POINTS ON LINES IN $\mathbb{F}_q^3$

SPUR FINAL PAPER, SUMMER 2013

ZIPEI NIE AND ANTHONY WANG

MENTOR: BEN YANG
PROJECT SUGGESTED BY LARRY GUTH
JULY 31, 2013

Abstract. In this paper we study an incidence problem in finite geometry. Suppose we are given a set X which is the union of some number of lines in $\mathbb{F}_q^3$. We choose a subset Y of X, such that for each line ℓ in X, at least half of its points are in Y. We show that $|Y|$ is always at least some fraction of $|X|$. Using the polynomial method and degree reduction, it was previously known that such a statement holds for large and small $|Y|$ (when $|Y| \geq q^2 \log q$ or $|Y| \leq q^2/\log q$). We close the gap by proving the statement for the remaining cases. We first note that such a statement holds for a set of points such that each point lies on at most two lines. We then show that there cannot be too many points with three or more lines lying on the zero set of a nonlinear polynomial, and use this to prove the statement in the remaining cases.

1. Introduction

The polynomial method has been used in recent years to prove results in incidence geometry. Dvir use the polynomial method to obtain a solution to the finite field Kakeya conjecture [2]. The solution to the Erdős distinct distances problem in the plane by Guth and Katz also utilized this method [5].

The idea of the polynomial method is to find a polynomial of minimal degree vanishing on a set of interest, and then use properties of polynomials to prove statements about the set.

Our problem concerns a set of lines in finite space $\mathbb{F}_q^3$. Letting X denote the set of points on these lines, we choose a subset Y such that for every line, at least half of its points are in Y. The question is then: is Y always a significant fraction of X? More precisely, our problem is the following:

Problem. *Suppose $\ell_1, \cdots, \ell_n$ are distinct lines in $\mathbb{F}_q^3$, and $X = \bigcup \ell_i$. For each $i = 1, \cdots, n$, let S_i be a subset of ℓ_i such that $|S_i| \geq \frac{q}{2}$. Let $|Y| = \bigcup S_i$. Does there exist a constant (independent of q) such that $|X| \leq c|Y|$.*

We answer this question in the affirmative. In a sense, this question is asking about whether a set of lines can have too many points lying on many lines. For example, if no two lines intersect, then each point is on exactly one line. In this situation, the set Y is always at least half of X, since we must take $\frac{q}{2}$ points from each line. In a similar fashion, we can show the following by just counting point-line incidences:

Proposition. *Let c be a constant. Suppose each point of X lies on at most c of the lines in the list ℓ_1, ℓ_2, $\cdots$, ℓ_n. Then $|X| \leq 2c|Y|$.*

Proof. Each point of Y lies on at most c lines, so the number point-line incidence pairs $\#(p, \ell)$ in Y is at most $c|Y|$. Each line of Y has at least $\frac{q}{2}$ points, so the number of point-line incidence pairs $\#(p, \ell)$ in Y is at least $\frac{nq}{2}$.

Each point of X lies on at least 1 line, so the number of point-line incidence pairs in X is at least $|X|$. Each line of X contains exactly q points, so the number of point-line incidence pairs in X is exactly nq.

We then get

$$|X| \leq nq = 2 \cdot \frac{nq}{2} \leq 2c|Y|,$$

as desired. □

The proof strategy for the problem is as follows. First, using the argument given by Guth in his polynomial course lecture notes [4, Lecture 12], we can use the polynomial method along with the probabilistic method to show the statement for sufficiently large and sufficiently small $|Y|$. In the remaining gap, this argument will show that there is a polynomial of low degree ($\lesssim \log q$) vanishing on all but a small subset of X containing at most $4|Y|$ elements.

To prove the statement in the gap, we are motivated by the previous proposition. Instead of bounding the number of lines than can pass through any point, we wish to show that there are not too many points with more than 2 lines through them. Modifying the proof of the proposition to take into account these points will then be enough to show the bound.

The proof that there are not too many points with many more than 2 lines through them follows some theory of incidence geometry in $\mathbb{R}^3$. Here, it is known

that for any given polynomial P, there exist a set of polynomials SP such that if a point lies on at least three lines in the zero set of P, then the polynomials in SP vanish at that point (see [4, Leture 14] for details). To get a similar argument to work in finite fields, there are several obstacles to overcome. Due to a lack of neighborhoods, we use the theory of formal power series to 'expand' P locally around any point. To get the special polynomials SP, we need to take derivatives of P. Due to problems with the characteristic being positive, we will need to change these derivatives to Hasse derivatives. Moreover, we will need to consider other sets of polynomials to deal with the obstacle related to the fact that in characteristic p, the derivative of x^p vanishes. Developing these tools for finite fields allows us to get enough control over the number of points with more than two lines to prove the desired bound.

The outline of the paper is as follows:

In Section 2, we give some preliminary results in order to use the polynomial method. Here bounds are given for the degree of the polynomial of smallest degree vanishing on a set of points or lines, the number of zeros a polynomial can have in n dimensional space over a finite field, the number of common zeros two polynomials can have in three dimensional space over a finite field, and a probability lemma needed to apply the probabilistic method. These are largely from lecture notes of Guth's course on the polynomial method [4].

In Section 3, we give in detail the arguments in [4, Lecture 12] showing that the statement holds for $|Y| \geq q^2 \log q$ and $|Y| \leq \frac{q^2}{\log q}$. Moreover, it is shown that there is a polynomial of degree $\lesssim \log q$ vanishing on all but a small subset of X in the case when $\frac{q^2}{\log q} \leq |Y| \leq q^2 \log q$.

Section 4 contains a largely self contained introduction to the theory of Hasse derivatives. The results we will need from this section include the product rule (Proposition 4.2.3(4)), the definition of a total derivative (Definition 4.2.9), and a result showing the nature of Hasse derivatives in prime characteristic (Corollary 4.3.3).

Section 5 develops the theory of n-flat points. This theory allows us to detect planes, which we then use to prove that a nonlinear irreducible polynomial P of small degree d in $\mathbb{F}_q[x, y, z]$ can have at most $\approx qd^3$ points with at least three lines through them lying on the zero set of P.

Section 6 finishes the proof of the main theorem. We first prove a statement for a set of lines lying on an irreducible polynomial of small degree, and then use this result to show the statement in all remaining cases.

Acknowledgements This research was done at MIT's undergraduate SPUR program in the summer of 2013. We would like to thank Ben Yang for being our project mentor and helping us in our research project. We would like to thank Larry Guth for suggesting the problem we worked on. We would like to thank Pavel Etingof for the suggestion of using formal power series, and Pavel Etingof and Jacob Fox for the weekly meetings. We would also like to thank Slava Gerovich for organizing the SPUR program.

2. Preliminaries

In this section, we prove some of the basic results needed for the use of the polynomial method. Most of this material is from Guth's lecture notes [4]; they are repeated here for completeness of exposition, and to set our notations.

2.1. **The degree of points and lines.**

Definition 2.1.1. Let S be a set of points in n dimensional space $\mathbb{F}^n$. The **degree** $\deg(S)$ of S is the minimal degree of the nonzero polynomials vanishing on S.

Let V_d be the vector space of polynomials with degree $\leq d$ in n variables.

Proposition 2.1.2. *For any finite set S, we have* $\deg(S) \leq n|S|^{\frac{1}{n}}$.

Proof. (From [4, Lecture 1, Pages 1-2]) The statement follows from a dimension counting argument. The evaluation map $E : V_d \to \mathbb{F}^{|S|}$ given by $p \mapsto (p(s))_{s\in S}$ is a linear map. The dimension of V_d is $\binom{d+n}{n} \geq \frac{d^n}{n!}$. For $d = (n!)^{\frac{1}{n}}|S|^{\frac{1}{n}} \leq n|S|^{\frac{1}{n}}$, the dimension of $V(d)$ is greater than $|S|$, so there is a polynomial of degree less than or equal to d vanishing on S. Thus, $\deg(S) \leq n|S|^{\frac{1}{n}}$. □

When the set S has some sort of algebraic structure, we can hope to do better than what we have for a generic set S. For example, we have the following:

Theorem 2.1.3. *For any L lines in $\mathbb{F}^n$, there is a nonzero polynomial of degree $\leq nL^{\frac{1}{n-1}}$ vanishing on every line.*

Proof. (From [4, Lecture 12, Page 1]) We wish to find a degree d such that $\dim V_d > (d+1)L$. If such a number exists, we can choose $d+1$ points on each line for a total of $(d+1)L$ points. Then, since $\dim V_d > (d+1)L$, we can find a polynomial of degree at most d vanishing on all $(d+1)L$ points. Because our polynomial is of degree at most d and vanishes on $d+1$ points of each line, it must vanish on each line.

Now to get $\dim V_d > (d+1)L$, note that $\dim V_d = \binom{d+n}{n} = \frac{1}{n!}(d+n)\cdots(d+1)$. Thus, $\frac{\dim V(d)}{d+1} \geq \frac{d^{n-1}}{n!}$. Taking $d \approx (n!)^{\frac{1}{n-1}}L^{\frac{1}{n-1}} \leq nL^{\frac{1}{n-1}}$ will work. □

We can explicitly calculate V_d to get a better degree in specific cases. For example, there is a quadratic vanishing on any 3 lines: here $V_2 = 10$, while $(2+1)3 = 9$.

2.2. **The Number of Zeros of a Polynomial.** For a function $f : \mathbb{F}_q^n \to \mathbb{F}_q$, let $Z(f)$ denote the set of zeros of f, that is, the set of $\mathbf{x} \in \mathbb{F}_q$ such that $f(\mathbf{x}) = 0$. For more than one function, $f_1, \cdots, f_k$, let $Z(f_1, \cdots, f_k)$ denote the common zeros of all k functions.

We have the following well-known bound on the number of zeros for polynomials in finite space.

Theorem 2.2.1. *Let $P : \mathbb{F}_q^n \to \mathbb{F}_q$ be a nonzero polynomial in n variables $x_1, \cdots, x_n$. The number of zeros of P is at most* $(\deg P)q^{n-1}$.

Proof. We use mathematical induction on the dimension of the space n.

When $n = 1$, this is just the statement that a d degree polynomial in one variable can have at most d roots.

Suppose we have shown the theorem for $n = k$. Let P be a polynomial of degree d in $k+1$ variables. We can write

$$P = \sum_{j=0}^{\ell} P_j(x_1, \cdots, x_k)x_{k+1}^j,$$

where P_j is a polynomial in k variables of degree at most $d-j$, and $P_\ell \neq 0$ where $\ell \leq d$. For any given $(x_1, \cdots, x_k)$, there are two cases:

(1) If $P_\ell(x_1, \cdots, x_k) \neq 0$, then there are at most ℓ values of x_{k+1} for which $P = 0$. Counting the total number of zeros possible from this case, we see that there are at most q^k elements of $\mathbb{F}_q^k$, so there are at most $q^k\ell$ zeros.
(2) If $P_\ell(x_1, \cdots, x_k) = 0$, then the worst that can happen is that all values of x_{k+1} will give 0. By the induction hypothesis, $P_\ell(x_1, \cdots, x_k) = 0$ for at most $\deg(P_\ell)q^{k-1} \leq (d-\ell)q^{k-1}$ values of $\mathbb{F}_q^k$. Thus, there are at most $(d-\ell)q^k$ zeros in this case.

We add to get that there are at most dq^k zeros, and the induction is done. □

What happens if we are looking at the common zeros of two polynomials? We have the following result:

Theorem 2.2.2. *Let P_1, P_2 be coprime polynomials $\mathbb{F}_q^3 \to \mathbb{F}_q$. The number of common roots of P_1 and P_2 does not exceed $3q(\deg P)(\deg Q)$.*

The proof is a modification of a proof of Bezout's theorem in the plane given in [4, Lecture 13, Pages 1-2]. First, we prove a very handy lemma (see [4, Lecture 13, Lemma 1.1]).

Lemma 2.2.3. *Let X be a finite set in $\mathbb{F}^n$, and $I(X)$ be the ideal of polynomials in $\mathbb{F}[x_1, \cdots, x_n]$ which vanish on X. Then $\dim(\mathbb{F}[x_1, \cdots, x_n]/I(X)) = |X|$.*

Note that X must be finite.

