2020

Solutions to Systems of Equations over Finite Fields

Rachel Petrik
University of Kentucky, rpetrik08@gmail.com
Digital Object Identifier: https://doi.org/10.13023/etd.2020.234

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Rachel Petrik, Student

Dr. David Leep, Major Professor

Dr. Peter Hislop, Director of Graduate Studies

SOLUTIONS TO SYSTEMS OF EQUATIONS OVER FINITE FIELDS

DISSERTATION

A dissertation submitted in partial
fulfillment of the requirements for
the degree of Doctor of Philosophy
in the College of Arts and Sciences
at the University of Kentucky

By
Rachel Louise Petrik
Lexington, Kentucky

Director: David B. Leep, Professor of Mathematics
Lexington, Kentucky
2020

ABSTRACT OF DISSERTATION

SOLUTIONS TO SYSTEMS OF EQUATIONS OVER FINITE FIELDS

This dissertation investigates the existence of solutions to equations over finite fields with an emphasis on diagonal equations. In particular:

1. Given a system of equations, how many solutions are there?

2. In the case of a system of diagonal forms, when does a nontrivial solution exist?

Many results are known that address (1) and (2), such as the classical Chevalley–Warning theorems. With respect to (1), we have improved a recent result of D.R. Heath–Brown, which provides a lower bound on the total number of solutions to a system of polynomials equations. Furthermore, we have demonstrated that several of our lower bounds are sharp under the stated hypotheses. With respect to (2), we have several improvements that extend known results. First, we have improved a result of James Gray by extending his theorem to a larger class of equations. Second, for particular degrees, number of forms, and finite fields, we have determined the minimal number of variables needed to guarantee the existence of a nontrivial solution. Third, there are many results, which address (2) for particular types of systems known as A-systems. We give a criterion that characterizes when a system of equations is an A-system. Finally, we have provided exposition that adds significantly more detail to two important papers by Tietäväinen.

KEYWORDS: Finite Fields, Diagonal Forms, Solutions to Equations

Rachel Louise Petrik

May 13, 2020

SOLUTIONS TO SYSTEMS OF EQUATIONS OVER FINITE FIELDS

By
Rachel Louise Petrik

David B. Leep
Director of Dissertation

Peter D. Hislop
Director of Graduate Studies

May 13, 2020
Date

"Families are the compass that guides us. They are the inspiration to reach great heights, and our comfort when we occasionally falter." —Brad Henry

For McCabe and Oslo

And for Mom, Dad, Matt, and Helena

The best family I could ask for.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, David Leep. Thank you for your attention to detail and thoughtful approach to presenting mathematics in an accessible way. I have learned so much working with you and am forever grateful to the guidance and support you provided throughout my time at UK.

Thank you also to the other members of my committee: Drs. Tim Gorringe, David Jensen, Uwe Nagel, and Cidambi Srinivasan. You took the time to listen to my presentations and read early drafts of this dissertation. Thank you for your insightful comments and questions.

I would like to extend a special thanks to Jared M. Smith and Oak Ridge National Lab for giving an algebraic number theorist (ME!) a chance to explore the amazing and daunting realm of cyber and information security research. I learned so much through this experience. I would also like to thank my fellow interns Berat, Luke, Raz, Saeed, and Terryl for their friendship, help, and patience as I explored computer science.

Thank you to everyone who encouraged me to pursue mathematics in graduate school. In particular, thank you to John Goldwasser, Kevin Milans, and David Miller. Without each of you, I certainly would not be here now.

Thank you to Rejeana Cassady, Christine Levitt, and Sheri Rhine for all of your help navigating all of the academic bureaucracy. Your assistance over the years has been truly invaluable.

I would have never made it this far without the support of my friends in the mathematics community. Darleen, the one who encouraged me to pursue new and exciting opportunities (and taught me how to cook); Deborah, the one who taught me to remember what's important (and also some great make-up skills); Nandita, the one who taught me how to be a confident mathematician; Julie, the one with the beautiful soul and great taste in wine; Devin, the one who encouraged me to explore therapy and was always down for a cup of coffee; Fulton, the one who was always there...always; Carolyn, the one who paved the way and gave wonderful advice.

To my friends outside of graduate school: Annie, my soul sister and the one who I skyped with every Monday. I would not have stayed sane without our weekly chats. Ashley, the one who always texted to make sure I was still alive and sent me cute pictures of Luna and Lefty.

Thank you to my mom, Lynn Petrik, for our walks along the bike path and her joy at all of my accomplishments. Thank you to my dad, James Petrik, for letting me take over your office to study for prelims and his last minute trips to Lexington. Thank you to my brother, Matt Petrik, for our late night runs to Sushi Blue and your listening ear. Thank you to my sister, Helena Petrik, for single handily furnishing my wardrobe and bringing me to tears with funny stories. In short, thank you all for your support, encouragement, and most of all, the Petrik steel.

Thank you to my goofy, wonderful, amazing, adorable, on-sale cat, Oslo! You are truly the best! I love you so much! (Also, you can fetch. Who knew!)

Finally, thank you to my amazing fiance, McCabe Olsen; the one who showed me that collaborative math could be fun and fulfilling (and made the best guacamole). Thank you for your encouragement, laughter, and sense of adventure. I would not have made it here without your unending support and guidance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction and Background

1.1 Introduction

The study of solutions to systems of polynomial equations is among the most classical questions in mathematics. Perhaps the most notable and well known example of this is the Fundamental Theorem of Algebra:

Theorem 1.1.1 (Fundamental Theorem of Algebra)**.** *Every non-zero, single-variable, degree n polynomial with complex coefficients has, counted with multiplicity, exactly n complex roots.*

The key behind this result is the fact that the field of complex numbers, $\mathbb{C}$, is algebraically closed. In general, one should not expect that an arbitrary polynomial of degree n over another field or ring has n solutions. Likewise, if one was to consider a polynomial equation in many variables, determining the number of solutions becomes a far more difficult problem. Furthermore, considering simultaneous solutions to systems of polynomials in multiple variables is even more daunting. For this reason, studying solutions to systems of polynomial equations over various fields is a rich and fruitful direction of research. The general topic of this dissertation is to considering solutions to systems of polynomial equations in many variables over finite fields.

To further motivate considering the problem over finite fields, one should note that a natural question to number theorists and, indeed, mathematicians in a variety of fields, is to consider integer solutions to polynomial equations. In this scenario, it is quite possible to have no solutions much less n solutions. A necessary condition for the existence of an integer solution is the existence of a solution over the finite field, $\mathbb{F}_p$, for every prime p. Subsequently, results which guarantee the existence or non-existence of solutions over finite fields are of interest. In the study of systems of polynomials over finite fields, there are typically three questions one can ask.

1. Can we find and describe all solutions?
2. How many solutions are there?
3. In the case of a system of homogeneous polynomials, when does a nontrivial solution exist?

Following a brief discussion on terminology and notation, we seek to address the second and third questions with the hope that this will provide insight to the first.

The goal of Chapter 2 is to study lower bounds on the number of solutions to systems of polynomial equations over finite fields. Given a general system of polynomials, a classical question in mathematics has been to determine the existence of a solution. For systems over finite fields, the keystone results on the existence of solutions are the *Chevalley–Warning theorems* of the 1930s. Over the years, mathematicians have sought to improve these results. The most recent of which are those of D.R. Heath–Brown [9]. We provide an improvement of all parts of this result.

In Chapter 3, we focus on the case of solutions to a single diagonal form of odd degree over finite fields. When restricting to the case of diagonal forms, one can obtain much deeper and more specialized results. Of particular interest are results on the number of variables necessary to guarantee the existence of a nontrivial solution. The focus of this Chapter is to consider two results of James Gray [6, 7]. The new contribution is a proof that both of Gray's results are, in fact, equivalent, as well as providing an improvement to one.

In Chapter 4, we introduce a generalization of odd degree diagonal forms called A-equations (A-systems). Throughout the literature, there are far more results which hold for diagonal forms and diagonal systems of odd degree that do not necessarily hold in the arbitrary degree case. Introduced by Tietäväinen [14], A-systems are systems of possibly even degree where some of the core proof techniques for the odd case still hold. The original contribution of this chapter is to provide a complete characterization of A-systems.

The focus of Chapter 5 is to approach systems of diagonal forms over finite fields in general. Sections 5.2 and 5.3 collate a number of known results in the area, as well as, providing more complete proofs than currently exist in the literature and some small improvements to some of these results. In Sections 5.4 and 5.5, we consider the computational question of finding the explicit lower bound on the number of variables needed to guarantee a nontrivial solution. In particular, for a system of r diagonal forms over $\mathbb{F}_q$ with specified degrees, we compute the largest integer such that there exists an anisotropic system in that many variables. In some special cases, we also compute the maximum of this value over all possible choices of q.

A more extended introduction is provided in the introduction of the remaining chapters.

1.2 Terminology and Notation

Let $\mathbb{F}_q$ denote the finite field of order $q = p^k$, where p is a prime and $k \in \mathbb{Z}_{>0}$. We say that a polynomial f in n variables is defined over $\mathbb{F}_q$ if $f \in \mathbb{F}_q[x_1, \ldots, x_n]$. Let $\deg(f) = d$.

We say that $f \in \mathbb{F}_q[x_1, \ldots, x_n]$ is a *(homogeneous) form of degree* d, for some $d \geq 0$, if $f \neq 0$ and each monomial in f has degree d.

Example 1.2.1. For example, $xy^2 + z^3 = 0$ is a homogeneous form of degree 3, but $x^2 + xy^2 + z^3 = 0$ is not a homogeneous form.

A *diagonal form* f of degree d is a form having the shape $f(x_1, \ldots, x_n) = \alpha_1 x_1^d + \cdots + \alpha_n x_n^d$. Notice that diagonal forms of degree d are homogeneous forms of degree d.

Example 1.2.2. For example, $x^3 + 2y^3 + 3z^3 = 0$ is a diagonal form of degree 3, but $xy^2 + z^3 = 0$ is not a diagonal form.

Assume that $f \in \mathbb{F}_q[x_1, \ldots, x_n]$. We say that $(a_1, \ldots, a_n) \in \mathbb{F}_q^n$ is a zero of f if $f(a_1, \ldots, a_n) = 0$. Suppose that $f \in \mathbb{F}_q[x_1, \ldots, x_n]$ is a homogeneous form of degree d, $d \geq 1$. Then it is clear that $f(0, \ldots, 0) = 0$. We say that $(a_1, \ldots, a_n) \in \mathbb{F}_q^n$ is a *nontrivial zero* of a homogeneous form f if $f(a_1, \ldots, a_n) = 0$ and $(a_1, \ldots, a_n) \neq (0, \ldots, 0)$.

Let $f \in \mathbb{F}_q[x_1, \ldots, x_n]$ be a homogeneous form of degree d, $d \geq 0$. We say that f is *isotropic* over $\mathbb{F}_q$ if f has nontrivial zero in $\mathbb{F}_q^n$, and that f is *anisotropic* over $\mathbb{F}_q$ if f is not isotropic over $\mathbb{F}_q$.

Example 1.2.3. Consider the homogeneous form $f = xy^2 + z^3$ over $\mathbb{F}_7$. A nontrivial zero of this form is $(2, 2, 3)$ as $f(2, 2, 3) = 35 \equiv 0 \bmod 7$. In particular, f is isotropic over $\mathbb{F}_7$.

To denote a system of polynomials, we write $\boldsymbol{f} = \{f_1, \ldots, f_r\}$, where $f_i \in \mathbb{F}_q[x_1, \ldots, x_n]$ for $1 \leq i \leq r$. Let $d_i = \deg(f_i)$ for $1 \leq i \leq r$. We define the degree of the system to be $d_1 + \cdots + d_r$. We say $\boldsymbol{f}$ is a *homogeneous system* if f_i is a homogeneous form for $1 \leq i \leq r$. Note that for a homogeneous system we do not require that $\deg(f_i) = \deg(f_j)$ for $i \neq j$.

Example 1.2.4. The following is a homogeneous system

$$\begin{aligned} f_1 &= xy^2 + z^3 \\ f_2 &= x^2 + 2y^2 + 5xy \\ f_3 &= 3x^7. \end{aligned}$$

We say $\boldsymbol{f}$ is a *system of diagonal forms* if f_i is a diagonal form for $1 \leq i \leq r$. Similarly, note that we do not require that $\deg(f_i) = \deg(f_j)$ for $i \neq j$.

Example 1.2.5. The following is a system of diagonal forms

$$\begin{aligned} f_1 &= x^3 + 2y^3 + 3z^3 \\ f_2 &= x^2 + 2y^2 + 5z^2 \\ f_3 &= 3x^7. \end{aligned}$$

Given a system of polynomials, $\boldsymbol{f} = \{f_1, \ldots, f_r\}$, we say that $(a_1, \ldots, a_n) \in \mathbb{F}_q^n$ is a zero of $\boldsymbol{f}$ if $f_i(a_1, \ldots, a_n) = 0$, $1 \leq i \leq r$. In the case where $\boldsymbol{f}$ is a system of forms, we say that such a zero is a nontrivial zero of $\boldsymbol{f}$ if $(a_1, \ldots, a_n) \neq (0, \ldots, 0)$. We say that a system of forms is *isotropic* over $\mathbb{F}_q$ if the system has a nontrivial common zero in $\mathbb{F}_q^n$. Otherwise, we say a system is *anisotropic* over $\mathbb{F}_q$.

We define $N(f; \mathbb{F}_q)$ to be the number of solutions of f over $\mathbb{F}_q$. Similarly, we define $N(\boldsymbol{f}; \mathbb{F}_q)$ to be the number of solutions of $\boldsymbol{f}$ over $\mathbb{F}_q$.

Let $\boldsymbol{f}$ be a system of r diagonal forms over $\mathbb{F}_q$ in n variables, where $\deg(f_i) = d_i$. Let $\vec{d} = (d_1, \ldots, d_r)$. We define $\Omega(r, \vec{d}, q)$ to be the minimal number such that if $n > \Omega(r, \vec{d}, q)$, then there exists a nontrivial solution to the system $\boldsymbol{f}$. In other words, $\Omega(r, \vec{d}, q)$ is the largest integer such that there exists an anisotropic system in that many variables of r diagonal forms of degrees $\vec{d}$ over $\mathbb{F}_q$. For simplicity, if $d := d_1 = \cdots = d_r$, we write $\Omega(r, d, q) = \Omega(r, \vec{d}, q)$. Again for simplicity, if $r = 1$, we write $\Omega(d, q) = \Omega(1, d, q)$.

For fixed values of r and $\vec{d}$, one more quantity of interest is $\max\limits_q \Omega(r, \vec{d}, q)$. We define $\max\limits_q \Omega(r, \vec{d}, q) = \max\{\Omega(r, \vec{d}, q) \mid q$ is an arbitrary prime power$\}$. As a result, if $n > \max\limits_q \Omega(r, \vec{d}, q)$, then any system of r diagonal forms of degrees $\vec{d}$ in n variables defined over $\mathbb{F}_q$ is isotropic independent of our choice of q.

Chapter 2 Minimal Number of Solutions to Systems of Equations

2.1 Introduction

The goal of this chapter is to study lower bounds on the number of solutions to systems of polynomial equations over finite fields. Given a general system of polynomials, a classical question in mathematics has been to determine the existence of a solution. For systems over finite fields, the keystone results on the existence of solutions are the *Chevalley–Warning theorems* of the 1930s.

Theorem 2.1.1 (Chevalley, [3]). *Let* $f_1, \ldots, f_r \in \mathbb{F}_q[x_1, \ldots, x_n]$, $\deg(f_i) = d_i$, $1 \leq i \leq r$. *Assume* $n > d_1 + d_2 + \cdots + d_r$ *and* $Z(\boldsymbol{f}, \mathbb{F}_q^n)$ *is nonempty. Then* $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq 2$.

In particular, if $f_1, \ldots, f_r$ are homogeneous forms, then the system has a nontrivial solution. Graphically, we can visualize this result with the following figure.

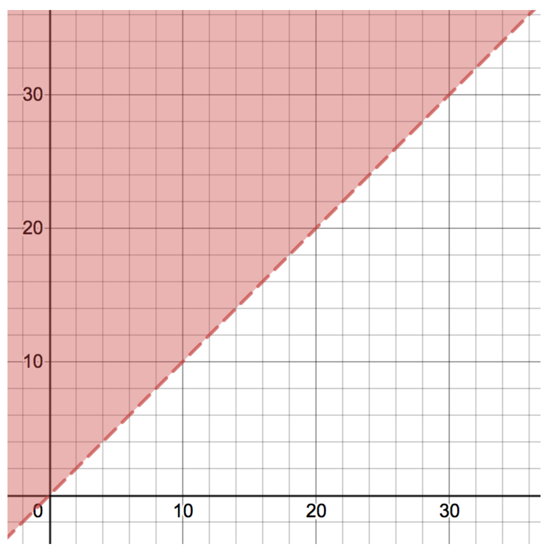


Figure 2.1: This figure illustrates the relationship between n and d needed to guarantee a nontrivial solution to a system of equations. The horizontal axis represents our degree d and the vertical axis represents the number of variables n. The shaded region is the order pairs of (d, n) that guarantee a nontrivial solution to our system.

Theorem 2.1.2 (Warning, [17]). *Let* $f_1, \ldots, f_r \in \mathbb{F}_q[x_1, \ldots, x_n]$, $\deg(f_i) = d_i$, $1 \leq i \leq r$. *Assume* $n > d_1 + d_2 + \cdots + d_r$*, then* $N(\boldsymbol{f}, \mathbb{F}_q^n) \equiv 0 \bmod p$.

In particular, if $f_1, \ldots, f_r$ are homogeneous forms, there is at least one solution, therefore at least $p \geq 2$ solutions.

Example 2.1.3. Consider the following system in $n = 7$ variables over $\mathbb{F}_5$.

$$f_1 = x_1^2 x_2 + 4x_3^3$$

$$f_2 = 2x_4^2 + x_5 + x_6 + x_7$$

Notice the total degree of the system is 5. Since $n = 7 > 5 = d_1 + d_2$ and $(0, 0, 0, 0, 0, 0, 0)$ is a solution, by Theorem 2.1.1, we know there is at least one more solution. In particular, a nontrivial solution. By Theorem 2.1.2, we know that the number of solutions is divisible by 5, so there are at least 5 solutions.

Over the years, mathematicians have sought to improve these results. Most notably are the following two improvements attributed to Warning [17] and Ax [1]. Warning showed that under the same hypotheses as Theorem 2.1.1, Chevalley's result could be improved.

Theorem 2.1.4 (Warning, [17]). *Let $f_1, \ldots, f_r \in \mathbb{F}_q[x_1, \ldots, x_n]$, $\deg(f_i) = d_i$, $1 \leq i \leq r$. Assume $n > d_1 + d_2 + \cdots + d_r$ and $Z(\boldsymbol{f}; \mathbb{F}_q^n)$ is nonempty. Then $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq q^{n-d}$.*

Similarly, Ax demonstrated that under the same hypotheses as Theorem 2.1.2, Warning's result could be improved.

Theorem 2.1.5 (Ax, [1]). *Let $f_1, \ldots, f_r \in \mathbb{F}_q[x_1, \ldots, x_n]$, $\deg(f_i) = d_i$, $1 \leq i \leq r$. Assume $n > d_1 + d_2 + \cdots + d_r$, then $N(\boldsymbol{f}, \mathbb{F}_q^n) \equiv 0 \bmod q$.*

Further improvements will require additional hypotheses, as examples exist that show the bound in Theorem 2.1.4 is best possible.

Example 2.1.6. Consider the polynomial

$$g = x_1^{p-1} + x_2^{p-1} + \cdots + x_{p-1}^{p-1}$$

over $\mathbb{F}_p$. Notice that g has no nontrivial solutions in $\mathbb{F}_p^{p-1}$ since $x^{p-1} = 1$ for all $x \in \mathbb{F}_p^*$. For any integer n where $n > p - 1$, consider

$$g = x_1^{p-1} + \cdots + x_{p-1}^{p-1} + 0x_p^{p-1} + \cdots + 0x_n^{p-1} \in \mathbb{F}_p[x_1, \ldots, x_n].$$

Then the zeros of g over $\mathbb{F}_p^n$ have $x_1 = \cdots = x_{p-1} = 0$, $x_p, \ldots, x_n$ arbitrary elements in $\mathbb{F}_p$. Thus, $N(g; \mathbb{F}_p^n) = p^{n-(p-1)}$, which is precisely the bound given by Theorem 2.1.4. In this case, the set of solutions forms a subspace of dimension $n - (p - 1)$ in $\mathbb{F}_p^n$.

One other class of examples that attains the lower bound given by Theorem 2.1.4 are norm forms. Let L/k be a finite algebraic extension of degree d. Let $N_{L/k} : L^* \to k^*$ be the norm map. Let $v_1, \ldots, v_d$ be a vector space basis of L over k. The polynomial $g(x_1, \ldots, x_d) = N_{L/k}(x_1 v_1 + \cdots + x_d v_d)$ is a homogeneous form of degree d in d variables defined over k. Assume $(a_1, \ldots, a_d) \neq (0, 0, \ldots, 0)$. Since $N_{L/k}(a_1 v_1 + \cdots + a_d v_d) \neq 0$, it follows that g is an anisotropic homogeneous form defined over k of degree d in d variables. This polynomial g is called a *norm form.* We can now consider $g \in \mathbb{F}_q[x_1, \ldots, x_d, x_{d+1}, \ldots, x_n]$. Notice that any solution must have $x_1 = x_2 = \cdots = x_d = 0$ since g is anisotropic in d variables. Since $x_{d+1}, \ldots, x_n$ do not appear in this form, we can set them equal to any of the q elements from $\mathbb{F}_q$ and still have a solution to g. This means $N_{\mathbb{F}_q}(g) = q^{n-d}$, which shows that Warning's bound is optimal. The same reasoning applies for any anisotropic form of degree d in d variables. In fact, the set of solutions forms a subspace of dimension $n - d$ in $\mathbb{F}_q^n$. In fact, virtually all known examples, where Theorem 2.1.4 is sharp, have the additional property that the set of solutions form an affine space of $\mathbb{F}_q^n$. Since it is not always convenient to assume that $\mathbf{0}$ is a solution.

Definition 2.1.7. An *affine space* is a coset of subspace. In particular, a subspace is an affine space. We say that two affine spaces are *parallel* if they are cosets of the same subspace. The *dimension* of an affine space is the dimension of the subspace.

When the solutions do not form an affine space, Heath-Brown, in 2011, [9] provided the following improvement.

Theorem 2.1.8 (Heath–Brown, [9], Theorem 2)**.** *Suppose that $n > d$ and that $Z(\boldsymbol{f}; \mathbb{F}_q^n)$, the collection of zeros over $\mathbb{F}_q^n$, is non-empty, and is not an affine space of $\mathbb{F}_q^n$. Then*

(i) For any q, we have $N(\boldsymbol{f}; \mathbb{F}_q^n) > q^{n-d}$;

(ii) If $q \geq 4$, we have $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq 2q^{n-d}$; and

(iii) For any q, we have $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq \dfrac{q^{n+1-d}}{(n+2-d)}$ provided that the polynomials $f_1, \ldots, f_r$ are homogeneous.

2.2 Main Result

In joint work with Leep, we show that each part of Theorem 2.1.8 can be improved.

Theorem 2.2.1 (Leep–Petrik)**.** *Suppose that $n > d$ and that $Z\boldsymbol{f}; \mathbb{F}_q^n)$, the collection of zeros over $\mathbb{F}_q^n$, is non-empty, and is not an affine space of $\mathbb{F}_q^n$. Then*

(i) For $q = 2$, $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq q^{n-d} + q$;

(ii) For $q \geq 3$, $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq 2q^{n-d}$;

(iii) For $q \geq 3$, $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq 2q^{n-d} + (q-2)q$ provided that the polynomials $f_1, \ldots, f_r$ are homogeneous;

(iv) For any q, $N(\boldsymbol{f}; \mathbb{F}_q^n) > \dfrac{q^{n+1-d}}{(n+2-d)}$ provided that the polynomials $f_1, \ldots, f_r$ are homogeneous.

Moreover, the bounds in (i), (ii), and (iii) are sharp under these hypotheses.

For the sake of completeness, we will present the proofs of the results required for Theorem 2.2.1.

Theorem 2.2.2 (Heath–Brown, [9], Theorem 1)**.** *With the notation above, we have*

$$N(L_1) \equiv N(L_2) \bmod q$$

for any two parallel affine spaces $L_1, L_2 \subseteq \mathbb{F}_q^n$ of dimension d or more.

Proof. Given a polynomial $f(x_1, \cdots, x_n)$ of total degree e, there are two reasonable ways of associating a form to it. One may take $f_-(x_1, \cdots, x_n)$ to be the homogeneous part of degree e, or one may define $f_+(x_0, \cdots, x_n) = x_0^e f(\frac{x_1}{x_0}, \cdots, \frac{x_n}{x_0})$. For a system $\boldsymbol{f}$, we define $\boldsymbol{f}_-$ and $\boldsymbol{f}_+$ by the above processes, using degree d_i for each polynomial f_i.

Clearly each zero of $\boldsymbol{f}$ produces exactly $q-1$ zeros of $\boldsymbol{f}_+$ with $x_0 \neq 0$. The zeros of $\boldsymbol{f}_+$ with $x_0 = 0$ correspond precisely to the zeros of $\boldsymbol{f}_-$.

Thus, $N(\boldsymbol{f}_+; \mathbb{F}_q^{n+1}) = (q-1)N(\boldsymbol{f}; \mathbb{F}_q^n) + N(\boldsymbol{f}_-; \mathbb{F}_q^n)$. In particular, if $n \geq d$, then (4) yields $q \mid N(\boldsymbol{f}_+; \mathbb{F}_q^{n+1})$ and hence $N(\boldsymbol{f}; \mathbb{F}_q^n) \equiv N(\boldsymbol{f}_-; \mathbb{F}_q^n) \pmod q$. Thus the value of $N(\boldsymbol{f}; \mathbb{F}_q^n)$ modulo q depends only on the leading homogeneous parts of the polynomials $f_1, \cdots, f_r$.

Let $L = (e_1, \cdots, e_k)$ be the affine space of dimension k, parallel to L_1 and L_2, but passing through the origin. Let $L_j = L + c_j$ for $j = 1, 2$. Then $N(\boldsymbol{f}; L_j) = N(g^{(j)}; \mathbb{F}_q^k)$, where

$$g_i^{(j)}(y_1, \cdots, y_k) = f_i(\sum_{l=1}^{k} y_l e_l + c_j) \text{ for } (j = 1, 2, 1 \leq i \leq r).$$

If L_1, L_2 have dimension strictly greater than d, then $k > d$. By Theorem 2.1.5, $q \mid N(g^{(j)}; \mathbb{F}_q^k)$. Since $N(g^{(j)}; \mathbb{F}_q^k) = N(\boldsymbol{f}; L_j)$, we find that $q \mid N(\boldsymbol{f}; L_j)$ for $j = 1, 2$. Thus $N(L_1) \equiv N(L_2) \bmod q$.

Assume L_1, L_2 have dimension equal to d. It could happen that the terms of degree d_i in $g_i^{(1)}$ all vanish, but this happens if and only if the corresponding terms in $g_i^{(2)}$ also vanish. In this situation, the total degree of each of the systems $(g_1^{(1)}, \cdots, g_r^{(1)})$ and $(g_1^{(2)}, \cdots, g_r^{(2)})$ will be m, where $m < d$. Since $d = k > m$, by (4), if follows that $q \mid N(g^{(j)}; \mathbb{F}_q^k)$. Thus, $q \mid N(\boldsymbol{f}; L_j)$ for $j = 1, 2$. Thus $N(L_1) \equiv N(L_2) \bmod q$.

It follows that we may assume that the leading homogeneous parts are the same. But since, the value of $N(\boldsymbol{f}; \mathbb{F}_q^n)$ modulo q depends only on the leading homogeneous parts of the polynomials $f_1, \cdots, f_r$. It follows that

$$N(\boldsymbol{f}; L_1) = N(g^{(1)}; \mathbb{F}_q^n) \equiv N(g_-^{(1)}; \mathbb{F}_q^n) \equiv N(g_-^{(2)}; \mathbb{F}_q^n) \equiv N(g^{(2)}; \mathbb{F}_q^n) = N(\boldsymbol{f}; L_2)$$

$\square$

Lemma 2.2.3 (Heath–Brown, [9], Lemma 1)**.** *Let $L_0 \subseteq \mathbb{F}_q^n$ be an affine space. Choose an affine space L of maximal dimension k such that $L \supseteq L_0$ and $N(\boldsymbol{f}; L) = N(\boldsymbol{f}; L_0)$. Suppose $L' \supset L$ is an affine space of dimension $k+1$ such that $N(\boldsymbol{f}; L')$ is minimal. Then*

$$N(\boldsymbol{f}; \mathbb{F}_q^n) \geq N(\boldsymbol{f}; L) + \frac{q^{n-k}-1}{q-1}(N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L)) \tag{2.1}$$

Proof. Note $\mathbb{F}_q^n$ is the disjoint union of L together with the sets $L^* \setminus L$, where L^* runs over all $(k+1)$-dimensional affine spaces containing L. There are $\dfrac{q^{n-k}-1}{q-1}$ such affine spaces L^*. Since $N(\boldsymbol{f}; L^* \setminus L) \geq N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L)$, we have the following

$$\begin{aligned} N(\boldsymbol{f}; \mathbb{F}_q^n) &= N(\boldsymbol{f}; L) + N(\boldsymbol{f}; \bigcup_{L^*}(L^* \setminus L)) \\ &= N(\boldsymbol{f}; L) + \sum_{L^*} N(\boldsymbol{f}; L^* \setminus L) \\ &\geq N(\boldsymbol{f}; L) + \frac{q^{n-k}-1}{q-1}(N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L)) \end{aligned}$$

□

For the next lemma, we will require the following definition.

Definition 2.2.4. A set of $t+1$ points in $\mathbb{F}_q^t$ are in *general position* if no $k+1$ of them lie in a k-dimensional subspace for $1 \leq k \leq t$.

Lemma 2.2.5 (Heath–Brown, [9], Lemma 2)**.** *Let $S \subseteq \mathbb{F}_q^t$ be a set containing $t+1$ points in general position. Then*

(i) If $q = 2$, and if there is no 2-plane $L \subseteq \mathbb{F}_q^t$ meeting S in exactly 3 points, then $S = \mathbb{F}_q^t$.

(ii) If $q \geq 3$, and if $l \subseteq S$ for every line l meeting S in at least two points, then $S = \mathbb{F}_q^t$.

(iii) If $q \geq 4$, and if $|S \cap l| \geq q-1$ for every line l meeting S in at least two points, then $\mathbb{F}_q^t \setminus S$ is contained in a hyperplane.

(iv) If $m \geq 2$ is an integer, and if $|S \cap l| \geq m+1$ for every line l meeting S in at least two points, then $|S| \geq \dfrac{m^{t+1}-1}{m-1}$.

Proof. (i) We will prove this using induction on t. The statement is trivially true for $t = 1$. When $t = 1$, $S \subseteq \mathbb{F}_2$ and $|S| \geq 2$. Since $|\mathbb{F}_2| = 2$, it follows that $S = \mathbb{F}_2$.

For our inductive hypothesis, assume S contains an affine space L_0 of dimension $t-1$ together with a point $P_0 \notin L_0$.

Now choose any point $P \notin L_0$ with $P \neq P_0$. We want to show $P \in S$. We will then be able to conclude that $S = \mathbb{F}_q^t$ as desired. Let $P_1 \in L_0$ and consider the 2-plane generated by P, P_0, and P_1. Since $q = 2$, this plane consists of the three generators together with a fourth point P_2, which must belong to L_0. Since $q = 2$, V_0 has only two cosets in $\mathbb{F}_q^t$. Thus $L_0 = V_0$ or $L_0 = V_0 + z$ for $z \in \mathbb{F}_q^t$. Thus $\mathbb{F}_q^t = L_0 \cup (L_0 + z)$. Since $P_0 \notin L_0$, $P_0 \in L_0 + z$. Similarly for P. Note $P_1 \in L_0$. Consider the 2-plane generated by P, P_0, P_1. This is a coset of a 2-dimensional vector space.

More explicitly, it is the coset $W + P$, where W is the 2-dimensional subspace generated by $P_0 - P$ and $P_1 - P$. Note that $P_2 = (P_0 - P) + (P_1 - P) + P = P_0 + P_1 - P$.

Since P_0, $P \notin L_0$ and $P_1 \in L_0$ and there are two cosets, it follows that $P_2 \in L_0$. Since P_0, P_1, and $P_2 \in S$, by our hypothesis $P \in S$ as well.

(ii) We will prove this using induction on t. The statement is trivially true for $t = 1$. Since $|\mathbb{F}_q| = q$ and $|l \cap S| = 2$, we have $l \subseteq S$. Since $|\mathbb{F}_q| = q$, it follows that $S = \mathbb{F}_q$.

For our inductive hypothesis, assume S contains an affine space L_0 of dimension $t-1$ together with a point $P_0 \notin L_0$.

Now choose any $P \notin L_0$, $P \neq P_0$. Suppose that the line l generated by P_0 and P meets L_0 at a point P_1. Then l meets S in at least 2 points, namely P_0 and P_1. Then by our hypothesis, l is contained in S, so $P \in S$.

This deals with all points P except those which lie on the hyperplane L_1, which is parallel to L_0 and which passes through P_0. We begin by fixing any point P_1 on L_0. We then consider the line l generated by P and P_1. Since $q \geq 3$, this line contains at least one additional point P_2, which cannot lie in L_1. ($P_2 \notin L_1$ because if $P_2 \in L_1$, then the line between P and P_2 would lie entirely in L_1).

Now to show $P_2 \in S$, we will repeat the initial argument. Since $P_2 \notin L-1$, the line generated by P_2 and P_0 meets L_0 at a point P_3. Then the line meets S in at least two points, namely P_0 and P_3. Thus the line is completely contained in S, so $P_2 \in S$. Since P_1, $P_2 \in S$, l meets S in at least two points, so $l \subseteq S$, which implies $P \in S$.

(iii) Let $S^c = \mathbb{F}_q^t \setminus S$. Our strategy will be to show that if S^c also contains $t+1$ points in general position, then $|S| > \frac{1}{2}q^t$ and $|S^c| > \frac{1}{2}q^t$, which will provide a contradiction. Observe that the hypothesis of part (iii) is symmetric between S and S^c, since S meets l in at least two points if and only if $|S^c \cap l| < q-1$.

We will prove this using induction on t. The statement is trivially true for $t = 1$. When $t = 1$, $S \subseteq \mathbb{F}_q$ and $|S| \geq 2$. Since $|S \cap l| \geq q-1$ for every line l meeting S in at least two points, it follows that $|S^c| \leq 1$. Thus S^c is contained in a hyperplane.

For our inductive hypothesis, suppose that for any affine space $L \subset \mathbb{F}_q^t$ of dimension $t-1$, either

- $R \cap L$ fails to contain t points in general position
- $R \cap L$ contains t points in general position

If the first, $R \cap L$ lies in a proper affine space of L. If the latter, then $L \setminus R = L \cap R^c$ lies in a proper affine space of L by induction. Thus either $L \cap R$ or $L \cap R^c$ lies in a proper affine space of L.

