THE GENERALIZED COMBINATORIAL LASOÑ–ALON–ZIPPEL–SCHWARTZ NULLSTELLENSATZ

LEMMA

GÜNTER ROTE

Abstract. We survey a few strengthenings and generalizations of the Combinatorial Nullstellensatz of Alon and the Schwartz–Zippel Lemma. These lemmas guarantee the existence of (a certain number of) nonzeros of a multivariate polynomial when the variables run independently through sufficiently large ranges.

Contents

1. Introduction 2
   1.1. The Quantitative and the Existence Conclusion 2
   1.2. Assumptions on the numbers \(d_i\) 2
   1.3. Applications 3
   1.4. Assumptions about the coefficient ring 4
   1.5. Comparison of the assumptions 4
   1.6. Tightness 4
   1.7. Existence conclusions in the literature 5
   1.8. Quantitative conclusions in the literature 6
   1.9. Comparison between the results 7
   1.10. Precursor results 7

2. Proof of Lemma X by division by a linear factor 8

3. Proof of Lemma Q 8

4. Largest total degree does not imply the Quantitative Conclusion 9

5. Proof of Lemma X by trimming 9

6. Weaker assumptions 10
   6.1. Successively largest sequences for the Quantitative Conclusion 10
   6.2. Weaker assumptions for the Existence Conclusion 11
   6.3. Connections between the assumptions 12
   6.4. Applications of the generalized results 12

7. Stronger constraints: The Generalized Alon–Füredi Theorem 12

8. Weaker quantitative conclusions for a maximal monomial 14

8.1. Additive increase of the bound 14

8.2. Hypergraph model 15

8.3. Bivariate polynomials 15

8.4. A puzzle 16

8.5. Solution of Problem 1 for finite fields, added June 6, 2023 16

9. What's in a name? 17

References 17

Appendix A. Proof of Lemma X via the coefficient formula 18
1. Introduction

1.1. The Quantitative and the Existence Conclusion. Consider a polynomial \( f \in K[x_1, \ldots, x_n] \) in \( n \) variables over a field or integral domain \( K \), and let \( S_1, \ldots, S_n \) be subsets of \( K \). We want to make statements about the nonzeros of \( f(x_1, \ldots, x_n) \) when the variables \( x_i \) run independently over the sets \( S_i \), under the assumption that these sets are sufficiently large, compared to certain parameters \( d_1, \ldots, d_n \) that are related to the degrees of the terms in \( f \). We may then derive a mere conclusion about the existence of a nonzero or a stronger statement about the number of nonzeros:

The Quantitative Conclusion. If \( |S_i| > d_i \) for all \( i = 1, \ldots, n \), then the number of tuples \( (x_1, \ldots, x_n) \in S_1 \times S_2 \times \cdots \times S_n \) such that \( f(x_1, \ldots, x_n) \neq 0 \) is at least

\[
(1) \quad \left( |S_1| - d_1 \right) \times \left( |S_2| - d_2 \right) \times \cdots \times \left( |S_n| - d_n \right)
\]

\[
= |S_1| \times |S_2| \times \cdots \times |S_n| \times \left( 1 - \frac{d_1}{|S_1|} \right) \left( 1 - \frac{d_2}{|S_2|} \right) \times \cdots \times \left( 1 - \frac{d_n}{|S_n|} \right).
\]

The product in the right half of the last line can be interpreted as a lower bound on the probability of getting a nonzero.

Since the product of the terms \( |S_i| - d_i \) is positive, an immediate consequence is

The Existence Conclusion. If \( |S_i| > d_i \) for all \( i = 1, \ldots, n \), then there exists a tuple of values \( (x_1, \ldots, x_n) \in S_1 \times S_2 \times \cdots \times S_n \) such that \( f(x_1, \ldots, x_n) \neq 0 \).

1.2. Assumptions on the numbers \( d_i \). These conclusions hold under a variety of different assumptions about the parameters \( d_1, \ldots, d_n \).

To describe these parameters, we recall a few standard definitions. A monomial is a product \( x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \) of powers of variables \( x_i \) (not including a coefficient from \( K \)). The degree of the monomial in the variable \( x_i \) is the exponent \( a_i \), and the total degree is the sum \( a_1 + \cdots + a_n \) of these exponents.

The monomials of a polynomial \( f \) are the monomials that have nonzero coefficients when the polynomial is written out in expanded form as a linear combination of monomials.

The (partial) degree of a polynomial \( f \) in the variable \( x_i \) (or the degree of \( x_i \) in \( f \)) is the largest exponent \( a_i \) for which \( x_i^{a_i} \) appears as a factor of a monomial of \( f \). The total degree of a polynomial is the largest total degree of any of its monomials. This is what is usually called the degree of the polynomial without further qualification.

A monomial of \( f \) is maximal if it does not divide another monomial of \( f \), see Figure 13.

Lemma X (Generalized Combinatorial Nullstellensatz, Lasoń 2010 [13, Theorem 2], Tao and Vu 2006 [21, Exercise 9.1.4, p. 332]). If \( x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \) is a maximal monomial of \( f \), then the Existence Conclusion holds.

The lexicographically largest monomial \( x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \) of \( f \) is defined in the usual sense, see Figure 14. \( a_1 \) is the largest exponent of \( x_1 \) in all monomials of \( f \), \( a_2 \) is the largest exponent of \( x_2 \) in all monomials that contain \( x_1^{a_1} \) as a factor, \( a_3 \) is the largest exponent of \( x_3 \) in all monomials that contain \( x_1^{a_1} x_2^{a_2} \) as a factor, and so on. Of course, we may get a different lexicographically
Lemma Q. If the lexicographically largest monomial of $f$ is $x_1^{d_1}x_2^{d_2}\ldots x_n^{d_n}$, then the Quantitative Conclusion holds.

1.3. Applications. Lemmas Q and X and their many relatives in the literature (to be discussed shortly) have numerous important applications to combinatorics and algorithms. The results with the Quantitative Conclusion are the basis for many randomized algorithms. The prime example is polynomial identity testing: Here one wants to check whether two polynomials are identical, or whether a given polynomial is identically zero. The polynomials are given by some algorithm that can evaluate them for specific values. Lemmas Q provides a randomized test for this property, provided

Figure 1. The forbidden monomials for the various assumptions are shown as grey regions, for $(d_1, d_2) = (4, 2)$. In the top row, $(e_1, e_2) = (1, 1)$ was chosen.
some a-priori bounds on the degree can be given. For more applications, see for example [16, Section 7].