Proof. If X is finite, the evaluation map $E : \mathbb{F}[x_1, \cdots, x_n] \to \mathrm{Fun}(X, \mathbb{F})$ is a surjective linear map with kernel $I(X)$. The claim follows by noting $\dim(\mathrm{Fun}(X, \mathbb{F})) = |X|$. □

Lemma 2.2.4. $I(\mathbb{F}_q^n) = (x_1^q - x_1, \cdots, x_n^q - x_n)$.

Proof. The proof can be gotten from dimension counting. Clearly $I(\mathbb{F}_q^n) \supset (x_1^q - x_1, \cdots, x_n^q - x_n)$. But $\mathbb{F}_q[x_1, \cdots, x_n]/(x_1^q - x_1, \cdots, x_n^q - x_n)$ has a basis given by the (images under the quotient map of the) monomials $x_1^{m_1} \cdots x_n^{m_n}$, where $0 \leq m_i \leq q-1$, so it has dimension q^n. This is exactly $|\mathbb{F}_q^n|$, so $I(\mathbb{F}_q^n)$ cannot be larger than $(x_1^q - x_1, \cdots, x_n^q - x_n)$. We then see that the evaluation map E applied to V_{3q-2} is surjective. □

Proof of Theorem 2.2.2. The ideal of polynomials vanishing on the common zeros of coprime polynomials P, Q in $\mathbb{F}_q^3$ must contain $(I(\mathbb{F}_q^3), P, Q)$. For our proof, the only information from $I(\mathbb{F}_q^3)$ we really need is that V_{3q-2} surjects onto $\mathrm{Fun}(\mathbb{F}_q^3, \mathbb{F}_q)$. Let $D = \dim P$, and $E = \dim Q$.

Let $I_d = (P) \cap V_d$, and $R_d = V_d/I_d$. Multiplication by P takes V_{d-D} bijectively onto I_d, so

$$\begin{aligned}\dim R_d = \dim V_d - \dim I_d = \dim V_d - \dim V_{d-D} &= \binom{d+3}{3} - \binom{d-D+3}{3} \\ &= \frac{1}{6}\left(D(3d^2+12d+11) - D^2(3d+6) + D^3\right) \\ &= \frac{1}{6}\left(3Dd^2 + (12D-3D^2)d + (11D - 6D^2 + D^3)\right).\end{aligned}$$

Let J_d be the image of $(Q) \cap V_d$ under the quotient map $\pi : V_d \to R_d$, and $S_d = V_d/J_d$. We claim that multiplication by $\pi(Q)$ takes R_{d-E} injectively into R_d. Indeed, suppose $r\pi(Q) = 0$ in R_d. Choosing a lift for the map, we can find a

polynomial P_1 in V_{d-E} such that $P_1Q = PP_2$ for some polynomial P_2. But P and Q are coprime, so P divides P_1, and thus $r = \pi(P_1) = 0$. Thus $\dim J_d \geq \dim V_{d-E}$, and

$$\begin{aligned}\dim S_d = \dim R_d - \dim J_d &\leq \dim R_d - \dim R_{d-E} \\ &= \frac{1}{6}\left(6dDE - 3DE^2 - 3D^2E + 12DE\right).\end{aligned}$$

Because the evaluation map from V_{3q-2} surjects onto the space $\mathrm{Fun}(Z(P,Q), \mathbb{F}_q)$ with kernel containing $I = (P,Q) \cap V_{3q-2}$, we have that $|Z(P,Q)| \leq \dim V_{3q-2}/I = \dim S_{3q-2} \leq 3qDE$, as desired. □

2.3. **Probability Lemma.** To apply the probabilistic method, we shall need a lemma in probability theory. Suppose you choose a subset of a set of n objects such that each object is chosen independently with probability p. You expect to choose np objects on average, and not to deviate too much from this (especially if n is large, by the central limit theorem). The lemma gives bounds on the probabilities of having more than $2np$ objects, and on having less than $\frac{1}{2}np$ objects in all situations.

Lemma 2.3.1. *Let X be a random variable with a binomial distribution $B(n,p)$. (That is, the number of objects you get if you choose each object in a set of n objects independently with probability p.) The mean is then $\mu = np$. If $p \leq \frac{1}{2}$, then*

1. *the probability that $X \geq 2\mu$ is at most $\exp(-\frac{1}{100}np)$.*
2. *the probability that $X \leq \frac{1}{2}\mu$ is at most $\exp(-\frac{1}{100}np)$.*

Proof. (From [4, Lecture 12, Page 6]) Let $X_i = 1$ if you choose the ith object, and 0 if you do not, so that $X = X_1 + \cdots + X_n$. Consider the random variable $Y = e^{\beta X}$ for some $\beta \in \mathbb{R}$. By independence, we have

$$\mathbb{E}e^{\beta X} = \mathbb{E}e^{\beta \sum_i X_i} = \mathbb{E}\prod_i e^{\beta X_i} = \prod_i \mathbb{E}e^{\beta X_i} = (1 - p + pe^{\beta})^n.$$

On the other hand, if $\beta \geq 0$, then $\mathbb{E}e^{\beta X} \geq e^{2\beta pn} P(X \geq 2\mu)$. Thus,

$$P(X \geq 2\mu) \leq \left(\frac{1 - p + pe^{\beta}}{e^{2\beta p}}\right)^n.$$

Plugging in $\beta = 1$, we get

$$P(X \geq 2\mu) \leq \left(\frac{1 + p(e-1)}{1 + 2p}\right)^n < e^{-\frac{1}{100}pn},$$

since $p((2 - \frac{1}{100} - (e-1))p - \frac{1}{50}) > 0$ for $p \in [0,1]$, so

$$1 + (e-1)p \leq 1 + (2 - \frac{1}{100})p - \frac{1}{50}p^2 \leq (1+2p)(1 - \frac{1}{100}p + \frac{1}{20000}p^2 - \cdots).$$

For the other inequality, we note that when $\beta \leq 0$, we get $\mathbb{E}e^{\beta X} \geq e^{\frac{1}{2}\beta pn} P(E \leq \frac{1}{2}\mu)$. As above, this gives

$$P(X \leq \frac{1}{2}\mu) \leq \left(\frac{1 - p + pe^{\beta}}{e^{\frac{1}{2}\beta p}}\right)^n.$$

Take $\beta = -\frac{1}{10}$, and use power series to find that

$$P(X \leq \frac{1}{2}\mu) \leq \left(\frac{1 - p + p(1 - \frac{1}{10} + \frac{1}{200} - \cdots)}{1 - \frac{1}{20}p + \cdots}\right)^n \leq \left(\frac{1 - \frac{19}{200}p}{1 - \frac{1}{20}p}\right)^n \leq e^{-\frac{1}{100}pn},$$

as $1-\frac{19}{200}p \leq 1-\frac{1}{20}p-\frac{1}{100}p+\frac{1}{2000}p^2+(1-\frac{1}{20}p)(\frac{1}{200000}p^2-\cdots) = (1-\frac{1}{20}p)e^{-\frac{1}{100}p}$. □

3. The Small and Large Cases

For quick reference, let us state our problem again.

Problem. *Suppose $\ell_1, \cdots, \ell_n$ are distinct lines in $\mathbb{F}_q^3$, and $X = \bigcup \ell_i$. For each $i = 1, \cdots, n$, let S_i be a subset of ℓ_i such that $|S_i| \geq \frac{q}{2}$. Let $|Y| = \bigcup S_i$. Does there exist a constant (independent of q) such that $|X| \leq c|Y|$.*

In this section, we will give the arguments showing that if $|Y| \leq \frac{q^2}{\log q}$ or if $|Y| \geq q^2 \log q$, then we can find a constant c independent of q such that $|X| \leq c|Y|$. The argument follows the one given in [4, Lecture 12, Proposition 2.1], but with more detailed bookkeeping.

3.1. **A Trivial Bound.**

Proposition 3.1.1. *In $\mathbb{F}_q^3$, we always have $|X| \leq q^3|Y|$.*

Proof. The set Y has at least 1 point, while the set X has at most q^3 points. □

The point of the trivial bound is to show that it suffices to prove $|X| \leq c|Y|$ for all but finitely many q.

3.2. **The two dimensional case.**

The proof when the lines are required to lie in $\mathbb{F}_q^2$ instead of $\mathbb{F}_q^3$ does not rely on the polynomial method, but merely the fact that there are not too many points in $\mathbb{F}_q^2$. As such, it carries over to the situations where $|X|$ is small in $\mathbb{F}_q^3$.

Theorem 3.2.1. *If the lines are required to lie in $\mathbb{F}_q^2$, we must have $|Y| \gtrsim |X|$.*

Proof. Let L denote the number of lines in X, which is the same as the number of lines in Y.

Case 1: $L \geq \frac{q}{2}$. First $|X| \leq q^2$, as there are only q^2 points. Take $\frac{q}{2}$ points from S_1. Take $\frac{q}{2}-1$ points from S_2 not already taken. Take $\frac{q}{2}-3$ points from S_3 not already taken, etc. We get that

$$|Y| \geq \frac{q}{2} + (\frac{q}{2}-1) + \cdots + 1 = \frac{1}{8}(q^2+2q).$$

Thus, $|Y| \geq \frac{1}{8}|X|$.

Case 2: $L < \frac{q}{2}$. This is similar to the above, except we terminate taking points from lines after L steps. We have that $|X| \leq qL$, and

$$|Y| \geq \frac{q}{2} + (\frac{q}{2}-1) + \cdots + (\frac{q}{2}-(L-1)) = \frac{qL}{2} - \frac{(L-1)L}{2} \geq \frac{qL}{2} - \frac{qL}{4} = \frac{qL}{4}.$$

Thus, $|Y| \geq \frac{1}{4}|X|$. □

Let us view the essential components of the above proof.

(1) Two lines intersect in at most one point.
(2) Each line of Y has at least $\frac{q}{2}$ points.
(3) $|X| \leq q^2$.
(4) Each line of X has q points.

The proof thus carries over whenever $|X| \leq cq^2$ in $\mathbb{F}_q^3$. We then get

Theorem 3.2.2. *If $|X| \leq cq^2$ for some constant c, then $c|Y| \gtrsim |X|$.*

3.3. **Degree Reduction.** We use the method of degree reduction and then apply the polynomial method. In degree reduction, the basic idea is to use the structure of the problem to get a better upper bound on the degree of a polynomial vanishing on the desired set. See [4, Lecture 12] for some examples of degree reduction.

In our problem, we have many lines, let us call this set $\mathcal{L}$. The idea is to find a small subset $\mathcal{L}_1$ of these lines which have a lot of incidences with the other lines. Then, we use the probabilistic method to find a subset $\mathcal{L}_2$ of $\mathcal{L}_1$ with the property that if a polynomial vanishes on all the lines of $\mathcal{L}_2$, then it must vanish on all the lines in $\mathcal{L}-\mathcal{L}_1$. We then use this polynomial to estimate the total number of points in $\mathcal{L}-\mathcal{L}_1$, and use the fact that $\mathcal{L}_1$ is "small" to get the rest.

Proposition 3.3.1. *There exists a constant c such that if $|Y| \leq \frac{q^2}{\log q}$ or $|Y| \geq q^2 \log q$, then $|X| \leq c|Y|$. Otherwise, there is a polynomial P with degree at most $c \log q$ vanishing on all but at most $\frac{4|Y|}{q}$ lines of Y.*

Proof. (Follows [4, Lecture 12, Proposition 2.1]) Let $\mathcal{L}$ denote the lines of Y (which are the same as the lines of X.) Choose an ordering for the lines in $\mathcal{L}$, so we have a list $\ell_1, \cdots, \ell_n$. We make a subset $\mathcal{L}_1$ of $\mathcal{L}$ as follows: first we put ℓ_1 in $\mathcal{L}_1$. Then, for each ℓ_i, we put ℓ_i into $\mathcal{L}_1$ if and only if it has at least $\frac{q}{4}$ points (in Y) not already in $\mathcal{L}_1$. Applying this procedure for each of $i = 2, \cdots, n$, we get a subset $\mathcal{L}_1$ of $\mathcal{L}$. As we add a line into $\mathcal{L}_1$ only if it contains at least $\frac{q}{4}$ points not already in $\mathcal{L}_1$, we get that $\frac{q}{4}|\mathcal{L}_1| \leq |Y|$, so $|\mathcal{L}_1| \leq \frac{4}{q}|Y|$. Moreover, every line not in $\mathcal{L}_1$ has at least $\frac{q}{4}$ points in $\mathcal{L}_1$.

