in (11.2). $\qquad\square$

Finally, we deduce the full statement of Theorem 1.2 from Proposition 11.3.

Proof of Theorem 1.2. We may assume without loss of generality that $0 < \delta < 1/2$. Let C_δ be a constant such that Proposition 11.3 holds (i.e., such that the statement in Theorem 1.2 holds under the additional assumption that $\sup_{z \in \mathbb{R}} \Pr[\zeta_i = z] \leq 1 - \delta$ for all $i = 1, \ldots, n$). Let us also write $\vec{\zeta} = (\zeta_1, \ldots, \zeta_n)$.

Let $J \subseteq [n]$ be the set of indices j for which $\Pr[\zeta_j = z_j] > 1 - \delta$ holds for some $z_j \in \mathbb{R}$ (such a z_j is unique if it exists). For each $j \in J$, let us represent the random variable ζ_j as $\zeta_j = \alpha_j + \xi_j \beta_j$, for a Rademacher random variable $\xi_j \in \{-1, 1\}$ and a random vector $(\alpha_j, \beta_j) \in \mathbb{R}^2$, as in Lemma 11.1. We do this in such a way that the random variables ξ_j and the random vectors (α_j, β_j) are all mutually independent for all $j \in J$. Note that we always have $\alpha_j + \beta_j = z_j$ for all $j \in J$ (by (b) in Lemma 11.1, recalling that $\Pr[\zeta = z_j] > 1 - \delta > 1/2$).

Now, our plan is to condition on arbitrary outcomes of (α_j, β_j) for $j \in J$, and prove the desired bound using the randomness of ξ_j for $j \in J$, and the randomness of ζ_i for $i \notin J$, applying Proposition 11.3.

For each $j \in J$ and each outcome of (α_j, β_j), the conditional distribution of ζ_j given our outcome of (α_j, β_j) is described by $\Pr[\zeta_j = \alpha_j + \beta_j \mid (\alpha_j, \beta_j)] = \Pr[\zeta_j = \alpha_j - \beta_j \mid (\alpha_j, \beta_j)] = 1/2$ if $\beta_j \neq 0$, and $\Pr[\zeta_j = z_j \mid (\alpha_j, \beta_j)] = 1$ if $\beta_j = 0$ (then $\alpha_j - \beta_j = \alpha_j + \beta_j = z_j$, so ζ_j is constant). Note that conditioning on outcomes of (α_j, β_j) for $j \in J$ does not change the distribution of ζ_i for $i \notin J$, and does not change the fact that the random variables $\zeta_1, \ldots, \zeta_n$ are independent.

For any outcome of $\vec{\beta}[J]$ (i.e., for any outcomes of β_j for $j \in J$), let $H_{\vec{\beta}[J]} \subseteq J$ be the set of indices $j \in J$ such that $\beta_j = 0$, and let $I_{\vec{\beta}[J]} = [n] \setminus H_{\vec{\beta}[J]}$ (i.e., $I_{\vec{\beta}[J]}$ is the subset of indices for which ζ_i "still has some randomness" after conditioning on the outcomes of (α_j, β_j) for $j \in J$). Note that we always have $[n] \setminus J \subseteq I_{\vec{\beta}[J]}$. Furthermore note that for any outcome of $((\alpha_j, \beta_j))_{j \in J}$, we have

$$\sup_{z \in \mathbb{R}} \Pr\left[\zeta_i = z \,\Big|\, ((\alpha_j, \beta_j))_{j \in J}\right] \leq 1 - \delta$$

for all $i \in I_{\vec{\beta}[J]}$ (here we are using that $\delta < 1/2$) and

$$\Pr\left[\zeta_i = z_i \,\Big|\, ((\alpha_j, \beta_j))_{j \in J}\right] = 1$$

for all $i \in H_{\vec{\beta}[J]}$.

Now, for any outcome of $\vec{\beta}[J]$ (which determines $H_{\vec{\beta}[J]}$ and $I_{\vec{\beta}[J]}$), we have $Q(\vec{\zeta}) = Q_{\vec{\beta}[J]}(\vec{\zeta}[I_{\vec{\beta}[J]}])$, where $Q_{\vec{\beta}[J]}$ is the polynomial in the entries of $\vec{\zeta}[I_{\vec{\beta}[J]}]$ obtained from $Q(\vec{\zeta})$ by substituting ζ_i with z_i for all $i \in H_{\vec{\beta}[J]}$ (recall that we always have $\zeta_i = z_i$ for all $i \in H_{\vec{\beta}[J]}$).