When applying the results with the Existence Conclusion, in particular the Combinatorial Nullstellensatz (Corollary X1), a nonzero solution of the polynomial at hand represents some combinatorial object whose existence should be guaranteed. See Alon [1] for a selection of applications.

The two application scenarios focus on different ends of the probability spectrum. In randomized algorithms, the “success probability” of finding a nonzero should ideally be close to 1, but a reasonable probability that decays only polynomially to zero is good enough. Then, by choosing larger sets $S_i$ or by repeating the experiment, the success probability can be amplified to any desired level. The precise probability bounds are not so important in this context.

On the other hand, when it comes to questions of existence, the success of the argument comes down to whether the probability of having a non-zero is non-zero or not. Here it is important to know the smallest values $d_i$ for which the Existence Conclusion holds.

1.4. Assumptions about the coefficient ring. To a lesser extent, the various results in the literature differ in the assumption about the underlying ring of coefficients. All results that we state (with the exception of Lemmas 7 and 8 in Appendix A which require $K$ to be a field) hold when $K$ is an integral domain, i.e., a commutative ring without zero divisors. We mention an even weaker condition under which the theorems hold: $K$ can be an arbitrary commutative ring, but none of the differences $x - y$ for $x, y \in S_i$ must be a zero divisor, see [18, Definition 2.8] or [3, Condition (D)].

1.5. Comparison of the assumptions. Figure 2 compares the strength of the various assumptions in these theorems, including some conditions that are defined in later sections.

The lexicographically largest condition of Lemma Q implies the maximality assumption of Lemma X but since the Quantitative Conclusion in Lemma Q is stronger than the Existence Conclusion in Lemma X, neither of the two results can be derived from the other. We will see in Section 4 that there is no common generalization.

While maximality is not sufficient to imply the Quantitative Conclusion, there are some weaker quantitative conclusions that one can derive under the maximality assumption, see Section 8.

The assumptions in Lemmas X and Q for the Existence or the Quantitative Conclusion are not the weakest assumptions in terms of the monomials of $f$ that we are aware of. The two boxes in the top row of Figure 1 and 2 correspond to some weakened assumptions, which we treat in Section 6.

1.6. Tightness. A simple family of polynomials shows that the bounds of Lemmas X and Q are tight: Select subsets $A_i \subset S_i$ of size $|A_i| = d_i$. Then the polynomial

$$\prod_{i=1}^{n} \prod_{a \in A_i} (x_i - a)$$
Figure 2. Relation between the assumptions on $d_1, \ldots, d_n$.

The Existence and/or some Quantitative Conclusion is indicated at the upper right corner of each box.

1.7. Existence conclusions in the literature. This is Alon’s original Combinatorial Nullstellensatz:

**Corollary X1** (Combinatorial Nullstellensatz, Alon 1999 [1, Theorem 1.2]).

If $x_1^{d_1}x_2^{d_2} \ldots x_n^{d_n}$ is a monomial of largest total degree, then the Existence Conclusion holds.

Alon derives Corollary X1 from a companion result, [1, Theorem 1.1] (which can be proved by the trimming procedure of Proposition 1 in Section 5). It states that, if the Existence Conclusion does not hold, and $f$ is zero on $S_1 \times S_2 \times \cdots \times S_n$, it can be represented in a certain way in the ideal generated by the polynomials $\prod_{a \in S_i}(x_i - a)$. This statement is analogous to Hilbert’s Nullstellensatz, and this justifies the name Combinatorial Nullstellensatz that Alon coined for these theorems. It is of interest in its own right, see [1, Section 9] or [4], but we will not pursue these connections.
1.8. Quantitative conclusions in the literature. The following bound follows by estimating the product $(1 - p_1)(1 - p_2)\ldots(1 - p_n)$ in (1) by the lower bound $1 - p_1 - p_2 - \cdots - p_n$.

**Corollary Q1** (Schwartz 1979 [19, 20, Lemma 1]). Under the assumptions of Lemma Q, i.e., if the lexicographically largest monomial of $f$ is $x_1^{d_1}x_2^{d_2}\ldots x_n^{d_n}$, the number of nonzeros is at least

$$|S_1 \times S_2 \times \cdots \times S_n| \cdot (1 - \frac{d_1}{|S_1|} - \frac{d_2}{|S_2|} - \cdots - \frac{d_n}{|S_n|}).$$

As a special case, when all sets $S_i$ are equal, we get:

**Corollary Q2** (The Schwartz–Zippel Lemma [1], Schwartz 1979 [19, 20, Corollary 1], see also [16, Theorem 7.2] or [21, Exercise 9.1.1, pp. 331–332]).

If $S_1 = S_2 = \cdots = S_n = S$ and the polynomial has total degree $d \geq 0$, then the number of nonzeros is at least

$$|S|^n \cdot (1 - \frac{d}{|S|}).$$

In other words, the probability of getting a zero of $f$ if the variables $x_i$ are uniformly and independently chosen from $S$ is at most $d/|S|$.

The probabilistic formulation with the upper bound $d/|S|$ on the probability of getting a zero is the common statement of this lemma. The same holds for the following statements, but for comparison, we formulate all theorems in terms of the number of nonzeros.

The following statement looks at the degree of $f$ in each variable $x_i$. It follows trivially from Lemma Q.

**Corollary Q3** (Generalized DeMillo–Lipton–Zippel Theorem [3, Thm. 4.6], Knuth 1997 [10, Ex. 4.6.1–16, p. 436]).

If $d_i$ is the degree of variable $x_i$ in $f$, the Quantitative Conclusion holds.

Note that $f$ does not have to contain the term $x_1^{d_1}x_2^{d_2}\ldots x_n^{d_n}$ in this case, but the powers occurring in the lexicographically largest monomial of $f$ are at most $d_i$.

As a special case, with a uniform bound on the degrees and all sets $S_i$ equal, we get:

**Corollary Q4** (Zippel 1979 [22, Theorem 1, p. 221]). Suppose that $f$ is not identically zero and the degree of each variable $x_i$ in $f$ is bounded by $d$, and $S_1 = S_2 = \cdots = S_n = S$. Then the number of nonzeros is at least

$$(|S| - d)^n = |S|^n \cdot (1 - d/|S|)^n.$$  

The following statement puts a stronger assumption on $d$:

**Corollary Q5** (DeMillo and Lipton 1978 [5, Inequality (1)]). If $f$ has total degree $d \geq 0$ and $S_1 = S_2 = \cdots = S_n = S = \{1, 2, \ldots, |S|\}$, then the number of nonzeros is at least

$$|S|^n \cdot (1 - d/|S|)^n.$$

---

Note that this has essentially the same assumptions as Corollary \ref{Q2} (only the assumption about the set \(S\) is more specialized), but a weaker conclusion.