Take d to be the smallest integer such that

$$d^2 \geq (16 \cdot 20 \cdot 10^4)^2 \cdot \max(1, \frac{|Y| \log q}{q^2}, \frac{|Y|^2}{q^4}).$$

For simplicity of notation, we let C denote the constant $16 \cdot 20 \cdot 10^4$. Note that $1 \leq \frac{|Y| \log q}{q^2}$ when $|Y| \geq \frac{q^2}{\log q}$, that $1 \leq \frac{|Y|^2}{q^4}$ when $|Y| \geq q^2$, and that $\frac{|Y| \log q}{q^2} \leq \frac{|Y|^2}{q^4}$ when $|Y| \geq q^2 \log q$. Thus, we can take

$$d \approx \begin{cases} C & \text{if } |Y| \leq \frac{q^2}{\log q}, \\ C \cdot \frac{|Y|^{\frac{1}{2}} (\log q)^{\frac{1}{2}}}{q} & \text{if } \frac{q^2}{\log q} \leq |Y| \leq q^2 \log q \\ C \cdot \frac{|Y|}{q^2} & \text{if } |Y| \geq q^2 \log q. \end{cases}$$

Now we use the probabilistic method. Choose a subset $\mathcal{L}_2$ of $\mathcal{L}_1$ by independently choosing each line in $\mathcal{L}_1$ with probability $p = \frac{1}{20} \frac{d^2}{|\mathcal{L}_1|}$.

First, we show that the proposition is true whenever this is not possible, i.e. $20|\mathcal{L}_1| \leq d^2$.

If $|\mathcal{L}_1| \leq \frac{1}{20} C^2$, then there is a polynomial P of degree at most $\widetilde{d} = \frac{3}{2\sqrt{5}} C$ vanishing on $\mathcal{L}_1$. As $\frac{q}{4}$ is asymptotically larger than the constant $\widetilde{d}$, we may assume that for each line in $\mathcal{L} - \mathcal{L}_1$, the polynomial vanishes on more than $\widetilde{d}$ points, so it vanishes on the whole line. Thus, the polynomial P vanishes on every line in $\mathcal{L}$. This implies that $|X| \leq \widetilde{d} q^2$, so a bound of the form $|X| \leq c|Y|$ is given by the same argument as in the 2 dimensional case (see Theorem 3.2.2).

If $|\mathcal{L}_1| \leq \frac{1}{20} C^2 \cdot \frac{|Y|^2}{q^4}$, then there exists a polynomial of degree at most $\frac{3C}{2\sqrt{5}} \frac{|Y|}{q^2}$ vanishing on all lines $\mathcal{L}_1$. If $\frac{q}{4} \geq \frac{3C}{2\sqrt{5}} \frac{|Y|}{q^2}$, then the polynomial must vanish on all

lines of $\mathcal{L}$. Thus, by Theorem 2.2.1, the number of points in X is at most $\frac{3C}{2\sqrt{5}}|Y|$. If $\frac{q}{4} \leq \frac{3C}{2\sqrt{5}}\frac{|Y|}{q^2}$, then $\frac{6C}{\sqrt{5}}|Y| \geq q^3$. Since $|X| \leq q^3$, we also have a bound in this case.

Now suppose that $|\mathcal{L}_1| \leq \frac{1}{20}C^2 \cdot \frac{|Y|\log q}{q^2}$. In this situation, we may assume that $\frac{q^2}{\log q} \leq |Y| \leq q^2 \log q$. There is a polynomial of degree $\frac{3C}{2\sqrt{5}}\frac{|Y|^{\frac{1}{2}}(\log q)^{\frac{1}{2}}}{q} \leq \frac{3C}{2\sqrt{5}}\log q$ vanishing on all of $\mathcal{L}_1$. Now, for all sufficiently large q, we have that $\frac{q}{4} \geq \frac{3C}{2\sqrt{5}}\log q$, so this polynomial vanishes on every line in $\mathcal{L}$, and the proposition holds.

We may now assume that $p = \frac{1}{20}\frac{d^2}{|\mathcal{L}_1|} < 1$, so that we may choose a subset $\mathcal{L}_2$ of $\mathcal{L}_1$ by choosing each line independently with probability p. The probability that there are more than $\frac{1}{10}d^2$ lines in $\mathcal{L}_2$ is at most $\exp(-\frac{1}{100}\cdot\frac{1}{20}\cdot d^2) \leq \exp(-\frac{1}{100}\cdot\frac{1}{20}\cdot C^2) = \exp(-\frac{1}{2000}\cdot(16\cdot 20\cdot 10^4)^2) \leq \exp(-10^7)$. As a polynomial of degree d vanishes on any $\frac{1}{10}d^2$ lines (Theorem 2.1.3), the probability that there is not a polynomial of degree d vanishing on $\mathcal{L}_2$ is at most $\exp(-10^7)$.

Each line ℓ in $\mathcal{L} - \mathcal{L}_1$ intersects $\mathcal{L}_1$ in at least $\frac{q}{4}$ points. Recalling that $|\mathcal{L}_1| \leq \frac{4}{q}|Y|$ and that $d \leq C\frac{|Y|^2}{q^2}$, the expected number of points on ℓ in $\mathcal{L}_2$ is then at least $\frac{qp}{4} = \frac{qd^2}{4\cdot 20|\mathcal{L}_1|} \geq \frac{q^2d^2}{16\cdot 20|Y|} \geq \frac{Cd}{16\cdot 20} = 10^4 d$. Recalling that $d^2 \leq \frac{|Y|\log q}{q^2}$, the probability that there are less than d points of $\mathcal{L}_2$ on ℓ is at most $\exp(-\frac{1}{100}\cdot\frac{qp}{4}) \leq \exp(-\frac{1}{100}\cdot\frac{q^2d^2}{16\cdot 20|Y|}) \leq \exp(-\frac{1}{100\cdot 16\cdot 20}C^2\log q) \leq q^{-10^7}$.

Now there are $q^2(q^2+q+1)$ total lines in $\mathbb{F}_q^3$. Thus, the probability that we can find an $\mathcal{L}_2$ such that there is a polynomial of degree d vanishing on $\mathcal{L}_2$ and such that every line in $\mathcal{L} - \mathcal{L}_1$ has more than d points of $\mathcal{L}_2$ on it is at least $1 - \exp(-10^7) - q^{-10^7+5} > 0$. Because such an event happens with nonzero probability, there exists a subset $\mathcal{L}_2$ of $\mathcal{L}_1$ with the desired properties. Thus, we can find a polynomial P with degree d vanishing on all lines of $\mathcal{L} - \mathcal{L}_1$.

We now finish the proof of the proposition.

If $|Y| \leq \frac{q^2}{\log q}$, then $d \approx C$. As P vanishes on all of $\mathcal{L} - \mathcal{L}_1$, the number of points in $\mathcal{L} - \mathcal{L}_1$ is at most dq^2. By an argument similar to the 2 dimensional case, there is a constant c such that the number of points in $\mathcal{L} - \mathcal{L}_1$ is at most $c|Y|$. Because $|\mathcal{L}_1| \leq \frac{4}{q}|Y|$, there are at most $4|Y|$ points in $\mathcal{L}_1$. Thus, the proposition works in this case.

When $\frac{q^2}{\log q} \leq |Y| \leq q^2\log q$, there is a polynomial of degree $d \approx C\cdot\frac{|Y|^{\frac{1}{2}}(\log q)^{\frac{1}{2}}}{q} \leq C\log q$ vanishing on $\mathcal{L} - \mathcal{L}_1$. As $|\mathcal{L}_1| \leq \frac{4}{q}|Y|$, the proposition is also proved in this case.

When $|Y| \geq q^2\log q$, we have a polynomial of degree $d \approx C\cdot\frac{|Y|}{q^2}$ vanishing on $\mathcal{L} - \mathcal{L}_1$. Thus, by Theorem 2.2.1, there are at most $C|Y|$ points in $\mathcal{L} - \mathcal{L}_1$. Combining this with the fact that there are at most $4|Y|$ points in $\mathcal{L}_1$, we get the desired conclusion. □

4. Hasse Derivatives

We wish to extend certain notions, such as being flat, from the real case to the finite field case. These notions are defined in terms of higher order derivatives. The naive definition of these higher order derivatives in nonzero characteristic is to compose taking the first derivative the proper number of times. However, this definition is unsatisfactory for our purposes. For example, the derivative of x^p in a field of characteristic p is given by $px^{p-1} = 0$, so using derivatives in this manner

will never detect the difference between $f(x)$ and $f(x) + x^p$. Moreover, taking p successive derivatives will always result in 0, so very limited information can be gotten from knowing all the higher derivatives at a point. These problems are related to the following: given a formal power series in one variable $P(x)$ over $\mathbb{R}$, we can Taylor expand about x to get

$$P(x+t) = \sum_{n=0}^{\infty} \frac{1}{n!} P^{(n)}(x) t^n = P(x) + P'(x)t + \frac{1}{2!} P''(x)t^2 + \cdots .$$

In characteristic p, dividing by $p!$ and higher factorials is forbidden as $p = 0$ is not invertible. These issues are dealt with using the Hasse derivatives, which are essentially defined so that a Taylor-like formula holds. For a reference on Hasse derivatives, see [6, Section 5.10] or [3].

4.1. **Hasse Derivatives in one variable.** Let A be a commutative ring with 1, and consider the A-algebra of polynomials in one variable $A[x]$.

Definition 4.1.1 (Hasse Derivative)**.** We define a sequence of A-linear operators $d^{(0)}$, $d^{(1)}$, $d^{(2)}$, $\cdots$ from $A[x]$ to $A[x]$ by setting

$$d^{(k)} x^n = \binom{n}{k} x^{n-k},$$

and then extending to $A[x]$ by linearity. That is,

$$d^{(k)} \sum_{i=0}^{n} a_i x^i = \sum_{i=0}^{n} a_i \binom{i}{k} x^{i-k} = \sum_{i=0}^{n-k} a_{i+k} \binom{i+k}{k} x^i.$$

The operator $d^{(k)}$ is called the **k-th Hasse derivative**.

Remark 4.1.2. When the characteristic of A is 0, the k-th Hasse derivative is a nonzero multiple of the k-th derivative: $d^{(k)}P = \frac{1}{k!} P^{(k)}$. In any characteristic, we can see that $k! d^{(k)} P = P^{(k)}$. Most notably, the first Hasse derivative is the same as the first derivative.

Note that for $k > n$, we have $d^{(k)} x^n = 0$, so the k-th Hasse derivative vanishes on polynomials of degree less than k. (The converse is not true, as $d^{(1)} x^n = 0$ in characteristic n.)

Our main reason to use the Hasse derivative is because a Taylor-series like formula holds:

Proposition 4.1.3. *(Taylor Expansion) Let $P(x) \in A[x]$. Then*

$$P(x+t) = \sum_{k=0}^{\infty} d^{(k)} P(x) t^k.$$

Note that $d^{(k)} P$ vanishes for $k > \deg P$, so the sum on the right is finite.

Proof. We just expand. Let $P(x) = a_0 + a_1 x + \cdots + a_n x^n$. Then,

$$\begin{aligned} P(x+t) &= \sum_{i=0}^{n} a_i (x+t)^i \\ &= \sum_{i=0}^{n} \sum_{k=0}^{i} a_i \binom{i}{k} x^{i-k} t^k \\ &= \sum_{k=0}^{n} \sum_{i=k}^{n} a_i \binom{i}{k} x^{i-k} t^k \\ &= \sum_{k=0}^{n} d^{(k)} P(x) t^k. \end{aligned}$$

The conclusion follows since $d^{(k)} P = 0$ for $k > \deg P$. □

Corollary 4.1.4. *If, for some $x \in A$, we have $d^{(k)} P(x) = 0$ for all $k > n$, then $\deg P \leq n$.*

Proof. Write $P(t) = P(x + (t - x))$ and Taylor expand. □

Remark 4.1.5. Corollary 4.1.4 gives us a method for overcoming the previously alluded to fact that in nonzero characteristic n, the kernel of $d^{(1)}$ consist of more than just the constant functions. That is, we cannot use the fact that the first derivative is identically zero to conclude that the original was constant. Instead, to prove something is constant, we need to verify that $d^{(1)}$, $d^{(2)}$, $\cdots$ all vanish at some point.