We claim that, if we condition on any outcomes of (α_j, β_j) for $j \in J$, then with respect to the resulting conditional probability space, $Q_{\vec{\beta}[J]}(\vec{\zeta}_i[I_{\vec{\beta}[J]}])$ satisfies the fixing box assumption in Theorem 1.2 (and therefore satisfies the assumptions of Proposition 11.3). Indeed, for any outcome of $((\alpha_j, \beta_j))_{j \in J}$, consider a fixing box $\prod_{i \in I_{\vec{\beta}[J]}} R_i$ for $Q_{\vec{\beta}[J]}(\vec{\zeta}[I_{\vec{\beta}[J]}])$, and suppose for the purpose of contradiction that there are fewer than m

40

indices $i \in I_{\vec{\beta}[J]}$ with $\Pr[\zeta_i \in R_i \,|\, ((\alpha_j, \beta_j))_{j \in J}] \le 1 - \delta$. Then, we can extend our fixing box for $Q_{\vec{\beta}[J]}$ to a fixing box $R_1 \times \cdots \times R_n$ for $Q(\zeta_1, \ldots, \zeta_n)$ by simply taking $R_j = \{z_j\}$ for all $j \in H_{\vec{\beta}[J]} \subseteq J$. Note that for each $j \in H_{\vec{\beta}[J]} \subseteq J$ we have $\Pr[\zeta_j \in R_j] = \Pr[\zeta_j = z_j] > 1 - \delta$. Also, for $j \in J \setminus H_{\vec{\beta}[J]} = J \cap I_{\vec{\beta}[J]}$ (i.e., for $j \in J$ with $\beta_j \ne 0$), we can only have $\Pr[\zeta_j \in R_j] \le 1 - \delta$ if $z_j \notin R_j$ (as $\Pr[\zeta_j = z_j] > 1 - \delta$), and in this case we also have $\Pr[\zeta_j \in R_j \,|\, (\alpha_j, \beta_j)] \le 1/2 \le 1 - \delta$ (as $\Pr[\zeta_j = z_j \,|\, (\alpha_j, \beta_j)] = \Pr[\zeta_j = \alpha_j + \beta_j \,|\, (\alpha_j, \beta_j)] = 1/2$). Finally, note that conditioning on $((\alpha_j, \beta_j))_{j \in J}$ does not change the distributions of ζ_i for $i \in [n] \setminus J \subseteq I_{\vec{\beta}[J]}$. So every index $i \in [n]$ with $\Pr[\zeta_i \in R_i] \le 1 - \delta$ has the property that $i \in I_{\vec{\beta}[J]}$ and $\Pr[\zeta_i \in R_i \,|\, ((\alpha_j, \beta_j))_{j \in J}] \le 1 - \delta$. Hence, there are fewer than m indices $i \in [n]$ with $\Pr[\zeta_i \in R_i] \le 1 - \delta$, contradicting our assumption on $Q(\zeta_1, \ldots, \zeta_n)$.

Given the above discussion, for any outcomes of (α_j, β_j) for $j \in J$, we can apply Proposition 11.3, to obtain

$$\sup_{z \in \mathbb{R}} \Pr\left[Q(\zeta_1, \ldots, \zeta_n) = z \,\Big|\, ((\alpha_j, \beta_j))_{j \in J} \right] \le \frac{C_\delta}{\sqrt{m}}.$$

The desired desired unconditional probability bound follows. $\qquad\square$

12 Concluding remarks

In this paper we have obtained essentially optimal bounds for the quadratic Littlewood–Offord problem. There are many interesting directions for further research.

Immediate generalisations. Theorem 1.1 is only about point concentration, and it would be interesting to prove a counterpart for small-ball concentration; i.e., under which assumptions on Q can we prove that

$$\sup_{z \in \mathbb{R}} \Pr[|Q(\xi_1, \ldots, \xi_n) - z| \le 1] \le O\left(\frac{1}{\sqrt{n}}\right)?$$

It would also be nice to prove a generalisation of Theorem 1.1 to polynomials of degree greater than 2. Generalising the conjecture of Nguyen and Vu, we believe the statement of Theorem 1.1 should hold whenever Q is a polynomial of bounded degree (note that without a bounded-degree assumption we cannot hope for sensible anticoncentration bounds; consider for example the *parity function* $Q(x_1, \ldots, x_n) = x_1 x_2 \ldots x_n$).