Let L_0 be the $(t-1)$-dimensional subspace generated by $P_1, \ldots, P_t$. When $L = L_0$, we must be in the second case, where $L \cap R^c$ lies in a proper affine space of L.

For every $P \in L_0 \cap R$, the line l generated by P and P_0 meets R in at least two points (namely P and P_0). Thus $|R \cap l| \geq q-1$ by hypothesis. Note for distinct choices of P the sets $l \setminus \{P_0\}$ are disjoint. Thus,

$$|R| \geq 1 + (q-2)|L_0 \cap R|. \tag{2.2}$$

Now suppose that every affine space L of dimension $t-1$, parallel to, but not equal to L_0 has the property that $L \cap R$ lies in a proper affine space of L. Then since $\mathbb{F}_q^t$ is a disjoint union of L_0 with the various spaces L, we see that $|R| \leq |L_0 \cap R| + (q-1)q^{t-2}$.

Comparing this with equation 2.2 yields the following

$$1 + (q-2)|L_0 \cap R| \leq |L_0 \cap R| + (q-1)q^{t-2}$$

$$1 + (q-3)|L_0 \cap R| \leq (q-1)q^{t-2}$$

$$(q-3)|L_0 \cap R| < (q-1)q^{t-2}$$

Now we will show $|L_0 \cap R| \geq q^{t-1} - q^{t-2}$. Recall $|L_0 \cap R^c| \leq q^{t-2}$. Since $(L_0 \cap R^c) \cup (L_0 \cap R) = L_0$ and $|L_0| = q^{t-1}$, we find

$$|L_0 \cap R^c| + |L_0 \cap R| = q^{t-1}$$

$$|L_0 \cap R| = q^{t-1} - |L_0 \cap R^c|$$

$$|L_0 \cap R| \geq q^{t-1} - q^{t-2}$$

Since $|L_0 \cap R| \geq q^{t-1} - q^{t-2}$, we find that $q < 4$ by the following computation:

$$(q-3)(q^{t-1} - q^{t-2}) \leq (q-3)|L_0 \cap R| < (q-1)q^{t-2}$$

$$q^t - 4q^{t-1} + 3q^{t-2} < q^{t-1} - q^{t-2}$$

$$q^{t-2}(q^2 - 4q + 3) < (q-1)q^{t-2}$$

$$(q^2 - 4q + 3) < (q-1)$$

$$q^2 - 5q + 4 < 0$$

$$(q-1)(q-4) < 0$$

We know $q > 1$, so $(q-1) > 0$. Thus $(q-4) < 0$, which implies $q < 4$.

This contradicts our initial assumption, thus there exists at least one affine space L_1 parallel but not equal to L_0 for which $L_1 \cap R^c$ is contained in a proper affine space of L_1.

If we pick any point $Q \in L_1 \cap R$ and count points of R on lines from Q to $L_0 \cap R$, then we will obtain at least $(q-2)|L_0 \cap R|$ points of R not lying in L_1, by the argument that established equation 2.2. Taking into account points in $L_1 \cap R$, we find that $|R| \geq (q-2)|L_0 \cap R| + |L_1 \cap R|$.

However, we have arranged that $L_0 \cap R^c$ and $L_1 \cap R^c$ are both contained in proper affine spaces, so that $|L_0 \cap R| \geq q^{t-1} - q^{t-2}$ and similarly for $|L_1 \cap R|$. It then follows that

$$\begin{aligned}
|R| &\geq (q-2)|L_0 \cap R| + |L_1 \cap R| \\
&\geq (q-2)(q^{t-1} - q^{t-2}) + (q^{t-1} - q^{t-2}) \\
&= (q-1)(q^{t-1} - q^{t-2}) \\
&= (q-1)^2 q^{t-2}
\end{aligned}$$

We will now show that $(q-1)^2 q^{t-2} > \frac{1}{2} q^t$. Consider the following string of inequalities. Note that the first inequality holds since $q \geq 4$.

$$\begin{aligned}
q(q-4) + 2 &> 0 \\
q^2 - 4q + 2 &> 0 \\
2q^2 - 4q + 2 &> q^2 \\
(q-1)^2 &> \frac{1}{2} q^2 \\
(q-1)^2 q^{t-2} &> \frac{1}{2} q^t
\end{aligned}$$

.

As explained above, this inequality leads to the assertion made in part (iii) of the lemma.

(iv) We will prove this using induction on t. The statement is trivially true for $t = 1$. When $t = 1$, $S \subseteq \mathbb{F}_2$ and $|S| \geq 2$. If $m \geq 2$ is an integer and if $|S \cap l| \geq m + 1$ for every line l meeting S in at least two points, then $|S| \geq \dfrac{m^2 - 1}{m - 1} = m + 1$.

For our inductive hypothesis, assume S contains an affine space L_0 of dimension $t - 1$ together with a point $P_0 \notin L_0$ such that $|L_0| \geq \dfrac{m^t - 1}{m - 1}$.

If $P \in L_0$, the line generated by P and P_0 contains at least two points of S, and hence contains at least $m + 1$ such points. For different points P the sets $l \setminus \{P_0\}$ are disjoints. Hence

$$
\begin{aligned}
|S| &\geq 1 + m|L_0| \\
&\geq 1 + m\Big(\frac{m^t - 1}{m - 1}\Big) \\
&= 1 + \frac{m^{t+1} - m}{m - 1} \\
&= \frac{m - 1 + m^{t+1} - m}{m - 1} \\
&= \frac{m^{t+1} - 1}{m - 1}
\end{aligned}
$$

□

The following result is given in [9] and is used frequently as a tool in the proofs of Theorem 2.1.8 and Theorem 2.2.1.

Lemma 2.2.6 (Heath–Brown, [9])**.** *Let L be an affine space of dimension d. If $n \geq d$ and $N(\boldsymbol{f}; L) = v$ where $1 \leq v \leq (q-1)$, then $N(\boldsymbol{f}; L^*) \geq v$ for every affine space L^* of dimension d parallel to L. As a result, $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq vq^{n-d}$.*

Proof. Let $N(\boldsymbol{f}; L) = v$ where $1 \leq v \leq q - 1$. Assume $N(\boldsymbol{f}; L^*) < v$ for some affine d-dimensional space parallel to L. By Theorem 2.2.2, $N(\boldsymbol{f}; L) \equiv N(\boldsymbol{f}; L^*) \pmod q$. ` This is a contradiction, since $1 \leq N(\boldsymbol{f}; L) - N(\boldsymbol{f}; L^*) \leq q - 1$.

Since we can cover $\mathbb{F}_q^n$ with d-dimensional affine spaces that lie in the same subspace, it follows that

$$
\begin{aligned}
N(\boldsymbol{f}; \mathbb{F}_q^n) &= N(\boldsymbol{f}; L) + \sum_{L^*} N(\boldsymbol{f}; L^*) \\
&= v + \sum_{L^*} N(\boldsymbol{f}; L^*) \\
&\geq v + v(q^{n-d} - 1) \\
&= vq^{n-d}
\end{aligned}
$$

□

The following result is require to prove the next lemma.

Lemma 2.2.7. *For $x > 0$, $v > 0$, $\left(1 + \frac{x}{v}\right)^v < e^x$.*

Proof. By setting $y = \frac{x}{v}$, it is sufficient to show that $(1 + y) < e^y$. Notice we get equality when $y = 0$. Taking the derivative of both sides, we know $1 < e^y$ for $y > 0$. Thus, we have the desired result. □

The following result is stated without proof in [9].

Lemma 2.2.8. *Suppose $v \in \mathbb{Z}$. Let $v \geq 2$ and $q \geq 2v+1$, then $\left\lfloor \frac{qv}{v+1} \right\rfloor^v > \frac{q^v}{v+1}$ with the exception of $v = 2$, $q = 5, 7$.*

Proof. First we will demonstrate that it is sufficient to show

$$v + 1 > \Big(1 + \frac{1}{q-1}\Big)^v \Big(1 + \frac{1}{v}\Big)^v$$

.

Consider

$$v + 1 > \Big(1 + \frac{1}{q-1}\Big)^v \Big(1 + \frac{1}{v}\Big)^v = \Big(\frac{q}{q-1}\Big)^v \Big(\frac{v+1}{v}\Big)^v.$$

Taking the v^{th} root of both sides yields

$$(v+1)^{\frac{1}{v}} > \Big(\frac{q}{q-1}\Big)\Big(\frac{v+1}{v}\Big).$$

Rearranging these terms through multiplication, we find

$$\frac{q}{(v+1)^{\frac{1}{v}}} < \frac{(q-1)v}{v+1} = \frac{qv}{v+1} - \frac{v}{v+1} \leq \left\lfloor \frac{qv}{v+1} \right\rfloor.$$

Assume $q \geq 2v + 1$. Then

$$1 + \frac{1}{q-1} \leq 1 + \frac{1}{2v} = 1 + \frac{\frac{1}{2}}{v}$$

$$\Big(1 + \frac{1}{q-1}\Big)^v \leq \Big(1 + \frac{\frac{1}{2}}{v}\Big)^v < e^{\frac{1}{2}}$$

Thus, by Lemma 2.2.7, we find

$$\begin{aligned}\Big(1 + \frac{1}{q-1}\Big)^v \Big(1 + \frac{1}{v}\Big)^v &< e^{\frac{1}{2}} e \\ &= e^{\frac{3}{2}} \\ &< 5 \\ &\leq v + 1 \text{ for } v \geq 4\end{aligned}$$

Now let $v = 3$. Thus, we find

$$\Big(1 + \frac{1}{q-1}\Big)^v \Big(1 + \frac{1}{v}\Big)^v < e^{\frac{1}{2}} \Big(\frac{4}{3}\Big)^3 = e^{\frac{1}{2}} \frac{64}{27} < 3.91 < 4 = v + 1.$$

Now let $v = 2$. Notice that

$$\Big(\frac{q}{q-1}\Big)^v \Big(\frac{v+1}{v}\Big)^v = \frac{9}{4}\Big(\frac{q}{q-1}\Big)^2$$

We want to know for what values of q is the following inequality satisfied.

$$\frac{9}{4}\left(\frac{q}{q-1}\right)^2 < 3$$

This inequality is satisfied if and only if

$$\left(\frac{q}{q-1}\right)^2 < \frac{4}{3}$$

This inequality is true if and only if $q \geq 8$.

$\square$

For convenience, we restate Heath–Brown's result, Theorem 2.1.8.

Theorem 2.1.8 (Heath-Brown, [9], Theorem 2)**.** Suppose that $n > d$ and that $Z(\boldsymbol{f}; \mathbb{F}_q^n)$ is non-empty, and is not an affine space of $\mathbb{F}_q^n$. Then

(i) For any q we have $N(\boldsymbol{f}; \mathbb{F}_q^n) > q^{n-d}$;

(ii) If $q \geq 4$ we have $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq 2q^{n-d}$;

(iii) For any q we have $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq \dfrac{q^{n+1-d}}{(n+2-d)}$ provided that the polynomials $f_1, \ldots, f_r$ are homogeneous.

We now present an improvement to Theorem 2.1.8 parts (i), (ii), and (iii), and introduce a new result. Notice that we have shown that part (ii) holds for $q = 3$ and thus part (i) is only relevant when $q = 2$. Furthermore, we have shown that part (iii) is never an optimal bound. Finally, we demonstrated sharpness of several of the improved bounds. The proof of Theorem 2.2.1 closely follows the ideas of Heath–Brown's proof.

Theorem 2.2.1 (Leep-Petrik)**.** Suppose that $n > d$ and that $Z(\boldsymbol{f}; \mathbb{F}_q^n)$ is non-empty, and is not an affine space of $\mathbb{F}_q^n$. Then

(i) For $q = 2$, we have $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq q^{n-d} + q$;

(ii) For $q \geq 3$, we have $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq 2q^{n-d}$;

(iii) For any q, we have $N(\boldsymbol{f}; \mathbb{F}_q^n) > \dfrac{q^{n+1-d}}{(n+2-d)}$ provided that the polynomials $f_1, \ldots, f_r$ are homogeneous.

(iv) For $q \geq 3$, we have $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq 2q^{n-d} + (q-2)q$ provided that the polynomials $f_1, \ldots, f_r$ are homogeneous;

Moreover, the bounds in (i), (ii), and (iv) are sharp under these hypotheses.

Proof. In this proof, we are required to check many cases, so we developed a convenient labeling to easily track where we are in the proof. Case IIB2(b) means we are in the proof of part(ii), Case B, Subcase 2, Subsubcase b.

Since we satisfy the hypothesis of Warning's corollary, we know that $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq q^{n-d}$. Since $Z(\boldsymbol{f}; \mathbb{F}_q^n)$ is not an affine space of $\mathbb{F}_q^n$, we know that $Z(\boldsymbol{f}; \mathbb{F}_q^n)$ contains a maximal set of $t \geq n+1-d$ points in general position. Let $\mathbb{F}_q^t$ be the space spanned by these points. Let $S = Z(\boldsymbol{f}; \mathbb{F}_q^t)$.

(i) **Case IA** Suppose the hypotheses of Lemma 2.2.5, parts (i),(ii) are satisfied. Then $S = \mathbb{F}_q^t$. Thus

$$N(\boldsymbol{f}; \mathbb{F}_q^n) \geq N(\boldsymbol{f}; \mathbb{F}_q^t) = |S| = q^t \geq q^{n+1-d} = q(q^{n-d}) \geq 2(q^{n-d}) = q^{n-d} + q^{n-d} \geq q^{n-d} + q$$

Then the only cases we have left to consider are the following:

Case IB Suppose the hypothesis of Lemma 2.2.5, part (i) is not satisfied. That is, if $q = 2$ and there exists a 2-plane $L_0 \subseteq \mathbb{F}_q^n$ with $N(\boldsymbol{f}; L_0) = 3$.

Case IC Suppose the hypothesis of Lemma 2.2.5, part(ii) is not satisfied. That is, if $q \geq 3$ and there is a line $L_0 \subseteq \mathbb{F}_q^n$ with $2 \leq N(\boldsymbol{f}; L_0) \leq q-1$.

Case IB Take L to be as in Lemma 2.2.3. Then $N(\boldsymbol{f}; L) = N(\boldsymbol{f}; L_0) = 3$. Note $\dim(L) = k \leq d$. For if $\dim(L) > d$ then $2 | N(\boldsymbol{f}; L)$ [1]. $\lightning$

Then Lemma 2.2.3 yields the following bound:

$$\begin{aligned}
N(\boldsymbol{f}; \mathbb{F}_q^n) &\geq N(\boldsymbol{f}; L) + \frac{q^{n-k}-1}{q-1}(N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L)) \\
&\geq N(\boldsymbol{f}; L) + \frac{q^{n-k}-1}{q-1} \\
&\geq 3 + \frac{q^{n-k}-1}{q-1} \\
&= 3 + \frac{q^{n-d}-1}{q-1} \\
&\geq 3 + 2^{n-d} - 1 \\
&= 2^{n-d} + 2
\end{aligned}$$

Case IC Take L to be as in Lemma 2.2.3. Then $2 \leq N(\boldsymbol{f}; L_0) = N(\boldsymbol{f}; L) \leq q-1$. Again notice $\dim(L) = k \leq d$. We will consider the following three cases:

Case IC(1) Suppose $k \leq d-2$. Then we can apply Lemma 2.2.3. This yields

$$\begin{aligned} N(\boldsymbol{f}; \mathbb{F}_q^n) &\geq N(\boldsymbol{f}; L) + \frac{q^{n-k}-1}{q-1}(N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L)) \\ &\geq \frac{q^{n+2-d}-1}{q-1} \\ &= q^{n+1-d} + q^{n-d} + \cdots + 1 \\ &\geq q^{n+1-d} + q^{n-d} \\ &\geq q^{n-d} + q \end{aligned}$$

Case IC(2) Suppose $k = d-1$. Then either:

(a) $N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L) = 1$ and thus $3 \leq N(\boldsymbol{f}; L') \leq q-1$, or

(b) $N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L) \geq 2$

Case IC2(a)) Suppose (a) holds. Since $\dim(L) = d-1$, this implies $\dim(L')$=d. Thus by Lemma 2.2.6, we have that

$$N(\boldsymbol{f}; \mathbb{F}_q^n) \geq 3q^{n-d} \geq q^{n-d} + q$$

Case IC2(b)) Suppose (b) holds. We may apply Lemma 2.2.3. By Lemma 2.2.3,

$$\begin{aligned} N(\boldsymbol{f}; \mathbb{F}_q^n) &\geq N(\boldsymbol{f}; L) + \frac{q^{n-k}-1}{q-1}(N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L)) \\ &\geq \frac{q^{n-k}-1}{q-1}(N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L)) \\ &\geq \frac{q^{n-(d-1)}-1}{q-1}(N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L)) \\ &= \frac{q^{n+1-d}-1}{q-1}(N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L)) \\ &= (q^{n-d} + q^{n-d-1} + \cdots + q + 1)(2) \\ &\geq 2q^{n-d} \\ &\geq q^{n-d} + q \end{aligned}$$

Case IC(3) Suppose $k = d$. By Lemma 2.2.6,

$$N(\boldsymbol{f}; \mathbb{F}_q^n) \geq vq^{n-d} = 2q^{n-d} \geq q^{n-d} + q$$

Remark: If we use Theorem 2.1.5, the result follows more easily because $q^{n-d}+q$ is the first integer larger that q^{n-d} that is divisible by q. However, the proof presented above uses more elementary results.

(ii) **Case IIA** Suppose the hypothesis of Lemma 2.2.5, part(ii) are satisfied. Then

$$N(\boldsymbol{f}; \mathbb{F}_q^n) \geq N(\boldsymbol{f}; \mathbb{F}_q^t) = |\mathbb{F}_q^t| = q^t \geq q^{n+1-d} \geq q(q^{n-d}) > 2q^{n-d}$$

Case IIB Suppose the hypothesis of Lemma 2.2.5, part (ii) is not satisfied. That is, if $q \geq 3$, there is a line $L_0 \subseteq \mathbb{F}_q^n$ with $2 \leq N(\boldsymbol{f}; L_0) \leq q-1$. Take L to be as in Lemma 2.2.3. Then $2 \leq N(\boldsymbol{f}; L) = N(\boldsymbol{f}; L_0) \leq q-1$. Note $\dim(L) = k \leq d$. Then there exist three subcases:

Case IIB(1) Suppose $k \leq d-2$. By Lemma 2.2.3,

$$\begin{aligned} N(\boldsymbol{f}; \mathbb{F}_q^n) &\geq N(\boldsymbol{f}; L) + \frac{q^{n-k}-1}{q-1}(N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L)) \\ &\geq \frac{q^{n-k}-1}{q-1} \geq \frac{q^{n-(d-2)}-1}{q-1} \\ &= \frac{q^{n+2-d}-1}{q-1} = q^{n+1-d} + q^{n-d} + \cdots + q + 1 \\ &> q^{n+1-d} = q(q^{n-d}) \\ &\geq 3q^{n-d} > 2q^{n-d} \end{aligned}$$

Case IIB(2) Suppose $k = d-1$. Then either

(a) $N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L) = 1$ and thus $3 \leq N(\boldsymbol{f}; L') \leq q$, or

(b) $N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L) \geq 2$

Case IIB2(a) Suppose (a) holds. Then $\dim(L') = d$. Since $3 \leq N(\boldsymbol{f}; L') \leq q$, we find that $N(\boldsymbol{f}; L^*) \geq 3$ for every affine space of dimension d parallel to L'. Thus, by Lemma 2.2.6, we have that $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq 3q^{n-d} > 2q^{n-d}$.

Case IIB2(b) Suppose (b) holds. Then Lemma 2.2.3 gives the following bound:

$$\begin{aligned} N(\boldsymbol{f}; \mathbb{F}_q^n) &\geq N(\boldsymbol{f}; L) + \frac{q^{n-k}-1}{q-1}(N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L)) \\ &\geq \frac{q^{n-k}-1}{q-1}(N(\boldsymbol{f}; L') - N(\boldsymbol{f}; L)) \\ &= \frac{q^{n+1-d}-1}{q-1} \times 2 \\ &= 2(q^{n-d} + q^{n-d-1} + \cdots + q + 1) \\ &> 2q^{n-d} \end{aligned}$$

Case IIB(3) Suppose $k = d$. Then $2 \leq N(\boldsymbol{f}; L) \leq q-1$. Then by Lemma 2.2.6, we find $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq 2q^{n-d}$.

(iii) **Case IIIA** Suppose the hypothesis of Lemma 2.2.5, part (iv) are satisfied. Applying Lemma 2.2.5, part (iv) with an integer $m \leq q-1$ to be chosen later yields,

$$\begin{aligned} N(\boldsymbol{f}; \mathbb{F}_q^n) &\geq \frac{m^{t+1}-1}{m-1} \\ &\geq \frac{m^{n+2-d}-1}{m-1} \\ &= m^{n+1-d} + m^{n-d} + \cdots + m + 1 \\ &> m^{n+1-d} \end{aligned}$$

Case IIIB Suppose the hypothesis of Lemma 2.2.5, part (iv) are not satisfied. That is, if $m \geq 2$ is an integer and there is a line $L_0 \subseteq \mathbb{F}_q^n$ with $2 \leq N(\boldsymbol{f}; L_0) \leq m$. Take L to be as in Lemma 2.2.3. Then $2 \leq N(\boldsymbol{f}; L) = N(\boldsymbol{f}; L_0) \leq m$. Note $\dim(L) = k \leq d$. Then we have the following 3 subcases:

Case IIIB(1) Suppose $k \leq d-2$. Then Lemma 2.2.3 yields the following bound:

$$\begin{aligned} N(\boldsymbol{f}; \mathbb{F}_q^n) &\geq \frac{q^{n+2-d}-1}{q-1}(N(\boldsymbol{f}; L') - m) \\ &= \frac{q^{n+2-d}-1}{q-1} \\ &= q^{n+1-d} + q^{n-d} + \cdots + q + 1 \\ &> q^{n+1-d} \end{aligned}$$

Case IIIB(2) Suppose $k = d-1$. Then either

(a) $N(\boldsymbol{f}; L') \geq q$, or

(b) $m + 1 \leq N(\boldsymbol{f}; L') \leq q - 1$

Case IIIB2(a) Suppose (a) holds. Then Lemma 2.2.3 gives the following bound:

$$\begin{aligned} N(\boldsymbol{f}; \mathbb{F}_q^n) &\geq \frac{q^{n-k}-1}{q-1}(N(\boldsymbol{f}; L') - m) \\ &\geq \frac{q^{n+1-d}-1}{q-1}(q-m) \\ &= (q-m)(q^{n-d} + q^{n-d-1} + \cdots + q + 1) \\ &> (q-m)q^{n-d} \end{aligned}$$

Case IIIB2(b) Suppose (b) holds. Then $\dim(L') = d$. Since $m+1 \leq N(\boldsymbol{f}; L') \leq q-1$, we find that $N(\boldsymbol{f}; L^*) \geq m+1$ for every affine space of dimension d parallel to L'. Thus by Lemma 2.2.6, we have that $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq (m+1)q^{n-d}$.

Case IIIB(3) Suppose $k = d$. In fact, we will show that this cannot occur. We will consider the following two cases:

Case IIIB3(a) Suppose $\mathbf{0} \in L$. Then $q-1 | N(\boldsymbol{f}; L) - 1$. This is because if $\boldsymbol{x} \in Z(\boldsymbol{f}; L)$, then every scalar multiple of $\boldsymbol{x}$ is also in $Z(\boldsymbol{f}; L)$. Since $1 \leq N(\boldsymbol{f}; L) - 1 \leq m-1 \leq q-2$, we reach a contradiction. ↯

Case IIIB3(b) Suppose $\mathbf{0} \notin L$. Consider the $(d+1)$-dimensional affine space $L' :=< L, \mathbf{0} >$. Here we find that $N(\boldsymbol{f}; L') = 1 + (q-1)N(\boldsymbol{f}; L)$. According to Ax's improvement, we have $q | N(\boldsymbol{f}; L')$. Thus $N(\boldsymbol{f}; L) \equiv 1 \bmod q$. ↯This is a contradiction since $2 \leq N(\boldsymbol{f}; L) \leq m$.

It follows that one of the inequalities must hold. Notice that the inequality produced in **Case IIIB(1)** is already good enough for the theorem and does not rely on the choice of m.

1. **Case IIIA**: $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq m^{n+1-d}$
2. **Case IIIB2(a)**: $N(\boldsymbol{f}; \mathbb{F}_q^n) > (q-m)q^{n-d}$
3. **Case IIIB2(b)**: $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq (m+1)q^{n-d}$

The required estimate now follows by choosing $m = \left\lfloor \dfrac{q(n+1-d)}{n+2-d} \right\rfloor$ and applying Lemma 2.2.8.

1. To prove inequality 1, we need to apply Lemma 2.2.8

$$N(\boldsymbol{f}; \mathbb{F}_q^n) \geq m^{n+1-d} = \left\lfloor \frac{q(n+1-d)}{n+2-d} \right\rfloor^{n+1-d} > \frac{q^{n+1-d}}{n+2-d}$$

2. To prove inequality 2, consider the following:

$$\begin{aligned} N(\boldsymbol{f}; \mathbb{F}_q^n) &\geq (q-m)q^{n-d} \\ &= \left(q - \left\lfloor \frac{q(n+1-d)}{n+2-d} \right\rfloor\right) q^{n-d} \\ &\geq \left(q - \frac{q(n+1-d)}{n+2-d}\right) q^{n-d} \\ &= q^{n+1-d} - \frac{q^{n+1-d}(n+1-d)}{n+2-d} \\ &= \frac{q^{n+1-d}}{n+2-d} \end{aligned}$$

3. To prove inequality 3, consider the following:

$$\begin{aligned}
N(\boldsymbol{f}; \mathbb{F}_q^n) &\geq (m+1)q^{n-d} \\
&= \left(\left\lfloor \frac{q(n+1-d)}{n+2-d} \right\rfloor + 1\right) q^{n-d} \\
&\geq \left(\frac{q(n+1-d)}{n+2-d}\right) q^{n-d} \\
&= \frac{q^{n+1-d}(n+1-d)}{n+2-d} \\
&\geq \frac{2q^{n+1-d}}{n+2-d} \\
&> \frac{q^{n+1-d}}{n+2-d}
\end{aligned}$$

It remains to check when $q = 2, 3, 4, 5, 7$.
When $q = 2$

$$N(\boldsymbol{f}; \mathbb{F}_2^n) > 2^{n-d} = \frac{2^{n+1-d}}{2} > \frac{2^{n+1-d}}{n+2-d}$$

When $q = 3$

$$N(\boldsymbol{f}; \mathbb{F}_3^n) > 3^{n-d} = \frac{3^{n+1-d}}{3} \geq \frac{3^{n+1-d}}{n+2-d}$$

When $q = 4$

$$N(\boldsymbol{f}; \mathbb{F}_4^n) \geq 2(4)^{n-d} = \frac{2(4)^{n+1-d}}{4} = \frac{4^{n+1-d}}{2} > \frac{4^{n+1-d}}{n+2-d}$$

When $q = 5$

$$N(\boldsymbol{f}; \mathbb{F}_5^n) \geq 2(5)^{n-d} = \frac{2(5)^{n+1-d}}{5} = \frac{2}{5}(5^{n+1-d}) > \frac{1}{n+2-d}(5^{n+1-d}) = \frac{5^{n+1-d}}{n+2-d}$$

When $q = 7$ and $v = n+1-d = 2$. Since $\boldsymbol{f}$ is a system of homogeneous forms, we know $N(\boldsymbol{f}; \mathbb{F}_7^n) \equiv 1 \mod 6$. Furthermore, by Lemma 2.2.6, we know $N(\boldsymbol{f}; \mathbb{F}_7^n) \geq 14$. Thus, $N(\boldsymbol{f}; \mathbb{F}_7^n) \geq 19 > \dfrac{7^2}{3}$.

(iv) Let $N = N(\boldsymbol{f}, \mathbb{F}_q^n)$ and let P be the number of projective zeros. We know that $N = P(q-1) + 1$. By Theorem 2.2.1, part (ii), we know $N \geq 2q^{n-d}$. By Theorem 2.1.5, we know $N = 2q^{n-d} + aq$ for some $a \in \mathbb{Z}_{\geq 0}$. If we consider these relationships mod $q-1$, we find $N \equiv 1 \mod q-1$ and $q \equiv 1 \mod q-1$. Thus, $2+a \equiv 1 \mod q-1$, which implies $a \equiv -1 \mod q-1$. Since $a \in \mathbb{Z}_{\geq 0}$, we know $a \geq q-2$, which gives the

desired result.

Now we will demonstrate that bounds (i), (ii) and (iv) are sharp under these hypotheses.
Sharpness of Part (i) Consider the polynomial

$$f(x_1, x_2, x_3, x_4) = x_1x_2 + x_3^2 + x_3x_4 + x_4^2$$

over $\mathbb{F}_2$. There are $2^3 - 2^2 + 2 = 6$ solutions to this equation. Since there are 6 solutions, we know that the set of solutions does not form a linear subspace of $\mathbb{F}_2^4$. The lower bound given in part (i) of the theorem is $q^{n-d} + q = 2^{4-2} + 2 = 6$.

Sharpness of Part (ii) Consider the polynomial

$$g(x, y, z) = xy + z^2 + 1$$

over $\mathbb{F}_3$. There are exactly $3 + 3 = 6$ solutions over $\mathbb{F}_3$. Since there are 6 solutions, we know that the set of solutions does not form a linear subspace of $\mathbb{F}_3^3$. The lower bound given in part (ii) of the theorem is $2q^{n-d} = 2(3)^{3-2} = 6$.

Sharpness of Part (iv) Consider the polynomial

$$h(x, y, z) = xy - z^2$$

over $\mathbb{F}_q$. There are exactly $(q - 1)q + q = q^2$ solutions to this equation. The lower bound given in part (iii) of the theorem is $2q^{n-d} + (q - 2)q = 2q + (q - 2)q = q^2$.

□

One might wonder how part (iii) of Theorem 2.2.1 compares to part (iv) of Theorem 2.2.1. In particular, one might ask if both results are necessary or if one is a subset of the other. Thus, we will take the time to discuss these two results. For convenience we have restated parts (iii) and (iv).

Theorem 2.2.1 (Leep-Petrik)**.** Suppose that $n > d$ and that $Z(\boldsymbol{f}; \mathbb{F}_q^n)$ is non-empty, and is not an affine space of $\mathbb{F}_q^n$. Then

(iii) For any q, we have $N(\boldsymbol{f}; \mathbb{F}_q^n) > \dfrac{q^{n+1-d}}{(n+2-d)}$ provided that the polynomials $f_1, \ldots, f_r$ are homogeneous.

(iv) For $q \geq 3$, we have $N(\boldsymbol{f}; \mathbb{F}_q^n) \geq 2q^{n-d} + (q-2)q$ provided that the polynomials $f_1, \ldots, f_r$ are homogeneous;

First observe that when $n - d = 1$, part (iii) reduces to $\frac{q^2}{3}$ and part (iv) reduces to q^2. Thus, when $n - d = 1$, part (iv) is a better, that is, larger lower bound. Secondly, when $n - d$ is "small" and q is "small", part (iv) will be a better lower bound. Otherwise, part (iii) is a better, that is, larger lower bound. This should make sense intuitively as the power of q appearing in part (iii) is larger than the power of q appearing in part (iv), so asymptotically, part (iii) is stronger as q gets large.

Chapter 3 Existence of Nontrivial Solutions for Diagonal Forms of Odd Degree

3.1 Introduction

The goal of this chapter is to study criteria on the number of variables needed to guarantee a nontrivial solution to an odd diagonal form. When restricting to the case of diagonal forms, one can obtain much deeper and more specialized results. Diagonal forms have been studied extensively in a variety of contexts. Of particular interest are results on the existence of nontrivial solutions. In other words, we wish to find bounds for or exact values of $\Omega(d, q)$ when d is odd. Some preliminary results in this direction are the following results. In [8], Gray showed the existence of nontrivial zeros in $\mathbb{F}_q$ of the following two forms

$$\alpha_1 x_1^p + \cdots + \alpha_p x_p^p$$

where $\alpha_i \in \mathbb{F}_q$, $p \geq 3$, and p is an odd prime, and

$$\alpha_1 x_1^p + \cdots + \alpha_{p-1} x_{p-1}^p$$

where $\alpha_i \in \mathbb{F}_q$, $p \geq 5$, and p is an odd prime. Moreover, Gray [6, 7] provides two lower bounds for the number of variables needed to guarantee the existence of a nontrivial solution.

Theorem 3.1.1 (Gray, [6]). *Let $d \mid q-1$, where d is an odd prime. The equation*

$$\alpha_1 x_1^d + \alpha_2 x_2^d + \cdots + \alpha_n x_n^d = 0$$

has a nontrivial solution over $\mathbb{F}_q$ if $n \geq d+4-(\lfloor 2\sqrt{d+2}\rfloor)$, where $\alpha_i \in \mathbb{F}_q^*$. That is,

$$\Omega(1, d, q) < d + 4 - (\lfloor 2\sqrt{d+2}\rfloor).$$

Theorem 3.1.2 (Gray, [7]). *Let $d \mid q-1$, where d is an odd prime. The equation*

$$\alpha_1 x_1^d + \alpha_2 x_2^d + \cdots + \alpha_n x_n^d = 0$$

has a nontrivial zero in $\mathbb{F}_q$ if $n \geq d+4-\Big(\Big\lfloor \frac{d+1}{2M}\Big\rfloor + 2M\Big)$, where $M = \Big\lfloor \frac{1+\sqrt{d+2}}{2}\Big\rfloor$ and $\alpha_i \in \mathbb{F}_q$. That is,

$$\Omega(1, d, q) < d + 4 - \Big(\Big\lfloor \frac{d+1}{2M}\Big\rfloor + 2M\Big).$$

In the following section, we will show that these results are equivalent and present an improvement to Theorem 3.1.1. We will conclude the following section with a graphical comparison of the improvement and Chevalley's Theorem (Theorem 2.1.1).

3.2 Results for Diagonal Forms of Odd Degree

We show that the previously mentioned bounds by Gray are equivalent with the following theorem.

Theorem 3.2.1 (Leep–Petrik)**.** *Theorem 3.1.1 and Theorem 3.1.2 are equivalent.*

Proof. In order to show Theorem 3.1.1 and Theorem 3.1.2 are equivalent, we need to show that

$$d + 4 - (\lfloor 2\sqrt{d+2} \rfloor) = d + 4 - \left(\left\lfloor \frac{d+1}{2M} \right\rfloor + 2M\right),$$

where $M = \left\lfloor \dfrac{1+\sqrt{d+2}}{2} \right\rfloor$. Rearranging terms it is sufficient to show

$$\lfloor 2\sqrt{d+2} \rfloor = \left\lfloor \frac{d+1}{2M} \right\rfloor + 2M, \tag{3.1}$$

where M is defined as above.