For example, an argument to show that in characteristic 0 it suffices to check $d^{(1)} P = 0$ via Hasse derivatives could proceed as follows: if $d^{(1)} P = 0$, then $d^{(2)} P = \frac{1}{2} d^{(1)} d^{(1)} P = \frac{1}{2} d^{(1)} 0 = 0$. Similarly $d^{(3)} P = \frac{1}{3} d^{(1)} d^{(2)} P = \frac{1}{3} d^{(1)} 0 = 0$. We can inductively show that $d^{(n)} P = 0$ for all $n \geq 1$. Letting $x = 0$ (or any other element of A) in the Taylor expansion shows that P is constant.

Remark 4.1.6. The Taylor expansion formula (Proposition 4.1.3) shows that we could have alternatively defined the k-th Hasse derivative as

$$d^{(k)} P(x) = [t^k] P(x + t),$$

where $[t^k] : A[x][t] \to A[x]$ gives the coefficient of t^k in an element of the polynomial ring $A[x][t]$. This alternate definition is essentially saying: we define the Hasse derivative so that the Taylor expansion formula holds. We prefer using this alternate definition in our proofs, as they would otherwise be a sequence of unmotivated calculations.

Proposition 4.1.7 (Composition of Hasse derivatives)**.** *We have:*

(1) $d^{(k_1)} d^{(k_2)} P = \binom{k_1+k_2}{k_1} d^{(k_1+k_2)} P$, *or more generally,*
(2) $d^{(k_1)} \cdots d^{(k_n)} P = \binom{k_1+\cdots+k_n}{k_1,\cdots,k_n} d^{(k_1+\cdots+k_n)} P$.

Proof. We prove (1) using the alternate definition given in Remark 4.1.6. Then (2) follows from (1) by induction.

We have

$$d^{(k_1)} d^{(k_2)} P(x) = d^{(k_1)} [t^{k_2}] P(x+t) = [s^{k_1} t^{k_2}] P(x+t+s).$$

Now $P(x+t+s) = \sum d^{(k)}P(x)(t+s)^k$. The terms of this sum are homogeneous in s and t, so the only one which could have a coefficient for the monomial $s^{k_1}t^{k_2}$ is the one with $k = k_1 + k_2$. Then,

$$d^{(k_1)}d^{(k_2)}P(x) = [s^{k_1}t^{k_2}](t+s)^{k_1+k_2}d^{(k_1+k_2)}P(x) = \binom{k_1+k_2}{k_1,k_2}d^{(k_1+k_2)}P(x),$$

as desired. □

Corollary 4.1.8. *The operators $d^{(0)}$, $d^{(1)}$, $\cdots$ commute.*

Proposition 4.1.9 (Product Rule)**.** *We have:*

(1) $d^{(k)}(P_1P_2) = \sum_{k_1+k_2=k} d^{(k_1)}P_1 \cdot d^{(k_2)}P_2$, *or, more generally,*
(2) $d^{(k)}(P_1 \cdots P_n) = \sum_{k_1+\cdots+k_n=k} d^{(k_1)}P_1 \cdots d^{(k_n)}P_n$.

Here, the k_i are constrained to be nonnegative.

Proof. These statements are almost immediate using the alternate definition in Remark 4.1.6. Note that statement (2) follows from statement (1) by induction.

We get

$$\begin{aligned} d^{(k)}(P_1P_2)(x) &= [t^k](P_1(x+t)P_2(x+t)) \\ &= \sum_{k_1+k_2=k} ([t^{k_1}]P_1(x+t))([t^{k_2}]P_2(x+t)) \\ &= \sum_{k_1+k_2=k} d^{(k_1)}P_1(x) \cdot d^{(k_2)}P_2(x), \end{aligned}$$

which was what we wanted. □

Proposition 4.1.10 (Chain Rule)**.** *We have the following formula for calculating the Hasse derivatives for a composition of two polynomials:*

$$d^{(k)}(P \circ Q) = \sum_{\sum_{i=1}^k ia_i = k} \binom{a_1 + \cdots + a_k}{a_1, \cdots, a_k} ((d^{(a_1+\cdots+a_k)}P) \circ Q) \cdot \prod_{j=1}^{k} (d^{(j)}Q)^{a_j}.$$

In particular,

$$d^{(1)}(P \circ Q) = ((d^{(1)}P) \circ Q) \cdot d^{(1)}Q.$$

Proof. Again, we use the alternate definition found in Remark 4.1.6.

Using the Taylor expansion formula twice, we get

$$\begin{aligned} d^{(k)}(P \circ Q)(x) &= [t^k]P(Q(x+t)) \\ &= [t^k]P\left(\sum_{j=0}^{\infty} d^{(j)}Q(x)t^j\right) \\ &= [t^k]\sum_{i=0}^{\infty} d^{(i)}P(Q(x))\left(\sum_{j=1}^{\infty} d^{(j)}Q(x)t^j\right)^i. \end{aligned}$$

Let $\mathbb{N}^\infty$ denote the set of sequences of natural numbers $a = (a_1, a_2, \cdots)$ which are eventually zero. Fix an i, and consider all sequences $a_1, a_2, \cdots$ whose sum is equal

to i. The binomial theorem for power series gives

$$\left(\sum_{j=1}^{\infty} d^{(j)}Q(x)t^j\right)^i = \sum_{\substack{a\in\mathbb{N}^\infty \\ \sum a_n = i}} \binom{i}{a_1, a_2, \cdots} \prod_j (d^{(j)}Q(x)t^j)^{a_j}$$

$$= \sum_{\substack{a\in\mathbb{N}^\infty \\ \sum a_n = i}} \binom{i}{a_1, a_2, \cdots} \left(\prod_j (d^{(j)}Q(x))^{a_j}\right) t^{\sum na_n}.$$

(We are ultimately allowed to use such a binomial theorem because for any given m, there are only finitely many sequences $a_1, a_2, \cdots$ in $\mathbb{N}^\infty$ with $\sum na_n = m$ and $\sum a_n = i$) Now, each sequence in $\mathbb{N}^\infty$ has a fixed sum. We can then rewrite the above series to get

$$d^{(k)}(P\circ Q)(x) = [t^k] \sum_{a\in\mathbb{N}^\infty} \binom{\sum a_n}{a_1, a_2, \cdots} (d^{(\sum a_n)}P)\circ Q(x) \left(\prod_j (d^{(j)}Q(x))^{a_j}\right) t^{\sum na_n}.$$

(We are allowed to do this, as for any given m, there are only finitely many sequences with $\sum na_n = m$.) Taking the coefficient of t^k then gives us the desired formula. □

Corollary 4.1.11. *Applying the chain rule to $P(ax+b)$ gives*

$$d^{(k)}(P(ax+b)) = a^k(d^{(k)}P)(ax+b).$$

4.2. **Hasse Derivatives in other Algebras.** Having done much of single variable differential calculus, we now move on to multiple variables and to power series. In an attempt to streamline the argument, we shall work somewhat abstractly at first so as to get partial derivatives, directional derivatives, total derivatives, etc. all in one go.

Definition 4.2.1. Let A be a ring with 1. Let B and R be A-algebras, and let $R[[t]]$ denote the formal power series in one variable t over R. For any homomorphism of A-algebras $\phi : B \to R[[t]]$ (so that $\phi(1) = 1$), we may define the Hasse derivatives $D^{(k)} : B \to R$ as

$$D^{(k)}b = [t^k]\phi(b).$$

Example 4.2.2 (Polynomials in One Variable)**.** For the Hasse derivatives in the polynomials of one variable, we have $B = R = A[x]$, and the homomorphism $\phi : A[x] \to A[x][[t]]$ is given by sending $x \mapsto x + t$.

Actually, it is sometimes better to view $B = R = A[x, dx]$, and then take $\phi : A[x, dx] \to A[x, dx][[t]]$ by sending $x \mapsto x + t\,dx$ and $dx \mapsto dx$. In this situation dx is just the name of a variable, and we will use the shorthand $dx^i := (dx)^i$. Then, our derivative on monomials becomes $D^{(k)}(x^n) = \binom{n}{k}x^{n-k}\,dx^k$. This slight change in notation becomes more important when we work in several variables and want to take a total derivative. In that situation, we want to be able to keep track of what happens in each of the different variables.

Proposition 4.2.3. *We have the following*

1. *The maps $D^{(k)} : B \to R$ are linear in A.*
2. *The map $D^{(0)} : B \to R$ is a homomorphism of A-algebras.*
3. *For $k = 1, 2, \cdots$, the maps $D^{(k)} : B \to R$ are zero on $A \subset B$, that is, $D^{(k)}(a) = 0$.*

(4) *(Product Rule)* $D^{(k)}(b_1b_2) = \sum_{k_1+k_2=k}(D^{(k_1)}b_1)(D^{(k_2)}b_2)$.

Proof. The first three items follow readily from the definition. Statement (4) follows just like in the one variable case:

$$\begin{aligned} D^{(k)}(b_1b_2) &= [t^k]\phi(b_1b_2) \\ &= [t^k]\phi(b_1)\phi(b_2) \\ &= \sum_{k_1+k_2=k}([t^{k_1}]\phi(b_1))([t^{k_2}]\phi(b_2)) \\ &= \sum_{k_1+k_2=k}(D^{(k_1)}b_1)(D^{(k_2)}b_2), \end{aligned}$$

thus giving us the product rule. □

Remark 4.2.4. A sequence of maps $D^{(0)}, D^{(1)}, \cdots$ from $B \to R$ satisfying statements (1)-(4) in Proposition 4.2.3 are called higher derivations from B to R over A of order ∞ (see [7, Definition 1.1]). It can be shown that higher derivations of order ∞ are in bijection with A-algebra homomorphisms $B \to R[[t]]$. (One way has been given by Proposition 4.2.3. To go the other way just define the map ϕ by $\phi(b) = \sum_{k=0}^{\infty} D^{(k)}b\,t^k$ and check it is an A-algebra homomorphism. See [7, Lemma 1.7] for details.) This guarantees that any sequence of derivatives we might define satisfying th product rule can be gotten from a map into a formal power series.

Proposition 4.2.5 (Expansion)**.** *We have:* $\phi(b) = \sum_{k=0}^{\infty} D^{(k)}(b)t^k$.

Definition 4.2.6 (Derivatives of power series in one variable)**.** Consider $B = R = A[[x]][dx]$, and let $\phi : B \to R[[t]]$ be given by the maps $x \mapsto x + t\,dx$ and $dx \mapsto dx$. (Note that $x+t\,dx \in (x,t)$ and $A[[x]][dx][[t]]$ is complete with respect to the topology generated by the ideal (x,t), so we may do this.) The resulting maps $d^{(k)}$ are the Hasse derivatives for power series.

Definition 4.2.7 (Partial Derivatives)**.** Let

$$B = R = A[[x_1, \cdots, x_n]][dx_1, \cdots, dx_n].$$

We define the partial derivatives $\partial_{x_i}^{(k)}$ with respect to the variable x_i using the map $\phi_i : B \to R[[t]]$ sending $x_i \mapsto x_i + t\,dx_i$ and fixing all other variables.

Definition 4.2.8 (Directional Derivatives)**.** Let

$$B = R = A[[x_1, \cdots, x_n]][du].$$

Let $u = (u_1, \cdots, u_n)$ be an ordered n-tuple of elements in A. We define the directional derivatives ∂_u^k in the direction of u to be the Hasse derivatives gotten from the map sending $x_i \mapsto x_i + u_i t\,du$ and $du \mapsto du$.

Definition 4.2.9 (Total Derivatives)**.** Let

$$B = R = A[[x_1, \cdots, x_n]][dx_1, \cdots, dx_n].$$

We define the total derivatives $D^{(k)}$ by using the map sending $x_i \mapsto x_i + t\,dx_i$ and $dx_i \mapsto dx_i$.

Remark 4.2.10. Looking at the total derivative case, we think of dx_i to be a small change in the x_i direction. Sending $x_i \mapsto x_i + t\,dx_i$ then changes the x_i coordinate slightly in that direction. The interpretation of t is as an algebraic object allowing us to count the degree of the newly introduced changes, and then applying $[t^k]$ gives us those changes of degree k.