1.9. **Comparison between the results.** The relation between the results in their published form is confusing. This is discussed at length in \cite[Section 4]{B} and in several blog posts.\footnote{https://anuragbishnoi.wordpress.com/2015/10/19/alon-furedi-schwartz-zippel-demillo-lipton-and-their-common-generalization/ https://rjlipton.wpcomstaging.com/2009/11/30/the-curious-history-of-the-schwartz-zippel-lemma/} Above, we have attempted to present them systematically in a logical order, irrespective of the historic development.

As mentioned in Section 1.3, the precise bounds for the Qualitative Conclusion are of minor importance for the applications, and researchers may prefer to state their results in a form that is more convenient to apply or easier to remember instead of the strongest form. Thus, the reason that Lemma \ref{Q} which is, among the statements with the Quantitative Conclusion discussed so far, the strongest and most general, was apparently not written down before is simply that nobody cared to do so.

1.10. **Precursor results.** We mention two precursor results: In the first edition of Knuth’s *Art of Computer Programming*, Vol. 2, there is a weaker, qualitative version of the Quantitative Conclusion:

**Corollary Q6** (Knuth 1969 \cite[Ex. 4.6.1–16, p. 379, solution on p. 546]{K}).

If \(f\) is not identically zero and \(S_1 = S_2 = \cdots = S_n = \{-N, -N + 1, \ldots, N - 1, N\}\), then the fraction of zeros of \(f\) in \(S_1 \times S_2 \times \cdots \times S_n\) goes to zero as \(N \to \infty\).

Øystein Ore, in 1922, already established the special case of the Schwartz–Zippel Lemma (Corollary \ref{Q2}) when the variables \(x_i\) run over all elements of a finite field.

**Corollary Q7** (Ore 1922 \cite[Theorem 6.13]{O}). If \(f \in \mathbb{F}_q[x_1, \ldots, x_n]\) is a polynomial of total degree \(d \geq 0\) over a finite field \(\mathbb{F}_q\) and \(S_1 = S_2 = \cdots = S_n = \mathbb{F}_q\), then the number of nonzeros is at least \((q - d)q^{n-1}\).

I have not been able to look are Ore’s work, and I am citing it according to \cite{O}.

1.11. **Proofs and extensions.** We give the very easy proofs of Lemmas \ref{X} and \ref{Q} in Sections 2 and 3, respectively. Another proof of Lemma \ref{X} which is based on the technique of trimming the polynomial, is given in Section 5.

It is the basis for the generalization of Lemma \ref{X} in Section 6.2. Yet another proof of Lemma \ref{X} is given in Appendix A.

In Section 7, we study the case where both the total degree and the individual degree of each variable is constrained: This is the Generalized Alon–Füredi Theorem of \cite{B}. The example in Section 4 shows that for a maximal \(x_1^{d_1}x_2^{d_2}\ldots x_n^{d_n}\), the Quantitative Conclusion in the form \((*)\) does not follow. In Section 8 we explore the question what quantitative statement we can nevertheless derive.
This question is wide open, and it leads to problems of extremal combinatorics and additive combinatorics. There are many other extensions of the Schwartz–Zippel Lemma or the Combinatorial Nullstellensatz. Among them, we mention a “multivariate” generalization with a quantitative conclusion [6], giving an upper bound on the number of zeros of \( f \) over \( S_1 \times S_2 \times \cdots \times S_n \), where the individual sets \( S_i \subset K^{\lambda_i} \) are themselves multidimensional, representing vectors or points or other geometric objects. This is used to derive incidence bounds in combinatorial geometry.

2. Proof of Lemma \( X \) by division by a linear factor

We sketch the proof of Lasoń [13, Theorem 2], which extends the very simple proof of the original Combinatorial Nullstellensatz (Corollary \( X \)) that was given by Michałek [15] in 2010.

Proof of Lemma \( X \). We use induction on \( d_1 + \cdots + d_n \). The base case \( d_1 + \cdots + d_n = 0 \) is obvious. Otherwise, assume w.l.o.g. that \( d_1 > 0 \). Pick an element \( a \in S_1 \) and divide \( f \) by \( x_1 - a \):

\[
(3) \quad f = q(x_1 - a) + r
\]

The remainder \( r \) is of degree 0 in \( x_1 \), i.e., it is a function \( r(x_2, \ldots, x_n) \) and does not depend on \( x_1 \). If \( r \) has a nonzero on \( S_2 \times \cdots \times S_n \), we obtain a nonzero of \( f \) by setting \( x_1 = a \). Suppose that \( r \) is zero on all of \( S_2 \times \cdots \times S_n \). Then we get a nonzero of \( f \) by finding a nonzero of \( q(x_1, x_2, \ldots, x_n) \) with \( x_1 \neq a \). The existence of such a nonzero in \((S_1 \setminus \{a\}) \times S_2 \times \cdots \times S_n\) is ensured by the inductive hypothesis: It is easy to check that \( x_1^{d_1-1}x_2^{d_2} \cdots x_n^{d_n} \) is indeed a maximal monomial of the quotient \( q \).

3. Proof of Lemma \( Q \)

Proof of Lemma \( Q \). The proof is by induction on \( n \). The induction basis for \( n = 1 \) is the elementary fact that a degree-\( d \) polynomial has at most \( d \) zeros. For \( n > 1 \), we write \( f \) in powers of \( x_1 \):

\[
(4) \quad f(x_1, \ldots, x_n) = \sum_{i=0}^{d_1} x_1^i h_i(x_2, \ldots, x_n)
\]

The sum contains in particular the nonzero term \( x_1^{d_1} h_{d_1}(x_2, \ldots, x_n) \). By definition, \( x_2^{d_2} \cdots x_n^{d_n} \) is the lexicographically largest monomial of \( h_{d_1} \). By induction, the number \( N \) of tuples \((x_2, \ldots, x_n) \in S_2 \times \cdots \times S_n\) for which \( h_{d_1}(x_2, \ldots, x_n) \neq 0 \) is at least

\[
N \geq (|S_2| - d_2) \cdots (|S_n| - d_n).
\]