Since M is the integer part of $\dfrac{1+\sqrt{d+2}}{2}$, we may write $\dfrac{1+\sqrt{d+2}}{2} = M + \epsilon$, where $0 \le \epsilon < 1$. Then we can find

$$\begin{aligned}
\frac{1+\sqrt{d+2}}{2} &= M + \epsilon \\
1 + \sqrt{d+2} &= 2M + 2\epsilon \\
\sqrt{d+2} &= 2M - 1 + 2\epsilon
\end{aligned}$$

Observe that since d is odd, $\epsilon \neq \frac{1}{2}$. For if $\epsilon = \frac{1}{2}$, $\sqrt{d+2} = 2M$, which is a contradiction since $\sqrt{d+2}$ must be odd.

Now consider $d + 1$.

$$\begin{aligned}
d + 1 &= (\sqrt{d+2} + 1)(\sqrt{d+2} - 1) \\
&= (2M + 2\epsilon)(2M - 2 + 2\epsilon) \\
&= 4M^2 - 4M + 2\epsilon(4M - 2) + 4\epsilon^2 \\
&= 4M^2 - 4M + 4\epsilon(2M - 1) + 4\epsilon^2
\end{aligned}$$

Thus,

$$d + 1 = 4M^2 - 4M + 4\epsilon(2M - 1) + 4\epsilon^2. \tag{3.2}$$

By Equation 3.2, we find that

$$\begin{aligned}
\frac{d+1}{2M} &= 2M - 2 + \frac{4\epsilon(2M-1) + 4\epsilon^2}{2M} \\
\left\lfloor \frac{d+1}{2M} \right\rfloor + 2M &= 4M - 2 + \left\lfloor \frac{4\epsilon(2M-1) + 4\epsilon^2}{2M} \right\rfloor.
\end{aligned}$$

Now that we have an expression for the left hand side of Equation 3.1, we will work to find an expression for the right hand side. Recall from above that $\sqrt{d+2} = 2M - 1 + 2\epsilon$. Using this, we find that

$$\begin{aligned} \lfloor 2\sqrt{d+2} \rfloor &= \lfloor 2(2M-1) + 4\epsilon \rfloor \\ &= \lfloor 4M - 2 + 4\epsilon \rfloor \\ &= 4M - 2 + \lfloor 4\epsilon \rfloor. \end{aligned}$$

Now notice that we want to show that

$$4M - 2 + \lfloor 4\epsilon \rfloor = 4M - 2 + \left\lfloor \frac{4\epsilon(2M-1) + 4\epsilon^2}{2M} \right\rfloor,$$

so it is sufficient to show

$$\lfloor 4\epsilon \rfloor = \left\lfloor \frac{4\epsilon(2M-1) + 4\epsilon^2}{2M} \right\rfloor.$$

By Equation 3.2, we have

$$4\epsilon(2M-1) + 4\epsilon^2 = d + 1 - 4M^2 + 4M \in \mathbb{Z}.$$

Using the division algorithm, we know we can write $\dfrac{4\epsilon(2M-1)+4\epsilon^2}{2M} = q + \dfrac{r}{2M}$, where $q, r \in Z$, $q \geq 0$, and $0 \leq r \leq 2M - 1$.

We now observe that $0 \leq 4\epsilon(1-\epsilon) < 1$. It is clear that $0 \leq 4\epsilon(1-\epsilon)$. Since $\epsilon \neq \frac{1}{2}$, we have $(2\epsilon - 1)^2 > 0$. Thus, $4\epsilon^2 - 4\epsilon + 1 > 0$ and by rearranging terms, we can conclude

$$\begin{aligned} 4\epsilon - 4\epsilon^2 &< 1 \\ 4\epsilon(1-\epsilon) &< 1. \end{aligned}$$

Since $0 \leq r \leq 2M - 1$ and $0 \leq 4\epsilon(1-\epsilon) < 1$, we know $0 \leq r + 4\epsilon(1-\epsilon) < 2M$. Thus,

$$
\begin{aligned}
\left\lfloor \frac{4\epsilon(2M-1)+4\epsilon^2}{2M} \right\rfloor &= \left\lfloor q + \frac{r}{2M} \right\rfloor \\
&= \left\lfloor q + \frac{r+4\epsilon(1-\epsilon)}{2M} \right\rfloor \\
&= \left\lfloor \frac{4\epsilon(2M-1)+4\epsilon^2}{2M} + \frac{4\epsilon(1-\epsilon)}{2M} \right\rfloor \\
&= \left\lfloor \frac{4\epsilon \cdot 2M}{2M} \right\rfloor \\
&= \lfloor 4\epsilon \rfloor .
\end{aligned}
$$

Therefore, we obtain the desired result.

□

Furthermore, we improved Theorem 3.1.1 to the following result. Notice we demonstrated that d simply needs to be an odd integer, not an odd prime integer, for the result to hold.

Theorem 3.2.2 (Leep–Petrik)**.** *Let $d \mid q-1$, where $q = p^k$ and d is an odd integer. The equation*

$$\alpha_1 x_1^d + \alpha_2 x_2^d + \cdots + \alpha_n x_n^d = 0 \tag{3.3}$$

has a nontrivial solution over $\mathbb{F}_q$ if $n \geq d+4-(\lfloor 2\sqrt{d+2} \rfloor)$, where $\alpha_i \in \mathbb{F}_q^*$. That is,

$$\Omega(1,d,q) < d+4-(\lfloor 2\sqrt{d+2} \rfloor).$$

The proof of Theorem 3.2.2 closely follows the ideas of Gray's proof.

Theorem 3.2.3 (Leep–Petrik)**.** *Let $d \mid q-1$, where $q = p^k$ and d is an odd integer. If there exists $\lambda \in \mathbb{F}_q$ such that $\lambda(\mathbb{F}_q^*)^d$ is a generator of $\mathbb{F}_q^*/(\mathbb{F}_q^*)^d$ and $1+\lambda \notin (\mathbb{F}_q^*)^d$ and $1+\lambda \notin \lambda(\mathbb{F}_q^*)^d$, then for $t = \lfloor 2\sqrt{d+2} \rfloor - 4$, the form*

$$e_0 \lambda^0 x_0^d + \cdots + e_{d-1} \lambda^{d-1} x_{d-1}^d = 0 \tag{3.4}$$

has a nontrivial solution in $\mathbb{F}_q$, where

$$e_i = \begin{cases} 0 & \text{for } i \in T = \{i_1, i_2, \ldots, i_t\}, 2 \leq i_1 < i_2 < \cdots < i_t = d-1 \\ 1 & \text{for } i \in T_0 = D \setminus T, D = \{0,1,2,\ldots,d-1\} \end{cases}$$

and where the initial block $\{e_0, e_1, \ldots, e_{i_1-1}\}$ of nonzero elements has a maximal length among the blocks of consecutive nonzero elements in $\{e_0, e_1, \ldots, e_{d-1}\}$.

Theorem 3.2.4 (Leep–Petrik)**.** *Let $d \mid q-1$, $q=p^k$, and d is an odd integer. Then there exists a $\lambda \in \mathbb{F}_q$ such that $\lambda(\mathbb{F}_q^*)^d$ is a generator of $\mathbb{F}_q^*/(\mathbb{F}_q^*)^d$ and $1+\lambda \notin (\mathbb{F}_q^*)^d$ and $1+\lambda \notin \lambda(\mathbb{F}_q^*)^d$. Furthermore, Theorem 3.2.2 is a consequence of Theorem 3.2.3.*

Proof. We may assume that no two coefficients α_i and α_j belong to the same coset of $\mathbb{F}_q^*$ modulo $(\mathbb{F}_q^*)^d$, for then $\alpha_i^{-1}\alpha_j = a^d$ for some nonzero element $a \in \mathbb{F}_q^*$ and $x_i = a$, $x_j = -1$, $x_h = 0$ $(h \neq i \neq j)$ provides a solution.

First notice that the number of variables between the two equations matches. That is, in Theorem 3.2.2, Equation 3.3 has at least $d+4-(\lfloor 2\sqrt{d+2}\rfloor)$ variables. In Theorem 3.2.3, we have an equation in d variables, where $t = \lfloor 2\sqrt{d+2}\rfloor - 4$ of the variables do not appear. Thus, Equation 3.4 has $d-t = d+4-(\lfloor 2\sqrt{d+2}\rfloor)$ variables.

The quotient $\mathbb{F}_q^*/(\mathbb{F}_q^*)^d$ is cyclic of order d. We want to show that there is a $\lambda \in \mathbb{F}_q^*$, $\lambda \notin (\mathbb{F}_q^*)^d$, $1+\lambda \notin (\mathbb{F}_q^*)^d$ and $1+\lambda \notin \lambda(\mathbb{F}_q^*)^d$ such that $\lambda(\mathbb{F}_q^*)^d$ generates all the cosets of $\mathbb{F}_q^*/(\mathbb{F}_q^*)^d$. Let

$$W = \{x \in \mathbb{F}_q^* \mid x = z^d - 1, z \in \mathbb{F}_q^*\}.$$

Then $|W| = \dfrac{q-1}{d} - 1$. Let

$$W^{-1} = \{x \in \mathbb{F}_q^* \mid xy = 1, y \in W\},$$

and let

$$V = (\mathbb{F}_q^*)^d \cup W \cup W^{-1}.$$

Then, if $d \geq 5$,

$$|V| \leq \frac{q-1}{d} + 2\Big(\frac{q-1}{d} - 1\Big) = \frac{3q-2d-3}{d} \leq \frac{3(q-1)}{5} - 2 < q-1.$$

This means that

$$|V^c| = (q-1) - |V| \geq \Big(1 - \frac{3}{d}\Big)q + 1 + \frac{3}{d}.$$

Recall the number of generators of $\mathbb{F}_q^*/(\mathbb{F}_q^*)^d$ is $\varphi(d)$. Since $d \geq 5$, $\varphi(d) \geq 4$. Thus,

$$\begin{aligned}
\Big(\frac{\varphi(d)}{d} - \frac{3}{d}\Big)(q-1) + 1 &\geq 0\\
\Big(\frac{\varphi(d)}{d} - \frac{3}{d}\Big)q + 1 + \frac{3}{d} - \frac{\varphi(d)}{d} &\geq 0\\
-\Big(\frac{3}{d}\Big)q + 1 + \frac{3}{d} + \Big(\frac{q-1}{d}\Big)\varphi(d) &\geq 0\\
q - \Big(\frac{3}{d}\Big)q + 1 + \frac{3}{d} + \Big(\frac{q-1}{d}\Big)\varphi(d) &\geq q\\
\Big(1 - \frac{3}{d}\Big)q + 1 + \frac{3}{d} + \Big(\frac{q-1}{d}\Big)\varphi(d) &\geq q\\
|V| + \Big(\frac{q-1}{d}\Big)\varphi(d) &\geq q\\
|V| + \{\text{Generators of } \mathbb{F}_q^*/(\mathbb{F}_q^*)^d \text{ in } \mathbb{F}_q^*\} &\geq q.
\end{aligned}$$

This means that there is an element $\lambda \in \mathbb{F}_q^*$ such that $\lambda(\mathbb{F}_q^*)^d$ generates all the cosets of $\mathbb{F}_q^*/(\mathbb{F}_q^*)^d$ and $\lambda \notin V$. Since $\lambda \notin V$, $\lambda \notin (\mathbb{F}_q^*)^d$ and $1+\lambda \notin (\mathbb{F}_q^*)^d$. Moreover, since V is closed under the operation of taking inverses $\lambda^{-1} \notin V$.

Furthermore, we can show $1+\lambda \notin \lambda(\mathbb{F}_q^*)^d$. Assume $1+\lambda \in \lambda(\mathbb{F}_q^*)^d$. Then $(1+\lambda)\lambda^{-1} = 1+\lambda^{-1} \in (\mathbb{F}_q^*)^d$. Thus, $\lambda^{-1} \in W \subset V$, which is a contradiction. Thus, $1+\lambda \notin \lambda(\mathbb{F}_q^*)^d$. Hence, $1+\lambda = \lambda^{i_0}a^d$ for some i_0 $(2 \le i_0 \le d-1)$ and for some $a \in \mathbb{F}_q^*$. Notice a must be nonzero, for otherwise $\lambda = -1 = (-1)^d \in (\mathbb{F}_q^*)^d$. Thus, for this particular generator λ, the cosets of $\mathbb{F}_q^*/(\mathbb{F}_q^*)^d$ are $\{\lambda^i(\mathbb{F}_q^*)^d\}$ for $0 \le i \le d-1$ and Equation 3.3 may be rewritten in the form of Equation 3.4 with exactly t zero coefficients.

Now consider the coefficients $e_i\lambda^i$ of Equation 3.4 written counterclockwise in cyclic order; there is a maximal block of consecutive coefficients of length i_1 and beginning with λ^r. Multiplication by λ^{d-r} preserves the cyclic order of the coset representatives, preserves the relative location of the zeros and after renumbering of the e's yields $e_0\lambda^0, e_1\lambda^1, \ldots, e_{i_1-1}\lambda^{i_1-1}$ as a maximal block of consecutive nonzero coefficients, while $e_{d-1}\lambda^{d-1} = 0$. Hence multiplication of Equation 3.4 by λ^{d-r} enables us to rewrite it in the desired form.

Now let $T = \{i \mid e_i = 0\}$. Then $T = \{i_1, i_2, \ldots, i_t\}$ for $2 \le i_1 < i_2 < \cdots < i_t = d-1$. Hence Equation 3.3 either has an obvious solution, or it can be rewritten to satisfy the hypotheses of Theorem 3.2.3. Hence, to establish Theorem 3.2.2, we need only complete the proof of Theorem 3.2.3. $\square$

We will need the following terminology for proving Theorem 3.2.3. The main structure we need to understand are naturally ordered subsequences of the ordered sequence $D = \{0, 1, 2, \ldots, d-1\}$. In particular, we need this to understand the index set of zero and non-zero coefficients in Equation 3.4, namely $T = \{i_1, i_2, \ldots, i_t\}$ for $2 \le i_1 < i_2 < \cdots < i_t = d-1$ and $T_0 = D \setminus T$.

Definition 3.2.5. The *length* of a sequence is the number of elements in the sequence.

Definition 3.2.6. If O is an ordered subsequence of D, a *distinguished sequence* in O is an ordered subsequence $\{j_1, j_2, \ldots, j_m\}$ of O such that

(i) $j_{l+1} - j_l = 1$ for $1 \le l \le m-1$

(ii) $j_m + 1 \notin O$.

Furthermore, we will call distinguished sequences in O, *O-sequences.* A *terminal* O-sequence is one that contains $i_t = d-1$.

Let $T_j = \{i \in T \mid i$ initiates a T-sequence of length $j\}$, and let us say that if $i \in T_0$, that is, $i \notin T$, then i initiates a T-sequence of length zero. One observation is that the initial T_0-sequences $\{0, 1, \ldots, i_1 - 1\}$ have maximal length among the T_0-sequences. Since d is finite, there exists a unique integer c such that $T_c \neq \varnothing$, $T_{c+1} = \varnothing$. If $T = \varnothing$, our convention yields $c = 0$. The sets $T_0, T_1, \ldots, T_c$ partition D.

Example 3.2.7. Let $d = 19$ and $T = \{4, 5, 6, 7, 13, 17, 18\}$. Then the family of T-sequences consists of

$$\{7\}, \{13\}, \{18\}, \{6, 7\}, \{17, 18\}, \{5, 6, 7\}, \{4, 5, 6, 7\}.$$

This means $\{18\}$ and $\{17, 18\}$ are terminal T-sequences. Using the above information, we can compute T_0, T_1, T_2, T_3, T_4, and T_5.

$$T_0 = \{0, 1, 2, 3, 6, 8, 10, 11, 12, 14, 15, 16\},$$

$$T_1 = \{7, 13, 18\},$$
$$T_2 = \{6, 17\},$$
$$T_3 = \{5\},$$
$$T_4 = \{4\},$$
$$T_5 = \varnothing$$

This means that for this particular d and T the value c defined above is 4. Notice that $T_0 \cup T_1 \cup T_2 \cup T_3 \cup T_4 = D$ and that the initial T_0-sequence is $\{0, 1, 2, 3, \}$ and has length $c = 4$.

Definition 3.2.8. We say that b can *appear effectively* at location i in Equation 3.4 if $b = \lambda^i a^d$ for some $a \in \mathbb{F}_q^*$ and some $i \in T_0$.

Lemma 3.2.9. *If d is an odd number, and if c is the maximum number of consecutive zero coefficients in Equation 3.4, then Equation 3.4 has a solution in $\mathbb{F}_q$, nontrivial in the $d - t$ effectively appearing variables, provided that $c + 1 < i_1$, that is, provided the first $c + 2$ coefficients are nonzero.*

Proof. Recall that $1+\lambda = \lambda^{i_0} a^d$ for $2 \leq i_0 \leq d-1$, $a \in \mathbb{F}_q^*$. If $i_0 \in T_j$, then $i_0+j \in T_0$. Otherwise if $i_0 + j \notin T_0$, then $i_0 + j \in T$, which contradicts $i_0 \in T_j$. So $i_0 + j \in T_0$. Also, if $i_0 \in T_j$, then $i_0 + j \leq d$ with equality holding only if i_0 belongs to a terminal sequence. This is because $i_0 \leq (d - 1) - (j - 1) \leq (d - j)$, which implies $i_0 + j \leq d$. We get equality when $i_0 = d - j$. Since there are $j - 1$ terms remaining past i_0, the last term in the sequence is $i_0 + j - 1 = d - j + j - 1 = d - 1$. Thus we get equality only if i_0 belongs to a terminal sequence. Then $\lambda^j + \lambda^{j+1} = \lambda^j(1 + \lambda) = \lambda^{i_0+j} a^d$ and $\lambda^j + \lambda^{j+1}$ appears effectively at location $i_0 + j$ when $i_0 + j < d$ and at location 0 when $i_0 + j = d$, since then $\lambda^{i_0+j} a^d = (\lambda a)^d$.

By assumption the initial T_0-sequence has length at least $c + 2$, in other words, it is $\{0, 1, 2, \ldots, c + 1, \ldots\}$. Then, since $j \leq c$, we see that j and $j + 1$, as well as $i_0 + j$, belong to T_0 and hence x_j, x_{j+1}, and x_{i_0+j} appear effectively in Equation 3.4. Let $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ where the 1 appears in the j^{th} position. Consider the following vector $-e_j - e_{j+1} + ae_{i_0+j}$. Consider evaluating Equation 3.4 at this d-tuple

$$\begin{aligned} -\lambda_j - \lambda_{j+1} + \lambda_{i_0+j}a^d &= -\lambda_j(1+\lambda) + \lambda_{i_0+j}a^d \\ &= -\lambda_j(\lambda^{i_0}a^d) + \lambda_{i_0+j}a^d \\ &= -\lambda^{i_0+j}a^d + \lambda_{i_0+j}a^d \\ &= 0 \end{aligned}$$

Therefore, $-e_j - e_{j+1} + ae_{i_0+j}$ is a non-trivial solution of Equation 3.4. Thus, we have found a solution that is non-trivial in the $d - t$ effectively appearing variables. □

Lemma 3.2.10. *If t is chosen so that $t + 4 < 2\sqrt{d+3}$, and if c is the maximum number of consecutive zero coefficients in Equation 3.4, then at least the first $c + 2$ coefficients in Equation 3.4 are nonzero, that is, when $c + 1 < i_1$.*

Proof. We note that $e_d - 1 = e_{i_t} = 0$ so that the presence of t zeros among the coefficients of Equation 3.4, with c of the zeros consecutively, leaves at most $t - c + 1$ separated blocks of consecutive nonzero coefficients. Suppose that the maximal length of these is less than $c+2$. Then we have, as the maximum possible number of nonzero coefficients,

$$(t - c + 1)(c + 1) \geq d - t.$$

Simplifying this inequality yields $c^2 - tc - 1 - 2t + d \leq 0$. Considering this as a quadratic equation in variable c. We find the discriminant to be $t^2 + 4 + 8t - 4d$ and know that it has a real zero if and only if

$$t^2 + 4 + 8t - 4d \geq 0,$$

which is the same as

$$(t + 4)^2 \geq 4(d + 3).$$

Taking the square root of both sides yields

$$t + 4 \geq 2\sqrt{d+3},$$

which contradicts our initial assumption on t. Thus, we conclude that for $t + 4 < 2\sqrt{d+3}$ the initial maximal block of consecutive nonzero coefficients has length at least $c + 2$. □

Before we proceed with the proof of Theorem 3.2.3, we need the following lemma.

Lemma 3.2.11. *Let $m \in \mathbb{Z}$. Assume $d \in \mathbb{Z}_{>0}$, d odd. If $m < 2\sqrt{d+3}$ and maximal, then $m = \lfloor 2\sqrt{d+2} \rfloor = \lfloor 2\sqrt{d+3} \rfloor$.*

Proof. Since $m < 2\sqrt{d+3}$, then $m^2 < 4(d + 3)$. Since d is odd, we know that $4(d + 3) \equiv 0 \mod 8$. We will now demonstrate that $m^2 \not\equiv 5, 6, 7 \mod 8$. First, assume m is odd. Then $m^2 = (2s + 1)^2 = 4s(s + 1) + 1$ for some $s \in \mathbb{Z}$. Since $4s(s + 1) \equiv 0 \mod 8$, we know $m^2 \equiv 4s(s + 1) + 1 \equiv 1 \mod 8$. Now, assume m is even. Then $m^2 = (2s)^2 = 4s^2$ for some $s \in \mathbb{Z}$. If s is odd, by previous argument, we know $s^2 \equiv 1 \bmod 8$. Thus, $m^2 \equiv 4 \bmod 8$. If s is even, then $m^2 \equiv 0 \bmod 8$. Since $m^2 \not\equiv 5, 6, 7 \mod 8$, $m^2 \leq 4(d+3) - 4$. Therefore $m^2 \leq 4d + 8 = 4(d+2)$. Thus $m \leq 2\sqrt{d+2}$. Since $m \in \mathbb{Z}$, $m \leq \lfloor 2\sqrt{d+2} \rfloor$. Since m is maximal among these integers, we know $m = \lfloor 2\sqrt{d+2} \rfloor$, which must also equal to $\lfloor 2\sqrt{d+3} \rfloor$ by definition of the floor function. $\square$

We should note that there is no weakening of Lemma 3.2.10, in replacing the condition $t+4 < 2\sqrt{d+3}$ by $t+4 = \lfloor 2\sqrt{d+2} \rfloor$, since by Lemma 3.2.11, the maximum integer less that $2\sqrt{d+3}$ is precisely $\lfloor 2\sqrt{d+2} \rfloor$, that is $\lfloor 2\sqrt{d+2} \rfloor = \lfloor 2\sqrt{d+3} \rfloor$.

For convenience, we restate Theorem 3.2.3 and then provide the proof.

Theorem 3.2.3. Let $d \mid q-1$, where $q = p^k$ and d is an odd integer. If there exists $\lambda \in \mathbb{F}_q$ such that $\lambda(\mathbb{F}_q^*)^d$ is a generator of $\mathbb{F}_q^*/(\mathbb{F}_q^*)^d$ and $1 + \lambda \notin (\mathbb{F}_q^*)^d$ and $1 + \lambda \notin \lambda(\mathbb{F}_q^*)^d$, then for $t = \lfloor 2\sqrt{d+2} \rfloor - 4$, the form

$$e_0\lambda^0 x_0^d + \cdots + e_{d-1}\lambda^{d-1}x_{d-1}^d = 0$$

has a nontrivial solution in $\mathbb{F}_q$, where

$$e_i = \begin{cases} 0 & \text{for } i \in T = \{i_1, i_2, \ldots, i_t\}, 2 \leq i_1 < i_2 < \cdots < i_t = d-1 \\ 1 & \text{for } i \in T_0 = D \setminus T, D = \{0, 1, 2, \ldots, d-1\} \end{cases}$$

and where the initial block $\{e_0, e_1, \ldots, e_{i_1 - 1}\}$ of nonzero elements has a maximal length among the blocks of consecutive nonzero elements in $\{e_0, e_1, \ldots, e_{d-1}\}$.

Proof. Since $t = \lfloor 2\sqrt{d+2} \rfloor - 4$, by Lemma 3.2.11, we have $t + 4 < 2\sqrt{d+3}$. By Lemma 3.2.10, we know that at least the first $c + 2$ coefficients in Equation 3.4 are nonzero. By Lemma 3.2.9, Equation 3.4 has a nontrivial solution in the $d-t$ effectively appearing variables, which gives the desired result. $\square$

Theorem 3.2.12 (Gray, [6])**.** *If $\mathbb{F}_q$ is a finite field and d and d_0 are odd numbers such that $d \geq d_0$, then Equation 3.3 has a nontrivial zero in $\mathbb{F}_q$ for $t = \lfloor 2\sqrt{d_0+2} \rfloor - 4$.*

This follows immediately from Theorem 3.1.1, since $t(d)$ is a non-decreasing function and $n = d - t(d) \leq d - t(d_0)$ for $d \geq d_0$. But Theorem 3.1.1 establishes the desired solution for Equation 3.3 in $d - t(d)$ variables and thus for $d - t(d_0)$ variables.

Finally, it is of interest to think about this result graphically and see how it compares to the classical result by Chevalley (Theorem 2.1.1). However, it is important to remember that Gray's result applies only to diagonal equations of odd degree. In Figure 3.1, the shaded region indicates ordered pairs that satisfy the relationship between n and d given by Theorem 3.2.3, however, the d value of the ordered pair must also be odd to guarantee a nontrivial solution. Notice we see a jagged behavior in this graph as a result of the floor function.

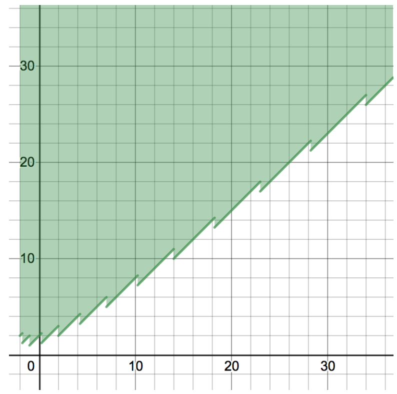


Figure 3.1: This figure illustrates the relationship between n and d needed to guarantee a nontrivial solution to a system of equations by Theorem 3.2.3. The horizontal axis represents our degree d and the vertical axis represents the number of variables n. The shaded region is the order pairs of (d, n) that may guarantee a nontrivial solution to our system.

Furthermore, it is of interest to compare the bounds given by Theorem 3.2.3 and Theorem 2.1.1. We should note that in terms of hypotheses, Theorem 2.1.1 holds for a larger class of functions, whereas Theorem 3.2.3 only holds for diagonal equations of odd degree. However, for diagonal equations of odd degree, we find that Theorem 3.2.3 is a stronger result.

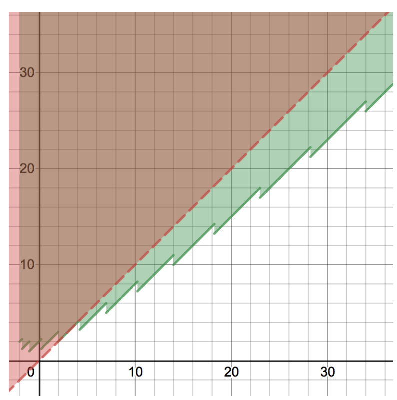


Figure 3.2: This figure illustrates the relationship between Theorem 2.1.1 (shaded in red) and Theorem 3.2.3 (shaded in green). The horizontal axis represents our degree d and the vertical axis represents the number of variables n. As we can see from the diagram, Theorem 3.2.3 improves Theorem 2.1.1 for diagonal equations of odd degree.

Chapter 4 Results on A-Systems

4.1 Introduction

Much of the existing machinery and results for diagonal forms only apply to the case where the degree of the form is odd. A common tool in many proof techniques relies on the fact that -1 is a d^{th} power over $\mathbb{F}_q$. In general, the case where the degree of the form is arbitrary is much more difficult because this property is not guaranteed. This motivates the definition of an A-equation or, more generally, an A-system as introduced by Tietäväinen in 1965 [14]. A system of polynomials is said to be an A-system if the following property holds. Let f_{ij} be a polynomial of degree c_{ij} over $\mathbb{F}_q$. For every $j = 1, \ldots, n$, there exist non-zero elements η_j and ξ_j of $\mathbb{F}_q$ such that $f_{ij}(\eta_j) = -f_{ij}(\xi_j)$, for every $i = 1, \ldots, r$. [14, Condition A, page 8].

Since we are restricting to homogeneous diagonal forms, the notion of A-system reduces to the following definition.

Definition 4.1.1 (Tietäväinen, [14]). Let $f_1, \ldots, f_r$ be homogeneous diagonal forms in $\mathbb{F}_q[x_1, \ldots, x_n]$ where $\deg(f_i) = d_i$. The above system is said to be an A-system if there exists an $\eta \in \mathbb{F}_q$ such that $\eta^{d_i} = -1$ for all $i = 1, \ldots, r$.

With this definition, we can make the following helpful observations. First, if $\operatorname{char}(\mathbb{F}_q) = 2$, then every system of homogeneous diagonal forms over $\mathbb{F}_q$ is an A-system by setting $\eta = 1$. Similarly, if d_i is odd for all $i = 1, \ldots, r$, then the system is an A-system by setting $\eta = -1$.

4.2 A Classification of A-Systems

In the case of A-systems, many results on the existence of nontrivial solutions are known (see, e.g., Section 4.4, Section 4.5, [14]). Consequently, it is of interest to develop criteria for determining if a given system of diagonal forms is an A-system. In joint work with Leep, we proved the following result in this direction.

Theorem 4.2.1 (Leep–Petrik). *Assume q is odd. Then $\boldsymbol{f}$ is an A-system if and only if $d_i = 2^k d_i'$ where $k \in \mathbb{Z}_{\geq 0}$ and d_i' odd and $2^{k+1} \mid q-1$.*

Note that since the case $\operatorname{char}(\mathbb{F}_q) = 2$ is already fully classified, assuming q odd is a reasonable assumption. Before proving this result, it may be helpful to see an example.

Example 4.2.2. Consider the system over $\mathbb{F}_{25}$.

$$x^{12} + 4y^{12} = 0$$

$$x^4 + 3y^4 + 2z^4 = 0$$

Since $d_1 = 2^2(3)$, $d_2 = 2^2(1)$, and $2^3 \mid 25 - 1$, the system is an A-system. Thus, there exists some $\eta \in \mathbb{F}_q$ such that $\eta^4 = \eta^{12} = -1$.

To prove Theorem 4.2.1, we will need the following lemmas.

Lemma 4.2.3 (Leep–Petrik). *If d_1 is even and d_2 is odd, then the system $\{f_1, f_2\}$ is not an A-system.*

Proof. Let $d = \gcd(d_1, d_2)$. We know d is odd because d_2 is odd. Assume $\eta^{d_1} = -1$ and $\eta^{d_2} = -1$. Consider $\eta^d = \eta^{rd_1 + sd_2} = (-1)^{r+s} = \pm 1$. Since $d_1 = 2kd$, we know $\eta^{d_1} = (\eta^d)^{2k} = (\pm 1)^{2k} = 1$ ↯. Thus there are no A-systems with both even and odd degree equations. □

Lemma 4.2.4 (Leep–Petrik). *If the system $\boldsymbol{f}$ is an A-system, then there exists a $k \in \mathbb{Z}_{\geq 0}$ such that $d_i = 2^k d_i'$, where d_i' is odd for all $i = 1, \ldots, r$.*

Proof. By Lemma 4.2.3, we know that all of the d_i are odd or all of the d_i are even. If all of the d_i are odd, we are done. Assume d_i is even for all $1 \leq i \leq r$. Since $f_1, \ldots, f_r$ is an A-system, there exists an element, $\eta \in \mathbb{F}_q^*$, such that $\eta^{d_i} = -1$ for all i. Notice that since $\eta^{d_i} = -1$ for all i, we know that $(\eta^2)^{\frac{d_i}{2}} = -1$ for all i. Let $y_i = x_i^2$ for all i, then $x_i^{d_i} = (y_i)^{\frac{d_i}{2}}$. This means, after performing the previously mentioned change of variable, that we have an A-system in y_i for $i = 1, \ldots, r$. Since we have an A-system, $\dfrac{d_i}{2}$ are either all even or all odd. Repeat this process until they are all odd. Let k represent the number of times we performed this process. This means that $2^k \mid d_i$ for all $i = 1, \ldots, r$, but $2^{k+1} \nmid d_i$ for all $i = 1, \ldots, r$. □

For convenience, we will restate Theorem 4.2.1.

Theorem 4.2.1 (Leep-Petrik). Assume q is odd. Then $\boldsymbol{f}$ is an A-system if and only if $d_i = 2^k d_i'$ where $k \in \mathbb{Z}_{\geq 0}$ and d_i' odd and $2^{k+1} \mid q - 1$.

Proof. Let $\mathbb{F}_q^* =< \alpha >$. Notice that $\alpha^{\frac{q-1}{2}} = -1$.
($\Leftarrow$) Choose $\eta = \alpha^{\frac{q-1}{2^{k+1}}}$.Then

$$\eta^{d_i} = \alpha^{d_i(\frac{q-1}{2^{k+1}})} = \alpha^{d_i'(\frac{q-1}{2})} = (-1)^{d_i'} = -1$$

for all $i = 1, \cdots, r$.
($\Rightarrow$) Since $\boldsymbol{f}$ is an A-system, Lemma 4.2.4 gives us $d_i = 2^k d_i'$ where $k \in \mathbb{Z}_{\geq 0}$ and d_i' odd. Since $\boldsymbol{f}$ is an A-system, there exists $\eta \in \mathbb{F}_q^*$ such that $\eta^{d_i} = -1$ for all $i = 1, \cdots, r$. Let $\eta = \alpha^\lambda$. Then $-1 = \eta^{d_i} = \alpha^{d_i \lambda}$. So $d_i \lambda = k_i(\frac{q-1}{2})$, where k_i is odd. Thus, $2^k d_i' \lambda = k_i(\frac{q-1}{2})$. Since k_i is odd, we know $2^k \mid \frac{q-1}{2}$. Therefore, $2^{k+1} \mid q - 1$. □

4.3 Preliminary Results

For Section 4.4 and Section 4.5, we will need the following preliminary results and notation to begin investigating A-equations and A-systems.