It is sometimes the case that techniques for point concentration can be straightforwardly adapted for small-ball concentration, and techniques for quadratic polynomials can be straightforwardly adapted for higher-degree polynomials (in particular, the previous bounds of Meka–Nguyen–Vu [37] and Kane [28], mentioned in the introduction, handle small ball concentration for polynomials of any bounded degree). The high-level strategy of the proof of Theorem 1.1 (as sketched in Section 2) makes sense in a very general context, but when trying to generalise Theorem 1.1 in the obvious ways, one runs into some subtle technical issues (see Remarks 2.1 and 2.2). We think it would be very interesting to investigate this further.

It would also be desirable to generalise our "geometric" theorem (Theorem 4.2) to geometric objects other than quadrics inside affine-linear subspaces. For example, we make the following conjecture (closely related to higher-degree generalisations of the Littlewood–Offord problem, and closely related to the directions in [17]).

Conjecture 12.1. *Let $0 \le d < r$ and q be integers. Let $\mathcal{Z} \subseteq \mathbb{R}^r$ be an algebraic variety of dimension d, with degree at most q. Consider vectors $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{R}^r$ such that for some positive integer t, one can form t disjoint bases from the vectors $\vec{a}_1, \ldots, \vec{a}_n$. Let $(\xi_1, \ldots, \xi_n) \in \{-1, 1\}^n$ be a sequence of independent Rademacher random variables. Then*

$$\Pr[\xi_1 \vec{a}_1 + \cdots + \xi_n \vec{a}_n \in \mathcal{Z}] \le \frac{C_{d,r,q}}{t^{(r-d)/2}}$$

for some $C_{d,r,q}$ only depending on d, r, q.

The Gotsman–Linial conjecture. The *Gotsman–Linial conjecture* is a conjecture in Boolean analysis which generalises nearly all polynomial Littlewood–Offord-type theorems. To state this conjecture we need to introduce some notation.

The i-th *influence* of a Boolean function $F : \{-1, 1\}^n \to \{-1, 1\}$ is defined as

$$\mathrm{Inf}_i(F) = \Pr[F(\xi_1, \ldots, \xi_{i-1}, \xi_i, \xi_{i+1}, \ldots \xi_n) \neq F(\xi_1, \ldots, \xi_{i-1}, -\xi_i, \xi_{i+1}, \ldots \xi_n)]$$

(i.e., the probability that changing the ith bit changes the output of the function). The *total influence* (also sometimes called the *average sensitivity*) $\mathrm{Inf}(F)$ of F is $\mathrm{Inf}_1(F) + \cdots + \mathrm{Inf}_n(F)$. For a function $Q : \{-1, 1\}^n \to \mathbb{R}$, the *threshold function* $F_Q : \{-1, 1\}^n \to \{-1, 1\}$ of Q is the Boolean function that detects whether $Q(x_1, \ldots, x_n) \geq 0$ or not. If Q is a degree-d polynomial, we say F_Q is a degree-d threshold function.

Gotsman and Linial [20] made a notorious conjecture on the highest total influence that an n-variable degree-d polynomial threshold function can have (namely, they conjectured that the total influence is maximised when $Q(x_1, \ldots, x_n)$ is a certain product of d terms of the form $(x_1 + \cdots + x_n + a)$). Unfortunately, the precise form of this conjecture has been falsified [4, 29] (it holds for the linear case $d = 1$ but already fails in the quadratic case $d = 2$).

Nonetheless, it still seems plausible that for every n-variable degree-d polynomial the total influence of threshold function can be bounded by $O(d\sqrt{n})$, or at least by $C_d\sqrt{n}$ for some constant C_d depending on d (conjectures in this direction are sometimes variously called the *weak* Gotsman–Linial conjecture). If such a bound were to hold, it would be an easy exercise to deduce Theorem 1.1 (and the generalisations discussed above). It would be interesting to investigate whether the techniques in this paper can be used to make progress on (at least the quadratic form of the) weak Gotsman–Linial conjecture.