Definition 4.2.11 (Derivatives in Multiple Dimensions)**.** Let B and R be the m-fold product of $A[[x_1, \cdots, x_n]][dx_1, \cdots, dx_n]$ with itself:

$$B = R = \prod_{i=1}^{m} A[[x_1, \cdots, x_n]][dx_1, \cdots, dx_n].$$

Note that $R[[t]] \cong \prod A[[x_1, \cdots, x_n]][dx_1, \cdots, dx_n][[t]]$ when we let $t = (t, \cdots, t)$. To get the partial derivatives $\partial_{x_i}^{(k)}$, use the map from B to $R[[t]]$ given by taking the n-fold product of the map ϕ_i in Definition 4.2.7 with itself. To get the total derivatives $D^{(k)}$, use the map which is the n-fold product of the map ϕ in Definition 4.2.9 with itself.

Next, we wish to show various facts about what happens when you take two derivatives in succession. For example, in the one variable case, we had the relation $d^{(k_1)}d^{(k_2)}P = \binom{k_1+k_2}{k_1} d^{(k_1+k_2)}P$. Recalling our method of proving this in Theorem 4.1.7, we see that we made use of two variables s and t. This suggests that the proofs will require us to "adjoin more than one variable". The following definition allows us to make such symbolic manipulations easier.

Definition 4.2.12. Let B be an A-algebra. Suppose we are given a homomorphism $\phi : B \to B[[t]]$ of A-algebras. Let C be a power series over B (in a finite number of variables).

Note that if we are given a homomorphism $\rho : R \to S$ of A-algebras, we can get a homomorphism $R[[t]] \to S[[t]]$ by first mapping $R \xrightarrow{\rho} S \hookrightarrow S[[t]]$ and then extending this to $R[[t]]$ by sending $t \mapsto t$. Thus, we can extend the map $\phi : B \to B[[t]]$ to a map $\phi : C \to C[[t]]$.

For an element $s \in C$ of order at least 1, let $\phi^s : C \to C$ denote the composition map of first applying $\phi : C \to C[[t]]$, and then sending t to s.

Remark 4.2.13. This is a rather technical definition, which is probably easier to explain in words. Basically, we have several maps $\phi_1, \phi_2 : B \to B[[t]]$. We can map an element $b \in B$ to a power series in the variable t using ϕ_1. Now we want to apply ϕ_2 to this power series, by applying ϕ_2 to all the coefficients (and not act on t). In the definition, the extension of ϕ from a map $B \to B[[t]]$ to a map $C \to C[[t]]$ serves the purpose of lining up the domains so we can do this (that is, use ϕ_2 on a power series in B without affecting the variables).

Now when we use ϕ_2 on an element of $B[[t]]$, we have two choices: either use a new variable s (so that ϕ_2 acts on the coefficients as a map $B \to B[[s]]$), or use the old variable t. The superscript notation ϕ_2^s and ϕ_2^t allows us to specify which choice we made.

Proposition 4.2.14. *Let ϕ_1 and ϕ_2 be A-algebra homomorphisms from B to $B[[t]]$ and $D_1^{(k)}$ and $D_2^{(k)}$ be the associated Hasse derivatives. Then*

$$D_2^{(k_2)} D_1^{(k_1)} b = [s^{k_2} t^{k_1}] \phi_2^s \circ \phi_1^t(b).$$

Proof. First,

$$D_2^{(k_2)} D_1^{(k_1)} b = D_2^{(k_2)} [t^{k_1}] \phi_1^t(b) = [s^{k_2}] \phi_2^s([t^{k_1}] \phi_1^t(b)).$$

Now note that $\phi_2^s(t^m) = t^m$ for all m, so $\phi_2^s[t^{k_1}] = [t^{k_1}]\phi_2^s$. Thus,

$$D_2^{(k_2)} D_1^{(k_1)} b = [s^{k_2} t^{k_1}] \phi_2^s \circ \phi_1^t(b),$$

as desired. □

Proposition 4.2.15. *Let $D_1^{(k)}$ and $D_2^{(k)}$ be the Hasse derivatives associated with the maps ϕ_1 and ϕ_2 from B to $B[[t]]$. We have*

(1) *If $\phi_1^s \circ \phi_1^t = \phi_1^{s+t}$, then the derivative $D_1^{(k)}$ has a composition law:*

$$D_1^{(k_1)} D_1^{(k_2)} = \binom{k_1+k_2}{k_1} D_1^{(k_1+k_2)}.$$

(2) *If $\phi_1^s \circ \phi_2^t = \phi_2^t \circ \phi_1^s$, then the Hasse derivatives commute:*

$$D_1^{(k_1)} D_2^{(k_2)} = D_2^{(k_2)} D_1^{(k_1)}.$$

Proof. For (1), we have,

$$D_1^{(k_1)} D_1^{(k_2)} b = [s^{k_1} t^{k_2}] \phi_1^s \circ \phi_1^t b = [s^{k_1} t^{k_2}] \phi_1^{s+t} b.$$

By the expansion formula, Proposition 4.2.5, we then get

$$[s^{k_1} t^{k_2}] \phi_1^{s+t} b = [s^{k_1} t^{k_2}] \sum_{i=0}^{\infty} D_1^{(i)} b \, (s+t)^i = \binom{k_1+k_2}{k_1} D_1^{(k_1+k_2)} b.$$

For (2), we get

$$D_1^{(k_1)} D_2^{(k_2)} b = [s^{k_1} t^{k_2}] \phi_1^s \circ \phi_2^t(b) = [s^{k_1} t^{k_2}] \phi_2^t \circ \phi_1^s(b) = D_2^{(k_2)} D_1^{(k_1)} b,$$

as desired. □

Corollary 4.2.16. *Applying these to the previously defined partial, directional, and total derivatives, we get*

(1) *Each set of partial derivatives $\partial_{x_i}^{(k)}$, each set of directional derivatives $\partial_u^{(k)}$, and the total derivatives $D^{(k)}$ satisfy the composition law given in Proposition 4.2.15.*
(2) *The partial derivatives $\partial_{x_i}^{(k)}$ and total derivatives $D^{(k)}$ all commute with each other.*

Proposition 4.2.17. *Let $D^{(k)}$, $\partial_1^{(k)}$, $\cdots$, $\partial_n^{(k)}$ be the Hasse derivatives associated with the maps $\phi, \phi_1, \cdots, \phi_n$ from B to $B[[t]]$. If $\phi^t = \phi_1^t \circ \cdots \circ \phi_n^t$, then*

$$D^{(k)} = \sum_{k_1+\cdots+k_n=k} \partial_1^{(k_1)} \cdots \partial_n^{(k_n)}.$$

Proof. We have

$$D^{(k)} b = [t^k] \phi^t(b) = [t^k] \phi_1^t \circ \cdots \circ \phi_n^t(b).$$

We note that $\phi_1^t \circ \cdots \circ \phi_n^t$ is the same as the composition of $\phi_1^{t_1} \circ \cdots \circ \phi_n^{t_n}$ with the map sending $t_i \mapsto t$. The monomials in $t_1, \cdots, t_n$ mapping to t^k are precisely those $t_1^{k_1} \cdots t_n^{k_n}$ with $k_1 + \cdots + k_n = k$. Thus,

$$D^{(k)} b = \sum_{k_1+\cdots+k_n=k} [t_1^{k_1} \cdots t_n^{k_n}] \phi_1^{t_1} \circ \cdots \circ \phi_n^{t_n}(b) = \sum_{k_1+\cdots+k_n=k} \partial_1^{(k_1)} \cdots \partial_n^{(k_n)}(b),$$

which was what we wanted. □

Corollary 4.2.18. *The total derivative can be written as a sum of partials:*

$$D^{(k)} = \sum_{k_1+\cdots+k_n=k} \partial_{x_1}^{(k_1)} \cdots \partial_{x_n}^{(k_n)}.$$

Remark 4.2.19. In fact, the partial derivative is of the form

$$\partial_{x_1}^{(k_1)} \cdots \partial_{x_n}^{(k_n)} P = P_{k_1 \cdots k_n} dx_1^{k_1} \cdots dx_n^{k_n},$$

for some element $P_{k_1 \cdots k_n} \in A[[x_1, \cdots, x_n]]$. Note that this is the only term in $D^{(k)}P$ with a nonzero $dx_1^{k_1} \cdots dx_n^{k_n}$ coefficient. Thus, if $D^{(k)}P = 0$, then $\partial_{x_1}^{(k_1)} \cdots \partial_{x_n}^{(k_n)} P = 0$ for all $k_1 + \cdots + k_n = k$.

4.3. **Prime Characteristic.** We now prove some statements which are only true in a prime characteristic p.

Lemma 4.3.1. *Suppose A has prime characteristic p, and $D^{(k)} : B \to R$ be any set of Hasse derivatives satisfying the composition rule $D^{(k_1)}D^{(k_2)} = \binom{k_1+k_2}{k_1} D^{(k_1+k_2)}$. Let m be a nonnegative integer, and $\overline{m_j \cdots m_0}$ denote the base p representation of m, so that $m = m_0 + m_1 p + \cdots + m_j p^j$ and $0 \leq m_i < p$. Then, $D^{(m)} = c_m D^{(m_j p^j)} \cdots D^{(m_1 p)} D^{(m_0)}$, where c_m is a nonzero constant in $\mathbb{Z}/p\mathbb{Z}$.*

Proof. As $D^{(m_j p^j)} \cdots D^{(m_0)} = \binom{m}{m_j p^j, \cdots, m_0} D^{(m)}$, it suffices to show that the number $\binom{m}{m_j p^j, \cdots, m_0}$ is not divisible by p. Let $v_p(m)$ denote the number of times the factor p appears in the prime factorization of a positive integer m. Let $\sigma_p(m) = m_0 + \cdots + m_j$ denote the sum of the digits in the base p representation of m. We see that

$$\begin{aligned} v_p(m!) &= \left\lfloor \frac{m}{p} \right\rfloor + \left\lfloor \frac{m}{p^2} \right\rfloor + \cdots \\ &= (m_1 + m_2 p + \cdots + m_j p^{j-1}) + (m_2 + \cdots + m_j p^{j-2}) + \cdots + (m_j) \\ &= m_1 + m_2(p+1) + \cdots + m_j(p^{j-1} + \cdots + 1) \\ &= \frac{1}{p-1}(m_0(1-1) + m_1(p-1) + m_2(p^2-1) + \cdots + m_j(p^j - 1)) \\ &= \frac{m - \sigma_p(m)}{p-1}. \end{aligned}$$

Then, it is easy to see

$$v_p(m!) = \frac{m - \sigma_p(m)}{p-1} = \sum_{i=0}^{j} \frac{m_i p^i - m_i}{p-1} = \sum_{i=0}^{j} v_p((m_i p^i)!) = v_p(\prod_{i=0}^{j} (m_i p^i)!),$$

thus implying $\binom{m}{m_j p^j, \cdots, m_0}$ is invertible in A. □

Lemma 4.3.2. *Suppose A has prime characteristic p, and let $D^{(k)} : B \to R$ be any set of Hasse derivatives satisfying the rule $D^{(k_1)}D^{(k_2)} = \binom{k_1+k_2}{k_1} D^{(k_1+k_2)}$. Let m and n be nonnegative integers with base p representations given by $m = \sum_{i=0}^{\infty} m_i p^i$ and $n = \sum_{i=0}^{\infty} n_i p^i$. If $D^{(m)}b = 0$, and $n_i \geq m_i$ for all i, then $D^{(n)}b = 0$.*

Proof. Let n_j denote the largest nonzero place of n in its base p expansion. Using the composition rule, we get

$$\prod_{i=0}^{j} \binom{n_i}{m_i} D^{(m)}b = D^{(n_j - m_j)} \cdots D^{(n_0 - m_0)} D^{(m)} b = 0.$$

The lemma follows when you note $n_i < p$, so $\binom{n_i}{m_i}$ is invertible for all i. □

Corollary 4.3.3. *Let $\mathbb{F}$ be a field with characteristic p, and let $f \in \mathbb{F}[[x, y]]$. Let $D^{(k)}$ denote the total derivatives in two variables. We have*

- *If $D^{(2)}f = 0$, then $D^{(m)}f = 0$ for all $m \not\equiv 0, 1 \mod p$. Moreover,*
$$D^{(p)}f = f_{p,0}(x,y)dx^p + f_{0,p}(x,y)dy^p$$
for some $f_{p,0}$ and $f_{0,p}$ in $\mathbb{F}[[x,y]]$. (Note, this is tautologically true in characteristic 2.)
- *If $D^{(2)}f = D^{(p)}f = \cdots = D^{(p^{k-1})}f = 0$, then $D^{(m)}f = 0$ for all $m \not\equiv 0, 1 \mod p^k$. Moreover,*
$$D^{(p^k)}f = f_{p^k,0}(x,y)(dx)^{p^k} + f_{0,p^k}(x,y)(dy)^{p^k}$$
for some $f_{p^k,0}$ and f_{0,p^k} in $\mathbb{F}[[x,y]]$.