Let V be the space of r-tuples over $\mathbb{F}_q$, where $q = p^k$. Let $\mathbf{a} = (a_1, a_2, \ldots, a_r)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_r)$ be elements of V and a, $b \in \mathbb{F}_q$. We will denote the 0-element $(0, 0, \ldots, 0)$ of V as $\mathbf{0}$. Define

$$\mathbf{a} + \mathbf{b} \coloneqq (a_1 + b_1, a_2 + b_2, \ldots, a_r + b_r)$$

$$a\mathbf{a} \coloneqq (aa_1, aa_2, \ldots, aa_r)$$

$$\mathbf{ab} \coloneqq a_1b_1 + a_2b_2 + \cdots + a_rb_r$$

Define the trace of a by the map

$$\mathrm{tr} : \mathbb{F}_{p^k} \to \mathbb{F}_p$$

where

$$a \mapsto a + a^p + \cdots + a^{p^{k-1}}$$

We define

$$e(a) \coloneqq e^{2\pi i \mathrm{tr}(a)/p}.$$

For $v \in \mathbb{Z}_{\geq 0}$, $d \in \mathbb{Z}_{>0}$, we define

$$e_d(v) \coloneqq e^{2\pi i v/d}.$$

Proposition 4.3.1 ([14], page 11)**.** *Let a, $b \in \mathbb{F}_q$. Then $e(a+b) = e(a)e(b)$*

Proof. First we will show that $\mathrm{tr}(a+b) = \mathrm{tr}(a) + \mathrm{tr}(b)$. By definition, we know

$$\mathrm{tr}(a+b) = (a+b) + (a+b)^p + \cdots + (a+b)^{p^{k-1}}$$

Since we are working over $\mathbb{F}_q$, which has characteristic p, we know that $(a+b)^{p^k} = a^{p^k} + b^{p^k}$ for $k \in \mathbb{Z}_{>0}$. Thus

$$\begin{aligned}
\mathrm{tr}(a+b) &= (a+b) + (a+b)^p + \cdots + (a+b)^{p^{k-1}} \\
&= a + b + a^p + b^p + \cdots + a^{p^{k-1}} + b^{p^{k-1}} \\
&= a + a^p + \cdots + a^{p^{k-1}} + b + b^p + \cdots + b^{p^{k-1}} \\
&= \mathrm{tr}(a) + \mathrm{tr}(b)
\end{aligned}$$

Now consider

$$\begin{aligned} e(a+b) &= e^{2\pi i \operatorname{tr}(a+b)/p} \\ &= e^{2\pi i (\operatorname{tr}(a)+\operatorname{tr}(b))/p} \\ &= e^{2\pi i \operatorname{tr}(a)/p} e^{2\pi i \operatorname{tr}(b)/p} \\ &= e(a)e(b) \end{aligned}$$

□

Proposition 4.3.2 ([14], Equation 8, page 11)**.** *Let $\boldsymbol{l}$, $\boldsymbol{a}$, and $\boldsymbol{b} \in V$. Then*

$$e(\boldsymbol{l}(\boldsymbol{a}+\boldsymbol{b})) = e(\boldsymbol{la})e(\boldsymbol{lb}) \tag{4.1}$$

for every element $\boldsymbol{l} \in V$.

Proof. By definition

$$\begin{aligned} e(\mathbf{l}(\mathbf{a}+\mathbf{b})) &= e^{2\pi i \operatorname{tr}(\mathbf{l}(\mathbf{a}+\mathbf{b}))/p} \\ &= e^{2\pi i \operatorname{tr}(\mathbf{la}+\mathbf{lb})/p} \\ &= e^{2\pi i (\operatorname{tr}(\mathbf{la})+\operatorname{tr}(\mathbf{lb}))/p} \\ &= e^{2\pi i \operatorname{tr}(\mathbf{la})/p} e^{2\pi i \operatorname{tr}(\mathbf{lb})/p} \\ &= e(\mathbf{la})e(\mathbf{lb}) \end{aligned}$$

□

Proposition 4.3.3 ([14], Equation 9, page 11)**.** *Let $a, b \in \mathbb{F}_q$. Then*

$$\sum_{a \in \mathbb{F}_q} e(ab) = \begin{cases} q & \text{if } b = 0 \\ 0 & \text{if } b \neq 0 \end{cases}$$

Proof. Assume $b = 0$. Then $e(0) = 1$. Since there are q elements that a runs over, it follows that $\sum_{a \in \mathbb{F}_q} e(0) = q$.

Now, assume $b \neq 0$. First notice that $\prod_{j=1}^{p-1}(x - e^{2\pi i j/p}) = x^{p-1} + x^{p-2} + \cdots + x + 1$. Notice that the coefficient on x^{p-2} is given by $-\sum_{j=1}^{p-1} e^{2\pi i j/p}$. This means that $-\sum_{j=1}^{p-1} e^{2\pi i j/p} = 1$, which implies that $\sum_{j=1}^{p-1} e^{2\pi i j/p} = -1$.

Consider the trace map. This is an additive group homomorphism. We know $\mathrm{im}(\mathrm{tr}) = \mathbb{F}_p$. By the First Isomorphism Theorem, we know that $|\ker(\mathrm{tr})| = p^{k-1}$. Thus, the cardinality of the preimage of every element in $\mathbb{F}_p$ under the trace map is the same. Since $b \neq 0$, we have $b\mathbb{F}_q = \mathbb{F}_q$. Then

$$\sum_{a \in \mathbb{F}_q} e(ab) = \sum_{a \in \mathbb{F}_q} e(a) = \sum_{a \in \mathbb{F}_q} e^{2\pi i \mathrm{tr}(a)/p} = \sum_{c=0}^{p-1} p^{k-1} e^{2\pi i c/p} = p^{k-1}(0) = 0.$$

Therefore,

$$\sum_{a \in \mathbb{F}_q} e(ab) = \begin{cases} q & \text{if } b = 0 \\ 0 & \text{if } b \neq 0. \end{cases}$$

$\square$

Proposition 4.3.4 ([14], Equation 10, page 12)**.** *Let $\boldsymbol{a}, \boldsymbol{b} \in V$. Then*

$$\sum_{\boldsymbol{a} \in V} e(\boldsymbol{ab}) = \begin{cases} q^r & \textit{if } \boldsymbol{b} = \boldsymbol{0} \\ 0 & \textit{if } \boldsymbol{b} \neq \boldsymbol{0} \end{cases}$$

Proof. Notice by applying Proposition 4.3.1, we can see that

$$e(\mathbf{ab}) = e(\sum_{j=1}^{r} a_j b_j) = \prod_{j=1}^{r} e(a_j b_j).$$

Now we will show

$$\sum_{\mathbf{a} \in V} e(\mathbf{ab}) = \prod_{j=1}^{r} \sum_{a \in \mathbb{F}_q} e(ab_j).$$

Consider the following

$$\begin{aligned} \sum_{\mathbf{a} \in V} e(\mathbf{ab}) &= \sum_{\mathbf{a} \in V} \prod_{j=1}^{r} e(a_j b_j) \\ &= \sum_{\mathbf{a} \in V} e(a_1 b_1) e(a_2 b_2) \dots e(a_r b_r) \\ &= \Big(\sum_{a \in \mathbb{F}_q} e(ab_1) \Big) \dots \Big(\sum_{a \in \mathbb{F}_q} e(ab_r) \Big) \\ &= \prod_{j=1}^{r} \sum_{a \in \mathbb{F}_q} e(ab_j). \end{aligned}$$

Applying Proposition 4.3.3, we find

$$\sum_{\mathbf{a} \in V} e(\mathbf{ab}) = \prod_{j=1}^{r} \sum_{a \in \mathbb{F}_q} e(ab_j) = \prod_{j=1}^{r} \begin{cases} q & \text{if } b_j = 0 \\ 0 & \text{if } b_j \neq 0 \end{cases} = \begin{cases} q^r & \text{if } \mathbf{b} = \mathbf{0} \\ 0 & \text{if } \mathbf{b} \neq \mathbf{0}. \end{cases}$$

$\square$

Lemma 4.3.5 ([14], Lemma 1, page 12)**.** *The inequality*

$$|\sum_{\xi \in \mathbb{F}_q} e(f(\xi))| \leq (c-1)q^{\frac{1}{2}}$$

holds on the assumption that f is a polynomial of degree c over $\mathbb{F}_q$ such that $f \neq g^p - g + \beta$ for every polynomial g over $\mathbb{F}_q$ and for every element β of $\mathbb{F}_q$. In particular, the inequality holds for all polynomials f with $\deg(f) = c$*, where $1 \leq c \leq p-1$.*

Lemma 4.3.5 is proved in [2].

Now consider the system

$$\begin{aligned} f_1 &= \alpha_{11}x_1^{d_1} + \alpha_{12}x_2^{d_1} + \cdots + \alpha_{1n}x_n^{d_1} \\ f_2 &= \alpha_{21}x_1^{d_2} + \alpha_{22}x_2^{d_2} + \cdots + \alpha_{2n}x_n^{d_2} \\ &\vdots \\ f_r &= \alpha_{r1}x_1^{d_r} + \alpha_{r2}x_2^{d_r} + \cdots + \alpha_{rn}x_n^{d_r} \end{aligned}$$

Let $\mathbf{g}_j(x_j) = (\alpha_{1j}x_j^{d_1}, \alpha_{2j}x_j^{d_2}, \ldots, \alpha_{rj}x_j^{d_r})$, where this represents the column corresponding to x_j. Notice the following lemma is the same as Theorem 5.3.1.

Lemma 4.3.6 ([14], Lemma 2, page 12)**.** *The number of solutions to the system* $\sum_{j=1}^{n} \alpha_{ij}x_j^{d_i} = 0$ *for $i = 1, \ldots, r$ is equal to*

$$N(\boldsymbol{f}; \mathbb{F}_q^n) = q^{n-r} + q^{-r} \sum_{\substack{\boldsymbol{l} \in \mathbb{F}_q^r \\ \boldsymbol{l} \neq \mathbf{0}}} \prod_{j=1}^{n} \sum_{\xi_j \in \mathbb{F}_q} e(\boldsymbol{l}\boldsymbol{g}_j(\xi_j))$$

.

Proof. Applying Proposition 4.3.2 and Proposition 4.3.4 yields

$$\begin{aligned} q^r N(\boldsymbol{f}; \mathbb{F}_q^n) &= \sum_{\xi_1 \in \mathbb{F}_q} \cdots \sum_{\xi_n \in \mathbb{F}_q} \sum_{\mathbf{l} \in \mathbb{F}_q^r} e(\mathbf{l} \sum_{j=1}^{n} \mathbf{g}_j(\xi_j)) \\ &= \sum_{\mathbf{l} \in \mathbb{F}_q^r} \sum_{\xi_1 \in \mathbb{F}_q} \cdots \sum_{\xi_n \in \mathbb{F}_q} \prod_{j=1}^{n} e(\mathbf{l}\mathbf{g}_j(\xi_j)) \\ &= \sum_{\mathbf{l} \in \mathbb{F}_q^r} \prod_{j=1}^{n} \sum_{\xi_j \in \mathbb{F}_q} e(\mathbf{l}\mathbf{g}_j(\xi_j)) \end{aligned}$$

Picking out the term where $\mathbf{l} = \mathbf{0}$ yields

$$q^r N(\boldsymbol{f}; \mathbb{F}_q^n) = q^n + \sum_{\substack{\mathbf{l} \in \mathbb{F}_q^r \\ \mathbf{l} \neq \mathbf{0}}} \prod_{j=1}^{n} \sum_{\xi_j \in \mathbb{F}_q} e(\mathbf{l}\mathbf{g}_j(\xi_j))$$

□

Lemma 4.3.7 (Tietäväinen, [14], Lemma 3, page 13)**.** *Let $f_1, \ldots, f_r$ denote homogeneous diagonal forms such that* $\deg(f_i) = d_i$. *Assume that* $\boldsymbol{f}$ *is an A-system. Then the system* $\boldsymbol{f}$ *has a nontrivial solution in* $\mathbb{F}_q^n$ *if* $2^n > q^r$.

Proof. Since $\boldsymbol{f}$ is an A-system, there exists $\eta \in \mathbb{F}_q$ such that $\eta^{d_i} = -1$ for all $i = 1, \ldots, r$. Let Δ be the collection of n-tuples with entries from $\{0, 1\}$. There are 2^n elements in Δ. Since there are r forms and each form can take on at most q values we know that there are q^r possible outputs for $\boldsymbol{f}(\Delta)$. If $2^n > q^r$, then by the Pigeonhole Principle, there exist $\boldsymbol{\delta}_1, \boldsymbol{\delta}_2 \in \Delta$ such that $\displaystyle\sum_{j=1}^{n} \mathbf{f}_j(\boldsymbol{\delta}_{1j}) = \sum_{j=1}^{n} \mathbf{f}_j(\boldsymbol{\delta}_{2j})$ and $\boldsymbol{\delta}_1 = (\delta_{11}, \ldots, \delta_{1n}) \neq (\delta_{21}, \ldots, \delta_{2n}) = \boldsymbol{\delta}_2$. Therefore, $\displaystyle\sum_{j=1}^{n} (\mathbf{f}_j(\boldsymbol{\delta}_{1j}) - \mathbf{f}_j(\boldsymbol{\delta}_{2j})) = 0$. Since $\delta_{1j}, \delta_{2j} \in \{0, 1\}$ for $j = 1, \ldots, n$, we know $\{(\delta_{1j}^{d_i}) - (\delta_{2j}^{d_i})\} \in \{0, 1, -1\}$. Then $\delta_{1j}^{d_i} - \delta_{2j}^{d_i} = \lambda_j^{d_i}$, where $\lambda_j \in \{0, 1, \eta\}$ because $\eta^{d_i} = -1$. Then $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ is a nontrivial solution to $\boldsymbol{f}$. $\square$

The following lemma is an extension of a result of Chowla [4].

Lemma 4.3.8 ([14], Lemma 4, page 13)**.** *If* γ *is a non-zero element of* $\mathbb{F}_q$ *and* $d \mid q-1$, *then*

$$\Big|\sum_{\xi \in \mathbb{F}_q} e(\gamma \xi^d)\Big| \leq (d-1) q^{\frac{1}{2}}.$$

Proof. Let ρ be a generator of the cyclic group $\mathbb{F}_q^*$. For $\alpha \in \mathbb{F}_q^*$, let $\operatorname{ind}(\alpha)$ be the integer such that $\rho^{\operatorname{ind}(\alpha)} = \alpha$. The equation $\xi^d = \zeta$ is solvable for a non-zero ζ of $\mathbb{F}_q$ if and only if $\operatorname{ind}(\zeta)$ is divisible by d. If $\operatorname{ind}(\zeta)$ is divisible by d, it has d solutions.

Let $d = 1$. We have $\displaystyle\sum_{\xi \in \mathbb{F}_q} e(\gamma \xi^d) = 0$ by Proposition 4.3.3 because $\gamma \neq 0$ and ξ^d runs over all elements of $\mathbb{F}_q$. Thus, when $d = 1$,

$$\Big|\sum_{\xi \in \mathbb{F}_q} e(\gamma \xi^d)\Big| = 0 = (d-1) q^{\frac{1}{2}}.$$

Hence we may assume that $d > 1$. Recall $e_d(v) = e^{2\pi i v/d}$ and define $U(k)$ to be

$$U(k) = \sum_{\substack{\zeta \in \mathbb{F}_q \\ \zeta \neq 0}} e_d(k \operatorname{ind} \zeta) e(\gamma \zeta)$$

Notice that when ζ is a d^{th} power, then $e_d(v) = e^{\frac{2\pi i v}{d}} = 1$ and when ζ is not a d^{th} power, then $\displaystyle\sum_{k=0}^{d-1} e_d(k \text{ ind } \zeta) = 0$.

When $k = 0$

$$\sum_{\substack{\zeta \in \mathbb{F}_q \\ \zeta \neq 0}} e_d(k \text{ ind}(\zeta)) e(\gamma \zeta) = \sum_{\substack{\zeta \in \mathbb{F}_q \\ \zeta \neq 0}} e_d(0) e(\gamma \zeta) = \sum_{\substack{\zeta \in \mathbb{F}_q \\ \zeta \neq 0}} e(\gamma \zeta) = -1$$

For $d > 1$, we have

$$\begin{aligned}
\sum_{\xi\in\mathbb{F}_q} e(\gamma\xi^d) &= 1 + \sum_{k=0}^{d-1}\sum_{\substack{\zeta\in\mathbb{F}_q\\ \zeta\neq 0}} e_d(k \text{ ind}(\zeta))e(\gamma\zeta) \\
&= 1 + (-1) + \sum_{k=1}^{d-1}\sum_{\substack{\zeta\in\mathbb{F}_q\\ \zeta\neq 0}} e_d(k \text{ ind}(\zeta))e(\gamma\zeta) \\
&= \sum_{k=1}^{d-1}\sum_{\substack{\zeta\in\mathbb{F}_q\\ \zeta\neq 0}} e_d(k \text{ ind}(\zeta))e(\gamma\zeta) \\
&= \sum_{k=1}^{d-1} U(k).
\end{aligned}$$

Moreover, for $k \not\equiv 0 \bmod d$,

$$\begin{aligned}
|U(k)|^2 &= \sum_{\substack{\xi\in\mathbb{F}_q\\ \xi\neq 0}} e_d(k \text{ ind } \xi)e(\gamma\xi)\sum_{\substack{\eta\in\mathbb{F}_q\\ \eta\neq 0}} e_d(-k \text{ ind } \eta)e(-\gamma\eta) \\
&= \sum_{\substack{\xi\in\mathbb{F}_q\\ \xi\neq 0}}\sum_{\substack{\eta\in\mathbb{F}_q\\ \eta\neq 0}} e_d(k \text{ ind } \xi\eta^{-1})e(\gamma(\xi-\eta))
\end{aligned}$$

Let $\zeta = \xi\eta^{-1}$.

$$|U(k)|^2 = \sum_{\substack{\zeta\in\mathbb{F}_q\\ \zeta\neq 0}} e_d(k \text{ ind } \zeta)\sum_{\substack{\eta\in\mathbb{F}_q\\ \eta\neq 0}} e(\gamma(\zeta-1)\eta)$$

Notice that when $\eta = 0$, $\sum_{\substack{\zeta\in\mathbb{F}_q\\ \zeta\neq 0}} e_d(k \text{ ind } \zeta) = 0$ so

$$= \sum_{\substack{\zeta\in\mathbb{F}_q\\ \zeta\neq 0}} e_d(k \text{ ind } \zeta)\sum_{\eta\in\mathbb{F}_q} e(\gamma(\zeta-1)\eta).$$

Using Proposition 4.3.3, we see that summation with respect to η gives q, for $\zeta = 1$, and 0, for $\zeta \neq 1$. Therefore, $|U(k)|^2 = q$ and $|U(k)| = q^{\frac{1}{2}}$.

Combining this with the equality $\sum_{\xi\in\mathbb{F}_q} e(\gamma\xi^d) = \sum_{k=1}^{d-1} U(k)$, we obtain

$$|\sum_{\xi\in\mathbb{F}_q} e(\gamma\xi^d)| \leq \sum_{k=1}^{d-1} |U(k)| \leq (d-1)q^{\frac{1}{2}}$$

□

Lemma 4.3.9. *If $\xi^d - \eta^d = 0$ and $d \mid q-1$, then there are $1 + d(q-1)$ solutions.*

Proof. First notice, if $\eta = 0$, then $\xi = 0$. Assume $\eta \neq 0$, then $\left(\frac{\xi}{\eta}\right)^d = 1$. There are exactly d values for $\frac{\xi}{\eta}$; one for each d^{th} root of 1. For each nonzero η, there are d values of ξ that will yield a solution. Since there are $(q-1)$ choices for η, there are $1 + d(q-1)$ solutions to $\xi^d - \eta^d = 0$. □

4.4 Results on A-Equations

In this section, we will give an overview of the results currently known for A-equations. It will become apparent that several of these results improve some previously known bounds giving more evidence of the usefulness of the classification of A-equations.

The theorem below is stated much more generally by Tietäväinen in [14], but when restricting to diagonal forms, the theorem is simplified to the following statement. It turns out that this result will ultimately be superceded by a later result for $d \geq 3$.

Theorem 4.4.1 (Tietäväinen, [14], Theorem 2, page 17)**.** *Assume $d \geq 2$ and $n \geq 2$. The A-equation $f = \sum_{j=1}^{n} \alpha_j x_j^d$, $\alpha_j \in \mathbb{F}_q^*$ has a non-trivial solution in $\mathbb{F}_q$ if either of the following conditions hold*

(i) $d = 2$, $n \geq 3$

(ii) $d \geq 3$, $n \geq 2\log_2(d-1) + 2$.

In particular, if $d \geq 3$, then $\Omega_A(d,q) < 2\log_2(d-1) + 2$.

Proof. *(i)* Let $d = 2$ and $n \geq 3$. By Chevalley's Theorem (Theorem 2.1.1,[3]), there is a non-trivial solution to f.

(ii) Now let $d \geq 3$ and $n \geq 2n\log_2(d-1) + 2$. Suppose the system has only the trivial solution in $\mathbb{F}_q$. By Lemma 4.3.7, this implies $2^n \leq q$. Combining this with $n - 2 \geq 2\log_2(d-1)$ we find

$$2^n \leq q$$

$$2^{n(2\log_2(d-1))} \leq q^{(n-2)}$$

$$2^{2n\log_2(d-1)} \leq q^{n-2}$$

$$(d-1)^{2n} \leq q^{n-2}$$

$$(d-1)^n \leq q^{\frac{1}{2}(n-2)}$$

Let $l \in \mathbb{F}_q^*$ and $\xi_j \in \mathbb{F}_q$. For every $j = 1, \ldots, n$, we know $l\alpha_j\xi_j^d$ is not zero unless $\xi_j = 0$. Hence, by Lemma 4.3.5, $|\sum_{\xi_j \in \mathbb{F}_q} e(l\alpha_j\xi_j^d)| \leq (d-1)q^{\frac{1}{2}}$. Therefore we have, by Lemma 4.3.6,

$$\begin{aligned}
N_{\mathbb{F}_q}(f) &= q^{n-1} + q^{-1}\sum_{l \in \mathbb{F}_q^*}\prod_{j=1}^{n}\sum_{\xi_j \in \mathbb{F}_q} e(l\alpha_j\xi_j^d) \\
&\geq q^{n-1} - q^{-1}(q-1)(d-1)^n q^{\frac{1}{2}n} \\
&\geq q^{n-1} - (q-1)q^{\frac{1}{2}(n-2)}(d-1)^n \\
&= q^{\frac{1}{2}(n-2)}\Big(q\big(q^{\frac{1}{2}(n-2)} - (d-1)^n\big) + (d-1)^n\Big)
\end{aligned}$$

Since

$$q^{\frac{1}{2}(n-2)} \geq (d-1)^n,$$

it follows that

$$N_{\mathbb{F}_q}(f) \geq q^{\frac{1}{2}(n-2)}(d-1)^n \geq (d-1)^{2n}$$

Since we know that $d \geq 3$, we have that $N_{\mathbb{F}_q}(f) > 1$, which is impossible. Thus, we have proven the desired result. □

For $\alpha \in \mathbb{F}_q$, define $S(\alpha) \coloneqq \sum_{\xi \in \mathbb{F}_q} e(\alpha\xi^d)$.

Lemma 4.4.2 (Tietäväinen, [14], Lemma 7, page 25)**.** *If ρ is a generator of the cyclic group $\mathbb{F}_q^*$ and $d \mid q-1$, then*

$$\sum_{j=0}^{d-1} |S(\rho^j)|^2 = (d-1)dq.$$

Proof. We have, by complex conjugation,

$$|S(\alpha)|^2 = \sum_{\xi \in \mathbb{F}_q} e(\alpha\xi^d)\sum_{\eta \in \mathbb{F}_q} e(-\alpha\eta^d) = \sum_{\xi \in \mathbb{F}_q}\sum_{\eta \in \mathbb{F}_q} e(\alpha(\xi^d - \eta^d)).$$

Observe when $\alpha = 0$,

$$|S(0)|^2 = \sum_{\xi \in \mathbb{F}_q}\sum_{\eta \in \mathbb{F}_q} e(0) = \sum_{\xi \in \mathbb{F}_q}\sum_{\eta \in \mathbb{F}_q} 1 = q^2.$$

By Lemma 4.3.9, the number of solutions of the equation $\xi^d - \eta^d = 0$ is $1 + d(q-1)$. Applying Proposition 4.3.3, we have

$$\sum_{\alpha\in\mathbb{F}_q}|S(\alpha)|^2=\sum_{\alpha\in\mathbb{F}_q}\sum_{\xi\in\mathbb{F}_q}\sum_{\eta\in\mathbb{F}_q}e(\alpha(\xi^d-\eta^d))=q+d(q-1)q.$$

Since ρ is a generator of the cyclic group $\mathbb{F}_q^*$, we know that

$$\sum_{j=0}^{q-2}|S(\rho^j)|^2=\sum_{\substack{\alpha\in\mathbb{F}_q\\ \alpha\neq 0}}|S(\alpha)|^2=\Big(\sum_{\alpha\in\mathbb{F}_q}|S(\alpha)|^2\Big)-|S(0)|^2=q+d(q-1)q-q^2=(d-1)(q-1)q$$

Moreover, for every $\eta\in\mathbb{F}_q^*$

$$S(\rho^j\eta^d)=\sum_{\xi\in\mathbb{F}_q}e(\rho^j\eta^d\xi^d)=\sum_{\zeta\in\mathbb{F}_q}e(\rho^j\zeta^d)=S(\rho^j).$$

Thus,

$$\Big(\frac{q-1}{d}\Big)\sum_{j=0}^{d-1}|S(\rho^j)|^2=\sum_{j=0}^{q-2}|S(\rho^j)|^2.$$

Therefore,

$$\sum_{j=0}^{d-1}|S(\rho^j)|^2=d(q-1)^{-1}\sum_{j=0}^{q-2}|S(\rho^j)|^2=(d-1)dq.$$

□

Lemma 4.4.3 (Tietäväinen, [14], Lemma 8, page 26)**.** *Let $E(0),\dots,E(d-1)$ be non-negative numbers. Let $\sum_{i=0}^{d-1}(E(i))^2=F$ and $E(d+i)=E(i)$ for every i. Let $k_1,\dots,k_n$ be distinct integers such that $0\le k_j\le d-1$, for every $j=1,\dots,n$, and let $2\le n\le d$. Then*

$$\sum_{h=0}^{d-1}\prod_{j=1}^{n}E(h+k_j)\le n^{1-\frac{1}{2}n}F^{\frac{1}{2}n}$$

Proof. It follows from the arithmetic – geometric mean inequality that

$$\sqrt[n]{\prod_{j=1}^{n}E(h+k_j)^2}\le n^{-1}\sum_{j=1}^{n}E(h+k_j)^2$$

Raising both sides to the $\frac{n}{2}$ power yields

$$\prod_{j=1}^{n}E(h+k_j)\le n^{-\frac{1}{2}n}\Big(\sum_{j=1}^{n}E(h+k_j)^2\Big)^{\frac{1}{2}n}$$

Furthermore,

$$\sum_{h=0}^{d-1}\sum_{j=1}^{n}E(h+k_j)^2=\sum_{j=1}^{n}\sum_{h=0}^{d-1}E(h+k_j)^2=nF,$$

and

$$0 \leq \sum_{j=1}^{n} E(h+k_j)^2 \leq F.$$

Dividing both equations by F we find that

$$\sum_{h=0}^{d-1} \left(\frac{\sum_{j=1}^{n} E(h+k_j)^2}{F} \right) = n$$

and

$$0 \leq \frac{\displaystyle\sum_{j=1}^{n} E(h+k_j)^2}{F} \leq 1.$$

Since $n \geq 2$ and

$$0 \leq \frac{\displaystyle\sum_{j=1}^{n} E(h+k_j)^2}{F} \leq 1,$$

we have

$$0 \leq \left(\frac{\sum_{j=1}^{n} E(h+k_j)^2}{F} \right)^{\frac{1}{2}n} \leq \frac{\displaystyle\sum_{j=1}^{n} E(h+k_j)^2}{F} \leq 1.$$

Thus,

$$\sum_{h=0}^{d-1} \left(\frac{\sum_{j=1}^{n} E(h+k_j)^2}{F} \right)^{\frac{1}{2}n} \leq \sum_{h=0}^{d-1} \frac{\sum_{j=1}^{n} E(h+k_j)^2}{F} = n$$

Multiplying by $F^{\frac{1}{2}n}$ yields

$$\sum_{h=0}^{d-1} (\sum_{j=1}^{n} E(h+k_j)^2)^{\frac{1}{2}n} \leq nF^{\frac{1}{2}n}.$$

Therefore,

$$\sum_{h=0}^{d-1} \prod_{j=1}^{n} E(h+k_j) \leq n^{-\frac{1}{2}n} \sum_{h=0}^{d-1} (\sum_{j=1}^{n} E(h+k_j)^2)^{\frac{1}{2}n} \leq n^{-\frac{1}{2}n}(nF^{\frac{1}{2}n}) = n^{1-\frac{1}{2}n} F^{\frac{1}{2}n}.$$

$\square$

Theorem 4.4.4 (Tietäväinen, [14], Theorem 7, page 27)**.** *If* $n \geq 3$ *and*

$$q \geq n^{-1} d(d-1)^{\frac{n}{n-2}}$$

then the A-equation $f = \sum_{j=1}^{n} \alpha_i x_i^d$ *has a nontrivial solution in* $\mathbb{F}_q$.

Proof. First observe that if $n > d$, then we have a nontrivial solution by Chevalley's Theorem (Theorem 2.1.1, [3]). Hence we may assume that $n \leq d$. Additionally, notice that we can reduce to the case where no two coefficients a_j are in the same set $\rho^i(\mathbb{F}_q^*)^d$, where ρ is a generator of the cyclic group $\mathbb{F}_q^*$ because in that case, the equation has a non-trivial solution in $\mathbb{F}_q$.

Applying Lemma 4.3.6, we know that the number of solutions to the equation is equal to

$$N(f;\mathbb{F}_q^n) = q^{n-1} + q^{-1}\sum_{\substack{\lambda\in\mathbb{F}_q\\ \lambda\neq 0}}\prod_{j=1}^{n} S(\lambda\alpha_j)$$

Let k_j be the least non-negative residue mod d of the index of α_j with respect to ρ. Notice that $k_1,\ldots,k_n$ are distinct integers such that $0 \leq k_j \leq d-1$ for all j. Simplifying $\sum_{\substack{\lambda\in\mathbb{F}_q\\ \lambda\neq 0}}\prod_{j=1}^{n} S(\lambda\alpha_j)$, we find

$$\sum_{\substack{\lambda\in\mathbb{F}_q\\ \lambda\neq 0}}\prod_{j=1}^{n} |S(\lambda\alpha_j)| = \sum_{h=0}^{q-2}\prod_{j=1}^{n}|S(\rho^{h+k_j})| = \Big(\frac{q-1}{d}\Big)\sum_{h=0}^{d-1}\prod_{j=1}^{n}|S(\rho^{h+k_j})|.$$

Recall that $S(\rho^{d+i}) = S(\rho^i)$ and, by Lemma 4.4.2, $\sum_{i=0}^{d-1}|S(\rho^i)|^2 = (d-1)dq$.

Applying Lemma 4.4.3 to $\Big(\frac{q-1}{d}\Big)\sum_{h=0}^{d-1}\prod_{j=1}^{n}|S(\rho^{h+k_j})|$, yields

$$\Big(\frac{q-1}{d}\Big)\sum_{h=0}^{d-1}\prod_{j=1}^{n}|S(\rho^{h+k_j})| \leq \Big(\frac{q-1}{d}\Big)n^{1-\frac{1}{2}n}\Big(\sum_{i=0}^{d-1}|S(\rho^i)|^2\Big)^{\frac{1}{2}n}$$

Combining all of this yields

$$\sum_{\substack{\lambda\in\mathbb{F}_q\\ \lambda\neq 0}}\prod_{j=1}^{n} |S(\lambda\alpha_j)| \leq \Big(\frac{q-1}{d}\Big)n^{1-\frac{1}{2}n}((d-1)dq)^{\frac{1}{2}n} = n^{1-\frac{1}{2}n}(q-1)d^{\frac{1}{2}n-1}(d-1)^{\frac{1}{2}n}q^{\frac{1}{2}n}.$$

Substituting back into the formula for the number of solutions yields

$$\begin{aligned}
N(f;\mathbb{F}_q^n) &\geq q^{n-1} - n^{1-\frac{1}{2}n}(q-1)d^{\frac{1}{2}n-1}(d-1)^{\frac{1}{2}n}q^{\frac{1}{2}n-1}\\
&\geq q^{\frac{1}{2}n-1}\bigg(q\Big(q^{\frac{1}{2}n-1} - n^{1-\frac{1}{2}n}d^{\frac{1}{2}n-1}(d-1)^{\frac{1}{2}n}\Big) + n^{1-\frac{1}{2}n}d^{\frac{1}{2}n-1}(d-1)^{\frac{1}{2}n}\bigg)\\
&\geq q^{\frac{1}{2}n-1}n^{1-\frac{1}{2}n}d^{\frac{1}{2}n-1}(d-1)^{\frac{1}{2}n}\\
&\geq (n^{-1}d)^{n-2}(d-1)^n
\end{aligned}$$

Since $n \geq 3$ and we have assumed that $n \leq d$, the final inequality implies that $N(f;\mathbb{F}_q^n) > 1$, which proves the desired result. $\square$

Theorem 4.4.5 (Tietäväinen, [14], Theorem 8, page 28)**.** *If* $n \geq 3$ *and*

$$2^n n \geq d(d-1)^{\frac{n}{n-2}},$$

then the A-equation $\sum_{j=1}^{n} \alpha_j x_j^d = 0$ *has a non-trivial solution in* $\mathbb{F}_q$.

Proof. By Lemma 4.3.7, if the A-equation has only the trivial solution in $\mathbb{F}_q$, then

$$2^n < q.$$

By Theorem 4.4.4, if the A-equation has only the trivial solution in $\mathbb{F}_q$, then

$$q < n^{-1} d(d-1)^{\frac{n}{n-2}}.$$

Combining these two results, we find that if the A-equation has only the trivial solution in $\mathbb{F}_q$, then

$$2^n < q < n^{-1} d(d-1)^{\frac{n}{n-2}}.$$

This implies

$$2^n n < d(d-1)^{\frac{n}{n-2}}.$$

Thus, if

$$2^n n \geq d(d-1)^{\frac{n}{n-2}},$$

the A-equation $\sum_{j=1}^{n} \alpha_j x_j^d = 0$ has a non-trivial solution in $\mathbb{F}_q$. □

Corollary 4.4.6. *If* $2^n n \geq d(d-1)^{\frac{n}{n-2}}$, *then* $1 + (2^n n)^{\frac{n-2}{2n-2}} > d$.

Proof. Since $2^n n \geq d(d-1)^{\frac{n}{n-2}}$, it follows

$$2^n n > (d-1)(d-1)^{\frac{n}{n-2}} = (d-1)^{\frac{2n-2}{n-2}}.$$

Thus,

$$(2^n n)^{\frac{n-2}{2n-2}} > d-1.$$

Therefore, $1 + (2^n n)^{\frac{n-2}{2n-2}} > d$.

□

Graphically, we can visualize this result with the following figure.