For a fixed \((x_2, \ldots, x_n)\) for which this case arises, \( f \) is a polynomial of degree \( d_1 \) in \( x_1 \). Therefore it has at most \( d_1 \) zeros, and at least \(|S_1| - d_1 \) nonzeros. Consequently, the number of nonzeros of \( f \) is at least

\[
(|S_1| - d_1)N \geq (|S_1| - d_1)(|S_2| - d_2) \cdots (|S_n| - d_n). \]

\[\square\]
4. Largest total degree does not imply the Quantitative Conclusion

We show that maximality (Lemma [X]) and not even largest total degree (Corollary [X1]) is not sufficient to derive the Quantitative Conclusion. A counterexample is the polynomial \( f(x_1, x_2) = x_1^2 - x_1x_2 + x_2^2 - 1 \), describing an ellipse in the plane, and the sets \( S_1 = S_2 = \{-1, 0, 1\} \), see Figure 3. The monomial \( x_1x_2 \) is a monomial of largest total degree, and the Quantitative Conclusion for \( d_1 = d_2 = 1 \) would predict at least \((|S_1| - d_1)(|S_2| - d_2) = 4\) nonzeros on \( S_1 \times S_2 \). However, there are only 3 nonzeros. (In fact, 3 is the smallest possible number of nonzeros for any polynomial for with \( x_1x_2 \) as maximal monomial, see Proposition 5 in Section 6.)

\[ \text{Figure 3. A quadratic bivariate polynomial with 6 zeros on a } 3 \times 3 \text{ grid} \]

5. Proof of Lemma [X] by trimming

The Combinatorial Nullstellensatz is a basic result, and it appears in a wide range of textbooks. Many of the proofs that I have seen in my (not very thorough) survey of the literature proceed in two steps along the following lines.

The first step reduces the polynomial \( f \) to a trimmed polynomial, whose degree in each variable is now less than \( |S_i| \), without changing the value of \( f \) on \( S_1 \times S_2 \times \cdots \times S_n \). After this reduction, one can apply any of the lemmas with the Quantitative Conclusion.

We include this proof because it lends itself to a generalization, Theorem 3 in Section 6.2.

The trimming procedure is described in the following statement:

**Proposition 1.** Let \( f \in K[x_1, \ldots, x_n] \) be a polynomial over a commutative ring \( K \), and let \( S_1, \ldots, S_n \subseteq K \) be sets.

Then \( f \) can be transformed into a polynomial \( \hat{f} \) with the following properties:

1. \( f \) and \( \hat{f} \) have the same values on \( S_1 \times S_2 \times \cdots \times S_n \).
2. In \( \hat{f} \), the degree in each variable \( x_i \) is less than \( |S_i| \).
3. If \( x_1^{e_1} \cdots x_n^{e_n} \) is a maximal monomial of \( f \) with \( e_i < |S_i| \) for all \( i \), then its coefficient remains unchanged by this transformation.

**Proof.** Let \( s_i = |S_i| \). The polynomials \( x_i^{s_i} \) and \( x_i^{s_i} - \prod_{a \in S_i} (x_i - a) \) have the same values for all \( x \in S_i \). Hence, we may successively replace \( x_i^{s_i} \) by the...
polynomial $x_i^{e_i} - \prod_{a \in S_i} (x_i - a)$, whose degree is smaller than $s_i$, and in this way, eliminate all powers of $x_i$ of degree $s_i$ or higher, without changing the value of $f$ on $S_1 \times S_2 \times \cdots \times S_n$. (Putting it differently, we divide $f$ by $\prod_{a \in S_i} (x_i - a)$ and take the remainder.)

If we do this for all variables, we arrive at a polynomial $\tilde{f}$ for which the degree in each variable $x_i$ is less than $s_i$.

To see Property 3, we observe that the modification, applied to a term $x_1^{e_1} x_2^{e_2} \cdots x_n^{e_n}$, only affects the coefficients of monomials $x_1^{b_1} x_2^{b_2} \cdots x_n^{b_n}$ with $b_i \leq e_i$ for all $i$. A monomial $x_1^{e_1} x_2^{e_2} \cdots x_n^{e_n}$ with $e_i < s_i$ for all $i$ is itself not subject to the trimming procedure, and if it is maximal, it has no monomials “above it” that could change its coefficient.

Since the degree $d_i$ in each variable $x_i$ is now less than $|S_i|$, we can apply Corollary Q3, which has an easy inductive proof along the lines of the proof of Lemma Q shown in Section 3, or we may pick a lexicographically largest monomial and apply Lemma Q directly.

5.1. Comparison of the proofs. It is instructive to compare the two proofs of Lemma X that we have seen. The trimming procedure is essentially a polynomial division, and it reduces the polynomial to a polynomial for which the Quantitative Conclusion holds. To prove the Quantitative Conclusion, one applies induction on the number of variables, as in the proof of Lemma Q (Section 3). The induction step is based on the fact that a univariate polynomial of degree $d$ has at most $d$ roots. This fact, finally, is proved by repeated division by a linear factor.

By contrast, the proof of Section 2, which goes back to Micha lek [15], puts the division by a linear factor at the very beginning. As we have seen, this makes the proof simple and direct.

In Appendix A we give another proof. It follows the suggested hint for the solution of Exercise 9.1.4 in Tao and Vu [21, p. 332], and it is the earliest proof of Lemma X. In contrast to the other proofs, it works only for fields.

6. Weaker assumptions

There is a way in which the respective assumptions of Lemma Q and Lemma X can be weakened. The two variations of the assumptions were developed independently, but they are remarkably similar in spirit, and the relation between them is analogous to the relation between lexicographically largest and maximal monomials. The assumptions are not easy to understand, and they are motivated mainly by the fact that the original proofs carry through with few changes.

6.1. Successively largest sequences for the Quantitative Conclusion. We define a more general notion than a lexicographically largest monomial, namely what we call a successively largest sequence $(d_1, \ldots, d_n)$ of exponents: Pick any monomial $x_1^{e_1} x_2^{e_2} \cdots x_n^{e_n}$ of $f$. We set $f_1$ to be the original polynomial $f_1(x_1, \ldots, x_n) = f(x_1, \ldots, x_n)$. For $j = 2, \ldots, n$, we inductively define $f_j(x_1, \ldots, x_n)$ as the coefficient of $x_j^{e_j-1}$ in $f_{j-1}(x_{j-1}, \ldots, x_n)$.

Finally, we let $d_j$ be the degree of $x_j$ in $f_j$, for $j = 1, \ldots, n$. 
Consider, for example, the polynomial $f(x_1, x_2) = x_1^4 + x_1^6x_2^9 + x_1x_2^2 + x_1x_2 + x_6$. Picking the term $x_1x_2$ leads to $f_2(x_2) = x_2^2 + x_2$, and thus a successively largest sequence $(d_1, d_2) = (7, 2)$. For the term $x_2^6$, we get $(d_1, d_2) = (7, 6)$. Figure 1A shows another example: $(d_1, d_2) = (4, 2)$ is a successively largest sequence with respect to the monomial $xy$.