Proof. The statements about the total derivative $D^{(m)}$ follow directly from Lemma 4.3.2. For the partial derivatives, note that $D^{(2)}f = \cdots = D^{(p^{k-1})}f = 0$ implies that $\partial_x^{(2)}f = \cdots = \partial_x^{(p^{k-1})}f = 0$ and $\partial_y^{(2)}f = \cdots = \partial_y^{(p^{k-1})}f = 0$, so $\partial_x^{(m)}f = \partial_y^{(m)}f = 0$ for $m \not\equiv 0, 1 \mod p^k$. □

5. Flat Points and Lines through them

In this section we carry over the notions critical points, flat points, et cetera in $\mathbb{R}^3$ over to the finite field case.

5.1. **Critical and Flat Points.** Let $\mathbb{F}$ be a field, and P be a polynomial in $\mathbb{F}[x,y,z]$ of degree d. Using the expansion formula for the total derivative, we may expand out $P(x+dx, y+dy, z+dz)$ as

$$\begin{aligned} P(x+dx,y+dy,z+dz) = D^{(0)}P(x,y,z) + D^{(1)}P(x,y,z,dx,dy,dz) \\ +D^{(2)}P(x,y,z,dx,dy,dz) \\ +\cdots + D^{(d)}P(x,y,z,dx,dy,dz), \end{aligned}$$

where $D^{(i)}P$ is a homogeneous polynomial of degree i in the variables dx, dy, dz.

Given any $\mathbb{F}$-algebra, A, and an element (a,b,c) in A^3, we let $D^{(i)}P(a,b,c)$ denote the homogeneous polynomial of degree i in dx, dy, dz inside the ring $A[dx,dy,dz]$ gotten by sending $x \mapsto a$, $y \mapsto b$ and $z \mapsto c$. We shall make the following definitions:

Definition 5.1.1. We say (a,b,c) is a **zero** of P if
$$D^{(0)}P(a,b,c) = P(a,b,c) = 0,$$
as an element of A.

Definition 5.1.2. We say (a,b,c) is a **critical point** of P if
$$D^{(1)}P(a,b,c) = P_x(a,b,c)dx + P_y(a,b,c)dy + P_z(a,b,c)dz = 0,$$
as an element of $A[dx,dy,dz]$, i.e. $P_x(a,b,c) = P_y(a,b,c) = P_z(a,b,c) = 0$.

Definition 5.1.3. We say (a,b,c) is a **flat point** of P if (a,b,c) is not a critical point, and $D^{(1)}P(a,b,c)$ divides $D^{(2)}P(a,b,c)$ in $A[dx,dy,dz]$. More generally, we say that (a,b,c) is an **n-flat point** of P if (a,b,c) is not critical, and
$$D^{(1)}P(a,b,c) | D^{(n)}P(a,b,c)$$
in $A[dx,dy,dz]$.

In the case of $F = A = \mathbb{R}$, there was a set of special polynomials SP given by

$$\partial_{e_i \times \nabla P(x,y,z)} \nabla P(x,y,z) \times \nabla P(x,y,z),$$

which vanishes if and only if (x, y, z) was a critical or a flat point in $\mathbb{R}^3$. Here, the e_i range over the three unit vectors in the directions of the x-, y-, and z-axes. We do not treat why these particular polynomials work here; see [4, Lecture 14, Section 4] for details. We do, however, show that there is a similar statement in our situation: that there are sets of polynomials SP^n which "detect" whether or not a point is n-flat.

Theorem 5.1.4. *Suppose $\mathbb{F}$ is a field, and $\mathbb{K}$ is a field extension of $\mathbb{F}$. Let P be a polynomial in $\mathbb{F}[x, y, z]$ of degree d. For every n, there exist a set SP^n of polynomials in $\mathbb{F}[x, y, z]$ such that $SP^n(a, b, c) = 0$ if and only if $(a, b, c) \in \mathbb{K}^3$ is either a critical point or a n-flat point. Moreover, each polynomial in SP^n has degree at most $(n + 1)d - 2n$.*

Definition 5.1.5. We shall call the set of polynomials SP^n the **n-special polynomials**.

To prove Theorem 5.1.4, note that our definition of n-flat is the condition that a linear polynomial $D^{(1)}P$ divides some higher degree polynomial $D^{(n)}P$. This suggests the use of the resultant. Given two polynomials f, g, the resultant $r(f, g)$ is essentially a polynomial in the coefficients of f and g which vanishes when f and g have a common factor. We develop enough of this theory to prove the desired statements. Chapter 2 of [6] gives some elimination theory. Our statements here slightly modify the arguments in [1] to get certain statements for integral domains.

Lemma 5.1.6. *Let k be a field. Let $f = a_1x + a_0$ be a linear polynomial, and $g = c_nx^n + \cdots + c_0$ be a polynomial of degree n in $k[x]$. There is a polynomial r with coefficients in $\mathbb{Z}$ such that $r(a_1, a_0, c_n, \cdots, c_0) = 0$ if and only if one of the following is true:*

1. *$f = 0$, (i.e. $a_1 = a_0 = 0$).*
2. *f divides g.*

Moreover, the polynomial r is homogeneous of degree n over a_1 and a_0 and homogeneous of degree 1 over c_n, $\cdots$, c_0.

Definition 5.1.7. We shall call this polynomial r the **resultant** of f and g.

Proof of Lemma 5.1.6. Let r be given by the determinant of the matrix

$$S = \begin{bmatrix} a_1 & 0 & 0 & \cdots & 0 & c_n \\ a_0 & a_1 & 0 & \cdots & 0 & c_{n-1} \\ 0 & a_0 & a_1 & \cdots & 0 & c_{n-2} \\ 0 & 0 & a_0 & \cdots & 0 & c_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_1 & c_1 \\ 0 & 0 & 0 & \cdots & a_0 & c_0 \end{bmatrix}.$$

(This matrix is the transpose of the Sylvester matrix of f and g.)

First, if $a_1 = a_0 = 0$, then the determinant is clearly zero. Next, if f divides g, then there is a polynomial, $h = b_{n-1}x^{n-1} + \cdots + b_0$ such that $f \cdot h = g$. Multiplying

out and equating coefficients, this implies that

$$\begin{bmatrix} a_1 & 0 & 0 & \cdots & 0 & c_n \\ a_0 & a_1 & 0 & \cdots & 0 & c_{n-1} \\ 0 & a_0 & a_1 & \cdots & 0 & c_{n-2} \\ 0 & 0 & a_0 & \cdots & 0 & c_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_1 & c_1 \\ 0 & 0 & 0 & \cdots & a_0 & c_0 \end{bmatrix} \begin{bmatrix} b_{n-1} \\ b_{n-2} \\ b_{n-3} \\ b_{n-4} \\ \vdots \\ b_0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.$$

Thus, there is a nontrivial solution to the linear equation $S\vec{v} = 0$, so the determinant of S is zero.

Next, suppose $r = 0$. Then, there is a nontrivial solution $\vec{v}_0$ to $S\vec{v} = 0$. If the last component of $\vec{v}_0$ is not zero, we can scale it so that the last component is -1. Then we get a linear equation as above, and we can take $b_{n-1}, \cdots, b_0$ to be the first n components of $\vec{v}_0$ to get a $(b_{n-1}x^{n-1} + \cdots + b_0)f = g$. If the last component were zero, then $a_1 = 0$, since otherwise we could solve to get $b_{n-1} = b_{n-2} = \cdots = b_0 = 0$, a contradiction to the fact that $\vec{v}_0$ is not the zero vector. If a_0 is not also zero, then clearly $f = a_0$ is a constant, so f divides g. Otherwise, we would have that $a_1 = a_0 = 0$.

Finally, as a determinant, the coefficients of r are in $\mathbb{Z}$. Moreover, the first n columns of S are homogeneous in a_1 and a_0, while the last column is homogeneous in $c_n, \cdots, c_0$, so the determinant is homogeneous of degree n over a_1 and a_0 and of degree 1 over $c_n, \cdots, c_0$. □

Lemma 5.1.8. *Let A be an integral domain. Let $f = a_1x + a_0$ and $g = c_nx^n + \cdots + c_0$ be polynomials in $A[x]$. If a_1 is invertible, then f divides g if and only if $r(a_1, a_0, c_n, \cdots, c_0) = 0$.*

Proof. Let K be the field of fractions of A. If f divides g in $A[x]$, then f divides g in $K[x]$, so $r(a_1, a_0, c_n, \cdots, c_0) = 0$.

Now suppose $r(a_1, a_0, c_n, \cdots, c_0) = 0$. Then by Lemma 5.1.6, we see that f divides g in $K[x]$. Let $(b_{n-1}x^{n-1} + \cdots + b_0)f = g$, where $b_i \in K$. From the equation

$$a_1b_{n-1} - c_n = 0,$$

and the fact that a_1 is invertible, we can conclude that b_{n-1} is in A. From the equation

$$a_1b_{n-2} + a_0b_{n-1} - c_{n-1} = 0$$

and the facts that a_0, b_{n-1}, c_{n-1} are in A and a_1 is invertible, we see that b_{n-2} is in A. Proceeding in this manner, we can inductively show that b_n, b_{n-1}, $\cdots$, b_0 are all in A, so f divides g in $A[x]$. □

Now, we have the tools to build the polynomials SP^n in Theorem 5.1.4.

Proof of Theorem 5.1.4. Note that

$$D^{(1)}P = P_x dx + P_y dy + P_z dz$$

and

$$D^{(n)}P = \sum_{n_1+n_2+n_3=n} P_{n_1n_2n_3} dx^{n_1} dy^{n_2} dz^{n_3}$$

are polynomials in $\mathbb{F}[x,y,z][dx,dy,dz]$. Also note that if the degree of P is d, then the degrees of P_x, P_y, and P_z is at most $d-1$, and the degrees of $P_{n_1n_2n_3}$ where $n_1+n_2+n_3=n$ is at most $d-n$.

We first view $D^{(1)}P$ and $D^{(n)}P$ as polynomials in dx. Expanding the determinant for the resultant along the last column shows it is homogeneous of degree n in dy and dz. We put the coefficients of each of the $n+1$ monomials $dy^n, dy^{n-1}dz, \cdots, dz^n$ to SP^n. The other polynomials in SP^n are gotten by repeating this procedure by viewing $D^{(1)}P$ and $D^{(n)}P$ as polynomials in dy, and then as polynomials in dz. Counting the degrees of the polynomials in SP^n, we see that they are at most $n(d-1)+d-n=(n+1)d-2n$.

Now if $D^{(1)}P(a,b,c)$ divides $D^{(n)}P(a,b,c)$ as polynomials in $\mathbb{K}[dx,dy,dz]$, then $D^{(1)}P(a,b,c)$ divides $D^{(n)}P(a,b,c)$ when viewed as polynomials in the variable dx, as polynomials in dy, and as polynomials in dz. Thus, we get that $SP^n(a,b,c)=0$. If $D^{(1)}P(a,b,c)=0$ as a polynomial in $\mathbb{K}[dx,dy,dz]$, then it is 0 as a polynomial in each of the variables separately. We therefore also get $SP^n(a,b,c)=0$ in this situation.