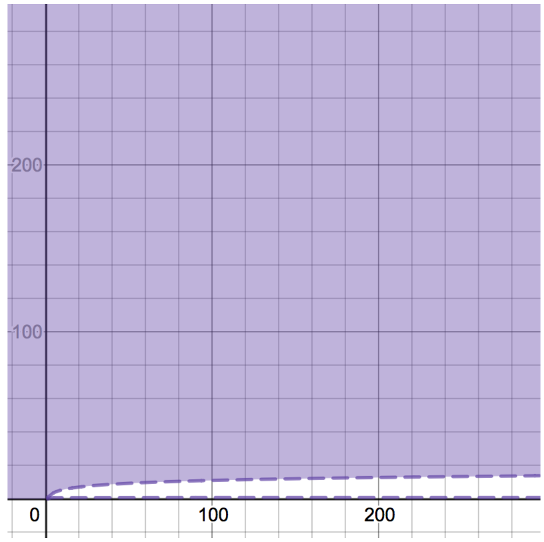
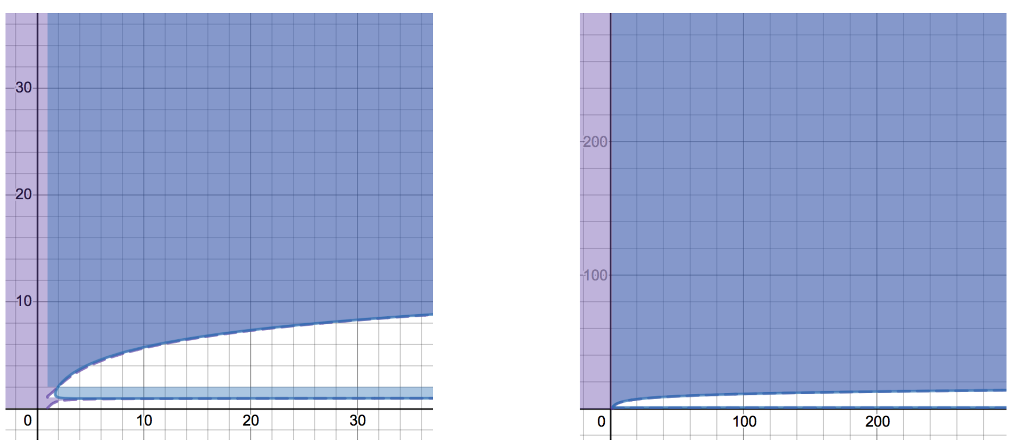


Figure 4.1: This figure illustrates the relationship between n and d needed to guarantee a nontrivial solution to a system of equations by Corollary 4.4.6. The horizontal axis represents our degree d and the vertical axis represents the number of variables n. The shaded region is the order pairs of (d, n) that guarantee a nontrivial solution to our A-equation.

While Corollary 4.4.6 is a bit easier to apply and generally nicer to work with, it is a worse bound than Theorem 4.4.5. Graphically, we can compare the two results with the following figures.



(a) Behavior for small values of d

(b) Behavior for large values of d

Figure 4.2: This figure compares the bounds given by Theorem 4.4.5 (blue region) and Corollary 4.4.6 (purple region). In fact, the bounds are so close, the graph does not differentiate between them.

Theorem 4.4.7 (Tietäväinen, [14], Theorem 9)**.** *If* $d \geq 2$,

$$\max_q \Omega_A(d, q) \leq \lceil 2\log_2(d) - \log_2\log_2(d) \rceil.$$

Proof. *Case I*: Suppose $d = 2$. Then $\lceil 2\log_2(d) - \log_2\log_2(d)\rceil = 2$. By Chevalley's Theorem (Theorem 2.1.1, [3]), if $n > 2$, the equation is isotropic. Thus, $\max_q \Omega_A(d,q) \le 2$.

Case II: Suppose $d = 3$. Then $\lceil 2\log_2(d) - \log_2\log_2(d)\rceil = 3$. By Chevalley's Theorem (Theorem 2.1.1, [3]), if $n > 3$, the equation is isotropic. Thus, $\max_q \Omega_A(d,q) \le 3$.

Case III: Suppose $d \ge 4$. Suppose that $n \ge 1 + 2\log_2(d) - \log_2\log_2(d)$. Then

$$\begin{aligned}
2^n &\ge 2^{1+2\log_2(d)-\log_2\log_2(d)} \\
&= 2(2^{2\log_2(d)})(2^{-\log_2\log_2(d)}) \\
&= 2d^2\Big(\frac{1}{\log_2(d)}\Big) \\
&= \frac{2d^2}{\log_2(d)}.
\end{aligned}$$

We will now show that

$$1 + 0.27\log_2(d) - \log_2\log_2(d) > 0.$$

This is equivalent to showing

$$1 + 0.27\log_2(d) > \log_2\log_2(d).$$

This is equivalent to showing

$$2^{1+0.27\log_2(d)} = 2d^{0.27} > \log_2(d) = 2^{\log_2\log_2(d)}.$$

Let $g(x) = 2x^{0.27} - \log_2(x)$. We will show that $g(x) > 0$ for all $x \ge 4$. By L'Hopital's rule, $\lim_{x\to\infty} g(x) = \infty$. Additionally, $g'(x) = \dfrac{0.54}{x^{0.73}} - \dfrac{1}{x\ln(2)}$. Let $x^* = (0.54\ln(2))^{-0.27}$. Notice that $g'(x^*) = 0$ and this x^* is unique. Furthermore, we can show $g(x^*) > 0$.

Consider the following

$$\begin{aligned}
g(x^*) &= 2((0.54\ln(2))^{-0.27})^{0.27} - \log_2((0.54\ln(2))^{-0.27}) \\
&= 2\Big(\frac{1}{0.54\ln(2)}\Big) + 0.27\log_2(0.54\ln(2)) \\
&> 5 + 0.27\log_2(0.54\ln(2)) \\
&> 5 + \log_2(0.54\ln(2)) \\
&> 0.
\end{aligned}$$

First we will show $g(4) > 0$.

$$\begin{aligned}
g(4) &= 2(4)^{0.27} - \log_2(4) \\
&= 2^{1.54} - 2 \\
&> 0
\end{aligned}$$

Suppose there exists an $a \in \mathbb{R}$ such that $g(a) < 0$ and $4 < a$. Since $\lim_{x\to\infty} g(x) = \infty$, there exists a $b \in \mathbb{R}$ such that $a < b$ and $g(b) > 0$. Since $[4, b]$ is a compact interval, we can find an absolute minimum on the set. By the Extreme Value Theorem, the absolute minimum occurs at one of the endpoints or a critical point in the interval. Since $g(4)$ and $g(b)$ are positive and $g(a)$ is negative, we know the absolute minimum cannot occur at either endpoint. Since $g(a) < 0$, we know that there is a critical point on the interval, which much be x^*. Since $g(x^*) > 0$, we reach a contradiction. Thus, there does not exist an $a \in \mathbb{R}$ such that $g(a) < 0$ and $4 < a$. Therefore, $g(x) > 0$ for all $x \geq 4$.

It follows

$$1 + 0.27\log_2(d) - \log_2\log_2(d) > 0.$$

Hence

$$\begin{aligned} n &> 1 + 2\log_2(d) - \log_2\log_2(d) - (1 + 0.27\log_2(d) - \log_2\log_2(d)) \\ &= 1.73\log_2(d). \end{aligned}$$

and futhermore,

$$2^n n > (1.73\log_2(d))\Big(\frac{2d^2}{\log_2(d)}\Big) = 3.46d^2.$$

On the other hand, we now show that

$$E\log_2(d) - 1 - \log_2(\log_2(d)) \geq 0$$

where

$$E = \begin{cases} 1 & \text{for } 4 \leq d \leq 7 \\ 0.87 & \text{for } d \geq 8. \end{cases}$$

Suppose $4 \leq d \leq 7$. Then $E = 1$. We want to show that

$$\log_2(d) - 1 - \log_2(\log_2(d)) \geq 0.$$

This is equivalent to showing

$$\log_2(d) - 1 \geq \log_2(\log_2(d)).$$

Exponentiating both sides yields

$$\begin{aligned} 2^{\log_2(d)-1} &\geq 2^{\log_2(\log_2(d))} \\ \frac{d}{2} &\geq \log_2(d). \end{aligned}$$

Thus it is equivalent to showing $\frac{d}{2} \geq \log_2(d)$. Consider $g(x) = \frac{x}{2} - \log_2(x)$. We will show that $g(4) = 0$ and $g'(x) \geq 0$ for $x \geq 4$, and conclude that $g(x) \geq 0$ for $4 \leq x \leq 7$.

$$\begin{aligned} g(4) &= 2 - \log_2(4) \\ &= 0. \end{aligned}$$

Now consider

$$g'(x) = \frac{1}{2} - \frac{1}{x\ln(2)}$$
$$= \frac{x\ln(2) - 2}{2x\ln(2)}.$$

Since $x\ln(2) - 2 > 0$ for $x \geq 4$, the result follows.

Suppose $d \geq 8$. Then $E = 0.87$. We want to show that

$$0.87\log_2(d) - 1 - \log_2(\log_2(d)) \geq 0.$$

This is equivalent to showing

$$0.87\log_2(d) - 1 \geq \log_2(\log_2(d)).$$

Exponentiating both sides yields

$$2^{0.87\log_2(d)-1} \geq 2^{\log_2(\log_2(d))}$$
$$\frac{d^{0.87}}{2} \geq \log_2(d).$$

Thus it is equivalent to showing $\frac{d^{0.87}}{2} \geq \log_2(d)$. Consider $g(x) = \frac{x^{0.87}}{2} - \log_2(x)$. We will show that $g(8) > 0$ and $g'(x) \geq 0$ for $x \geq 8$, and conclude that $g(x) > 0$ for $x \geq 8$.

$$g(8) = \frac{8^{0.87}}{2} - \log_2(8)$$
$$> 3.05252 - 3$$
$$> 0.$$

Now consider

$$g'(x) = \frac{0.87}{2x^{0.13}} - \frac{1}{x\ln(2)} = \frac{0.87x\ln(2) - 2x^{0.13}}{2x^{1.13}\ln(2)} = \frac{0.87x^{0.87}\ln(2) - 2}{2x\ln(2)}.$$

Since $0.87x^{0.87}\ln(2) - 2 > 0$ for $x \geq 8$, the result follows.

Hence

$$n \geq 1 + 2\log_2(d) - \log_2(\log_2(d)) - (E\log_2(d) - 1 - \log_2(\log_2(d)))$$
$$= 2 + 2\log_2(d) - E\log_2(d)$$

Thus, $n - 2 \geq (2 - E)\log_2(d)$.

Furthermore, substituting in the assigned values for E yields

$$\frac{2}{n-2} \leq \frac{2}{(2-E)\log_2(d)} \leq \begin{cases} \dfrac{2}{\log_2(d)} & \text{for } 4 \leq d \leq 7 \\[2ex] \dfrac{1.77}{\log_2(d)} & \text{for } d \geq 8. \end{cases}$$

Furthermore, we will show that

$$(d-1)^{\frac{2}{n-2}} < d^{\frac{2}{n-2}} \leq \begin{cases} 4 & \text{for } 4 \leq d \leq 7 \\ \\ 3.42 & \text{for } d \geq 8. \end{cases}$$

Suppose $4 \leq d \leq 7$. From the previous inequalities, it follows that $\dfrac{2}{n-2} \leq \dfrac{2}{\log_2(d)}$. Notice that showing

$$d^{\frac{2}{\log_2(d)}} \leq 4$$

is equivalent to showing

$$\log_2(d^{\frac{2}{\log_2(d)}}) \leq \log_2(4).$$

Simplifying both sides yields

$$2 = \frac{2}{\log_2(d)} \log_2(d) \leq \log_2(4) = 2,$$

which yields the desired result.

Now suppose $d \geq 8$. From the previous inequalities, it follows that $\dfrac{2}{n-2} \leq \dfrac{1.77}{\log_2(d)}$. Notice that showing

$$d^{\frac{1.77}{\log_2(d)}} \leq 3.42$$

is equivalent to showing

$$\log_2(d^{\frac{1.77}{\log_2(d)}}) \leq \log_2(3.42).$$

Simplifying both sides yields

$$1.77 = \frac{1.77}{\log_2(d)} \log_2(d) \leq \log_2(3.42) \leq 1.774,$$

which yields the desired result.

Therefore, we will show that

$$(d-1)^{\frac{n}{n-2}} < \begin{cases} 3.44d & \text{for } 4 \leq d \leq 7 \\ \\ 3.42d & \text{for } d \geq 8. \end{cases}$$

Suppose $4 \leq d \leq 7$. Since $d \leq 7$, it follows that $0.14d < 1$, which is equivalent to $d-1 < 0.86d$. Consider the following inequalities

$$(d-1)^{\frac{n}{n-2}} < (d-1)(d^{\frac{2}{n-2}}) < 0.86d(4) = 3.44d,$$

which yields the desired result.

Suppose $d \geq 8$. Consider the following inequalities

$$(d-1)^{\frac{n}{n-2}} < d(d^{\frac{2}{n-2}}) < d(3.42) = 3.42d,$$

which yields the desired result.

Combining this with $2^n n > 3.46d^2$, we find that $2^n n > d(d-1)^{\frac{n}{n-2}}$ for $d \geq 4$. By Theorem 4.4.5, the A-equation is isotropic. $\square$

Graphically, we can visualize this result with the following figure.

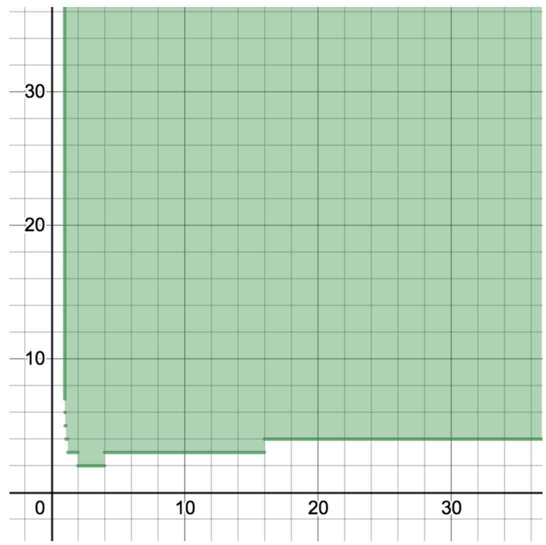
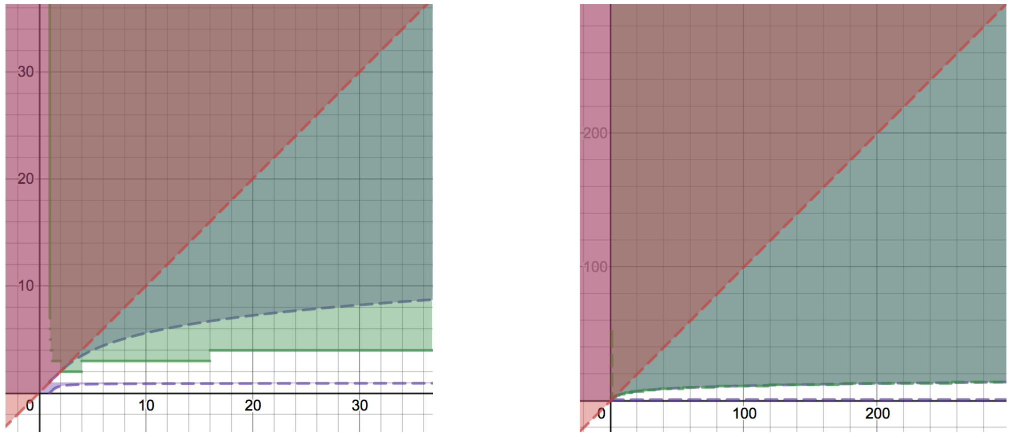


Figure 4.3: This figure illustrates the relationship between n and d needed to guarantee a nontrivial solution to a system of equations by Theorem 4.4.7. The horizontal axis represents our degree d and the vertical axis represents the number of variables n. The shaded region is the order pairs of (d, n) that guarantee a nontrivial solution to our A-equation.

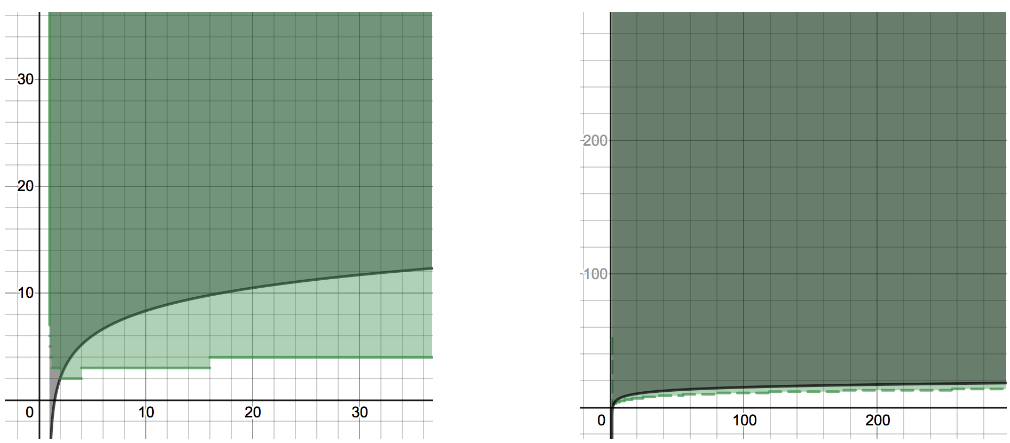
Furthermore, it is of interest to compare the bounds given by Theorem 4.4.7 and Corollary 4.4.6 with the bound given by Chevalley (Theorem 2.1.1.



(a) Behavior for small values of d (b) Behavior for large values of d

Figure 4.4: These figures illustrates the relationship between Corollary 4.4.6 (purple region), Theorem 4.4.7 (green region), and Theorem 2.1.1 (red region). The horizontal axis represents our degree d and the vertical axis represents the number of variables n. We can see that the results given by Theorem 4.4.7 are the strongest for A-equations.

In the case of diagonal forms, we realized that Theorem 4.4.1 and Theorem 4.4.7 were incredibly similar results. Graphically, we can visualize them with the following figure.



(a) Behavior for small values of d (b) Behavior for large values of d

Figure 4.5: These figures compares the bounds given by Theorem 4.4.1 (black region) and Theorem 4.4.7 (green region). We can see that Theorem 4.4.7 is a stronger result.

In fact, for $d \geq 3$, Theorem 4.4.7 is a stronger result than Theorem 4.4.1. The following lemma and proposition prove the previous statement.

Lemma 4.4.8. *If $d \geq 3$, then* $2\log_2(d) - \log_2(\log_2(d)) \leq 2\log_2(d-1) + 1$.

Proof. The statement is equivalent to proving $\dfrac{d^2}{\log_2(d)} \leq 2(d-1)^2$. It is sufficient to show $(1+\frac{1}{d-1})^2 = \dfrac{d^2}{(d-1)^2} \leq 2\log_2(d) = \log_2(d^2)$.

Suppose $d = 3$. Then we satisfy the desired inequality since $\dfrac{9}{4} < 3 < \log_2(9)$. Notice that $(1+\frac{1}{d-1})^2$ is a decreasing function and $\log_2(d^2)$ is an increasing function. Therefore, since the inequality held for $d = 3$, the inequality is true for $d > 3$. $\square$

Proposition 4.4.9. *Let $d \geq 3$. Suppose $\Omega_A(d,q) \leq \lceil 2\log_2(d) - \log_2(\log_2(d))\rceil$. Then*

(i) $\Omega_A(d,q) < 2\log_2(d-1)+2$,

(ii) If $n \geq 2\log_2(d-1)+2$, then the A-equation $\sum_{j=1}^{n} \alpha_j x_j^d$ is isotropic over $\mathbb{F}_q$.

Proof. (i) By Lemma 4.4.8, for $d \geq 3$, we have $2\log_2(d) - \log_2(\log_2(d)) \leq 2\log_2(d-1)+1$. Thus, $\lceil 2\log_2(d) - \log_2(\log_2(d))\rceil < 2\log_2(d-1)+2$. By hypothesis,

$$\Omega_A(d,q) \leq \lceil 2\log_2(d) - \log_2(\log_2(d))\rceil < 2\log_2(d-1)+2.$$

(ii) If $n \geq 2\log_2(d-1)+2 > \Omega_A(d,q)$, then the A-equation $\sum_{j=1}^{n} \alpha_j x_j^n$ is isotropic. $\square$

4.5 Results on A-Systems

In this section, we will give an overview of the results currently known for A-systems. It will become apparent that several of these results improve some previously known bounds giving more evidence of the usefulness of the classification of A-systems.

Theorem 4.5.1 (Tietäväinen, [14], Part of Theorem 3, page 19)**.** *Let $d_i = p^{k_i}b_i$, where $p \nmid b_i$ and let $s_i = p^{k-k_i}$. If $\sum_{j=1}^{n} \alpha_{ij} x_j^{d_i} = 0$, $i = 1,\ldots,r$, is an A-system, then $\sum_{j=1}^{n} \alpha_{ij}^{s_i} x_j^{b_i} = 0$, $i = 1,\ldots,r$, is an A-system. Additionally, if $\sum_{j=1}^{n} \alpha_{ij}^{s_i} x_j^{b_i} = 0$, $i = 1,\ldots,r$, has a nontrivial zero over $\mathbb{F}_q$, then $\sum_{j=1}^{n} \alpha_{ij} x_j^{d_i} = 0$, $i = 1,\ldots r$, has a nontrivial zero over $\mathbb{F}_q$.*

Proof. Since $\alpha^q = \alpha$, for all $\alpha \in \mathbb{F}_q$. We may assume $d_i < q$ for all $i = 1,\ldots,r$. Thus

$$p^{k_i} \leq p^{k_i}b_i < q = p^k$$

Therefore, $k_i < k$. If $\sum_{j=1}^{n} \alpha_{ij} x_j^{d_i} = 0$, $i = 1, \ldots, r$, is an A-system, then there exists $\eta \in \mathbb{F}_q^*$ such that $\eta^{d_i} = -1$ for all $i = 1, \ldots, r$. Notice

$$(\eta^{b_i})^{p^{k_i}} = \eta^{d_i} = -1 = (-1)^{p^{k_i}}$$

for all $i = 1, \ldots, r$. Since $\varphi : \mathbb{F}_q \to \mathbb{F}_q$, $x \mapsto x^p$ is an automorphism, it follows that $\eta^{b_1} = \eta^{b_2} = \cdots = \eta^{b_r} = -1$. Thus, $\sum_{j=1}^{n} \alpha_{ij}^{s_i} x_j^{b_i}$ is an A-system.

Now let $(\lambda_1, \ldots, \lambda_n)$ be a nontrivial solution to $\sum_{j=1}^{n} \alpha_{ij}^{s_i} x_j^{b_i}$, $i = 1, \ldots, r$. Consider

$$\sum_{j=1}^{n} \alpha_{ij} \lambda_j^{d_i} = \sum_{j=1}^{n} \alpha_{ij}^{p^k} \lambda_j^{d_i} = \sum_{j=1}^{n} (\alpha_{ij}^{s_i} \lambda_j^{b_i})^{p^{k_i}} = (\sum_{j=1}^{n} \alpha_{ij}^{s_i} \lambda_j^{b_i})^{p^{k_i}} = 0^{p^{k_i}} = 0.$$

Thus $(\lambda_1, \ldots, \lambda_n)$ is also a nontrivial solution of $\sum_{j=1}^{n} \alpha_{ij} x_j^{d_i} = 0$, $i = 1, \ldots, r$. $\square$

Theorem 4.5.2 (Tietäväinen, [14], Theorem 4, page 20)**.** *Assume $d \geq 2$. The A-system $\sum_{j=1}^{n} \alpha_{ij} x_j^d = 0$ for $i = 1, \ldots, r$ has a non-trivial solution in $\mathbb{F}_q$ if*

$$n \geq 2r(r + D)$$

where $D = \max(\log_2(d-1), 1)$.

Proof. If $d = 2$, we know that $D = 1$. Since $r \geq 1$, we find

$$n \geq 2r(r+1) > 2r.$$

By Chevalley's Theorem (Theorem 2.1.1, [3]), the A-system $\sum_{j=1}^{n} \alpha_{ij} x_j^d = 0$, $i = 1, \ldots, r$ has a non-trivial solution in $\mathbb{F}_q$. Therefore, we may assume that $d \geq 3$, which implies that $D = \log_2(d-1)$.

First consider the case $r = 1$. When $r = 1$,

$$n \geq 2(1 + \log_2(d-1)) = 2\log_2(d-1) + 2.$$

By Theorem 4.4.1, the A-equation $\sum_{j=1}^{n} \alpha_{1j} x_j^d = 0$ has a non-trivial solution in $\mathbb{F}_q$.

Assume that this theorem is true for systems of $r - 1$ equations where $r \geq 2$. We will show that it is true for systems of r equations.

Suppose by way of contradiction that $\sum_{j=1}^{n} \alpha_{ij} x_j^d = 0$, $i = 1, \ldots, r$ has only the trivial solution in $\mathbb{F}_q$ and that $n \geq 2r(r + D)$. By Lemma 4.3.7, $2^n \leq q^r$, which implies $2^{\frac{n}{r}} \leq q$.

First suppose that $q < 2^{2r}(d-1)^2$. Then

$$2^{2r+2D} \leq 2^{\frac{n}{r}} \leq q < 2^{2r}(d-1)^2$$

$$2^{2D} < (d-1)^2$$

$$2D < 2\log_2(d-1)$$

$$D < \log_2(d-1)$$

Since $D = \log_2(d-1)$, we reach a contradiction ↯.

Now suppose $q \geq 2^{2r}(d-1)^2$. Taking the square root of both sides yields

$$q^{\frac{1}{2}} \geq 2^r(d-1) > (d-1).$$

Thus, we have $q > (d-1)q^{\frac{1}{2}}$.

Suppose that $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_r) \in \mathbb{F}_q^r \setminus \mathbf{0}$. Then at least one λ_i, say λ_u, is non-zero. Therefore, the system $\sum_{j=1}^{n} \alpha_{ij} x_j^d = 0$ for $i = 1, \ldots, r$ is equivalent to the following system

$$\begin{cases} \sum_{j=1}^{n} \alpha_{ij} x_j^d = 0 \text{ for } i = 1, \ldots, u-1, u+1, \ldots, r \\ \sum_{i=1}^{r} \lambda_i f_i = 0 \end{cases}$$

Let $\mathbf{g}_j(x_j) = (\alpha_{1j} x_j^d, \alpha_{2j} x_j^d, \ldots, \alpha_{rj} x_j^d)$. We will show that $\boldsymbol{\lambda} \mathbf{g}_j(x_j) = \sum_{i=1}^{r} \lambda_i \alpha_{ij} x_j^d$, for $j = 1, \ldots, n$, is identically zero for at most $t := \lceil 2(r-1)(r-1+D) \rceil - 1$ values of j.

Suppose that $\boldsymbol{\lambda} \mathbf{g}_j(x_j) = \sum_{i=1}^{r} \lambda_i \alpha_{ij} x_j^d$ is identically zero for $t+1$ values of j, say $j = 1, \ldots, t+1$. Then we can rewrite the system above as

$$\begin{cases} \sum_{j=1}^{n} \alpha_{ij} x_j^d = 0 \text{ for } i = 1, \ldots, u-1, u+1, \ldots, r \\ \sum_{j=t+2}^{n} \sum_{i=1}^{r} \lambda_i \alpha_{ij} x_j^d = 0 \end{cases}$$

Let $x_{t+2} = \cdots = x_n = 0$. Then our system becomes $\sum_{j=1}^{t+1} \alpha_{ij} x_j^d = 0$ for $i = 1, \ldots, u-1, u+1, \ldots, r$. By the induction hypothesis, this system has a non-trivial solution $(x_1, \ldots, x_{t+1})$ if $t+1 \geq 2(r-1)(r-1+D)$. Since $t = \lceil 2(r-1)(r-1+D) \rceil - 1$, we know

$$t + 1 = \lceil 2(r-1)(r-1+D) \rceil \geq 2(r-1)(r-1+D).$$

Thus, the system has a non-trivial solution $\lightning$. Thus, we know $\boldsymbol{\lambda}\mathbf{g}_j(x_j) = \sum_{i=1}^{r} \lambda_i \alpha_{ij} x_j^d$ is identically zero for at most $t = \lceil 2(r-1)(r-1+D) \rceil - 1$ values of j.

By Lemma 4.3.6, the number of solutions of the system $\sum_{j=1}^{n} \alpha_{ij} x_j^d = 0$ for $i = 1, \ldots, r$ is equal to

$$N(\boldsymbol{f}; \mathbb{F}_q^n) = q^{n-r} + q^{-r} \sum_{\substack{\boldsymbol{\lambda} \in \mathbb{F}_q^r \\ \boldsymbol{\lambda} \neq \mathbf{0}}} \prod_{j=1}^{n} \sum_{\xi_j \in \mathbb{F}_q} e(\boldsymbol{\lambda}\mathbf{g}_j(\xi_j)).$$

If $\boldsymbol{\lambda}\mathbf{g}_j$ is identically zero for some j, then $\sum_{\xi_j \in \mathbb{F}_q} e(\boldsymbol{\lambda}\mathbf{g}_j(\xi_j)) = q$.

In the other cases it follows that $\boldsymbol{\lambda}\mathbf{g}_j(x_j)$ satisfies the assumption of Lemma 4.3.8, which tells us that $\left| \sum_{\xi_j \in \mathbb{F}_q} e(\boldsymbol{\lambda}\mathbf{g}_j(\xi_j)) \right| \leq (d-1) q^{\frac{1}{2}}$.

Therefore,

$$
\begin{aligned}
N(\boldsymbol{f};\mathbb{F}_q^n) &\geq q^{n-r} - q^{-r}\Big|\sum_{\substack{\boldsymbol{\lambda}\in\mathbb{F}_q^r\\ \boldsymbol{\lambda}\neq\mathbf{0}}}\prod_{j=1}^{n}\sum_{\xi_j\in\mathbb{F}_q} e(\boldsymbol{\lambda}\mathbf{g}_j(\xi_j))\Big| \\
&= q^{n-r} - q^{-r}(q^r-1)q^t(d-1)^{(n-t)}q^{\frac{1}{2}(n-t)} \\
&= q^{n-r} - (q^r-1)(d-1)^{n-t}q^{\frac{1}{2}(n+t-2r)} \\
&= q^{\frac{1}{2}(n+t-2r)}\Big(q^{\frac{1}{2}(n-t)} - (q^r-1)(d-1)^{(n-t)}\Big) \\
&= q^{\frac{1}{2}(n+t-2r)}\Big(q^r\big(q^{\frac{1}{2}(n-t-2r)} - (1-q^{-r})(d-1)^{(n-t)}\big)\Big) \\
&= q^{\frac{1}{2}(n+t-2r)}\Big(q^r\big(q^{\frac{1}{2}(n-t-2r)} - (d-1)^{(n-t)} + q^{-r}(d-1)^{(n-t)}\big)\Big) \\
&= q^{\frac{1}{2}(n+t-2r)}\Big(q^r\big(q^{\frac{1}{2}(n-t-2r)} - (d-1)^{(n-t)}\big) + (d-1)^{(n-t)}\Big) \\
&> q^{\frac{1}{2}(n-t-2r)}(d-1)^{n-t}
\end{aligned}
$$

Since $2(r-1)(r-1+D)+1 > \lceil 2(r-1)(r-1+D)\rceil$, we can provide a lower bound on $n-t$

$$
\begin{aligned}
n-t &> 2r(r+D) - (\lceil 2(r-1)(r-1+D)\rceil - 1) \\
&> 2r(r+D) - 2(r-1)(r-1+D) \\
&= 4r + 2D - 2
\end{aligned}
$$

We will now show that $2^{2r}(d-1)^2 = (d-1)^{2+\frac{2r}{D}}$. It is sufficient to show that

$$2r = \frac{2r}{D}\log_2(d-1),$$

which is true because $D = \log_2(d-1)$. Therefore, $q \geq 2^{2r}(d-1)^2 = (d-1)^{2+\frac{2r}{D}}$.

Since $r \geq 2$, it follows

$$\frac{2r}{D} > \frac{2r}{r-1+D} = \frac{4r}{2(r-1+D)} > \frac{4r}{n-t-2r}$$

Therefore,

$$q^{n-t-2r} > ((d-1)^{(2+\frac{2r}{D})})^{(n-t-2r)}$$

$$> ((d-1)^{(2+\frac{4r}{n-t-2r})})^{(n-t-2r)}$$

$$= (d-1)^{2(n-t-2r)+4r}$$

$$= (d-1)^{2(n-t)}$$

Combining this with our lower bound on $N(\boldsymbol{f}; \mathbb{F}_q^n)$, we obtain

$$N(\boldsymbol{f}; \mathbb{F}_q^n) > q^{\frac{1}{2}(n+t-2r)}(d-1)^{n-t} > (d-1)^{2(n-t)} > 1$$

which is a contradiction ↯. Thus, our theorem has been proved.

$\square$

Corollary 4.5.3 (Tietäväinen, [14], Theorem 3, page 19)**.** *Let $d_i = p^{k_i} b_i$, where $p \nmid b_i$. Then the A-system $\sum_{j=1}^{n} \alpha_{ij} x_j^{d_i} = 0$, $1 \le i \le r$, has a nontrivial solution if*

$$n \ge \min\Big(1 + \sum_{i=1}^{r} b_i, 2r(r+B)\Big),$$

where $B = \max(\log_2(b-1), 1)$ and $b = \max b_i$.

Proof. Let $s = p^{k-k_i}$. By Theorem 4.5.1, $\sum_{j=1}^{n} (\alpha_{ij})^s x_j^{b_i} = 0, 1 \le i \le r$, is an A-system. By Chevalley's Theorem (Theorem 2.1.1, [3]), $\sum_{j=1}^{n} (\alpha_{ij})^s x_j^{b_i} = 0, 1 \le i \le r$, has a nontrivial solution when $n \ge 1 + \sum_{i=1}^{r} b_i$. By Theorem 4.5.2, $\sum_{j=1}^{n} (\alpha_{ij})^s x_j^{b_i} = 0, 1 \le i \le r$, has a nontrivial solution if $n \ge 2r(r+B)$, where $B = \max(\log_2(b-1), 1)$ and $b = \max b_i$. By Theorem 4.5.1, $\sum_{j=1}^{n} \alpha_{ij} x_j^{d_i} = 0, 1 \le i \le r$, has a nontrivial solution if $\sum_{j=1}^{n} (\alpha_{ij})^s x_j^{b_i} = 0, 1 \le i \le r$, has a nontrivial solution. Thus if $n \ge 1 + \sum_{i=1}^{r} b_i$ or $n \ge 2r(r+B)$, $\sum_{j=1}^{n} \alpha_{ij} x_j^{d_i} = 0, 1 \le i \le r$, has a nontrivial solution. $\square$

Lemma 4.5.4 (Tietäväinen, [14], Lemma 5, page 23)**.** *Suppose that there exist non-zero elements σ and τ of $\mathbb{F}_q$ such that $\sigma^d - \tau^d = 1$. Then the A-system $\sum_{j=1}^{n} \alpha_{ij}x_j^d = 0$ for $i = 1, \dots, r$ has a non-trivial solution in $\mathbb{F}_q$ if $3^n > q^r$.*

Proof. Since $3^n > q^r$, two of the 3^n vectors $(\sum_{j=1}^{n} \alpha_{1j}\delta_j^d, \dots, \sum_{j=1}^{n} \alpha_{rj}\delta_j^d)$, where $\delta_j = 0, 1,$ or σ, are equal. Hence $\sum_{j=1}^{n} \alpha_{ij}\delta_{1j}^d = \sum_{j=1}^{n} \alpha_{ij}\delta_{2j}^d$ for $i = 1, \dots, r$, where $\delta_{kj} = 0, 1,$ or σ, for $k = 1, 2$, and $(\delta_{11}, \dots, \delta_{1n}) \neq (\delta_{21}, \dots, \delta_{2n})$. Therefore, $\sum_{j=1}^{n} \alpha_{ij}\epsilon_j = 0$ for $i = 1, \dots, r$ where $\epsilon_j = \delta_{1j}^d - \delta_{2j}^d \in \{0, \pm 1, \pm\sigma^d, \pm\tau^d\}$, and $(\epsilon_1, \dots, \epsilon_n) \neq (0, \dots, 0)$.