Inverse theory. For both the linear and quadratic Littlewood–Offord problems, one cannot hope for a general bound stronger than $O(1/\sqrt{n})$. However, it is natural to investigate assumptions under which one can prove stronger bounds.

In the linear case (studying random variables of the form $X = a_1\xi_1 + \cdots + a_n\xi_n$), this has been an enormously successful direction of research. Early highlights include Stanley's solution of the so-called *Erdős–Moser problem* [46] (introducing tools from algebraic topology to prove an optimal bound under the assumption that all of the coefficients $a_1, \ldots, a_n$ are distinct), and a paper of Halász [21] which introduced Fourier-analytic methods to prove stronger bounds when $a_1, \ldots, a_n$ are in a certain sense "additively unstructured". Perhaps most famously, Tao and Vu [51] proved a so-called *Inverse Littlewood–Offord theorem*, which shows that $\sup_{z \in \mathbb{R}} \Pr[X = z]$ is *extremely* small (smaller than n^{-C} for any constant C) unless $a_1, \ldots, a_n$ have very special additive structure (roughly speaking, most of $a_1, \ldots, a_n$ lie inside a *generalised arithmetic progression*). This had a number of important applications in random matrix theory [48, 49, 51]. The inverse theory of the linear Littlewood–Offord theorem is now essentially complete, thanks to *optimal inverse theorems* of Nguyen and Vu [38] and Rudelson and Vershynin [45] (see also [50]) that give a precise quantification of the extent to which anticoncentration is controlled by the additive structure of $a_1, \ldots, a_n$.

In the quadratic case (studying random variables of the form $X = Q(\xi_1, \ldots, \xi_n)$ for a quadratic polynomial Q), much less is known. An analogue of the original Tao–Vu inverse theorem was proved by Nguyen [39], but it is not clear how an optimal inverse theorem should even be formulated. Such a theorem would have to somehow incorporate the additive information that is relevant for the linear Littlewood–Offord problem, in addition to algebraic considerations (e.g., whether Q factorises into linear factors or not). See [5, 27, 32] for some progress and conjectures related to algebraic aspects of the inverse theory of the quadratic Littlewood–Offord problem.

References

[1] L. Addario-Berry and L. Eslava, *Hitting time theorems for random matrices*, Combin. Probab. Comput. **23** (2014), no. 5, 635–669.

[2] N. Alon, D. Hefetz, M. Krivelevich, and M. Tyomkyn, *Edge-statistics on large graphs*, Combin. Probab. Comput. **29** (2020), no. 2, 163–189.

[3] T. D. Browning and A. Gorodnik, *Power-free values of polynomials on symmetric varieties*, Proc. Lond. Math. Soc. (3) **114** (2017), no. 6, 1044–1080.

[4] B. Chapman, *The Gotsman–Linial conjecture is false*, Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, PA, 2018, pp. 692–699.

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

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

[7] K. P. Costello and V. Vu, *On the rank of random sparse matrices*, Combin. Probab. Comput. **19** (2010), no. 3, 321–342.

[8] K. P. Costello and V. H. Vu, *The rank of random graphs*, Random Structures Algorithms **33** (2008), no. 3, 269–285.

[9] E. Croot, *Fourier proof of the classical Littlewood–Offord inequality*, `https://ecroot.math.gatech.edu/littlewood_offord.pdf.`, 2011.

[10] V. De la Pena and E. Giné, *Decoupling: from dependence to independence*, Springer Science & Business Media, 1999.

[11] W. Doeblin, *Sur les sommes d'un grand nombre de variables aléatoires indépendantes*, Bull. Sci. Math **63** (1939), no. 2, 23–32.

[12] W. Doeblin and P. Lévy, *Calcul des probabilités. sur les sommes de variables aléatoires indépendantes'a dispersions bornées inférieurement*, CR Acad. Sci **202** (1936), 2027–2029.

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

[14] A. Ferber, V. Jain, and Y. Zhao, *On the number of Hadamard matrices via anti-concentration*, Combin. Probab. Comput. **31** (2022), no. 3, 455–477.

[15] A. Ferber, M. Kwan, A. Sah, and M. Sawhney, *Singularity of the k-core of a random graph*, Duke Math. J. **172** (2023), no. 7, 1293–1332.