Note that $x_1^{d_1}x_2^{d_2} \ldots x_n^{d_n}$ is not necessarily a monomial of $f$. As with the lexicographically largest monomial, this notion depends on the chosen order of the variables.

**Theorem 2** (Knuth 1998 [10], Answer to Ex. 4.6.1–16, pp. 674–675). For a successively largest sequence $d_1, \ldots, d_n$, the Quantitative Conclusion holds.

**Proof.** The proof of Lemma Q goes through with straightforward adaptations. We proceed by induction on $n$. We write $f$ in powers of $x_1$ as in (4):

$$f(x_1, \ldots, x_n) = \sum_{i=0}^{d_1} x_1^i h_i(x_2, \ldots, x_n)$$

By assumption, the sum contains the nonzero term $x_1^{e_1} f_2(x_2, \ldots, x_n)$. By definition, $(d_2, \ldots, d_n)$ is a successively largest sequence for $f_2$.

For a fixed tuple $(x_2, \ldots, x_n)$ with $f_2(x_2, \ldots, x_n) \neq 0$, $f$ is a nonzero polynomial of degree at most $d_1$ in $x_1$. In contrast to the case of Lemma Q the degree can be smaller than $d_1$, but the conclusion that $f$ has hat most $d_1$ zeros remains valid. The argument finishes in the same way as for Lemma Q.

Knuth [10, p. 675] mentions further ideas of strengthening the bound, and points out the significance in the context of sparse polynomials.

6.2. Weaker assumptions for the Existence Conclusion.

**Theorem 3** (Schauz 2008 [13], Theorem 3.2(ii)). Assume $|S_i| > d_i \geq e_i$ for $i = 1, \ldots, n$, and assume that $x_1^{e_1} \ldots x_n^{e_n}$ is a monomial of $f$. If $f$ contains no other monomial $x_1^{e_1'} \ldots x_n^{e_n'}$ with $e_i' = e_i$ or $e_i' > d_i$ for each $i = 1, \ldots, n$, then the Existence Conclusion holds.

Figure 1B illustrates this condition. In the terminology of Schauz, the tuple $(e_1, \ldots, e_n)$ is called a “$(d_1, \ldots, d_n)$-leading multi-index”. The term $x_1^{d_1} \ldots x_n^{d_n}$ is not required to appear in $f$.

Theorem 3 may be stronger than Lemma X. For example, for the polynomial

$$f(x_1, x_2) = x_1^4x_2^8 + x_1x_2^2 + x_1^6x_2^2,$$

which is a sparser variant of the polynomial in Figure 1B, we may take $(e_1, e_2) = (1, 1)$ and $(d_1, d_2) = (4, 2)$.

The forbidden exponent pairs can be written concisely as $\{e_1, d_1 + 1, d_1 + 2, d_1 + 3, \ldots\} \times \{e_2, d_2 + 1, d_2 + 2, d_2 + 3, \ldots\}$, except $(e_1, e_2)$ itself.

**Proof of Theorem 3**. The proof by trimming from Section 3 goes through: Observe that trimming a monomial $x_1^{e_1}x_2^{e_2} \ldots x_n^{e_n}$ creates monomials in which the powers $x_1^{e_1}$ with $c_i < |S_i|$ are unchanged. Only the powers $x_1^{c_i}$ with $c_i \geq |S_i|$ are replaced by smaller powers. Thus, the monomials $x_1^{e_1} \ldots x_n^{e_n}$ that
are excluded by the assumption of Theorem 3 are precisely those monomials whose trimming process could affect the chosen monomial $x_1^{e_1} \ldots x_n^{e_n}$. □

Schauz showed the stronger statement that the coefficient of $x_1^{e_1} \ldots x_n^{e_n}$ can be represented in terms of the values of $f$ on $S_1 \times S_2 \times \cdots \times S_n$, thus generalizing the coefficient formula (14) in Appendix A. For further information and more references, see [4].

6.3. Connections between the assumptions. There is a connection between Theorems 2 and 3: The assumptions of the first theorem imply the assumptions of the second. In particular, if $(d_1, \ldots, d_n)$ is a successively largest degree sequence with respect to the monomial $x_1^{e_1} \ldots x_n^{e_n}$, then the assumptions of Theorem 3 hold.

Looking at the top two rows of Figure 1, one can notice some general pattern: The conditions for the Quantitative Conclusion in the left column (lexicographically largest monomial, successively largest sequence) depend on the ordering of the variables, whereas the conditions for the Existence Conclusion in the right column (maximal monomial, the $(d_1, \ldots, d_n)$-leading multi-index of Theorem 3) are insensitive to the variable order.

One can observe (and prove) the following curious connection between the forbidden monomials, which are shown as shaded regions of Figure 1: The forbidden terms for $x_1^{d_1} x_2^{d_2} \ldots x_n^{d_n}$ being a maximal monomial can be obtained as the intersection of the forbidden terms for being a lexicographically largest monomial over all $n!$ orderings of the variables.

The same relation holds between a successively largest sequence (Theorem 2) and the condition of Theorem 3, if the defining monomial $x_1^{e_1} \ldots x_n^{e_n}$ is held fixed.

6.4. Applications of the generalized results. In the applications of the Combinatorial Nullstellensatz or the Schwartz–Zippel Lemma and its relatives, the degree bounds on the polynomial $f$ are derived a priori, and not by looking at a particular polynomial that is explicitly given. Thus, the added generality offered by Theorems 2 and 3 is only academic and of little practical use. Even for the Generalized Combinatorial Nullstellensatz (Lemma X), we are not aware of a convincing application for which the classic Combinatorial Nullstellensatz (Corollary X1) would not suffice.

Such an application was indeed given by Lasoń [13, Theorem 4], but it appears somewhat fabricated. The polynomial can be obtained from some homogeneous polynomial $h(x_1, \ldots, x_n)$ by replacing each variable $x_i$ by some polynomial $f_i(x_i)$ (and adding some linear terms). In a homogeneous polynomial, every monomial is both maximal and of maximum total degree, but after the modification, the terms acquire different degrees, and Corollary X1 no longer applies.

7. Stronger constraints: The Generalized Alon–Füredi Theorem

Bishnoi, Clark, Potukuchi, and Schmitt [3] give a precise bound on the minimum number of nonzeros when, in addition to a bound $d_i$ on the degree of each variable $x_i$, the total degree $d$ is specified. The bound is not explicit:
It is formulated in terms of an optimization problem of minimizing the product of variables $y_i$ under linear constraints.