Since $\sum_{j=1}^{n} \alpha_{ij}x_j^d = 0$ is an A-system, there exists an element η of $\mathbb{F}_q$ such that $\eta^d = -1$. Consequently, $\epsilon_j = \chi_j^d$ where $\chi_j \in \{0, 1, \eta, \sigma, \eta\sigma, \tau, \eta\tau\}$. Therefore, the A-system $\sum_{j=1}^{n} \alpha_{ij}x_j^d = 0$ has the non-trivial solution $(\chi_1, \dots, \chi_n)$ in $\mathbb{F}_q$. $\square$

Lemma 4.5.5 (Tietäväinen, [14], Lemma 6, page 23)**.** *Assume $q \equiv 1 \bmod 3d$. Then the A-system $\sum_{j=1}^{n} \alpha_{ij}x_j^d = 0$ for $i = 1, \dots r$ has a non-trivial solution in $\mathbb{F}_q$ if $3^n > q^r$.*

Proof. Let $\mathbb{F}_q^* = \langle \rho \rangle$. Then $\rho^{q-1} = \rho^{3md} = 1$, where $m = \frac{q-1}{3}$. Thus,

$$\rho^{3md} - 1 = (\rho^{md} - 1)(\rho^{2md} + \rho^{md} + 1) = 0.$$

Since $\rho^{md} \neq 1$, we have $\rho^{2md} + \rho^{md} + 1 = 0$. Since our system is an A-system, we know there exists an $\eta \in \mathbb{F}_q$ such that $\eta^d = -1$. Then we can rewrite $\rho^{2md}+\rho^{md}+1 = 0$ as $(\eta\rho^{2m})^d - (\rho^m)^d = 1$. By Lemma 4.5.4 with $\sigma = \eta\rho^{2m}$, $\tau = \rho^m$, the result follows. $\square$

Theorem 4.5.6 (Tietäväinen, [14], Theorem 6, page 24)**.** *Assume $d \geq 2$. The A-system $\sum_{j=1}^{n} \alpha_{ij}x_j^d = 0; i = 1, \dots r$ has a non-trivial solution in every finite field $\mathbb{F}_q$, $q \equiv 1 \bmod 3d$ if*

$$n \geq 2r(r + D'),$$

where $D' = \max(\log_3(d-1), 1)$.

The proof of Theorem 4.5.6 follows the same structure as the proof of Theorem 4.5.2. The proof of Theorem 4.5.6 uses Lemma 4.5.5 instead of Lemma 4.3.7. For this reason, we will omit the proof of Theorem 4.5.6.

Chapter 5 Existence of Nontrivial Solutions for Diagonal Forms

5.1 Introduction

The focus of this chapter is to approach systems of diagonal forms of arbitrary degree over finite fields. Sections 5.2 and 5.3 collate a number of known results in the area, as well as, providing more complete proofs than currently exist in the literature and some small improvements to some of these results. In Sections 5.4 and 5.5, we consider the computational question of finding the explicit lower bound on the number of variables needed to guarantee a nontrivial solution. In particular, for a system of r diagonal forms over $\mathbb{F}_q$ with specified degrees, we compute the largest integer such that there exists an anisotropic system in that many variables. In some special cases, we also compute the maximum of this value over all possible choices of q.

5.2 General Diagonal Forms

While Gray considered the case of an odd prime degree equation, Tietäväinen was able to make improvements to Chevalley's result (Theorem 2.1.1) for a single equation of arbitrary degree. In the case of a single equation Chevalley's result showed that if $n \geq d+1$, we were guaranteed a nontrivial solution, whereas, with the exception of few forms, Tietäväinen showed that if $n \geq \dfrac{d+3}{2}$, we are guaranteed a nontrivial solution. This result, asymptotically, halves the bound given by Chevalley. Tietäväinen stated that a particular case of the proof could be verified using known results, but did not present the full argument. We will present the full proof here.

Theorem 5.2.1 (Tietäväinen, [16], Theorem 2)**.** *Let $\mathbb{F}_q$ be the finite field of p^k elements and suppose that $d \mid q-1$. Assume $(d,k) \neq (p-1,1)$. If $n \geq \dfrac{d+3}{2}$, then the equation*

$$\alpha_1 x_1^d + \cdots + \alpha_n x_n^d = 0$$

has a non-trivial solution in $\mathbb{F}_q$.

Graphically, we can visualize the result in the following figure.

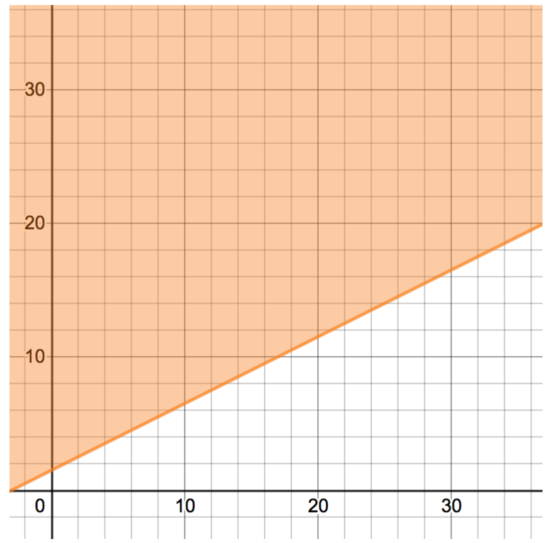


Figure 5.1: This figure illustrates the relationship between n and d needed to guaranteed a nontrivial solution to a single diagonal equation. The horizontal axis represents our degree d and the vertical axis represents the number of variables n. The shaded region is the ordered pairs (d, n) that guarantee a nontrivial solution to our equation.

Furthermore, it is of interest to compare the bounds given by Theorem 5.2.1 with Chevalley's result.

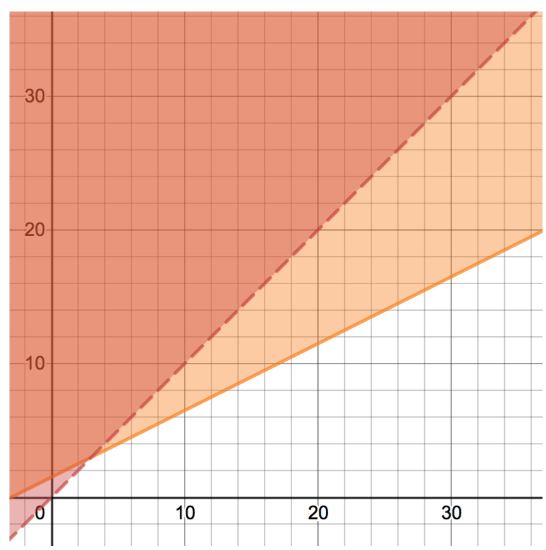


Figure 5.2: This figure illustrates the relationship between Theorem 5.2.1 (orange region) and Theorem 2.1.1 (red region). For a single equation, we can see the Theorem 5.2.1 almost halves the bound given in Theorem 2.1.1.

The proof of Theorem 5.2.1 follows the ideas of Tietäväinen's proof, which requires the following notation and lemmas.

In the case of a single form, we can always reduce to the case where $d \mid q-1 = p^k - 1$. Let A be a subset of $\mathbb{F}_q$. Define the following

$$H(A) := \{\eta \in \mathbb{F}_q \mid A + \eta = A\}$$

$$Q_w = Q_w(\alpha_1, \dots, \alpha_w) := \{\eta \mid \eta = \sum_{j=1}^{w} \alpha_j \xi_j^d, \xi_j \in \mathbb{F}_q\}$$

Notice that $H(A)$ is an additive subgroup of $\mathbb{F}_q$, and A is the union of some additive cosets of $\mathbb{F}_q$ modulo $H(A)$. Furthermore, there is an integer l_w such that $|Q_w^*| = \dfrac{l_w(q-1)}{d}$.

Definition 5.2.2. If $H(A) \neq \{0\}$, A is said to be *periodic.* If $H(A) = \{0\}$, A is *aperiodic.*

Definition 5.2.3. An element γ of Q_{s+1} has a *unique representation* in Q_{s+1} if it has only one representation as $\gamma = \alpha + \beta$ with $\alpha \in Q_s$ and $\beta \in \alpha_{s+1}(\mathbb{F}_q)^d$.

Lemma 5.2.4 (Tietäväinen, [16], Lemma 1)**.** *If $\alpha_1, \dots, \alpha_{w+1}$ are non-zero elements of $\mathbb{F}_q$ and $\dfrac{q-1}{d} \nmid p^v - 1$ for $1 \leq v < k$, then $l_{w+1} \geq \min(1 + l_w, d)$.*

For a proof, see [15], proof of Lemma 1.

Lemma 5.2.5 (Tietäväinen, [16], Lemma 2)**.** *If $d \leq \dfrac{q-1}{2}$, then $\displaystyle\sum_{\alpha \in Q_w} \alpha = 0$.*

Proof. Since $d \leq \dfrac{q-1}{2}$, we know $|(\mathbb{F}_q^*)^d| \geq 2$. Thus, there exists an element $\beta \in (\mathbb{F}_q^*)^d$ such that $\beta \neq 1$. Clearly $\beta Q_w = Q_w$. Let $\displaystyle\sum_{\alpha \in Q_w} \alpha = \delta$. Then

$$\beta\delta = \sum_{\alpha \in Q_w} \beta\alpha = \sum_{\gamma \in Q_w} \gamma = \delta$$

Thus $\beta\delta - \delta = 0$, which implies $(\beta - 1)\delta = 0$. Since $\beta \neq 1$, we know $\delta = 0$. $\square$

Lemma 5.2.6 (Tietäväinen, [16], Lemma 3)**.** *If A is a periodic subset of $\mathbb{F}_q$ and $p > 2$, then $\displaystyle\sum_{\alpha \in A} \alpha = 0$*

Proof. Since $H(A)$ is an additive subgroup of $\mathbb{F}_q$, $H(A) \neq \{0\}$, and $p > 2$, we can conclude $2H(A) = H(A)$. Let $\displaystyle\sum_{\alpha \in H(A)} \alpha = \delta$. Then

$$2\delta = \sum_{\alpha \in H(A)} 2\alpha = \sum_{\beta \in H(A)} \beta = \delta$$

Thus $2\delta - \delta = 0$, which implies $\delta = 0$.

Let $B = \gamma + H(A)$ be an additive coset of $\mathbb{F}_q$ modulo $H(A)$. Then

$$\sum_{\alpha \in B} \alpha = |H(A)|\gamma + \sum_{\alpha \in H(A)} \alpha = |H(A)|\gamma$$

Since p divides $|H(A)|$, we know $|H(A)|\gamma = 0$. Since A is the union of some additive cosets of $\mathbb{F}_q$ modulo $H(A)$, we obtain the desired result. $\square$

Lemma 5.2.7 (Tietäväinen, [16], Lemma 4)**.** *Suppose that* $\frac{q-1}{d} \nmid p^v - 1$ *for* $1 \leq v < k$ *and that* A *is a non-empty periodic subset of* $\mathbb{F}_q$ *such that* $\alpha A = A$ *for every element* $\alpha \in (\mathbb{F}_q^*)^d$*. Then* $A = \mathbb{F}_q$*.*

Proof. Since A is periodic, $H(A) \neq \{0\}$ and thus $|H(A)| = p^v$ where $v \geq 1$. If $\beta \in H(A)^*$, then

$$\beta\alpha + A = \beta\alpha + \alpha A = \alpha(\beta + A) = \alpha A = A$$

for every element $\alpha \in (\mathbb{F}_q^*)^d$. This means that $\beta(\mathbb{F}_q^*)^d$ is a subset of $H(A)^*$. Thus $H(A)^*$ is the union of some cosets of the multiplicative group $\mathbb{F}_q^*$ modulo $(\mathbb{F}_q^*)^d$. This implies that $\frac{q-1}{d}$ divides $|H(A)^*| = p^v - 1$. By the assumptions of this lemma, $v = k$ and thus $H(A) = \mathbb{F}_q$, which implies $A = \mathbb{F}_q$. $\square$

Lemma 5.2.8 (Tietäväinen, [16], Lemma 5)**.** *Let* A *and* B *be subsets of* $\mathbb{F}_q$ *satisfying*

$$|A + B| \leq |A| + |B| - 2$$

Then $A + B$ *is periodic.*

Lemma 5.2.8 is due to Kneser [12].

Lemma 5.2.9 (Tietäväinen, [16], Lemma 6)**.** *Let* A *and* B *be subsets of* $\mathbb{F}_q$ *such that*

$$|A + B| = |A| + |B| - 1$$

Let $\gamma_1, \ldots, \gamma_t$ *denote all the elements in* $|A + B|$ *having only one representation as* $\gamma_j = \alpha_j + \beta_j$*, where* $\alpha_j \in A$ *and* $\beta_j \in B$*. Then*

(i) If $t = 0$*, then* $A + B$ *is either periodic or can be made periodic by adding one element.*

(ii) If $t = 1$*, then* $A + B$ *is either periodic or can be made periodic by deleting one element.*

(iii) If $t \geq 3$*, then either* $\alpha_1 = \cdots = \alpha_t$ *or* $\beta_1 = \cdots = \beta_t$*. Moreover, every element in* $A + B$*, other than* γ_i*, has at least* t *representations as the sum of an element of* A *and an element of* B*.*

Lemma 5.2.9 is a special case of Theorem 6.1 in a paper by Kemperman [11].

Lemma 5.2.10 (Tietäväinen, [16], Lemma 7)**.** *Let v be a factor of k such that $p^v - 1$ is divisible by $\frac{q-1}{d}$. Then the equation*

$$\alpha_1 x_1^d + \cdots + \alpha_n x_n^d = 0$$

has a non-trivial solution in $\mathbb{F}_q$ if $n \geq 1 + \frac{kd(p^v - 1)}{v(p^k - 1)}$.

Lemma 5.2.10 follows from Theorem 5.3.4, which is Theorem 5 from [14] (Note that the proof of Theorem 5.3.4 uses only material prior to this result).

Theorem 5.2.11 (Tietäväinen, [16], Theorem 1)**.** *Let $\mathbb{F}_q$ be the finite field of p^k elements where p is an odd prime. Suppose that $\alpha_1, \ldots, \alpha_n$ are nonzero elements of $\mathbb{F}_q$, $d \mid q-1$, $\frac{q-1}{d} \nmid p^v - 1$ for $1 \leq v < k$, and $d < \frac{q-1}{2}$. Then the form $\alpha_1 x_1^d + \cdots + \alpha_n x_n^d$ represents either all the elements or at least $\frac{(2n-1)(q-1)}{d} + 1$ elements of $\mathbb{F}_q$.*

Proof. First notice that $Q_1 = \alpha \mathbb{F}_q^d$. Thus, $|Q_1^*| = \frac{q-1}{d}$, so $l_1 = 1$. Since $l_1 = 1$, we will show the statement of the theorem is equivalent to the following:

$$l_{s+1} \geq \min(2 + l_s, d) \text{ for } s = 1, \ldots, n-1.$$

Notice when $l_{s+1} = d$ for some s, it follows that $|Q_{s+1}^*| = q - 1$ and thus

$$\alpha_1 x_1^d + \cdots + \alpha_{s+1} x_{s+1}^d$$

represents all the elements of $\mathbb{F}_q$. Therefore, $\alpha_1 x_1^d + \cdots + \alpha_n x_n^d$ represents all the elements of $\mathbb{F}_q$. Otherwise if $l_{s+1} = 2 + l_s$ for all s, then $l_n = (2n-1)$. This means that $|Q_n^*| = \frac{(2n-1)(q-1)}{d}$ and thus $\alpha_1 x_1^d + \cdots + \alpha_n x_n^d$ represents at least $\frac{(2n-1)(q-1)}{d} + 1$ elements of $\mathbb{F}_q$.

By Lemma 5.2.4, we know that $l_{s+1} \geq \min(1 + l_s, d)$. Notice from the discussion above, if $l_{s+1} = d$ for some s, then $\alpha_1 x_1^d + \cdots + \alpha_n x_n^d$ represents all the elements of $\mathbb{F}_q$ and the result holds.

Now suppose $l_{s+1} = 1 + l_s$. We will show that this implies either $l_{s+1} = 2 + l_s$ or $l_{s+1} = d$. Let t be the number of elements in Q_{s+1} that have unique representations in Q_{s+1}. If γ has a unique representation in Q_{s+1}, then so does every element in the set $\gamma(\mathbb{F}_q^*)^d$. Hence $t \equiv 1 \bmod \frac{q-1}{d}$ if 0 has only the trivial representation in Q_{s+1} and $t \equiv 0 \bmod \frac{q-1}{d}$ if 0 has at least one non-trivial representation in Q_{s+1}.

Case I: Suppose Q_{s+1} is periodic. Then $l_{s+1} = d$ by Lemma 5.2.7. Thus, $\alpha_1 x_1^d + \cdots + \alpha_n x_n^d$ represents all the elements of $\mathbb{F}_q$.

Case II: Suppose Q_{s+1} is aperiodic and $t = 0$. Then by Lemma 5.2.9, Q_{s+1} can be made periodic by adding one element. Let β be this element. Then $\sum_{\alpha \in Q_{s+1}} \alpha = -\beta$

by Lemma 5.2.6. On the other hand, $\sum_{\alpha \in Q_{s+1}} \alpha = 0$ by Lemma 5.2.5. Thus $\beta = 0$ ↯. This is impossible, since $0 \in Q_{s+1}$.

Case III: Suppose Q_{s+1} is aperiodic and $t = 1$. Then by Lemma 5.2.9, Q_{s+1} can be made periodic by deleting one element. By Lemma 5.2.6 and Lemma 5.2.5, this element is 0. Therefore Q^*_{s+1} is periodic ↯. This is impossible by Lemma 5.2.7.

Case IV: Suppose Q_{s+1} is aperiodic and $t = 2$. This case is impossible since $\frac{q-1}{d} > 2$.

Case V: Suppose Q_{s+1} is aperiodic and $t \geq 3$. Let $\delta_1, \ldots, \delta_t$ be the uniquely represented elements in Q_{s+1}. Then, by Lemma 5.2.9, either there exists some unique element $\beta \in Q_s$ such that

$$\delta_j = \beta + \alpha_{s+1}\gamma_j^d \text{ for } j = 1, \ldots, t$$

or there exists some unique element $\gamma^d \in \mathbb{F}_q^d$ such that

$$\delta_j = \beta_j + \alpha_{s+1}\gamma^d \text{ for } j = 1, \ldots, t$$

where the $\beta_j \in Q_s$.

Suppose there exists some unique element $\beta \in Q_s$ such that

$$\delta_j = \beta + \alpha_{s+1}\gamma_j^d \text{ for } j = 1, \ldots, t.$$

Let $\epsilon^d \notin \{0, 1\}$. Then $\epsilon^d \delta_j$ has the unique representation

$$\epsilon^d \delta_j = \epsilon^d \beta + \alpha_{s+1}(\gamma_j \epsilon)^d$$

in Q_{s+1}. Consequently, $\epsilon^d \delta_j$ is one of the elements $\delta_1, \ldots, \delta_t$ and therefore $\epsilon^d \beta = \beta$. Hence $\beta = 0$. It follows from this that the set of uniquely represented elements, U, in Q_{s+1} is $\alpha_{s+1}\mathbb{F}_q^d$ or $\alpha_{s+1}(\mathbb{F}_q^*)^d$. If $U = \alpha_{s+1}\mathbb{F}_q^d$, then Q_s^* is periodic by an observation on page 64 of [13]. Thus by Lemma 5.2.7, $Q_s^* = \mathbb{F}_q$, which is impossible.

If $U = \alpha_{s+1}(\mathbb{F}_q^*)^d$, then

$$Q_s^* + \alpha_{s+1}(\mathbb{F}_q^*)^d = Q_s$$

and, by Lemma 5.2.8, Q_s is periodic. Thus, by Lemma 5.2.7, $Q_s = \mathbb{F}_q$. So $Q_{s+1} = \mathbb{F}_q$ ↯. This is impossible because Q_s is aperiodic.

Suppose now there exists some unique element $\gamma^d \in \mathbb{F}_q^d$ such that

$$\delta_j = \beta_j + \alpha_{s+1}\gamma^d \text{ for } j = 1, \ldots, t$$

where the $\beta_j \in Q_s$. Then the product $\epsilon^d \delta_j$, where $\epsilon^d \notin \{0, 1\}$, has a unique representation

$$\epsilon^d \delta_j = \epsilon^d \beta_j + \alpha_{s+1}\epsilon^d \gamma^d$$

in Q_{s+1} and therefore $\alpha_{s+1}\epsilon^d\gamma^d = \alpha_{s+1}\gamma^d$. Therefore, $\gamma = 0$. If 0 has only the trivial representation, then by [13], there exists a non-zero element α of $\mathbb{F}_q$ such that the difference set $Q_{s+1} \setminus \alpha\mathbb{F}_q^d$ is periodic. This is impossible by Lemma 5.2.7. For simplicity, let $m = \dfrac{q-1}{d}$. Thus there exists a positive integer g such that $t = gm$, $g \leq l_s$. Now, by Lemma 5.2.9, those elements in Q_{s+1} not uniquely represented in Q_{s+1} have at least gm representations. As there are $(1 + l_s m)(1 + m)$ sums of an element of Q_s and an element of $\alpha_{s+1}\mathbb{F}_q^d$, we get

$$(1 + (1 - l_s g)m)gm + gm \leq (1 + l_s m)(1 + m)$$

which simplifies to

$$m^2(g - 1 - m^{-1})(g - l_s - m^{-1}) \geq 0$$

. Since $g < l_s + m^{-1}$, we have the inequality $g \leq 1 + m^{-1}$. Thus $g = 1$. Suppose that $\alpha(\mathbb{F}_q^*)^d$ is the subset of Q_s which consists of exactly those elements which have a unique representation in Q_{s+1}. Then

$$Q_s + \alpha_{s+1}(\mathbb{F}_q^*)^d = Q_{s+1} \setminus \alpha(\mathbb{F}_q^*)^d$$

and thus, by Lemma 5.2.8, $Q_{s+1} \setminus \alpha(\mathbb{F}_q^*)^d$ is periodic. This is impossible by Lemma 5.2.7.

Now we have show that the equation $l_{s+1} = 1 + l_s$ implies the equation $l_{s+1} = k$. Thus, we have shown the desired result. □

The following lemma is required to prove the main result of this section, Theorem 5.2.1. We suspect that Tietäväinen knew the result to be true, but Tietäväinen did not present this result explicitly.

Lemma 5.2.12 (Leep–Petrik)**.** *Let $\mathbb{F}_q$ be the finite field of p^k elements and suppose that $d \mid q - 1$. Assume $(d, k) \neq (p - 1, 1)$. If $n \geq \dfrac{d+3}{2}$, then the A-equation*

$$\alpha_1 x_1^d + \cdots + \alpha_n x_n^d = 0$$

has a non-trivial solution in $\mathbb{F}_q$.

Proof. Suppose $d = 2$. Since $n \geq \dfrac{d+3}{2}$, $n \geq \dfrac{5}{2}$. This implies $n \geq 3$. Since $\max_q \Omega_A(1, 2, q) = 2$ by Table 5.3, the equation is isotropic.

Suppose $d = 3$. Since $n \geq \dfrac{d+3}{2}$, $n \geq 3$. Since $\max_q \Omega_A(1, 3, q) = 2$ by Table 5.3, the equation is isotropic.

Suppose $d = 4$. Since $n \geq \dfrac{d+3}{2}$, $n \geq \dfrac{7}{2}$. This implies $n \geq 4$. Since $\max_q \Omega_A(1, 4, q) = 2$ by Table 5.3, the equation is isotropic.

Suppose $d = 5$. Since $n \geq \dfrac{d+3}{2}$, $n \geq 4$. Since $\max_q \Omega_A(1, 5, q) = 3$ by Table 5.3, the equation is isotropic.

Suppose $d = 6$. Since $n \geq \dfrac{d+3}{2}$, $n \geq \dfrac{9}{2}$. This implies $n \geq 5$. Since $\max_q \Omega_A(1, 6, q) = 3$ by Table 5.3, the equation is isotropic.

Suppose $d = 7$. Since $n \geq \dfrac{d+3}{2}$, $n \geq 5$. Since $\max_q \Omega_A(1,7,q) = 3$ by Table 5.3, the equation is isotropic.

Suppose $d \geq 8$. By Theorem 4.4.7, if $n \geq 1 + \lceil 2\log_2(d) - \log_2(\log_2(d))\rceil$, then the A-equation $\alpha_1 x_1^d + \cdots + \alpha_n x_n^d$ is isotropic. We want to show that

$$n \geq \frac{d+3}{2} \geq 1 + \lceil 2\log_2(d) - \log_2(\log_2(d))\rceil.$$

First we will demonstrate that it is sufficient to show

$$\frac{d+3}{2} > 1 + 2\log_2(d) - \log_2(\log_2(d)).$$

Suppose this has been shown. Then

$$n \geq \frac{d+3}{2} > 1 + 2\log_2(d) - \log_2(\log_2(d)).$$

Thus,

$$n \geq \lceil 1 + 2\log_2(d) - \log_2(\log_2(d))\rceil = 1 + \lceil 2\log_2(d) - \log_2(\log_2(d))\rceil.$$

Thus, it is sufficient to show

$$\frac{d+3}{2} > 1 + 2\log_2(d) - \log_2(\log_2(d)).$$

Observe

$$\frac{d+3}{2} > 1 + 2\log_2(d) - \log_2(\log_2(d))$$

if and only if

$$\frac{d+1}{2} > 2\log_2(d) - \log_2(\log_2(d)) = \log_2(d^2) - \log_2(\log_2(d)) = \log_2\left(\frac{d^2}{\log_2(d)}\right).$$

Since $d \geq 8$, it follows that $\log_2(d) \geq 3$. Since $\log_2\left(\dfrac{d^2}{3}\right) \geq \log_2\left(\dfrac{d^2}{\log_2(d)}\right)$, it is sufficient to show

$$\frac{d+1}{2} > \log_2\left(\frac{d^2}{3}\right).$$

Thus, it is sufficient to show

$$d + 1 > 2\log_2\left(\frac{d^2}{3}\right) = \log_2\left(\frac{d^4}{9}\right).$$

To prove this last inequality, let $g(x) = x + 1 - \log_2\left(\dfrac{x^4}{9}\right)$. We will show that $g(8) > 0$ and that $g(x)$ is increasing for $x \geq 8$ and conclude that $g(x) > 0$ for all $x \geq 8$.

First observe

$$g(8) = 8 + 1 - \log_2\left(\frac{8^4}{9}\right) = 9 - (4\log_2(8) - \log_2(9)) > 9 - (12 - 3) = 0.$$

Now compute $g'(x)$.

$$\begin{aligned} g'(x) &= 1 - \frac{1}{\frac{x^4}{9}\ln(2)}\left(\frac{4x^3}{9}\right) \\ &= 1 - \frac{9}{x^4\ln(2)}\left(\frac{4x^3}{9}\right) \\ &= 1 - \frac{4x^3}{x^4\ln(2)} \\ &= 1 - \frac{4}{x\ln(2)} \\ &= \frac{x\ln(2)-4}{x\ln(2)} \end{aligned}$$

Since $x\ln(2) > 0$ for $x > 1$ and $x\ln(2) - 4 > 0$ for $x > 6$, it follows $g'(x) > 0$ for $x \geq 8 > 6$. Thus, $d + 1 > \log_2\left(\frac{d^4}{9}\right)$ for $d \geq 8$. Thus, $\alpha_1 x_1^d + \cdots + \alpha_n x_n^d = 0$ is isotropic for $d \geq 8$.

Therefore, $\alpha_1 x_1^d + \cdots + \alpha_n x_n^d = 0$ is isotropic. □

Remark: It is clear that Theorem 4.4.7 is a stronger result than Lemma 5.2.12 when $d \geq 8$, though one may find that Lemma 5.2.12 is a simpler result to apply. However, for the values $d = 2, 3, 4, 5, 6, 7$, we will use Table 5.1 to understand how the results compare.

We are now ready to prove the main result. For convenience, we restate Theorem 5.2.1.

Theorem 5.2.1 (Tietäväinen, [16], Theorem 2)**.** Let $\mathbb{F}_q$ be the finite field of p^k elements and suppose that $d \mid q-1$. Assume $(d,k) \neq (p-1,1)$. If $n \geq \dfrac{d+3}{2}$, then the equation

$$\alpha_1 x_1^d + \cdots + \alpha_n x_n^d = 0$$

has a non-trivial solution in $\mathbb{F}_q$.

Proof. We may assume that $\alpha_1, \ldots, \alpha_n$ are non-zero, for if $\alpha_j = 0$, then the equation has a non-trivial solution, namely $x_j = 1$, $x_i = 0$ for $i \neq j$. The case $d = 1$ is trivial. We shall now assume $d \geq 2$.

Table 5.1: In this table, we have computed the different bounds given by Lemma 5.2.12 and Theorem 4.4.7 for $d = 2, 3, 4, 5, 6, 7$. The results in blue (i.e. $d = 3, 5, 7$) indicate the values of d where Lemma 5.2.12 is a stronger result. The results in black (i.e. $d = 2, 4, 6$) indicate the values of d where Lemma 5.2.12 and Theorem 4.4.7 provide the same bound.

	$\frac{d+3}{2}$	$1 + \lceil 2\log_2(d) - \log_2(\log_2(d)) \rceil$
$d = 2$	$\frac{5}{2}$	3
$d = 3$	3	4
$d = 4$	$\frac{7}{2}$	4
$d = 5$	4	5
$d = 6$	$\frac{9}{2}$	5
$d = 7$	5	6

Case I: Suppose that all the assumptions of Theorem 5.2.11 are satisfied. Then by Theorem 5.2.11,

$$\alpha_1 x_1^d + \cdots + \alpha_{n-1} x_{n-1}^d$$

represents either all the elements of $\mathbb{F}_q$ or represents at least

$$\frac{(2(n-1)-1)(q-1)}{d} + 1 = \frac{(2n-3)(q-1)}{d} + 1$$

elements of $\mathbb{F}_q$. Since $n \geq \dfrac{d+3}{2}$, we know that

$$\frac{(2n-3)(q-1)}{d} + 1 \geq q.$$

Therefore, $\alpha_1 x_1^d + \cdots + \alpha_{n-1} x_{n-1}^d$ represents all the elements of $\mathbb{F}_q$. In particular, there exist elements $\xi_1, \ldots, \xi_{n-1}$ in $\mathbb{F}_q$ such that

$$\alpha_1 x_1^d + \cdots + \alpha_{n-1} x_{n-1}^d = -\alpha_n.$$

Hence the equation has the non-trivial solution $(\xi_1, \ldots, \xi_{n-1}, 1)$.

Case II: Suppose that there exists an integer v such that $1 \leq v < k$ and $\dfrac{q-1}{d} \mid p^v - 1$. Now $k \geq 2$ and thus $q \geq 4$. Additionally, we will next show that we may assume that v is a divisor of k. First note that $\gcd(p^k - 1, p^v - 1) = p^{\gcd(k,v)} - 1$. Since $\frac{p^k-1}{d} \mid p^k - 1$ and $\frac{p^k-1}{d} \mid p^v - 1$, we know $\frac{p^k-1}{d} \mid p^{\gcd(k,v)-1}$. Thus, one such integer that satisfies the properties of v is $\gcd(k, v)$ and so we may choose $v = \gcd(k, v)$. With this choice of v, we have $v \mid k$. Thus, since $k > v$ and $v \mid k$, $k \geq 2v$.

Suppose first that $q = p^k = 4$. Since $n \geq 3$, the equation has, by Lemma 5.2.10, a non-trivial solution in $\mathbb{F}_q$.

Suppose now that $q = p^k > 4$. If $p^v = 2$, then $p = 2$, $v = 1$, and $k \geq 3$. Thus

$$\frac{q-1}{p^v-1} = 2^k - 1 > 2k = \frac{2k}{v}.$$

If $p^v \geq 3$, then

$$\frac{q-1}{p^v-1} = 1 + p^v + \cdots + p^{k-v} \geq 1 + 3\Big(\frac{k}{v} - 1\Big) \geq \frac{2k}{v}$$

since $k \geq 2v$, which implies $\dfrac{k}{v} \geq 2$. Thus, in every case,

$$v(q-1) \geq 2k(p^v - 1)$$

and consequently by the assumption that $n \geq \dfrac{d+3}{2}$ we find that

$$n \geq \frac{d+3}{2} \geq \frac{d}{\frac{v(q-1)}{k(p^v-1)}} + \frac{3}{2} = \frac{kd(p^v-1)}{v(q-1)} + \frac{3}{2}$$

By Lemma 5.2.10, this inequality implies that the equation has a non-trivial solution in $\mathbb{F}_q$.

Case III: Suppose $p = 2$. Since $p = 2$, $\mathrm{char}(\mathbb{F}_q) = 2$ and thus $\alpha_1 x_1^d + \cdots + \alpha_n x_n^d$ is an A-equation. By Lemma 5.2.12, we have the desired result.

Case IV: Suppose $d \geq \dfrac{q-1}{2}$. First suppose $d = \dfrac{q-1}{2}$. Let $\mathbb{F}_q^* =< \beta >$. Then $\eta = \beta^d = -1$. Then -1 is a dth power in $\mathbb{F}_q$ and thus, $\alpha_1 x_1^d + \cdots + \alpha_n x_n^d = 0$ is an A-equation. By Lemma 5.2.12, we have the desired result.

Now suppose $d > \dfrac{q-1}{2}$. Since $d \mid q-1$, it follows that $d = q - 1 = p^k - 1$. Since the case $d = p-1$, $k = 1$ is excluded, it follows that $k > 1$. Since $\dfrac{q-1}{d} = 1$, it follows that $\frac{q-1}{d} \mid p - 1$. Since $k \geq 2$ and $\dfrac{q-1}{d} \mid p-1$, this case is a special case of **Case II** and thus, has been proved.