Suppose $SP^n(a,b,c)=0$. If $D^{(1)}P(a,b,c)=0$, then (a,b,c) is a critical point, and we are done. Otherwise, one of $P_x(a,b,c)$, $P_y(a,b,c)$ or $P_z(a,b,c)$ is not zero. Without loss of generality, suppose it is $P_x(a,b,c)$. By Lemma 5.1.8, we see that $D^{(1)}P(a,b,c)$ divides $D^{(n)}P(a,b,c)$ as polynomials in $\mathbb{K}[dy,dz][dx]$, which means the same is true in $\mathbb{K}[dx,dy,dz]$. □

Corollary 5.1.9. *Suppose A is a $\mathbb{F}$-algebra and an integral domain. Let $(a,b,c)\in A^3$. If $SP^n(a,b,c)=0$ and at least one of $P_x(a,b,c)$, $P_y(a,b,c)$ or $P_z(a,b,c)$ is invertible, then $D^{(1)}P(a,b,c)$ divides $D^{(n)}P(a,b,c)$.*

Proof. The proof is the same as the last part of the proof of Theorem 5.1.4. □

Definition 5.1.10. Let P be a polynomial in $\mathbb{F}[x,y,z]$. We say that P is **totally flat** if P divides every polynomial in SP^2. We say that P is **totally n-flat** if P divides every polynomial in SP^n.

It is clear from the definition and Theorem 5.1.4 that every point in the zero set of an totally n-flat polynomial P is either critical or n-flat.

5.2. **Lines on the Zero Set of a Polynomial.** Let $\mathbb{F}$ be a field, and P be a polynomial with degree d less than the number of elements in $\mathbb{F}$. Suppose $(x_0,y_0,z_0)\in\mathbb{F}^3$ is on the zero set of P, which we denote by $Z(P)$. We wish to find the number of lines through (x_0,y_0,z_0) and contained in $Z(P)$. Looking in the direction $(dx,dy,dz)\neq 0$, we see that P vanishes on the line $(x_0,y_0,z_0)+(dx,dy,dz)t$ if and only if

$$\begin{aligned} P(x_0+t\,dx,y_0+t\,dy,z_0+t\,dz) &= D^{(0)}P(x_0,y_0,z_0)+D^{(1)}P(x_0,y_0,z_0,dx,dy,dz)t \\ &\quad +D^{(2)}P(x_0,y_0,z_0,dx,dy,dz)t^2 \\ &\quad +\cdots+D^{(d)}P(x_0,y_0,z_0,dx,dy,dz)t^d \\ &= 0, \end{aligned}$$

for every $t\in\mathbb{F}$. Since the degree d of P is less than the number of elements in $\mathbb{F}$, this happens if and only if $D^{(i)}P(x_0,y_0,z_0)(dx,dy,dz)=0$ for all i. Now the direction (dx,dy,dz) is only determined up to a multiplicative constant, and $D^{(i)}P(x_0,y_0,z_0)$ is a homogeneous polynomial of degree i in the variables dx, dy, and dz. We can then conclude:

Proposition 5.2.1. *Let P be a polynomial in $\mathbb{F}[x,y,z]$ with degree less than the number of points in $\mathbb{F}$. The number of lines through $(x_0,y_0,z_0) \in \mathbb{F}^3$ on the zero set of P is equal to the number of common zeros of $D^{(0)}(x_0,y_0,z_0)$, $\cdots$, $D^{(d)}P(x_0,y_0,z_0)$ in the projective plane $\mathbb{F}P^2$.*

Remark 5.2.2. We can interpret the meaning of some of these derivatives being equal to zero:

- $D^{(0)}P(x_0,y_0,z_0) = P(x_0,y_0,z_0)$, so this is zero if and only if (x_0,y_0,z_0) is a zero of P.
- $D^{(1)}P(x_0,y_0,z_0) = P_x(x_0,y_0,z_0)dx + P_y(x_0,y_0,z_0)dy + P_z(x_0,y_0,z_0)dz$. This is saying that either (x_0,y_0,z_0) is critical, or all lines through it lying on P must be on the tangent plane.
- $D^{(d)}P(x_0,y_0,z_0)$ is actually a polynomial in $\mathbb{F}[dx,dy,dz]$ which does not depend on the value of (x_0,y_0,z_0). This polynomial vanishes if and only if P vanishes at infinity in the direction (dx,dy,dz).

Lemma 5.2.3. *A point $(x_0,y_0,z_0) \in \mathbb{F}^3$ which is not critical or flat can be on at most 2 distinct lines in $Z(P)$. (Here, we assume $\deg P < |\mathbb{F}|$.)*

Proof. If (x_0,y_0,z_0) is not critical, then $D^{(1)}P(x_0,y_0,z_0)$ is not zero. If (x_0,y_0,z_0) is not flat, then $D^{(1)}P(x_0,y_0,z_0)$ does not divide $D^{(2)}P(x_0,y_0,z_0)$. But then $D^{(1)}P(x_0,y_0,z_0)$ and $D^{(2)}P(x_0,y_0,z_0)$ are coprime and have degrees 1 and 2 with respect to the variables dx, dy and dz, so by Bezout's theorem, they can have at most 2 common points in the projective plane. □

Corollary 5.2.4. *If a noncritical point (x_0,y_0,z_0) is on 3 or more distinct lines in $Z(P)$, then it is flat.*

To further study the number of lines through a given noncritical point, we look at the power series expansion about that point. Indeed, if P is not critical at (x_0,y_0,z_0), we can translate so that the point becomes the origin by replacing P with $P(x-x_0,y-y_0,z-z_0)$. Without loss of generality, suppose $P(0,0,0) = 0$, and $P_z(0,0,0) \neq 0$. By the theory of power series, we can write

$$P(x,y,z) = (z - f(x,y))Q(x,y,z),$$

where $f(x,y)$ is a power series in x and y with $f(0,0) = 0$, and Q is in $\mathbb{F}[[x,y]][z]$ satisfying $Q(0,0,0) = P_z(0,0,0) \neq 0$. We can then see that $Q(0,0,f(0,0)) \neq 0$, so $Q(x,y,f(x,y))$ is a unit in the ring of power series $\mathbb{F}[[x,y]]$.

The following proposition relates the flatness properties of P with the Hasse derivatives of f.

Lemma 5.2.5. *Let $\mathbb{F}$ be a field with characteristic p. We have the following:*

- *P is totally flat, then $D^{(2)}f = 0$.*
- *If P is totally flat, totally p-flat, $\cdots$, and totally p^k-flat, then $D^{(p^k)}f = 0$. (Note: the totally flat and totally p-flat conditions are redundant in characteristic 2.)*

Proof. We use Corollary 5.1.9 by sending $x \mapsto x$, $y \mapsto y$ and $z \mapsto f(x,y)$ in $\mathbb{F}[[x,y]]$. Since $P_z(0,0,f(0,0)) = P_z(0,0,0) \neq 0$, we set that the power series $P_z(x,y,f(x,y))$ has nonzero constant term and is therefore invertible in $\mathbb{F}[[x,y]]$.

Now suppose P is totally flat. Then P divides every polynomial in SP^2. Because $P = (z-f(x,y))Q(x,y,z)$ as power series, $z-f(x,y)$ divides every element of SP^2.

Thus, $SP^2(x, y, f(x, y)) = 0$ in $\mathbb{F}[[x, y]]$. We can then apply Corollary 5.1.9 to conclude that $D^{(1)}P(x, y, f(x, y))$ divides $D^{(2)}P(x, y, f(x, y))$ as formal power series. Similarly, if P is totally n-flat, then $D^{(1)}P(x, y, f(x, y))$ divides $D^{(n)}P(x, y, f(x, y))$.

Now using the product rule on $P = (z - f(x, y))Q$, we get

$$D^{(1)}P(x, y, z) = D^{(1)}(z - f(x, y))Q + (z - f(x, y))D^{(1)}Q,$$

and

$$D^{(2)}P(x, y, z) = D^{(2)}(z - f(x, y))Q + D^{(1)}(z - f(x, y))D^{(1)}Q + (z - f(x, y))D^{(2)}Q.$$

Plugging in $z = f(x, y)$, we get that

$$D^{(1)}P(x, y, f(x, y)) = (dz - D^{(1)}f(x, y))Q(x, y, f(x, y)),$$

and

$$\begin{aligned} D^{(2)}P(x, y, f(x, y)) = &-D^{(2)}f(x, y)Q(x, y, f(x, y)) \\ &+(dz - D^{(1)}f(x, y))D^{(1)}Q(x, y, f(x, y)). \end{aligned}$$

Noting again that $Q(x, y, f(x, y))$ is a unit in $\mathbb{F}[[x, y]]$, we conclude that if the power series $D^{(1)}P(x, y, f(x, y))$ divides the power series $D^{(2)}P(x, y, f(x, y))$, then

$$dz - D^{(1)}f(x, y) \mid D^{(2)}f(x, y).$$

But $D^{(2)}P(x, y, f(x, y))$ has no dz term, so this is only possible if $D^{(2)}f(x, y)$ is zero.

The proof of the second point is similar, and uses induction. Supposing P is totally flat, totally p-flat, $\cdots$ and totally p^{k-1}-flat, we get that $D^{(2)}f = D^{(p)}f = \cdots = D^{(p^{k-1})}f = 0$. By Corollary 4.3.3, all the derivatives $D^{(2)}f$, $D^{(3)}f$, $\cdots$, $D^{(p^k-1)}f$ are zero. Thus, using the product rule, we get

$$D^{(p^k)}P = (-D^{(p^k)}f)Q + (dz - D^{(1)}f)D^{(p^k-1)}Q + (z - f(x, y))D^{(p^k)}Q.$$

Plugging in $z = f(x, y)$ and then noting that because P is totally p^k-flat, we have $D^{(1)}P(x, y, f(x, y))$ divides $D^{(p^k)}P(x, y, f(x, y))$. Because $Q(x, y, f(x, y))$ is a unit,

$$dz - D^{(1)}f(x, y) | D^{(p^k)}f(x, y).$$

But $D^{(p^k)}f(x, y)$ has no dz dependence, so it must be 0. □

Remark 5.2.6. If the characteristic of $\mathbb{F}$ were 0, then the first point still holds. Now $D^{(2)}f = 0$ implies $D^{(n)}f = \frac{1}{\binom{n}{2}}D^{(2)}f = 0$ for $n \geq 2$. Thus, f is a linear polynomial, so P looks locally like a plane, that is, P factors as $(z - f(x, y))Q$ in $\mathbb{F}[x, y, z]$. The lemma gives a method of doing something similar in characteristic p, as the following corollary shows.

Corollary 5.2.7. *Suppose P has degree d, and let p^k be the largest power of p less than or equal to d. If P is totally flat, totally p-flat, $\cdots$, and totally p^k-flat, then f is linear in x and y, and P factors as $(z + ax + by)Q(x, y, z)$ in the polynomials.*

Proof. Note that $D^{(n)}P = 0$ for $n > d$. Thus, we have that P is automatically p^ℓ-flat for powers p^ℓ of p greater than d. Using the lemma, we see that $0 = D^{(2)}f = D^{(p)}f = D^{(p^2)}f = \cdots$, which in turns implies that $D^{(n)}f = 0$ for all $n \geq 2$. Thus, we see that f is a linear polynomial, and the result follows. □

Lemma 5.2.8. *Let $\mathbb{F}_q$ be a finite field with characteristic $p > 0$, and P be a polynomial with degree $d < q$. Suppose (x_0, y_0, z_0) is not a critical point of P. Then,*

- *If (x_0, y_0, z_0) is on 3 or more distinct lines in $Z(P)$, then it is a flat point.*
- *If P is totally flat and (x_0, y_0, z_0) is on 2 or more distinct lines in $Z(P)$, then it is a p-flat point.*
- *If P is totally flat, totally p-flat, $\cdots$, and totally p^{k-1}-flat and (x_0, y_0, z_0) is on 2 or more distinct lines in $Z(P)$, then it is a p^k-flat point.*

Proof. Translate so that (x_0, y_0, z_0) is at the origin, and assume without loss of generality that $P_z(0,0,0)$ is not zero. The first point is Corollary 5.2.4.