[16] J. Fox, M. Kwan, and L. Sauermann, *Anti-concentration for subgraph counts in random graphs*, Ann. Probab. **49** (2021), no. 3, 1515–1553.

[17] J. Fox, M. Kwan, and H. Spink, *Geometric and o-minimal Littlewood–Offord problems*, Ann. Probab. **51** (2023), no. 1, 101–126.

[18] J. Fox and L. Sauermann, *A completion of the proof of the edge-statistics conjecture*, Adv. Comb. (2020), Paper No. 4, 52.

[19] M. Glasgow, M. Kwan, A. Sah, and M. Sawhney, *The exact rank of sparse random graphs*, arXiv preprint arXiv:2303.05435.

[20] C. Gotsman and N. Linial, *Spectral properties of threshold functions*, Combinatorica **14** (1994), no. 1, 35–50.

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

[22] D. R. Heath-Brown, *The density of rational points on curves and surfaces*, Ann. of Math. (2) **155** (2002), no. 2, 553–595.

[23] R. Howard, *Estimates on the concentration function of sets in $\mathbb{R}^d$: Notes on lectures of Oskolkov*, https://people.math.sc.edu/howard/Notes/concentration.pdf, 2000.

[24] S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[25] L. Jones, *On the distribution of sums of vectors*, SIAM J. Appl. Math. **34** (1978), no. 1, 1–6.

[26] T. Juškevičius, *The sharp form of the Kolmogorov–Rogozin inequality and a conjecture of Leader–Radcliffe*, arXiv preprint arXiv:2201.09861.

[27] D. Kane, *A structure theorem for poorly anticoncentrated polynomials of Gaussians and applications to the study of polynomial threshold functions*, Ann. Probab. **45** (2017), no. 3, 1612–1679.

[28] D. M. Kane, *The correct exponent for the Gotsman–Linial conjecture*, Comput. Complexity **23** (2014), no. 2, 151–175.

[29] H. W. Kim, C. Maldonado, and J. Wellens, *On graphs and the Gotsman–Linial conjecture for $d = 2$*, arXiv preprint arXiv:1709.06650.

[30] A. Kolmogorov, *Sur les propriétés des fonctions de concentrations de M. P. Lévy*, Ann. Inst. H. Poincaré **16** (1958), 27–34.

[31] M. Kwan, A. Sah, L. Sauermann, and M. Sawhney, *Anticoncentration in Ramsey graphs and a proof of the Erdős–McKay conjecture*, Forum Math. Pi **11** (2023), Paper No. e21.

[32] M. Kwan and L. Sauermann, *An algebraic inverse theorem for the quadratic Littlewood-Offord problem, and an application to Ramsey graphs*, Discrete Anal. (2020), Paper No. 12, 34.

[33] M. Kwan, B. Sudakov, and T. Tran, *Anticoncentration for subgraph statistics*, J. Lond. Math. Soc. (2) **99** (2019), no. 3, 757–777.

[34] I. Leader and A. J. Radcliffe, *Littlewood-Offord inequalities for random variables*, SIAM J. Discrete Math. **7** (1994), no. 1, 90–101.

[35] 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.

[36] A. Martinsson, F. Mousset, A. Noever, and M. Trujić, *The edge-statistics conjecture for $\ell \ll k^{6/5}$*, Israel J. Math. **234** (2019), no. 2, 677–690.

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

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

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

[40] 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.

[41] A. M. Odlyzko, *On subspaces spanned by random selections of ± 1 vectors*, J. Combin. Theory Ser. A **47** (1988), no. 1, 124–133.

[42] J. Pila, *Density of integral and rational points on varieties*, no. 228, 1995, Columbia University Number Theory Seminar (New York, 1992), pp. 4, 183–187.

[43] A. Razborov and E. Viola, *Real advantage*, ACM Trans. Comput. Theory **5** (2013), no. 4, Art. 17, 8.

[44] 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.

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

[46] R. P. Stanley, *Weyl groups, the hard Lefschetz theorem, and the Sperner property*, SIAM J. Algebraic Discrete Methods **1** (1980), no. 2, 168–184.

[47] T. Tao and V. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2006.

[48] T. Tao and V. Vu, *On the singularity probability of random Bernoulli matrices*, J. Amer. Math. Soc. **20** (2007), no. 3, 603–628.

[49] 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.

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

[51] 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.