**Theorem 4** (The Generalized Alon–Füredi Theorem, Bishnoi et al. [3]). Let $f$ be a polynomial of total degree $d$, whose degree in each variable $x_i$ is at most $d_i$, where $d_i < |S_i|$. Then $f$ has at least $N$ nonzeros on $S_1 \times S_2 \times \cdots \times S_n$, where $N$ is the optimum value of the following minimization problem:

\[
\begin{align*}
\text{(5)} & \quad \text{minimize} & & y_1 y_2 \cdots y_n \\
\text{(6)} & \quad \text{subject to} & & |S_i| - d_i \leq y_i \leq |S_i|, \quad \text{for } i = 1, \ldots, n \\
\text{(7)} & \quad \sum_{i=1}^n y_i = |S_1| + \cdots + |S_n| - d
\end{align*}
\]

**Figure 4.** Forbidden monomials for the Generalized Alon–Füredi Theorem, for $d_1 = 5, d_2 = 4, d = 7$. For an example with $|S_1| = |S_2| = 8$, the optimal value $N = y_1 y_2 = 18$ is achieved by $(y_1, y_2) = (3, 6)$.

Figure 4 illustrates the assumptions. They combine the constraints of Figure 4a and 4b.

**Proof.** The theorem can be derived from Lemma Q. The optimization problem (5–7) can be interpreted as looking for a lexicographically largest monomial $x_1^{e_1} x_2^{e_2} \cdots x_n^{e_n}$ that is consistent with the assumptions of the theorem and for which Lemma Q gives the weakest bound.

To start the formal proof, note first that the optimum value $N$ of (5–7) does not change if we turn (7) into an inequality:

\[
\sum_{i=1}^n y_i \geq |S_1| + \cdots + |S_n| - d
\]

This is easily seen as follows: Take a solution $(y_1, \ldots, y_n)$ satisfying (6) and (7). The assumptions of the theorem imply $d \leq \sum_{i=1}^n d_i$. Therefore, as long as the inequality (7) is strict, one can always find a variable $y_i$ that is not at its lower bound, i.e., $y_i > |S_i| - d_i$. We can therefore reduce this variable, reducing the product $y_1 \cdots y_n$.

The proof is now straightforward: Let $x_1^{e_1} x_2^{e_2} \cdots x_n^{e_n}$ be the lexicographically largest monomial of $f$. By the assumptions on $f$, $e_i \leq d_i$ and
By Lemma Q, the number of nonzeros is at least
\[ (|S_1| - e_1)(|S_2| - e_2) \cdots (|S_n| - e_n) = y_1 y_2 \cdots y_n, \]
which is at least the minimum value \( N \) of (5) under (6) and (7).

Bishnoi et al. [3] proved Theorem [1] directly by induction on \( n \). They showed that the bound is tight for all combinations of values \( d_i \) and \( |S_i| \) to which the theorem applies. They also derived the Generalized DeMillo–Lipton–Zippel Theorem (Corollary Q3) from it.

In the (original) Alon–Füredi Theorem [2, Theorem 5], the degrees \( d_i \) in the individual variables are not constrained, and there is an important difference: It is assumed that \( f \) has at least one nonzero on \( S_1 \times S_2 \times \cdots \times S_n \). Because of this extra assumption, the Alon–Füredi Theorem is not a straightforward corollary of the Generalized Alon–Füredi Theorem, see [3, Sections 2.2–2.3]. In the constraints defining the bound \( N \), the lower bound in (6) is replaced by \( y_i \geq 1 \). As a consequence, in contrast to Theorem [1], it is easy to solve the optimization problem: Starting from the lower bound \( y_1 = \cdots = y_n = 1 \), consider the variables \( y_i \) in order of decreasing sizes \( |S_i| \) and greedily enlarge each \( y_i \) value to its upper bound \( |S_i| \) until (7) is fulfilled.

8. **Weaker quantitative conclusions for a maximal monomial**

We have seen in Section 4 that for a maximal monomial, or even for a monomial of largest total degree, the Quantitative Conclusion in the form (1) does not hold. Can we still say something about the number of nonzeros beyond the fact that it is at least 1, which is the trivial consequence of the Existence Conclusion?

8.1. **Additive increase of the bound.** A very weak quantitative conclusion is given by the following statement.

**Proposition 5.** If \( x_1^{d_1} x_2^{d_2} \cdots x_n^{d_n} \) is a maximal monomial, then the number of nonzeros over the grid \( S_1 \times \cdots \times S_n \), with \( |S_i| > d_i \) for all \( i \), is at least
\[ 1 + (|S_1| - (d_1 + 1)) + (|S_2| - (d_2 + 1)) + \cdots + (|S_n| - (d_n + 1)). \]

In other words, at each step of increasing \( |S_i| \) above the lower bound \( d_i + 1 \) that is necessary for the Existence Conclusion, the guaranteed number of nonzeros increases by 1.

For example, with \( (d_1, d_2) = (1, 1) \) and \( |S_1| = |S_2| = 3 \), we conclude that there must be at least 3 nonzeros. Thus, the ellipse example of Section 4 cannot be improved by choosing a different grid \( S_1 \times S_2 \) of the same size.

A version of Proposition [5] was stated in 2022 by Knuth for the restricted case that \( x_1^{d_1} x_2^{d_2} \cdots x_n^{d_n} \) is a monomial of largest total degree [11, Ex. MPR–114, p. 23, answer on p. 388]. The proof goes through without changes
when \( x_1^{d_1} x_2^{d_2} \ldots x_n^{d_n} \) is a maximal monomial and we base the argument on Lemma \( \text{X} \) instead of Corollary \( \text{X1} \).

**Proof of Proposition 5.** We can eliminate any chosen nonzero \((x_1, \ldots, x_n)\) from \( S_1 \times S_2 \times \cdots \times S_n \) by removing \( x_j \) from \( S_j \), for an arbitrary \( j \). (This may eliminate additional nonzeros.)

Thus, if there were fewer than the claimed number of nonzeros, we could eliminate them by successively removing an element from some \( S_j \) while keeping \( |S_j| \geq d_j + 1 \). Eventually we would arrive at a grid on which \( f \) is identically zero, contradicting Lemma \( \text{X} \). \( \Box \)

**8.2. Hypergraph model.** Stronger asymptotic bounds can be obtained by using tools from extremal combinatorics. It is natural to associate an \( n \)-partite \( n \)-uniform hypergraph to the zeros of an \( n \)-variate polynomial over a grid \( S_1 \times \cdots \times S_n \). The hypergraph contains the hyperedge \((x_1, \ldots, x_n)\) whenever \( f(x_1, \ldots, x_n) = 0 \). The Existence Conclusion then says that the hypergraph contains no complete subhypergraph \( K^{(r)}(d_1 + 1, \ldots, d_n + 1) \). What does this last statement alone (without regarding the algebraic origin of the hypergraph) imply about the number of nonzeros in \( S_1 \times \cdots \times S_n \)?