For the second point, if P is totally flat, we can write

$$D^{(p)}P = (-D^{(p)}f)Q + (dz - D^{(1)}f)D^{(p-1)}Q + (z - f(x,y))D^{(p)}Q.$$

Recall that $Q(0,0,0) = P_z(0,0,0) \in \mathbb{F}_q$ is a nonzero constant. Also, note that in a power series $\mathbb{F}_q[[x,y,z]][dx,dy,dz]$, we are always allowed to send $(x,y,z) \mapsto (0,0,0)$ to get an element of $\mathbb{F}_q[dx,dy,dz]$. Thus,

$$D^{(p)}P(0,0,0) = -(D^{(p)}f(0,0))Q(0,0) + (dz - f_x(0,0)dx - f_y(0,0)dy)D^{(p-1)}Q(0,0).$$

Recalling that

$$D^{(1)}P(x,y,z) = (dz - D^{(1)}f(x,y))Q(x,y,z) + (z - f(x,y))D^{(1)}Q(x,y,z),$$

We get $D^{(1)}(0,0,0) = (dz - f_x(0,0)dx - f_y(0,0)dy)Q(0,0)$. Thus, the common roots of $D^{(1)}P(0,0,0)$ and $D^{(p)}P(0,0,0)$ are the common roots of $dz - f_x(0,0)dx - f_y(0,0)dy$ and $D^{(p)}f(0,0)$.

Now because P is flat, we have that $D^{(2)}f = 0$, so $D^{(p)}f(0,0) = f_{x^p}(0,0)dx^p + f_{y^p}(0,0)dy^p$ by Corollary 4.3.3. Using the Frobenius automorphism, we see that

$$D^{(p)}f(0,0) = ((f_{x^p}(0,0))^{\frac{q}{p}}dx + (f_{y^p}(0,0))^{\frac{q}{p}}dy)^p.$$

If this is not zero, then $dz - f_x(0,0)dx - f_y(0,0)dy$ and $D^{(p)}f(0,0)$ can have at most 1 common zero in the projective plane. But this is impossible if $(0,0,0)$ has two or more lines in $Z(P)$ through it. Hence, $D^{(p)}f(0,0) = 0$, and we can conclude that $D^{(1)}P(0,0,0)$ divides $D^{(p)}P(0,0,0)$.

The proof of the last is similar to the proof of the second point, except we write

$$D^{(p^k)}f(0,0) = ((f_{x^{p^k}}(0,0))^{\frac{q}{p^k}}dx + (f_{y^{p^k}}(0,0))^{\frac{q}{p^k}}dy)^{p^k},$$

where $p^k \leq d < q$. (In the case $p^k > d$, every point is p^k flat, so the statement is vacuously true.) □

We can now prove a theorem estimating the number of points in $Z(P)$ lying on at least three lines in $Z(P)$.

Theorem 5.2.9. *Let $\mathbb{F}_q$ be a finite field, and $P \in \mathbb{F}_q[x,y,z]$ be an irreducible polynomial with degree d. Suppose P is not zero and not linear. Then, there are at most $3qd^2(d-1)$ points with at least 3 lines on $Z(P)$.*

Proof. Let p be the characteristic of $\mathbb{F}_q$, and let p^k be the largest power of p less than d. The polynomial P cannot be totally flat, totally p-flat, $\cdots$, totally p^k-flat, since that means it factors $P = (z + ax + by)Q$, a contradiction.

If P is not totally flat, then since P is irreducible, it does not divide some element of SP^2. But the degree of the polynomials in SP^2 is at most $(2+1)d - 2 \cdot 2 \leq$

$(d+1)d - 2d = d^2 - d$. Thus, using Theorem 2.2.2, we see that there are at most $qd^2(d-1)$ flat points. Since all points with three lines are flat, we are done in this case.

If P is totally flat, but not totally p-flat, then P does not divide some element of SP^p. The degree of the polynomials in SP^p is at most $(p+1)d - 2 \cdot p \leq d^2 - d$, so there are at most $qd^2(d-1)$ points which are p-flat. As P is totally flat, every point with three lines must be totally p-flat, so we are also done in this case.

In a similar manner, we must have that P is totally flat, totally p-flat, $\cdots$, totally $p^{\ell-1}$-flat, but not totally p^ℓ-flat for some $\ell \leq k$. Then P does not divide some polynomial in SP^{p^ℓ}, so there are at most $qd(p^\ell+1)d - 2 \cdot p^\ell \leq qd^2(d-1)$ points which have three lines through them. □

6. Proof of the Theorem

In this section we finish proof of the main theorem of our paper. Recall that Proposition 3.3.1 leaves only the gap $\frac{q^2}{\log q} \leq |Y| \leq q^2 \log q$.

6.1. **Irreducible Polynomials.**

Theorem 6.1.1. *Let P be an irreducible polynomial in $\mathbb{F}_q[x,y,z]$, with degree d at most $c \log q$ for some constant c. Suppose X is a union of lines $\ell_1, \cdots, \ell_n$ on the zero set $Z(P)$ of P. Let $S_i \subset \ell_i$ with $|S_i| \geq \frac{q}{2}$, and let $Y = \bigcup S_i$. Then there exists a constant C such that $|X| \leq C|Y|$.*

Proof. The case when P is linear is the same as the statement in the two dimensional case. We thus assume P is not linear.

Using Proposition 3.3.1, we may assume that $|Y| \geq \frac{q^2}{\log q}$.

We look at the number of point-line incidence pairs $\#(p,\ell)$ in Y. We break this up into three types of points:

(1) The normal points *np*, which are points which have at most 2 lines through them.
(2) The special points *sp*, which are noncritical points with at least 3 lines.
(3) The critical points *cp*. We divide the lines through these points as the critical lines *cl*, which are lines all of whose points are critical points, and the non-critical lines *ncl*, which are those lines through critical points that are not critical lines.

We then get

$$\#(p,\ell) = \#(np,\ell) + \#(sp,\ell) + \#(cp,cl) + \#(cp,ncl).$$

We look at each of these terms in turn:

(1) There are at most $|Y|$ normal points in Y, and these can have at most 2 lines through them, so $\#(np,\ell) \leq 2|Y|$.
(2) By Theorem 5.2.9, there are at most $3qd^3$ special points. The lines through a noncritical point must all lie on the tangent plane. It is known that two coprime polynomials of degrees d_1, d_2 less than q in $\mathbb{F}_q[x,y,z]$ can have at most $d_1 d_2$ common lines (see [4, Lecture 13, Theorem 4.1] for details). Thus, a plane and an nonlinear irreducible polynomial of degree $d < q$ can have at most d lines in common, so we can bound $\#(sp,\ell)$ by $3qd^4$.

(3) A critical line is a line on which P, P_x, P_y and P_z all vanish. Because P_x has degree less than P, and P is irreducible, P and P_x are coprime unless $P_x = 0$. Now if $P_x = P_y = P_z = 0$, then P is a polynomial in x^p, y^p, and z^p, so P is a perfect p-th power by the Frobenius automorphism. This contradicts the fact that P is irreducible. Thus P is coprime with one of P_x, P_y or P_z, so there are at most d^2 critical lines. We thus get that $\#(cp, cl) \leq qd^2$.

(4) A noncritical line can have at most $d-1$ critical points, because if the polynomials P_x, P_y, P_z of degree $d-1$ vanish on d points, they must vanish on the whole line. Letting L denote the total number of lines, we see that $\#(cp, ncl) \leq dL$.

Now each line must have at least $\frac{q}{2}$ points on it, so we have

$$\frac{qL}{2} \leq \#(p, l) \leq 2|Y| + 3qd^4 + qd^2 + dL.$$

Moving the dL term to the left side, we get

$$(\frac{q}{2} - d)L \leq 2|Y| + 3qd^4 + qd^2.$$

Recall that $d \leq c \log q$. For large enough q, the left side is greater than 0. Multiplying by $\frac{q}{\frac{q}{2}-d}$, we get

$$qL \leq \frac{q}{\frac{q}{2} - d}(2|Y| + 3qd^4 + qd^2).$$

Each line of X has q points, so $|X| \leq qL$. The limit as $q \to \infty$ of $\frac{q}{\frac{q}{2}-d}$ is 2, so we can bound this from above by a constant. Recalling that $|Y| \geq \frac{q^2}{\log q}$, we see that as $q \to \infty$, the term $2|Y|$ dominates $3qc^4(\log q)^4$ and $qc^2(\log q)^2$. Therefore

$$|X| \leq qL \leq \frac{q}{\frac{q}{2} - d}(2|Y| + 3qd^4 + qd^2) \leq C|Y|,$$

for some constant C, as desired. □

6.2. **The General Case.**

Theorem 6.2.1. *Suppose $\ell_1, \cdots, \ell_n$ are distinct lines in $\mathbb{F}_q^3$, and $X = \bigcup \ell_i$. For each $i = 1, \cdots, n$, let S_i be a subset of ℓ_i such that $|S_i| \geq \frac{q}{2}$. Let $|Y| = \bigcup S_i$. Then there exists a constant c independent of q such that $|X| \leq c|Y|$.*

Proof. By Proposition 3.3.1, we only need to prove the case where $\frac{q^2}{\log q} \leq |Y| \leq q^2 \log q$. In this situation, we know there is a polynomial P with degree at most $C \log q$ which vanishes on all but $\frac{4}{q}|Y|$ lines, which contain at most $4|Y|$ points.

Factor P into a product of irreducibles $P = P_1 \cdots P_n$. We may assume all the irreducibles are distinct, since we can remove any repeats to get a polynomial with lower degree vanishing on the same lines.

Let ℓ be any line on $Z(P)$. We claim that ℓ is on P_i for some i. If not, then P_1 can contribute at most $\deg P_1$ zero points, P_2 can contribute at most $\deg P_2$ zero points, etc. In total, the line ℓ can have at most $\deg P = C \log q$ points which are zero on P, which contradicts the fact that all of its points are zero on P.

Let $\mathcal{L}$ denote the lines on $Z(P)$ which are in X. Let $\mathcal{L}_i$ denote the lines of $\mathcal{L}$ on P_i, let X_i denote the set of points in $\mathcal{L}_i$, and let Y_i denote the set $Y \cap X_i$. By

Theorem 6.1.1, there is a constant K such that $|X_i| \leq K|Y_i|$ for all i. Then

$$|X| - 4|Y| \leq \sum_{i=1}^{n} |X_i| \leq K \sum_{i=1}^{n} |Y_i|.$$

Now $\sum_{i=1}^{n} |Y_i|$ is the sum of $|Y|$ and the number of times a point repeatedly counted. We thus get

$$\sum_{i=1}^{n} |Y_i| \leq |Y| + \sum_{i \neq j} |Y_i \cap Y_j|.$$

The points $Y_i \cap Y_j$ are common zeros of the coprime polynomials P_i and P_j, so there are at most $3q(\deg P_i)(\deg P_j)$ of them. Expanding $(\sum P_i)^2$ shows that $\sum_{i=neqj} (\deg P_i)(\deg P_j) \leq (\deg P)^2 \leq C^2(\log q)^2$. Thus,

$$|X| - 4|Y| \leq K|Y| + C^2 K q(\log q)^2.$$

As $|Y| \geq \frac{q^2}{\log q}$, we may conclude that there is a constant c such that

$$|X| \leq c|Y|,$$

which is what we set out to prove. □

References

[1] Christian Bruun. *The Sylvester Matrix and Resultants*. From Math 499: VIGRE Computational Algebraic Geometry Junior Seminar Fall 2007 at Rice University. Available at `http://math.rice.edu/~cbruun/vigre/vigreHW9.pdf`.

[2] Zeev Dvir. *On the size of Kakeya sets in finite fields*. Available at `http://arxiv.org/abs/0803.2336`.

[3] Felix Fontein. *The Hasse derivative* on Felix' Math Place blog. Available at `http://math.fontein.de/2009/08/12/the-hasse-derivative/`.

[4] Larry Guth. *Polynomial Method Course Notes*. Available at `http://math.mit.edu/~lguth/PolynomialMethod.html`. Notes typed by Larry Guth, Adam Hesterberg, Laszlo Lovasz, Rik Sengupta, Sean Simmons, Yufei Zhao, Gaku Liu, Yi Sun, Chiheon Kim, Andrey Grinshpun, Ben Yang, Efrat Shaposhnik, Sam Elder.

[5] Larry Guth, Nets Katz. *On the Erdos distinct distance problem in the plane*. Available at `http://arxiv.org/abs/1011.4105`.

[6] J.W.P. Hirschfeld, G. Korchmáros, F. Torres. *Algebraic Curves over a Finite Field*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, 2008.

[7] Paul Vojta. *Jets via Hasse-Schmidt Derivations*. Diophantine Geometry, Proceedings, Edizioni della Nomale, Pisa, 2007, pp. 335-361. Available at `http://arxiv.org/abs/math/0407113`.