Therefore, we have proved the desired result. $\square$

Two remarks about the assumptions of Theorem 5.2.11. The assumption that $\dfrac{q-1}{d} \nmid p^v - 1$ for $1 \leq v < k$ is a necessary assumption of Theorem 5.2.11. Assume that there exists a $v \in \mathbb{Z}$ such that $1 \leq v \leq k$ and $\frac{q-1}{d} \mid p^v - 1$. If Theorem 5.2.11 holds, then $x_1^d + \cdots + x_n^d$ represents either q elements or at least $\dfrac{(2n-1)(q-1)}{d}$ elements of $\mathbb{F}_q$. By [15], we may assume $v \mid k$. Since $\dfrac{q-1}{d} \mid p^v - 1$, it follows that $q - 1 \mid d(p^v - 1)$. Thus, elements of $\mathbb{F}_q^d$ are solutions to the equation $x^{p^v} = x$. Thus

$\mathbb{F}_q^d \subseteq \mathbb{F}_{p^v}$. Thus $x_1^d + \cdots + x_n^d$ only represents elements of $\mathbb{F}_{p^v}$. This means that $x_1^d + \cdots + x_n^d$ cannot represent all of $\mathbb{F}_q$. Furthermore, choosing $n > \dfrac{(p^v-1)d}{2(q-1)} + \dfrac{1}{2}$ means $x_1^d + \cdots + x_n^d$ represents at least $\dfrac{(2n-1)(q-1)}{d} + 1 > p^v$. Thus, Theorem 5.2.11 does not hold.

The assumption $d < \dfrac{q-1}{2}$ is also necessary for the assumptions of Theorem 5.2.11. Consider $x_1^d + x_2^d = 0$ over $\mathbb{F}_p$ for $p > 5$. Let $d = \frac{p-1}{2}$. If Theorem 5.2.11 holds, $x_1^d + x_2^d$ represents either p elements or at least 7 elements. However, since $\mathbb{F}_p^d = \{0, 1, -1\}$, it follows that $x_1^d + x_2^d$ only represents the following five elements $\{0, 1, -1, 2, -2\}$. Thus, Theorem 5.2.11 does not hold.

5.3 Systems of Diagonal Forms

In the previous two sections, we considered results for the existence of nontrivial solutions to a single diagonal form. A natural generalization is to consider results for a system of diagonal forms. There are a few classical results, which will be presented in this section.

The following result is due to André Weil [18].

Theorem 5.3.1 ([10], Chapter 8, page 103)**.** *Let $\mathbb{F}_q$ denote the finite field of cardinality q. Consider a diagonal form $f = \alpha_1 x_1^d + \cdots + \alpha_n x_n^d$ where $q \equiv 1 \bmod d$ and $\alpha_i \in \mathbb{F}_q^*$ for $1 \leq i \leq n$. Let $N_{\mathbb{F}_q}(f)$ denote the number of affine zeros of f in $\mathbb{F}_q$. Then*

$$|N_{\mathbb{F}_q}(f) - q^{n-1}| \leq M(n,d)(q-1)q^{\frac{n}{2}-1},$$

where $M(n,d) := \dfrac{(d-1)^n + (-1)^n(d-1)}{d}$. In particular, if $N_{\mathbb{F}_q}(f) \geq 2$, then f has a nontrivial zero in $\mathbb{F}_q$.

Theorem 5.3.1 can be generalized to hold for systems of diagonal forms all of degree d. We will need the following lemma to prove this generalization.

Lemma 5.3.2. *Let $f_1, \ldots, f_r$ be forms of degree d over $\mathbb{F}_q$ in n variables that are linearly independent. Then*

$$q^n + \sum_{(c_1,\ldots,c_r)\neq\vec{0}} N\Big(\sum_{i=1}^r c_i f_i; \mathbb{F}_q\Big) = q^r N(\boldsymbol{f}; \mathbb{F}_q) + q^{r-1}(q^n - N(\boldsymbol{f}; \mathbb{F}_q)).$$

Proof. Let $f_1, \ldots, f_r$ be forms of degree d over $\mathbb{F}_q$ in n variables that are linearly independent. Let $(a_1, \ldots, a_n) \in \mathbb{F}_q^n$. We will obtain the desired result using the following function $\displaystyle\sum_{i=1}^r c_i f_i$ where $c_i \in \mathbb{F}_q$. Consider the linear map

$$\varphi_{(a_1,\ldots,a_n)} : \mathbb{F}_q^r \to \mathbb{F}_q$$

where

$$(c_1, \dots, c_r) \mapsto \sum_{i=1}^{r} c_i f_i(a_1, \dots, a_n).$$

Notice if $(a_1, \dots, a_n)$ is a common zero of the f_i for $1 \le i \le r$, then $\ker(\varphi_{(a_1,\dots,a_n)}) = \mathbb{F}_q^r$ and these $(a_1, \dots, a_n)$ are counted by $N(\boldsymbol{f}; \mathbb{F}_q)$. Consider the case where $(a_1, \dots, a_n)$ is not a common zero, then $(a_1, \dots, a_n) \neq (0, \dots, 0)$. Without loss of generality, assume $f_1(a_1, \dots, a_n) \neq 0$. Then choose $c_2 = \cdots = c_r = 0$ and let c_1 run over all of $\mathbb{F}_q$. This means $\mathrm{im}(\varphi_{(a_1,\dots,a_n)}) = \mathbb{F}_q$, so $\dim(\mathrm{im}(\varphi_{(a_1,\dots,a_n)})) = 1$. By the First Isomorphism Theorem, we know $\dim(\ker(\varphi_{(a_1,\dots,a_n)})) = r - 1$. And those $(a_1, \dots, a_n)$ are counted by $q^n - N(\boldsymbol{f}; \mathbb{F}_q)$.

We will demonstrate the above equality by counting the total number of times $(a_1, \dots, a_n)$ appears on the left hand side. First notice that the q^n term on the left hand side of the equation corresponds to when $(c_1, \dots, c_r) = (0, \dots, 0)$. If $(a_1, \dots, a_n)$ is a common zero of all the f_i, then the left hand side counts $(a_1, \dots, a_n)$, q^r times. Since there are $N(\boldsymbol{f}; \mathbb{F}_q)$ such points, the common zeros are counted $q^r N(\boldsymbol{f}; \mathbb{F}_q)$ times. If $(a_1, \dots, a_n)$ is not a common zero of all the f_i, then the left hand side counts $(a_1, \dots, a_n)$, q^{r-1} times. Since there are $q^n - N(\boldsymbol{f}; \mathbb{F}_q)$ such points, the not common zeros are counted $q^{r-1}(q^n - N(\boldsymbol{f}; \mathbb{F}_q))$ times. Thus we have that

$$q^n + \sum_{(c_1,\dots,c_r)\neq\vec{0}} N(\sum_{i=1}^{r} c_i f_i; \mathbb{F}_q) = q^r N(\boldsymbol{f}; \mathbb{F}_q) + q^{r-1}(q^n - N(\boldsymbol{f}; \mathbb{F}_q))$$

□

Theorem 5.3.3. *Let $f_1, \dots, f_r$ be diagonal forms of degree d over $\mathbb{F}_q$ in n variables that are linearly independent where $q \equiv 1 \bmod d$. Then*

$$|N(\boldsymbol{f}; \mathbb{F}_q) - q^{n-r}| \le \frac{M(n,d) q^{\frac{n}{2}-1}(q^r - 1)}{q^{r-1}},$$

where $M(n,d) = \dfrac{(d-1)^n + (-1)^n(d-1)}{d}$.

Proof. By Lemma 5.3.2,

$$q^n + \sum_{(c_1,\dots,c_r)\neq\vec{0}} N(\sum_{i=1}^{r} c_i f_i; \mathbb{F}_q) = q^r N(\boldsymbol{f}; \mathbb{F}_q) + q^{r-1}(q^n - N(\boldsymbol{f}; \mathbb{F}_q)).$$

Simplifying the above expression yields

$$\sum_{(c_1,\dots,c_r)\neq\vec{0}} N(\sum_{i=1}^{r} c_i f_i; \mathbb{F}_q) = (q^r - q^{r-1}) N(\boldsymbol{f}; \mathbb{F}_q) + q^{n+r-1} - q^n.$$

Using Weil's result (Theorem 5.3.1, [18]), we know

$$|N(\sum_{i=1}^{r} c_i f_i; \mathbb{F}_q) - q^{n-1}| \le M(n,d) q^{\frac{n}{2}-1}(q-1).$$

Applying Weil's result (Theorem 5.3.1, [18]) $q^r - 1$ times, we find

$$|\sum_{(c_1,\ldots,c_r)\neq\vec{0}} N(\sum_{i=1}^{r} c_i f_i; \mathbb{F}_q) - (q^r-1)q^{n-1}| \;\; \leq M(n,d)q^{\frac{n}{2}-1}(q-1)(q^r-1),$$

$$|(q^r - q^{r-1})N(\boldsymbol{f}; \mathbb{F}_q) + q^{n+r-1} - q^n - (q^r-1)q^{n-1}| \;\; \leq M(n,d)q^{\frac{n}{2}-1}(q-1)(q^r-1),$$

$$|(q^r - q^{r-1})N(\boldsymbol{f}; \mathbb{F}_q) - (q^n - q^{n-1})| \;\; \leq M(n,d)q^{\frac{n}{2}-1}(q-1)(q^r-1),$$

$$|q^{r-1}N(\boldsymbol{f} : \mathbb{F}_q) - q^{n-1}| \;\; \leq M(n,d)q^{\frac{n}{2}-1}(q^r-1),$$

$$|N(\boldsymbol{f}; \mathbb{F}_q) - q^{n-r}| \;\; \leq M(n,d)q^{\frac{n}{2}-1}\frac{q^r-1}{q^{r-1}}.$$

□

The second result that has improved existing bounds on the number of variables needed to guarantee the existence of a nontrivial solution is due to Tietäväinen in 1965.

Theorem 5.3.4 (Tietäväinen, [14], Theorem 5, page 21)**.** *Assume that $k = mu$ where m and u are positive integers. Then the system $\sum_{j=1}^{n} \alpha_{ij} x_j^d = 0$ for $i = 1, \ldots, r$, $\alpha_{ij} \in \mathbb{F}_q$ has a non-trivial solution in $\mathbb{F}_q = \mathbb{F}_{p^k}$ if $n \geq 1 + b^{-1}dru$ where $b = (d,h)$ and $h = \dfrac{(p^k-1)}{(p^m-1)}$.*

Proof. Let $\mathbb{F}_q^* = \langle \rho \rangle$, that is, ρ is an element of order $p^k - 1$. Then $\mathbb{F}_{p^m} = \langle \rho^h \rangle$ since ρ^h has order $p^m - 1$. Since $b \mid h$, we know $\frac{h}{b} \in \mathbb{Z}$. Thus, $(\mathbb{F}_{p^m}^*)^{\frac{d}{b}} = \langle \rho^{\frac{hd}{b}} \rangle \subseteq (\mathbb{F}_q^*)^d$. Thus, $(\mathbb{F}_{p^m})^{\frac{d}{b}} \subseteq (\mathbb{F}_q^d)$.

We can write $\mathbb{F}_q$ as $\mathbb{F}_{p^m}[\theta]$ for some $\theta \in \mathbb{F}_q$. We can write α_{ij} as follows

$$\alpha_{ij} = \sum_{s=0}^{u-1} \alpha_{ijs}\theta^{s-1}, \text{ where } \alpha_{ijs} \in \mathbb{F}_{p^m}.$$

Consider the system $\sum_{j=1}^{n} \alpha_{ijs} y_j^{b^{-1}d} = 0$ for $i = 1, \ldots, r$; $s = 0, \ldots, u-1$. This is a system of ru diagonal equations each of degree $b^{-1}d$ with coefficients in $\mathbb{F}_{p^m}$. It is sufficient to find a solution to this system where each y_j lies in $\mathbb{F}_{p^m}$. By Chevalley's result (Theorem 2.1.1), if $n \geq 1 + b^{-1}dru$, then there exists a non-trivial solution $(\eta_1, \ldots, \eta_n)$

with $\eta_j \in \mathbb{F}_{p^m}$ for $j = 1, \ldots, n$. Since $(\mathbb{F}_{p^m})^{\frac{d}{b}} \subseteq (\mathbb{F}_q^d)$, there exists $\xi_1, \ldots, \xi_n \in \mathbb{F}_q$ not all zero such that $\xi_j^d = \eta_j^{b^{-1}d}$. Thus, $(\xi_1, \ldots, \xi_n)$ is a non-trivial solution of $\sum_{j=1}^{n} \alpha_{ij} x_j^d$ for $i = 1, \ldots, r$. $\square$

Notice that when we are over a prime field Theorem 5.3.4 gives us precisely the bound given by Chevalley (Theorem 2.1.1, [3]). However, when we are over $\mathbb{F}_q$, where $\mathbb{F}_q$ is not a prime field, we obtain some improvements. Consider the system $\sum_{j=1}^{n} \alpha_{ij} x_j^d$ for $i = 1, \ldots, r$ in $\mathbb{F}_{64} = \mathbb{F}_{2^6}$.

$$n \geq \begin{cases} 1 + r & \text{for } d = 1, m = 6 \\ 1 + 2r & \text{for } d = 3 \text{ or } d = 9, m = 3 \\ 1 + 3r & \text{for } d = 7 \text{ or } d = 21, m = 2 \\ 1 + 6r & \text{for } d = 63, m = 1 \end{cases}$$

5.4 Particular Values of $\mathbf{\Omega(r, d, q)}$

A classical question that all of the results in Chapter 5 seek to answer is what is the minimal number of variables needed to guarantee a nontrivial solution. In this section, we want to collect all of the known values of $\Omega(r, d, q)$. Let $f_1, \ldots, f_r$ be homogeneous diagonal forms all of degree d over $\mathbb{F}_q[x_1, \ldots, x_n]$.

Theorem 5.4.1 (Leep–Petrik)**.** *Let $d' = \gcd(d, q-1)$. Then $\Omega(r, d, q) = \Omega(r, d', q)$.*

Proof. We know $\Omega(r, d', q) \leq \Omega(r, d, q)$ since every d^{th} power is a d'^{th} power.

We will now show $\Omega(r, d, q) \leq \Omega(r, d', q)$. Let $n = \Omega(r, d', q)$. Consider $\sum_{j=1}^{n+1} a_{ij} x_j^d = 0$ for $1 \leq i \leq r$. Let $A = (a_{ij})$ be the $r \times (n+1)$ matrix given by coefficients from our system. Consider the following matrix equation

$$A \begin{bmatrix} x_1^d \\ x_2^d \\ \vdots \\ x_{n+1}^d \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

We know that there exists a nonzero $\boldsymbol{c} = \begin{bmatrix} c_1^{d'} \\ c_2^{d'} \\ \vdots \\ c_{n+1}^{d'} \end{bmatrix}$ such that

$$A \begin{bmatrix} c_1^{d'} \\ c_2^{d'} \\ \vdots \\ c_{n+1}^{d'} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

since $\Omega(r, d', q) = n$. Since $d' = \gcd(d, q-1)$, we know that $c_i^{d'} = m_i^d$ for all i where $m_i \in \mathbb{F}_q$. Since $\boldsymbol{c}$ is a nonzero vector, we know that $\boldsymbol{m} = \begin{bmatrix} m_1^d \\ m_2^d \\ \vdots \\ m_{n+1}^d \end{bmatrix}$ is nonzero vector.

Thus, $\begin{bmatrix} m_1^d \\ m_2^d \\ \vdots \\ m_{n+1}^d \end{bmatrix}$ is a nontrivial solution to our system, so $\Omega(r, d, q) \leq \Omega(r, d', q)$.

Therefore, $\Omega(r, d, q) = \Omega(r, d', q)$. □

By Theorem 5.4.1, if $f_1, \ldots, f_r$ are homogeneous diagonal forms all of degree d over $\mathbb{F}_q[x_1, \ldots, x_n]$, we will assume, without loss of generality, that $d \mid q-1$.

Theorem 5.4.2.

$$\Omega(r_1, \vec{d_1}, q) + \Omega(r_2, \vec{d_2}, q) \leq \Omega(r_1 + r_2, (\vec{d_1}, \vec{d_2}), q) \leq \sum_{i=1}^{r_1} d_{1i} + \sum_{i=1}^{r_2} d_{2i}$$

where $(\vec{d_1}, \vec{d_2}) = (d_{11}, d_{12}, \ldots, d_{1r_1}, d_{21}, d_{22}, \ldots, d_{2r_2})$.

Proof. We know that there exists an anisotropic system of r_1 forms in $\Omega(r_1, \vec{d_1}, q)$ variables over $\mathbb{F}_q$ and that there exists an anisotropic system of r_2 forms in $\Omega(r_2, \vec{d_2}, q)$ variables over $\mathbb{F}_q$. If we choose the variables to be disjoint, then we have an anisotropic system with $r_1 + r_2$ forms in $\Omega(r_1, \vec{d_1}, q) + \Omega(r_2, \vec{d_2}, q)$ variables. Thus $\Omega(r_1, \vec{d_1}, q) + \Omega(r_2, \vec{d_2}, q) \leq \Omega(r_1 + r_2, (\vec{d_1}, \vec{d_2}), q)$.

By Chevalley's result (Theorem 2.1.1), we know that

$$\Omega(r_1 + r_2, (\vec{d_1}, \vec{d_2}), q) \leq \sum_{i=1}^{r_1} d_{1i} + \sum_{i=1}^{r_2} d_{2i}$$

□

Theorem 5.4.3. *If* $d = 1$, *then* $\Omega(r, 1, q) = r$.

The proof of Theorem 5.4.3 uses standard results from linear algebra.

Theorem 5.4.4. *If* $d = 2$, *then* $\Omega(r, 2, q) = 2r$.

Proof. By Chevalley's result (Theorem 2.1.1), we know $\Omega(r, 2, q) \leq 2r$. Since $d \mid q-1$ and $d = 2$, it follows that q is odd. Since q is odd, we know $|\mathbb{F}_q^*/(\mathbb{F}_q^*)^2| = 2$. Let u be a quadratic non-residue. The form $x^2 - uy^2 = 0$ is anisotropic over $\mathbb{F}_q$. We have now shown that $\Omega(1, 2, q) \geq 2$. Thus by Theorem 5.4.2, $\Omega(r, 2, q) \geq 2r$. Therefore, $\Omega(r, 2, q) = 2r$. □

Theorem 5.4.5. *If* $d = p - 1$ *and* $q = p$, *then* $\Omega(r, p-1, p) = r(p-1)$.

Proof. We know that $\Omega(r, p-1, p) \leq r(p-1)$ by Chevalley's result (Theorem 2.1.1). Now we need to show that there is an anisotropic system of r forms with degree $p-1$ over $\mathbb{F}_p$. Consider the following form

$$g = x_1^{p-1} + x_2^{p-1} + \cdots + x_{p-1}^{p-1}.$$

We know g is anisotropic over $\mathbb{F}_p$. We have now shown that $\Omega(1, p-1, q) \geq p-1$. Thus by Theorem 5.4.2, $\Omega(r, p-1, p) \geq r(p-1)$. Therefore, $\Omega(r, p-1, p) = r(p-1)$. $\square$

Theorem 5.4.6.

(i) If $d = 3$ *and* $q = 4$*, then* $\Omega(r, 3, 4) = 2r$.

(ii) If $d = 2^k - 1$ *and* $q = 2^k$*, then* $\Omega(r, 2^k - 1, 2^k) = kr$.

(iii) If $d = \dfrac{p^k - 1}{p - 1}$ *and* $q = p^k$*, then* $\Omega\Big(r, \dfrac{p^k-1}{p-1}, p^k\Big) = kr$.

Proof. (iii) First notice that parts (i) and (ii) of this Theorem are simply special cases of part (iii). Thus, it is sufficient to provide only the proof for part (iii).

We know that there exists a k-dimensional vector space basis of $\mathbb{F}_{p^k}/\mathbb{F}_p$ with elements $\omega_1, \ldots, \omega_k$. This implies that $\omega_1 x_1^d + \omega_2 x_2^d + \cdots + \omega_k x_k^d = 0$ is anisotropic over $\mathbb{F}_{p^k}$. If we repeat this form r times in disjoint variables, we have an anisotropic system of r forms in kr variables. Thus $\Omega\Big(r, \dfrac{p^k-1}{p-1}, p^k\Big) \geq kr$.

Let $f_1 = (a_{1,1}\omega_1 + \cdots + a_{1,k}\omega_k)x_1^d + \cdots + (a_{kr+1,1}\omega_1 + \cdots + a_{kr+1,k}\omega_k)x_{kr+1}^d$. Notice $f_1 = 0$ if and only if $\displaystyle\sum_{i=1}^{kr+1} a_{i,1}x_i^d = 0, \ldots, \sum_{i=1}^{kr+1} a_{i,kr+1}x_i^d = 0$.

Since $N_{\mathbb{F}_{p^k}^*/\mathbb{F}_p^*}$, the norm map, is surjective when $x \mapsto x x^p \ldots x^{p^{k-1}} = x^{\frac{p^k-1}{p-1}}$, we can reduce this to $\displaystyle\sum_{i=1}^{kr+1} a_{i,1}y_i = 0, \ldots, \sum_{i=1}^{kr+1} a_{i,kr+1}y_i = 0$ where $y_i \in \mathbb{F}_p$ and $a_{i,j} \in \mathbb{F}_p$. Since we have kr linear equations over $\mathbb{F}_p$, linear algebra tells us that $kr+1$ variables guarantees a nontrivial solution. Thus $\Omega\Big(r, \dfrac{p^k-1}{p-1}, p^k\Big) < kr+1$. Therefore, $\Omega\Big(r, \dfrac{p^k-1}{p-1}, p^k\Big) = kr$. $\square$

Remark: Notice that Theorem 5.4.6, part (iii) is a special case of Theorem 5.3.4 with $m = 1$, $u = k$, $h = \frac{p^k-1}{p-1} = d$, and $b = \gcd(d, h) = d$.

Theorem 5.4.7. *Assume* $d = 3$ *and* $q = 7$.

(i) $2r \leq \Omega(r, 3, 7) < \frac{r\ln(7)}{\ln(2)} < 2.81r$.

(ii) If $r = 1$*, then* $\Omega(3, 7) = 2$.

(iii) If $r = 2$*, then* $\Omega(2, 3, 7) = 5$.

(iv) If $r = 3$, then $7 \le \Omega(3,3,7) \le 8$.

Proof. (i) By Lemma 4.3.7, $\Omega(r,3,7) < \frac{r\ln(7)}{\ln(2)}$, which implies that $\Omega(r,3,7) < \frac{r\ln(7)}{\ln(2)} < 2.81r$. Notice that $(\mathbb{F}_7)^3 = \{0,1,-1\}$. Consider the form $x^3 - 2y^3 = 0$. This is an anisotropic form in 2 variables over $\mathbb{F}_7$. Thus, $\Omega(1,3,7) \ge 2$. By Theorem 5.4.2, $\Omega(r,3,7) \ge 2r$.
(ii) By (i), we know that $2 \le \Omega(3,7) < 2.81$. Thus, $\Omega(3,7) = 2$.
(iii) By (i), we know that $4 \le \Omega(2,3,7) \le 6$. We now show $\Omega(2,3,7) = 5$. Furthermore, if we consider the following system of equations, one can check that it is anisotropic over $\mathbb{F}_7$.

$$\begin{aligned} f_1 &= x_1^3 \qquad + x_3^3 + 2x_4^3 + 3x_5^3 \\ f_2 &= \qquad x_2^3 + 3x_3^3 + 5x_4^3 + 3x_5^3 \end{aligned}$$

Thus, we know that $\Omega(2,3,7) \ge 5$. By (i), we know $\Omega(2,3,7) < \frac{r\ln(7)}{\ln(2)} < 5.7$. Thus $\Omega(2,3,7) = 5$.
(iv) By (i), we know that $6 \le \Omega(3,3,7) < 2.81(3) < 9$. Thus $6 \le \Omega(3,3,7) \le 8$. Since $\Omega(1,3,7) = 2$ and $\Omega(2,3,7) = 5$, by Theorem 5.4.2, we know $\Omega(3,3,7) \ge 7$. $\square$

One problem of interest is to compute the value of $\Omega(3,3,7)$. While it may seem like this problem should be a simple computation, it is in fact much more complex. Even with clear reductions and simplifications, solving the problem by brute-force is not possible with our computational resources. Consequently, it is necessary to prove theoretical results which exclude large classes of equations so as to make the computation feasible. Interestingly, this problem has connections to the plus-minus Davenport Constant of finite groups. In particular, $\Omega(3,3,7) + 1 = D_{\pm}(C_7^3)$.

Theorem 5.4.8. *Assume $d = 3$ and $q = 7$.*

(i) If r is even, then $2.5r \le \Omega(r,3,7) < 2.81r$.

(ii) If r is odd, then $2.5r - \frac{1}{2} \le \Omega(r,3,7) < 2.81r$.

Proof. (i) Let $r = 2\ell$ for some $\ell \in \mathbb{Z}_{\ge 0}$. Since we now that $\Omega(2,3,7) = 5$, by Theorem 5.4.2, we have $\Omega(r,3,7) = \Omega(2\ell,3,7) \ge 5\ell = \frac{5}{2}r = 2.5r$. By Theorem 5.4.7, part (i), we know that $\Omega(r,3,7) < 2.81r$.
(ii) Let $r = 2\ell+1$ for some $\ell \in \mathbb{Z}_{\ge 0}$ Since we know $\Omega(1,3,7) = 2$ and $\Omega(2\ell,3,7) \ge 5\ell$, by Theorem 5.4.2, we know $\Omega(r,3,7) = \Omega(2\ell+1,3,7) \ge 5\ell + 2 = \frac{5}{2}r - \frac{1}{2} = 2.5r - \frac{1}{2}$. By Theorem 5.4.7, we know $\Omega(r,3,7) < 2.81r$. $\square$

5.5 Particular Values of $\max\limits_q \Omega(d,q)$ and $\max\limits_q \Omega_A(d,q)$

Another question of interest is given a single equation, what is the number of variables needed to guarantee the existence of a nontrivial solution over any finite field, $\mathbb{F}_q$. In particular, what is $\max\limits_q \Omega(d,q)$ for various d. The following table summarizes the known results.

Table 5.2: The results for $d \neq 13$ can be found in [14] in Section 21, pg 32-34. For completeness, the proofs have been included in Appendix A.

	$\max\limits_{q} \Omega(d,q)$	$\max\limits_{p \text{ prime}} \Omega(d,p)$	$\max\limits_{q} \Omega_A(d,q)$	$\max\limits_{p \text{ prime}} \Omega_A(d,p)$
$d=3$	2	2	2	2
$d=5$	3	3	3	3
$d=7$	3	3	3	3
$d=9$	4	4	4	4
$d=11$	4	4	4	4
$d=13$	4	4	4	4
$d=2$	2	2	2	2
$d=4$	≤ 4	≤ 4	2	2
$d=6$	≤ 6	≤ 6	3	3
$d=8$	≤ 8	≤ 8	4	4

While Tietäväinen computed the values for $d = 3, 5, 7, 9, 11$, we were able to determine $\max\limits_{q} \Omega(13, q)$.

Proposition 5.5.1 (Leep–Petrik)**.** *Let $d = 13$. Then $\sum_{j=1}^{n} a_j x_j^{13} = 0$ has a nontrivial solution over every $\mathbb{F}_q$ if $n \geq 5$. In other words,*

$$\max_{q} \Omega(13, q) = \max_{p \text{ prime}} \Omega(13, p) = 4.$$

Proof. First we will show that $\sum_{j=1}^{5} a_j x_j^{13} = 0$ has a nontrivial solution over $\mathbb{F}_q$ for all q. Then we will find a form $\sum_{j=1}^{4} a_j x_j^{13} = 0$ that is anisotropic over $\mathbb{F}_q$ for some q.

By Lemma 4.3.7, $\sum_{j=1}^{5} a_j x_j^{13} = 0$ has a nontrivial solution if $q < 2^5 = 32$. By Theorem 4.4.4, $\sum_{j=1}^{5} a_j x_j^{13} = 0$ has a nontrivial solution if $q \geq \frac{1}{5}(13)(12)^{\frac{5}{3}}$, that is, when $q \geq 164$. Since $13 \mid q-1$, the remaining fields to check are $\mathbb{F}_{53}$, $\mathbb{F}_{79}$, $\mathbb{F}_{131}$, and $\mathbb{F}_{157}$. Since $79 \equiv 1 \mod 39$ and $157 = 1 \bmod 39$ and $79 < 3^5$ and $157 < 3^5$, by Lemma 4.5.5, $\sum_{j=1}^{5} a_j x_j^{13} = 0$ has a nontrivial solution over $\mathbb{F}_{79}$ and $\mathbb{F}_{157}$. It remains to check $\mathbb{F}_{53}$ and $\mathbb{F}_{131}$. Using a computer code found in Appendix B, $\sum_{j=1}^{5} a_j x_j^{13} = 0$ has a nontrivial solution over $\mathbb{F}_{53}$ and $\mathbb{F}_{131}$.

Using a computer code found in Appendix B (Example B.0.1),

$$x_1^{13} + 4x_2^{13} + 64x_3^{13} + 125x_4^{13} = 0$$

is anisotropic over $\mathbb{F}_{131}$. Thus,

$$\max_q \Omega(13, q) = \max_{p \text{ prime}} \Omega(13, p) = 4.$$

□

One more question of interest is when are the bounds for A-equations sharp. We may use our chart from before to help analyze this. By Lemma 5.2.12, it follows that $\Omega_A(d, q) \leq \lceil \frac{d+1}{2} \rceil$.

Table 5.3: Exploring sharpness of results on A-equations. In Column 3, we have the bound given by Theorem 4.4.7 and in Column 4, we have the bound given by Lemma 5.2.12. You can see that the only time Theorem 4.4.7 is sharp is when $d = 2$. However, Lemma 5.2.12 is sharp for $d = 2, 3, 5$.

	$\max_q \Omega_A(d, q)$	$\max_{p \text{ prime}} \Omega_A(d, p)$	$\lceil 2\log_2(d) - \log_2(\log_2(d)) \rceil$	$\lceil \frac{d+1}{2} \rceil$
$d = 3$	2	2	3	2
$d = 5$	3	3	4	3
$d = 7$	3	3	5	4
$d = 9$	4	4	5	5
$d = 11$	4	4	6	6
$d = 13$	4	4	6	7
$d = 2$	2	2	2	2
$d = 4$	2	2	3	3
$d = 6$	3	3	4	4
$d = 8$	4	4	5	5

Appendix A Proofs for $\max_{\mathbf{q}} \boldsymbol{\Omega}(\mathbf{d}, \mathbf{q})$ and $\max_{\mathbf{q}} \boldsymbol{\Omega}_{\mathbf{A}}(\mathbf{d}, \mathbf{q})$

In this appendix, we will present the proofs for determining the values of $\max_q \Omega(d,q)$ for $d = 3, 5, 7, 9, 11$ and the proofs for determining the values of $\max_q \Omega_A(d,q)$ for $d = 2, 4, 6, 8$. We will need a few extra results before we can proceed.

Lemma A.0.1 (Tietäväinen, [14], Lemma 9)**.** *If $2d+1$ is a prime, then*

$$\max_{p \text{ } prime} \Omega(1,d,p) = \max_{p \text{ } prime} \Omega_A(1,d,p) \geq \lceil \log_2(d+1) \rceil.$$

Proof. Let $q = 2d+1$, which means $(\mathbb{F}_q)^d = (\mathbb{F}_q)^{\frac{q-1}{2}} = \{0, 1, -1\}$. We will show that the equation $\sum_{j=1}^{n} 2^{j-1} x_j^d$ has only the trivial solution in $\mathbb{F}_q$ when $2^n - 1 < 2d+1$.

Assume $2^n - 1 < 2d+1$. Suppose $\sum_{j=1}^{n} 2^{j-1} c_j^d = 0$, where $c_j \in \mathbb{F}_q$. Then $c_j^d \in \{0, 1, -1\}$. Let $\epsilon_j \in \{0, 1, -1\}$, $1 \leq j \leq n$. Since $|\sum_{j=1}^{n} \epsilon_j 2^{j-1}| \leq \sum_{j=1}^{n} 2^{j-1} = 2^n - 1 < 2d + 1 = q$, it follows that there would be an equation $\sum_{j=1}^{n} \epsilon_j 2^{j-1} = 0$ in $\mathbb{Z}$.

Next suppose $\sum_{j=1}^{n} 2^{j-1} c_j^d = 0$ as an equation over $\mathbb{Z}$. Let J be the smallest j such that $c_j \neq 0$. Notice $2^{J-1} \mid 2^{J-1} c_J^d$, but $2^J \nmid 2^{J-1} c_J^d$. However, $2^J \mid \sum_{j=J+1}^{n} 2^{j-1} c_j^d$ and $\sum_{j=J+1}^{n} 2^{j-1} c_j^d = -2^{J-1} c_j^d$, which is a contradiction ↯. Thus $c_j = 0$ for all $j = 1, \ldots, n$. Thus, the equation $\sum_{j=1}^{n} 2^{j-1} x_j^d$ has only the trivial solution in $\mathbb{F}_q$.

Rearranging this inequality demonstrates the equation has only the trivial solution when

$$2^n - 1 < 2d + 1$$
$$2^{n-1} < d + 1$$
$$n < \log_2(d+1) + 1.$$

Since $\lceil \log_2(d+1) \rceil < \log_2(d+1) + 1$, we can take $n = \lceil \log_2(d+1) \rceil$ and therefore there exists an anisotropic form in n variables over $\mathbb{F}_{2d+1}$. This means that

$$\max_{p \text{ prime}} \Omega(d,p) = \max_{p \text{ prime}} \Omega_A(d,p) \geq \lceil \log_2(d+1) \rceil.$$

$\square$

We will also need the following observations. Let ρ be a generator of the cyclic group $\mathbb{F}_q^*$. Let I_v be an index set $\{i_1, \ldots, i_v\}$, where $0 \leq i_j \leq d-1$ for all $j = 1, \ldots, v$. Assume $I_v \subseteq I_{v+1}$ and define $Q_v := Q_v(I_v) = \sum_{j \in I_v} \rho^{i_j} \mathbb{F}_q^d$. Clearly, $Q_v \subseteq Q_{v+1}$.

If $\alpha_1, \ldots, \alpha_v \in \mathbb{F}_q^*$, let $R(\alpha_1, \ldots, \alpha_v) = \{\eta \in \mathbb{F}_q \mid \eta = \sum_{j=1}^{v} \alpha_j \xi_j^d, \xi_j \in \mathbb{F}_q\}$. We can see that $R(\alpha_1, \ldots, \alpha_v) = Q_v$ if $\alpha_j \in \rho^{i_j}(\mathbb{F}_q^*)^d$ because the cosets $\alpha_j(\mathbb{F}_q^*)^d = \rho^{i_j}(\mathbb{F}_q^*)^d$ are equal. Furthermore, if $\eta \in Q_v \setminus \{0\}$, so is the set $\eta(\mathbb{F}_q^*)^d$. Hence $Q_v \setminus \{0\}$ is a union of cosets of $\mathbb{F}_q^*$ modulo $(\mathbb{F}_q^*)^d$. Thus, $|Q_v| = 1 + l_v(\frac{q-1}{d})$ for some $l_v \in \mathbb{Z}_{\geq 0}$. Clearly, $l_{v+1} \geq l_v$.