This is a question from extremal (hyper-)graph theory.

We can apply the following result of Erdős from 1964 [7, Corollary, p. 188].

**Proposition 6.** Consider the family of \( n \)-partite \( n \)-uniform hypergraphs that contain no complete \( K^{(n)}(l, \ldots, l) \), for some \( l \geq 2 \).

Then there is a threshold \( s_0(n, l) \) such that in every hypergraph of the family with at least \( s \) vertices in each color class, for \( s > s_0(n, l) \), the edge density is at most

\[
\frac{(3n)^n}{s^{1/n-1}}.
\]

(In the original statement in [7], our \( n \) is denoted by \( r \), which adheres better to the conventions of hypergraphs, and our \( s \) is denoted by \( n \).)

We translate this to our setting: If \( x_1^{d_1} x_2^{d_2} \ldots x_n^{d_n} \) is a maximal monomial of \( f \), Lemma \( \text{X} \) implies that the hypergraph corresponding to the zeros does not contain a complete \( K^{(r)}(l, \ldots, l) \), with \( l = 1 + \max\{d_1, \ldots, d_n\} \). We conclude that the density of zeros in \( S_1 \times S_2 \times \cdots \times S_n \) is bounded by (8) if \( s := \min\{|S_1|, \ldots, |S_n|\} \) is big enough. This is good enough for the property that is essential for the applications: The probability of hitting a zero goes to 0 as the size of all sets \( S_i \) is increased. However, the convergence is very slow.

**8.3. Bivariate polynomials.** For a polynomial of \( n = 2 \) variables, we are in the setting of bipartite graphs, where the classic result of Kövári, Sós, and Turán [8] applies. In particular, if \( x_1^{d_1} x_2^{d_2} \) is a maximal monomial, then the bipartite graph with \( |S_1| + |S_2| \) vertices that models the zeros on \( S_1 \times S_2 \) contains no complete bipartite subgraph \( K_{d_1+1, d_2+1} \). Assuming \( s = |S_1| = |S_2| \), we conclude from the Kövári–Sós–Turán Theorem that such a graph has at most \( O(s^{2-1/l}) \) edges, where \( l = \min\{d_1, d_2\} + 1 \). Note that, in contrast to the case of hypergraphs above, we use \( \min\{d_1, d_2\} \) and not \( \max \). Hence the density of zeros is

\[
O(1/\sqrt{s}).
\]
The bound of the Kővári–Sós–Turán Theorem is known to be tight for several small values of \( l \) in the combinatorial setting, where all we know is that the bipartite subgraph \( K_{d_1+1,d_2+1} \) is forbidden. This completely ignores the origin of the problem from the polynomial \( f \). Can a polynomial with such a large fraction \( \Theta(1/s^{1/l}) \) of zeros on an \( s \times s \) grid be constructed?

8.4. **A puzzle.** The first nontrivial example is \((d_1,d_2) = (1,1)\), i.e., \( xy \) should be a maximal monomial. Such a polynomial, after suitable scaling, has the form

\[
\begin{equation}
\tag{9}
f(x, y) = -xy + P(x) + Q(y),
\end{equation}
\]

where \( P(x) \) and \( Q(y) \) are polynomials of arbitrarily high degree.

Let us denote the elements that we substitute for \( x \) by \( S_1 = \{a_1, \ldots, a_s\} \), with distinct elements \( a_i \), and similarly for the values \( S_2 = \{b_1, \ldots, b_s\} \) that we substitute for \( y \). Let \( u_i = P(a_i) \) and \( v_j = Q(b_j) \) be the corresponding values of the polynomials. Then the zeros of \( f \) on \( S_1 \times S_2 \) are the index pairs \((i,j)\) with

\[
a_i b_j = u_i + v_j \quad (1 \leq i, j \leq s).
\]

We can thus reformulate our question as follows:

**Problem 1.** Let \( s \) be fixed.

Find two sequences of \( a_1, \ldots, a_s \) and \( b_1, \ldots, b_s \) of distinct numbers, and two sequences \( u_1, \ldots, u_s \) and \( v_1, \ldots, v_s \) of not necessarily distinct numbers, such that the multiplication table of the first two sequences agrees with the addition table of the last two sequences in as many positions \((i,j)\) as possible:

\[
a_i b_j = u_i + v_j
\]

For example, the following multiplication and addition tables, which are derived from the ellipse example of Section 4, have 6 coinciding entries:

\[
\begin{array}{ccc}
\times & 1 & 3 & 5 \\
6 & 6 & 18 & 30 \\
7 & 7 & 21 & 35 \\
8 & 8 & 24 & 40 \\
\end{array}
\quad \quad \quad \quad
\begin{array}{ccc}
+ & 1 & 17 & 29 \\
1 & 2 & 18 & 30 \\
6 & 7 & 23 & 35 \\
7 & 8 & 24 & 36 \\
\end{array}
\]

The question has now become a problem of additive combinatorics. It is clear that Problem 1 is not more restricted than asking for the zeros of (9):

We can find an interpolating polynomial \( P \) and \( Q \) for any values \( a_i \) and \( u_i \), or \( b_i \) and \( v_i \), respectively, since the degree of \( P \) and \( Q \) is not bounded.

As discussed above, the bipartite graph that models the zeros of \( f \) contains no \( K_{2,2} \); this can also be shown directly from the definition of an addition and multiplication table. Hence the number of zeros is \( O(s^{3/2}) \). Can this bound be achieved, asymptotically, or does the algebra imply a sharper upper bound? Is there a construction with a superlinear number of zeros?

8.5. **Solution of Problem 1 for finite fields, added June 6, 2023.**

Recently, Alexey Gordeev (private communication) has informed me that he has a solution of Problem 1 in finite fields. Specifically, for any \( m > 1 \) and any prime \( p \), he constructs an \( m \)-variate polynomial over the field \( \mathbb{F}_p^m \) for which \( x_1 x_2 \ldots x_m \) is a maximal monomial, and for which the fraction of zeros among the \( s^m = (p^m)^m \) \( m \)-tuples is \( \Theta(1/p) = \Theta(1/s^{1/m}) \).
9. What’s in a name?

In the late 1970’s, the first randomized primality tests were discovered. Randomized algorithms were gaining popularity, and their usefulness was recognized. It is thus no coincidence that various forms of the Schwartz–Zippel Lemma were discovered independently, as the topic was “in the air”. The papers of Schwartz and Zippel were even presented at the same conference in 1979 and published back to back in the proceedings volume [19, 22].