Let $\sum_{j=1}^{n} \alpha_j x_j^d = 0$ be an A-equation. Since $\alpha_j \in \mathbb{F}_q^*$, we may choose i_j's such that $\alpha_j \in \rho^{i_j}(\mathbb{F}_q^*)^d$ for every $j = 1, \ldots, n$. If $l_{n-1} = l_n$, then $\rho^{i_n}(\mathbb{F}_q^*)^d \subseteq Q_{n-1}$. In particular, $\alpha_n \in Q_{n-1}$. Thus, $\alpha_n = \sum_{j=1}^{n-1} \alpha_j \xi_j^d$ for some $\xi_j \in \mathbb{F}_q$ for $j = 1, \ldots, n-1$, which implies that the equation has a nontrivial solution, $(\xi_1, \ldots, \xi_{n-1}, \eta)$, where $\eta^d = -1$. If $l_{n-1} = d$, then $l_{n-1} = l_n$. Therefore, if $l_{n-1} = d$ for every index set I_{n-1} in $\mathbb{F}_q$, then every A-equation has a nontrivial solution in $\mathbb{F}_q$. Since we may divide $\sum_{j=1}^{n} \alpha_j x_j^d$ by α_1, we can always restrict to sequences with $\alpha_1 = 1$, that is, $\alpha_1 \in (\mathbb{F}_q^*)^d$.

Lemma A.0.2 (Tietäväinen, [14], Lemma 10)**.** *If p is a prime and $d < \frac{1}{2}(q-1)$, then*

$$l_{v+1} \geq \min(l_v + 2, d).$$

This lemma has been proved in [5].

Lemma A.0.3 (Tietäväinen, [14], Lemma 11)**.** *If p is a prime, $d < \frac{1}{2}(q-1)$, and*

$$l_2(\{0, i_2\}) \geq d - 2(n-3)$$

whenever i_2 is one of the integers $1, 2, \ldots, \lfloor \frac{d}{n-1} \rfloor$, then the A-equation $\sum_{j=1}^{n} \alpha_j x_j^d = 0$ has a non-trivial solution in $\mathbb{F}_q$.

Proof. As observed above, if $l_{n-1} = d$ for every index set I_{n-1} in $\mathbb{F}_p$, then the A-equation has a nontrivial solution in $\mathbb{F}_p$. Furthermore, by Lemma A.0.2, $l_{n-1} \geq \min(l_{n-2} + 2, d)$. Since $l_2 \geq d - 2(n-3)$, applying Lemma A.0.2 inductively, we know $l_{n-1} \geq d - 2(n-3) + 2(n-3) = d$.

Now we just need to show that we need only check the inequality for $i_2 \in \{1, 2, \ldots, \lfloor \frac{d}{n-1} \rfloor\}$. Observe that we are allowed to perform the following manipulations on the index set:

1. Permute the members of I_v.

2. Addition modulo d of the fixed integer r to every member of I_v (This corresponds to multiplying the A-equation by ρ^r).

Furthermore, we may assume that the i_j's are non-equal and each is not equal to 0. Therefore, the members of I_{n-1} are $n-1$ non-equal elements of the cycle $(0, 1, \ldots, d-1)$. Thus, there exist two elements of I_{n-1} such that the distance between them in the cycle is less than or equal to $\lfloor \frac{d}{n-1} \rfloor$. Consequently, by applying manipulations 1 and 2, we can transform I_{n-1} to a form such that the first term is 0 and the second term is one of the integers $\{1, 2, \ldots, \lfloor \frac{d}{n-1} \rfloor\}$. □

Proposition A.0.4. *Let $d = 3$. Then $\sum_{j=1}^{n} \alpha_j x_j^3 = 0$ has a nontrivial solution over every $\mathbb{F}_q$ if $n \geq 3$. In other words,*

$$\max_q \Omega(3, q) = \max_{p \text{ prime}} \Omega(3, p) = 2.$$

Proof. Since $2d + 1 = 7$ is prime, by Lemma A.0.1, it follows that

$$\max_{p \text{ prime}} \Omega_A(1, 3, q) \geq \lceil \log_2(3 + 1) \rceil = 2.$$

Suppose $n = 3$. By Theorem 4.4.5, $\sum_{j=1}^{n} \alpha_j x_j^3 = 0$ has a nontrivial solution if

$$2^3(3) \geq 3(3-1)^{\frac{3}{1}} = 3(2)^3.$$

Since the inequality is always satisfied when $n = 3$ and $d = 3$, $\sum_{j=1}^{n} \alpha_j x_j^3 = 0$ has a nontrivial solution over $\mathbb{F}_q$ for all q. Thus,

$$\max_q \Omega(3, q) = \max_{p \text{ prime}} \Omega(3, p) = 2.$$

□

Proposition A.0.5. *Let $d = 5$. Then $\sum_{j=1}^{n} \alpha_j x_j^5 = 0$ has a nontrivial solution over every $\mathbb{F}_q$ if $n \geq 4$. In other words,*

$$\max_q \Omega(5, q) = \max_{p \text{ prime}} \Omega(5, p) = 3.$$

Proof. Since $2d + 1 = 11$ is prime, by Lemma A.0.1, it follows that

$$\max_{p \text{ prime}} \Omega_A(3, q) \geq \lceil \log_2(5 + 1) \rceil = 3.$$

Suppose $n = 4$. We will show that for any $q \equiv 1 \bmod 5$, $\sum_{j=1}^{n} \alpha_j x_j^5 = 0$ has a nontrivial solution. By Lemma 4.3.7, $\sum_{j=1}^{n} \alpha_j x_j^5 = 0$ has a nontrivial solution if $q < 2^4 = 16$. By

Lemma 4.5.5, since $16 \equiv 1 \bmod 15$ and $16 < 3^4 = 81$, $\sum_{j=1}^{n} \alpha_j x_j^5 = 0$ has a nontrivial solution. Furthermore, by Theorem 4.4.4, $\sum_{j=1}^{n} \alpha_j x_j^5 = 0$ has a nontrivial solution if

$$q \geq \frac{1}{4}(5)(5-1)^{\frac{4}{2}} = 20.$$

Thus, $\sum_{j=1}^{n} \alpha_j x_j^5 = 0$ has a nontrivial solution for $q \leq 16$ and for $q \geq 20$. Since there are no primes or prime powers in between 16 and 20 congruent to 1 mod 5, $\sum_{j=1}^{n} \alpha_j x_j^5 = 0$ has a nontrivial solution for all q. Thus,

$$\max_{q} \Omega(5,q) = \max_{p \text{ prime}} \Omega(5,p) = 3.$$

□

Proposition A.0.6. *Let $d = 7$. Then $\sum_{j=1}^{n} \alpha_j x_j^7 = 0$ has a nontrivial solution over every $\mathbb{F}_q$ if $n \geq 4$. In other words,*

$$\max_{q} \Omega(7,q) = \max_{p \text{ prime}} \Omega(7,p) = 3.$$

Proof. Suppose $n = 4$. We will show that for any $q \equiv 1 \bmod 7$, $\sum_{j=1}^{n} \alpha_j x_j^7 = 0$ has a nontrivial solution. By Lemma 4.3.7, $\sum_{j=1}^{n} \alpha_j x_j^7 = 0$ has a nontrivial solution if $q < 2^4 = 16$. Furthermore, by Theorem 4.4.4, $\sum_{j=1}^{n} \alpha_j x_j^7 = 0$ has a nontrivial solution if

$$q \geq \frac{1}{4}(7)(7-1)^{\frac{4}{2}} = 63.$$

Thus, $\sum_{j=1}^{n} \alpha_j x_j^7 = 0$ has a nontrivial solution for $q < 16$ and for $q \geq 63$. It remains to check all primes and prime powers between 16 and 62 that are congruent to 1 mod 7. Thus, it remains to check the case $q = 29$ and $q = 43$.

Assume $q = 29$. Since 2 is a generator of $\mathbb{F}_{29}^*$, we can compute the cosets $2^m(\mathbb{F}_{29}^*)^7$ for $m = 0, \ldots, 6$.

$$2^0(\mathbb{F}_{29}^*)^7 = \{1, 12, 17, 28\}$$
$$2^1(\mathbb{F}_{29}^*)^7 = \{2, 24, 5, 27\}$$

$$2^2(\mathbb{F}_{29}^*)^7 = \{4, 19, 10, 25\}$$
$$2^3(\mathbb{F}_{29}^*)^7 = \{8, 9, 20, 21\}$$
$$2^4(\mathbb{F}_{29}^*)^7 = \{16, 18, 11, 13\}$$
$$2^5(\mathbb{F}_{29}^*)^7 = \{3, 7, 22, 26\}$$
$$2^6(\mathbb{F}_{29}^*)^7 = \{6, 14, 15, 23\}$$

Observe that these sets form a partition of $\mathbb{F}_{29}^*$. We will now compute $2^0(\mathbb{F}_{29}^*)^7 + 2^1(\mathbb{F}_{29}^*)^7$ and $2^0(\mathbb{F}_{29}^*)^7 + 2^2(\mathbb{F}_{29}^*)^7$.

$$2^0(\mathbb{F}_{29}^*)^7 + 2^1(\mathbb{F}_{29}^*)^7 = 2^0(\mathbb{F}_{29}^*)^7 \cup 2^1(\mathbb{F}_{29}^*)^7 \cup 2^2(\mathbb{F}_{29}^*)^7 \cup 2^5(\mathbb{F}_{29}^*)^7 \cup 2^6(\mathbb{F}_{29}^*)^7$$

$$2^0(\mathbb{F}_{29}^*)^7 + 2^2(\mathbb{F}_{29}^*)^7 = 2^0(\mathbb{F}_{29}^*)^7 \cup 2^1(\mathbb{F}_{29}^*)^7 \cup 2^2(\mathbb{F}_{29}^*)^7 \cup 2^3(\mathbb{F}_{29}^*)^7 \cup 2^4(\mathbb{F}_{29}^*)^7 \cup 2^5(\mathbb{F}_{29}^*)^7$$

Hence $l_2(\{0, i_2\}) \geq 5$, for $i_2 = 1, 2$. Thus, by Lemma A.0.3, $\sum_{j=1}^{n} \alpha_j x_j^7 = 0$ has a non-trivial solution in $\mathbb{F}_{29}$.

Assume $q = 43$. By Lemma 4.5.5, since $q = 1 + 3(2)(7)$ and $43 < 3^4$, we know that $\sum_{j=1}^{4} \alpha_j x_j^7 = 0$ has a nontrivial solution over $\mathbb{F}_{43}$. Thus $\max_q \Omega(7, q) \leq 3$.

On the other hand, the equation

$$x_1^7 + 2x_2^7 + 8x_3^7 = 0$$

has only the trivial solution in $\mathbb{F}_{29}$ since $2^3(\mathbb{F}_{29})^7 \not\subseteq 2^0(\mathbb{F}_{29})^7 + 2^1(\mathbb{F}_{29})^7$. Thus $\max_{p \text{ prime}} \Omega(7, p) \geq 3$.

Therefore,

$$\max_q \Omega(7, q) = \max_{p \text{ prime}} \Omega(7, p) = 3.$$

□

Proposition A.0.7. *Let $d = 9$. Then $\sum_{j=1}^{n} \alpha_j x_j^9 = 0$ has a nontrivial solution over every $\mathbb{F}_q$ if $n \geq 5$. In other words,*

$$\max_q \Omega(9, q) = \max_{p \text{ prime}} \Omega(9, p) = 4.$$

Proof. By Lemma A.0.1, since $2(9) + 1 = 19$ is prime, we know that for $n \geq 1 + \lceil \log_2(9 + 1) \rceil = 5$, $\sum_{j=1}^{5} \alpha_j x_j^9 = 0$ has a nontrivial solution over prime fields. Suppose

$n = 5$. We will show that for any $q \equiv 1 \bmod 9$, $\sum_{j=1}^{5} \alpha_j x_j^9 = 0$ has a nontrivial solution. By Lemma 4.3.7, $\sum_{j=1}^{5} \alpha_j x_j^9 = 0$ has a nontrivial solution if $q < 2^5 = 32$. By Theorem 4.4.4, $\sum_{j=1}^{5} \alpha_j x_j^9 = 0$ has a nontrivial solution if $q \geq \frac{1}{5}(9)(9-1)^{\frac{5}{3}}$, thus when $q \geq 58$. Thus, it remains to check all primes and prime powers between 32 and 58 that are congruent to 1 mod 9. The only prime or prime power in that range congruent to 1 mod 9 is 37. Thus, we simply need to show that $\sum_{j=1}^{5} \alpha_j x_j^9 = 0$ has a nontrivial solution over $\mathbb{F}_{37}$.

Assume $q = 37$. Since 2 as a generator of $\mathbb{F}_{37}^*$, we can compute the cosets $2^m(\mathbb{F}_{37}^*)^9$ for $m = 0, \ldots, 8$.

$$\begin{aligned}
2^0(\mathbb{F}_{37}^*)^9 &= \{1, 31, 36, 6\} \\
2^1(\mathbb{F}_{37}^*)^9 &= \{2, 25, 35, 12\} \\
2^2(\mathbb{F}_{37}^*)^9 &= \{4, 13, 33, 24\} \\
2^3(\mathbb{F}_{37}^*)^9 &= \{8, 26, 29, 11\} \\
2^4(\mathbb{F}_{37}^*)^9 &= \{16, 15, 21, 22\} \\
2^5(\mathbb{F}_{37}^*)^9 &= \{32, 30, 5, 7\} \\
2^6(\mathbb{F}_{37}^*)^9 &= \{27, 23, 10, 14\} \\
2^7(\mathbb{F}_{37}^*)^9 &= \{17, 9, 20, 28\} \\
2^8(\mathbb{F}_{37}^*)^9 &= \{34, 18, 3, 19\}
\end{aligned}$$

Observe that these sets form a partition of $\mathbb{F}_{37}^*$. We will now compute $2^0(\mathbb{F}_{37}^*)^9 + 2^1(\mathbb{F}_{37}^*)^9$ and $2^0(\mathbb{F}_{37}^*)^9 + 2^2(\mathbb{F}_{37}^*)^9$.

$$2^0(\mathbb{F}_{37}^*)^9 + 2^1(\mathbb{F}_{37}^*)^9 = 2^0(\mathbb{F}_{37}^*)^9 \cup 2^1(\mathbb{F}_{37}^*)^9 \cup 2^2(\mathbb{F}_{37}^*)^9 \cup 2^3(\mathbb{F}_{37}^*)^9 \cup 2^8(\mathbb{F}_{37}^*)^9$$

$$2^0(\mathbb{F}_{37}^*)^9 + 2^2(\mathbb{F}_{37}^*)^9 = 2^0(\mathbb{F}_{37}^*)^9 \cup 2^1(\mathbb{F}_{37}^*)^9 \cup 2^2(\mathbb{F}_{37}^*)^9 \cup 2^5(\mathbb{F}_{37}^*)^9 \cup 2^6(\mathbb{F}_{37}^*)^9 \cup 2^8(\mathbb{F}_{37}^*)^9$$

Hence $l_2(\{0, i_2\}) \geq 5$, for $i_2 = 1, 2$. Thus, by Lemma A.0.3, $\sum_{j=1}^{n} \alpha_j x_j^9 = 0$ has a non-trivial solution in $\mathbb{F}_{37}$. Thus $\max_q \Omega(9, q) \leq 4$.

On the other hand, the equation

$$x_1^9 + 2x_2^9 + 4x_3^9 + 8x_4^9 = 0$$

has only the trivial solution in $\mathbb{F}_{19}$. Thus $\max\limits_{p \text{ prime}} \Omega(9,p) \geq 4$.

Therefore,

$$\max_q \Omega(9,q) = \max_{p \text{ prime}} \Omega(9,p) = 4.$$

$\square$

Proposition A.0.8. *Let $d = 11$. Then $\sum\limits_{j=1}^{n} \alpha_j x_j^9 = 0$ has a nontrivial solution over every $\mathbb{F}_q$ if $n \geq 5$. In other words,*

$$\max_q \Omega(11,q) = \max_{p \text{ } prime} \Omega(11,p) = 4.$$

Proof. By Lemma A.0.1, since $2(11)+1 = 23$ is prime, we know that for $n \geq 1 + \lceil \log_2(11+1) \rceil = 5$, $\sum\limits_{j=1}^{5} \alpha_j x_j^{11} = 0$ has a nontrivial solution over prime fields. Suppose $n = 5$. We will show that for any $q \equiv 1 \bmod 11$, $\sum\limits_{j=1}^{5} \alpha_j x_j^{11} = 0$ has a nontrivial solution. By Lemma 4.3.7, $\sum\limits_{j=1}^{5} \alpha_j x_j^{11} = 0$ has a nontrivial solution if $q < 2^5 = 32$. By Theorem 4.4.4, $\sum\limits_{j=1}^{5} \alpha_j x_j^{11} = 0$ has a nontrivial solution if $q \geq \frac{1}{5}(11)(11-1)^{\frac{5}{3}}$, thus when $q \geq 103$. Thus, it remains to check all primes and prime powers between 32 and 103 that are congruent to 1 mod 11. This means we need to verify that $\sum\limits_{j=1}^{5} \alpha_j x_j^{11} = 0$ is isotropic over $\mathbb{F}_{67}$ and $\mathbb{F}_{89}$.

Assume $q = 67$. By Lemma 4.5.5, since $67 = 1 + 3(2)(11)$ and $67 < 3^5$, $\sum\limits_{j=1}^{5} \alpha_j x_j^{11} = 0$ has a nontrivial solution over $\mathbb{F}_{67}$.

Now assume $q = 89$. Since 3 is a generator of $\mathbb{F}_{89}^*$, we can compute the cosets $3^m(\mathbb{F}_{89}^*)^{11}$ for $m = 0, \ldots, 10$.

$$3^0(\mathbb{F}_{89}^*)^{11} = \{1, 37, 55, 34, 88, 77, 12, 52\}$$
$$3^1(\mathbb{F}_{89}^*)^{11} = \{3, 22, 76, 13, 86, 53, 36, 67\}$$
$$3^2(\mathbb{F}_{89}^*)^{11} = \{9, 66, 50, 39, 80, 70, 19, 23\}$$
$$3^3(\mathbb{F}_{89}^*)^{11} = \{27, 20, 61, 28, 62, 32, 57, 69\}$$
$$3^4(\mathbb{F}_{89}^*)^{11} = \{81, 60, 5, 84, 8, 7, 82, 29\}$$
$$3^5(\mathbb{F}_{89}^*)^{11} = \{65, 2, 15, 74, 24, 21, 68, 87\}$$
$$3^6(\mathbb{F}_{89}^*)^{11} = \{17, 6, 45, 44, 72, 63, 26, 83\}$$

$$3^7(\mathbb{F}_{89}^*)^{11} = \{51, 18, 46, 43, 38, 11, 78, 71\}$$
$$3^8(\mathbb{F}_{89}^*)^{11} = \{64, 54, 49, 40, 25, 33, 56, 35\}$$
$$3^9(\mathbb{F}_{89}^*)^{11} = \{14, 73, 58, 31, 75, 10, 79, 16\}$$
$$3^{10}(\mathbb{F}_{89}^*)^{11} = \{42, 41, 85, 4, 47, 30, 59, 48\}$$

Observe that these sets form a partition of $\mathbb{F}_{89}^*$. We will now compute $3^0(\mathbb{F}_{89}^*)^{11} + 3^1(\mathbb{F}_{89}^*)^{11}$ and $3^0(\mathbb{F}_{89}^*)^{11} + 3^2(\mathbb{F}_{89}^*)^{11}$.

$$\begin{aligned} 3^0(\mathbb{F}_{89}^*)^{11} + 3^1(\mathbb{F}_{89}^8)^{11} = 3^0(\mathbb{F}_{89}^*)^{11} &\cup 3^1(\mathbb{F}_{89}^*)^{11} \cup 3^2(\mathbb{F}_{89}^*)^{11} \cup 3^5(\mathbb{F}_{89}^*)^{11} \\ &\cup 3^8(\mathbb{F}_{89}^*)^{11} \cup 3^9(\mathbb{F}_{89}^*)^{11} \cup 3^{10}(\mathbb{F}_{89}^*)^{11} \end{aligned}$$

$$\begin{aligned} 3^0(\mathbb{F}_{89}^*)^{11} + 3^2(\mathbb{F}_{89}^*)^{11} = 3^0(\mathbb{F}_{89}^*)^{11} &\cup 3^1(\mathbb{F}_{89}^*)^{11} \cup 3^2(\mathbb{F}_{89}^*)^{11} \cup 3^3(\mathbb{F}_{89}^*)^{11} \cup 3^4(\mathbb{F}_{89}^*)^{11} \\ &\cup 3^5(\mathbb{F}_{89}^*)^{11} \cup 3^7(\mathbb{F}_{89}^*)^{11} \cup 3^8(\mathbb{F}_{89}^*)^{11} \cup 3^9(\mathbb{F}_{89}^*)^{11} \end{aligned}$$

Hence $l_2(\{0, i_2\}) \geq 7$, for $i_2 = 1, 2$. Thus, by Lemma A.0.3, $\sum_{j=1}^{n} \alpha_j x_j^{11} = 0$ has a non-trivial solution in $\mathbb{F}_{89}$. Thus $\max_q \Omega(11, q) \leq 4$.

On the other hand, the equation

$$x_1^{11} + 2x_2^{11} + 4x_3^{11} + 8x_4^{11} = 0$$

has only the trivial solution in $\mathbb{F}_{23}$. Thus, $\max_{p \text{ prime}} \Omega(11, p) \geq 4$.

Therefore,

$$\max_q \Omega(11, q) = \max_{p \text{ prime}} \Omega(11, p) = 4.$$

□

Now, we will compute values of $\max_q \Omega_A(1, d, q)$ for $d = 2, 4, 6, 8$.

Proposition A.0.9. *Let $d = 2$. Then the A-equation $\sum_{j=1}^{n} \alpha_j x_j^2 = 0$ has a nontrivial solution over every $\mathbb{F}_q$ if $n \geq 3$. In other words,*

$$\max_q \Omega_A(2, q) = \max_{p \text{ prime}} \Omega_A(2, p) = 2.$$

Proof. By Theorem 4.4.7, we know that

$$\max_q \Omega_A(2, q) \leq \lceil 2 \log_2(2) - \log_2(\log_2(2)) \rceil = 2.$$

By Lemma A.0.1, we know that

$$\max_{p \text{ prime}} \Omega_A(2, p) \geq \lceil \log_2(2 + 1) \rceil = 2.$$

Since $\max_q \Omega_A(2,q) \leq \max_{p \text{ prime}} \Omega_A(2,p)$, it follows that

$$\max_q \Omega_A(2,q) = \max_{p \text{ prime}} \Omega_A(2,p) = 2.$$

$\square$

Proposition A.0.10. *Let $d = 4$. Then the A-equation $\sum_{j=1}^{n} \alpha_j x_j^4 = 0$ has a nontrivial solution over every $\mathbb{F}_q$ if $n \geq 3$. In other words,*

$$\max_q \Omega_A(4,q) = \max_{p \text{ prime}} \Omega_A(4,p) = 2.$$

Proof. First, we will show that $\max_{p \text{ prime}} \Omega_A(4,p) \geq \max_{p \text{ prime}} \Omega_A(2,p)$. Consider the A-equation $\sum_{j=1}^{n} \alpha_j x_j^4$. Let $(\xi_1, \ldots, \xi_n)$ be a nontrivial solution of that A-equation. Let $y_j = x_j^2$. Then we can rewrite the A-equation as $\sum_{j=1}^{n} \alpha_j y_j^2$. Notice that this new A-equation is a diagonal form of degree 2 and has the nontrivial solution $(\xi_1^2, \ldots, \xi_n^2)$. Thus, if there exists a nontrivial solution to $\sum_{j=1}^{n} \alpha_j x_j^4$, then there exists a nontrivial solution to $\sum_{j=1}^{n} \alpha_j y_j^2$. Thus, $\max_{p \text{ prime}} \Omega_A(4,p) \geq \max_{p \text{ prime}} \Omega_A(2,p)$. Since $\max_{p \text{ prime}} \Omega_A(2,p) = 2$, it follows $\max_{p \text{ prime}} \Omega_A(4,p) \geq 2$.

By Theorem 4.2.1, $\sum_{j=1}^{n} \alpha_j x_j^4 = 0$ is an A-equation if and only if $d \mid \frac{q-1}{2}$. Thus, we need only show that $\sum_{j=1}^{n} \alpha_j x_j^4 = 0$ has a nontrivial solution over $\mathbb{F}_q$ for all $q \equiv 1 \bmod 8$.

Suppose $n = 3$. By Theorem 4.4.4, $\sum_{j=1}^{3} \alpha_j x_j^4 = 0$ has a nontrivial solution if $q \geq \frac{1}{3}(4)(4-1)^3 = 36$. By Lemma 4.3.7, $\sum_{j=1}^{3} \alpha_j x_j^4 = 0$ has a nontrivial solution if $q < 2^3 = 8$. It remains to check the primes and prime powers between 8 and 36 that are congruent to 1 mod 8. Thus it remains to check $q = 9$ and $q = 25$.

Let $q = 9$. By Theorem 5.3.4, letting $m = 1$, we find that $\sum_{j=1}^{3} \alpha_j x_j^4 = 0$ has a nontrivial solution over $\mathbb{F}_9$.

Let $q = 25$. By Lemma 4.5.5, since $25 = 1 + 3(2)(4)$, $\sum_{j=1}^{3} \alpha_j x_j^4 = 0$ has a nontrivial solution if $q < 3^3$. Since $25 < 27$, there exists a nontrivial solution over $\mathbb{F}_{25}$. Thus $\max_q \Omega(4,q) < 3$.

Therefore,

$$\max_{q} \Omega(4,q) = \max_{p \text{ prime}} \Omega(4,p) = 2.$$

□

Proposition A.0.11. *Let $d = 6$. Then the A-equation $\sum_{j=1}^{n} \alpha_j x_j^6 = 0$ has a nontrivial solution over every $\mathbb{F}_q$ if $n \geq 4$. In other words,*

$$\max_{q} \Omega_A(6,q) = \max_{p \text{ prime}} \Omega_A(6,p) = 3.$$

Proof. By Theorem 4.2.1, $\sum_{j=1}^{n} \alpha_j x_j^6 = 0$ is an A-equation if and only if $d \mid \frac{q-1}{2}$. Thus we need only show that $\sum_{j=1}^{n} \alpha_j x_j^6 = 0$ has a nontrivial solution over $\mathbb{F}_q$ for all $q \equiv 1 \bmod 12$. By Lemma A.0.1, since $2(6)+1 = 13$ is prime, we know that $\max_{p \text{ prime}} \Omega_A(6,p) \geq \lceil \log_2(6+1) \rceil = 3$. Furthermore, by Theorem 4.4.7, $\max_{q} \Omega_A(6,q) \leq \lceil 2\log_2(6) - \log_2(\log_2(6)) \rceil = 4$.

Suppose $n = 4$. By Lemma 4.3.7, $\sum_{j=1}^{4} \alpha_j x_j^6 = 0$ has a nontrivial solution over $\mathbb{F}_q$ when $q < 2^4 = 16$. By Theorem 4.4.4, $\sum_{j=1}^{4} \alpha_j x_j^6 = 0$ has a nontrivial solution over $\mathbb{F}_q$ when $q \geq \frac{1}{4}(6)(6-1)^2 = 37.5$, that is, when $q \geq 38$. It remains to check the primes and prime powers between 16 and 37 that are congruent to 1 mod 12. Thus it remains to check $q = 25$ and $q = 37$.

Assume $q = 25$. By Theorem 5.3.4, letting $m = 1$, we find that $\sum_{j=1}^{4} \alpha_j x_j^6 = 0$ has a nontrivial solution over $\mathbb{F}_{25}$.

Assume $q = 37$. By Lemma 4.5.5, since $37 = 1 + 3(2)(6)$, $\sum_{j=1}^{4} \alpha_j x_j^6 = 0$ has a nontrivial solution if $q < 3^4$. Since $37 < 81$, there exists a nontrivial solution over $\mathbb{F}_{37}$. Thus, $\max_{q} \Omega_A(6,q) < 4$.

Therefore,

$$\max_{q} \Omega_A(6,q) = \max_{p \text{ prime}} \Omega_A(6,p) = 3.$$

□

Proposition A.0.12. *Let $d = 8$. Then the A-equation $\sum_{j=1}^{n} \alpha_j x_j^8 = 0$ has a nontrivial solution over every $\mathbb{F}_q$ if $n \geq 5$. In other words,*

$$\max_{q} \Omega_A(8,q) = \max_{p \text{ prime}} \Omega_A(8,p) = 4.$$

Proof. By Theorem 4.2.1, $\sum_{j=1}^{n} \alpha_j x_j^8 = 0$ is an A-equation if and only if $d \mid \frac{q-1}{2}$. Thus we need only show that $\sum_{j=1}^{n} \alpha_j x_j^8 = 0$ has a nontrivial solution over $\mathbb{F}_q$ for all $q \equiv 1 \bmod 16$.

By Lemma A.0.1, since $2(8) = 1 = 17$ is prime $\max_{p \text{ prime}} \Omega_A(8,p) \geq \lceil \log_2(8+1) \rceil = 4$.

Suppose $n = 5$. By Lemma 4.3.7, $\sum_{j=1}^{5} \alpha_j x_j^8 = 0$ has a nontrivial solution over $\mathbb{F}_q$ for $q < 2^5 = 32$. By Theorem 4.4.4, $\sum_{j=1}^{5} \alpha_j x_j^8 = 0$ has a nontrivial solution over $\mathbb{F}_q$ for $q \geq \frac{1}{5}(8)(8-1)^{\frac{5}{3}}$, that is, $q \geq 41$. Since there are no primes or prime powers between 32 and 41 that are congruent to 1 mod 16, it follows that $\max_q \Omega_A(8,q) < 5$.

Therefore,

$$\max_q \Omega_A(8,q) = \max_{p \text{ prime}} (8,p) = 4.$$

$\square$

Appendix B Computational Component

Example B.0.1. Checking for Anisotropic Forms over $\mathbb{F}_p$ for Fixed p

This code takes as inputs: d, n, p, and ρ, where ρ is a fixed primitive element of $\mathbb{F}_p$.

```
import numpy as np
from itertools import product

A = [\rho^i] for i=1, \dots, d-1

C = (A tuple containing the elements of F_p^d)
D = list(product(C, repeat = n-1))

for j_1 in range(d-1):
  for j_2 in range(d-1):
    \vdots
      for j_{n-1} in range(d-1):
        if j_1 < j_2 < \hdots < j_{n-1}:
          B = (A[j_1], A[j_2], \dots, A[j_{n-1}])
          I = np.eye(1)
        else:
          continue
        J = np.matrix([B])
        K = np.hstack([I,J])
        count = 0
        for l in D:
          Z = K*np.transpose([l])
          if Z[0] % p == 0:
            count += 1
        if count == 1:
          print(K)
print("Done")
```

Bibliography

[1] James Ax. Zeroes of polynomials over finite fields. *Amer. J. Math.*, 86:255–261, 1964.

[2] L. Carlitz and S. Uchiyama. Bounds for exponential sums. *Duke Math. J.*, 24:37–41, 1957.

[3] C. Chevalley. Démonstration d'une hypothèse de M. Artin. *Abh. Math. Sem. Univ. Hamburg*, 11(1):73–75, 1935.

[4] S. Chowla. On the congruence $\sum_{i=1}^{s} a_i {x_i}^k \equiv 0(mod\ p)$. *J. Indian Math. Soc. (N.S.)*, 25:47–48, 1961.

[5] S. Chowla, H. B. Mann, and E. G. Straus. Some applications of the Cauchy-Davenport theorem. *Norske Vid. Selsk. Forh. Trondheim*, 32:74–80, 1959.

[6] James F. Gray. Diagonal forms of odd degree over a finite field. *Michigan Math. J.*, 7:297–301, 1960.

[7] James F. Gray. Diagonal forms of odd degree over a finite field. *Abstract 564-86, Notices of the Amer. Math. Soc.*, 6, Dec. 1959.

[8] James Francis Gray. *Diagonal forms of prime degree.* ProQuest LLC, Ann Arbor, MI, 1959. Thesis (Ph.D.)–University of Notre Dame.

[9] D. R. Heath-Brown. On Chevalley-Warning theorems. *Uspekhi Mat. Nauk*, 66(2(398)):223–232, 2011.

[10] Kenneth Ireland and Michael Rosen. *A classical introduction to modern number theory*, volume 84. Springer Science & Business Media, 2013.

[11] J. H. B. Kemperman. On small sumsets in an abelian group. *Acta Math.*, 103:63–88, 1960.

[12] Martin Kneser. Ein Satz über abelsche Gruppen mit Anwendungen auf die Geometrie der Zahlen. *Math. Z.*, 61:429–434, 1955.

[13] D. J. Lewis. Diagonal forms over finite fields. *Norske Vid. Selsk. Forh. Trondheim*, 33:61–65, 1960.

[14] Aimo Tietäväinen. On the non-trivial solvability of some equations and systems of equations in finite tields. *Ann. Acad. Sci. Fenn. Ser. A I No.*, 360:38, 1965.

[15] Aimo Tietäväinen. On pairs of additive equations. *Ann. Univ. Turku. Ser. A I No.*, 112:7, 1967.

[16] Aimo Tietäväinen. On diagonal forms over finite fields. *Ann. Univ. Turku. Ser. A I No.*, 118:10, 1968.

[17] Ewald Warning. Bemerkung zur vorstehenden Arbeit von Herrn Chevalley. *Abh. Math. Sem. Univ. Hamburg*, 11(1):76–83, 1935.

[18] André Weil. Numbers of solutions of equations in finite fields. *Bull. Amer. Math. Soc.*, 55:497–508, 1949.

VITA

Rachel Louise Petrik

EDUCATION

University of Kentucky, Lexington, KY

Ph.D. Mathematics (Expected) May 2020
Advisor: David Leep

M.A. Mathematics May 2016

West Virginia University, Morgantown, WV

B.A. Mathematics May 2014

Summa Cum Laude

AWARDS & FELLOWSHIPS

External Awards:

- AMS JMM Graduate Student Travel Grant Jan. 2020
- NSF Mathematics Science Graduate Internship Summer 2018, Summer 2019

Internal Awards:

- University of Kentucky Outstanding Teaching Award, University of Kentucky Apr. 2018
- Mathematics Summer Research Fellowship, University of Kentucky Summer 2017
- College of Arts and Sciences Outstanding Teaching Assistant Award, University of Kentucky Apr. 2017
- Kentucky Opportunity Fellowship, University of Kentucky Aug. 2014–May 2015
- Scholarship of Distinction Level 1, West Virginia University Aug. 2010 – May 2014

PUBLICATIONS

- R. Petrik, B. Arik, J. M. Smith, ”Towards Architecture and OS-independent Malware Detection via Memory Forensics,” ACM Conference on Computer and Communications Security (ACM CCS), 2018. https://www.osti.gov/servlets/purl/1486945