The name Schwartz–Zippel Lemma stuck, despite the accumulation of sibilant consonants, and despite the priority of DeMillo and Lipton [5]. A blog post of Richard Lipton from 2009 proposed various possible reasons for this fact. We add to this discussion by speculating that the poor typesetting quality of the Information Processing Letters at the time may have contributed to the fact that the paper [5] was not sufficiently received. In addition, the quirk with the capital letter in the middle of the family name might have caused some insecurity and uneasiness. In the title of this note, we honor the tradition of omitting DeMillo and Lipton.

We have seen that Lason’s generalization of Alon’s Combinatorial Nullstellensatz was predated by an exercise in a textbook, but he must nevertheless credited for bringing the statement of Lemma X to the published journal literature. The major reason for including his name is the rhyme.

References


Appendix A. Proof of Lemma X via the coefficient formula

This proof follows the hint of Tao and Vu [21, Exercise 9.1.4, p. 332] and works out their exercise, see also Lasoń [13, Section 3]. Essentially the same proof, for the original Combinatorial Nullstellensatz (Corollary X1), was given by Kouba [12] in 2009.

As an intermediate result, we get a formula [14] for the coefficient of $x_1^{d_1}x_2^{d_2}\ldots x_n^{d_n}$ in terms of the values of $f$ on $S_1 \times S_2 \times \cdots \times S_n$ (the Coefficient Formula of Lasoń [13, Theorem 3]).

We emphasize that, in contrast to other statements in this note, the following proof supposes that the coefficient ring is a field (and we call it $\mathbb{F}$).

We start with a preparatory lemma:

Lemma 7. Let $\mathbb{F}$ be a field. For a finite nonempty set $S \subseteq \mathbb{F}$, there is a function $g_S : S \rightarrow \mathbb{F}$ with the following property:

\[
\sum_{x \in S} g_S(x)x^k = \begin{cases} 
0, & \text{for } k = 0, 1, \ldots, |S| - 2 \\
1, & \text{for } k = |S| - 1
\end{cases}
\]

Proof. The equations (10)–(11) form a system of $|S|$ linear equations in the $|S|$ unknowns $u_j = g_S(a_j)$ for $a_j \in S = \{a_1, a_2, \ldots, a_{|S|}\}$. The coefficient matrix is a Vandermonde matrix, and hence the system has a unique solution. (The situation is the same as in Lagrange interpolation, except that the coefficient matrix is transposed.)
The solutions $u_j$ can actually be obtained explicitly as the quotient of two Vandermonde determinants:

$$u_j = g_S(a_j) = 1 / \prod_{k \neq j} (a_j - a_k) \quad \square$$

**Proof of Lemma X.** It is no loss of generality to assume $|S_i| = d_i + 1$. Take the functions $g_{S_i}$ for $i = 1, \ldots, n$, and multiply them together:

$$\tilde{g}(x_1, \ldots, x_n) := g_{S_1}(x_1)g_{S_2}(x_2) \cdots g_{S_n}(x_n)$$

Continuing to follow the suggested procedure of Tao and Vu [21, Exercise 9.1.4], we consider the quantity

$$\tilde{F} := \sum_{x_1 \in S_1} \sum_{x_2 \in S_2} \cdots \sum_{x_n \in S_n} f(x_1, \ldots, x_n)\tilde{g}(x_1, \ldots, x_n),$$

and we want to show that $\tilde{F} \neq 0$. Let us see how the transformation from $f$ to $\tilde{F}$ affects the monomials $x_1^{a_1} \cdots x_n^{a_n}$ of $f$:

$$\sum_{x_1 \in S_1} \sum_{x_2 \in S_2} \cdots \sum_{x_n \in S_n} x_1^{a_1} \cdots x_n^{a_n} g_{S_1}(x_1)g_{S_2}(x_2) \cdots g_{S_n}(x_n)$$

Equation (14), in connection with (12) and (13), is called the coefficient formula.

This expression vanishes whenever $a_i < d_i$ for some $i$, by (10). The only monomial of $f$ that is not annihilated in this way is the maximal monomial $x_1^{d_1}x_2^{d_2} \cdots x_n^{d_n}$. For this monomial, the term (15) becomes 1, by (11). Therefore $\tilde{F}$ as given by (14) is equal to the coefficient of $x_1^{d_1}x_2^{d_2} \cdots x_n^{d_n}$ in $f$, expressing it in terms of the values of $f$ on the grid $S_1 \times S_2 \times \cdots \times S_n$. Accordingly, (14), in connection with (12) and (13), is called the coefficient formula.

By the assumption of Lemma X, $x_1^{d_1}x_2^{d_2} \cdots x_n^{d_n}$ appears in $f$, and thus its coefficient $\tilde{F} \neq 0$. Therefore, by (14), there must be an $(x_1, x_2, \ldots, x_n) \in S_1 \times S_2 \times \cdots \times S_n$ with $f(x_1, \ldots, x_n) \neq 0$. \quad \square

We conclude with a few remarks. The hint of Tao and Vu [21, Exercise 9.1.4] actually suggests to prove a more general version of Lemma 7:

**Lemma 8.** For a set $S$ with $|S| > d$, there is a function $g_{S,d} : S \to \mathbb{R}$ with the following property:

$$\sum_{x \in S} g_{S,d}(x)x^k = \begin{cases} 0, & \text{for } k = 0, 1, \ldots, d - 1 \\ 1, & \text{for } k = d \end{cases}$$

This can be derived by applying Lemma 7 to an arbitrary subset $S' \subseteq S$ of size $|S'| = d + 1$ and setting $g_{S,d}(x) = 0$ for $x \notin S'$. We have instead chosen to simplify the proof by assuming $|S| = d + 1$.

Tao and Vu [21, Exercise 9.1.4] formulate their exercise “for a field whose characteristic is 0 or greater than max $d_i$.” I don’t see how the characteristic of the field comes into play.

Since we are constructing some sort of interpolating function $g$, which depends on solving a system of equations, this proof depends on $F$ being a field (or at least, a ring in which all nonzero differences $a - a'$ for $a, a' \in S_i$
are units). Under some weaker algebraic conditions (see Section 1.4), it is still true that the coefficient of $x_1^{d_1}x_2^{d_2}\ldots x_n^{d_n}$ in $f$ is uniquely determined by the values of $f$ at the points $(x_1, x_2, \ldots, x_n) \in S_1 \times S_2 \times \cdots \times S_n$ [18]. Statement 2.8(v), see also [4].