On a Generalization of Rédei's Theorem

András Gács

ELTE, Budapest

gacs@cs.elte.hu

Abstract In 1970 Rédei and Megyesi proved that a set of p points in $AG(2,p)$, p prime, is a line, or it determines at least $\frac{p+3}{2}$ directions. In '81 Lovász and Schrijver characterized the case of equality. Here we prove that the number of determined directions cannot be between $\frac{p+5}{2}$ and $2\frac{p-1}{3}$. The upper bound obtained is one less than the smallest known example.

1. Introduction

In this paper we consider subsets of points of $AG(2,q)$, the desarguesian affine plane of order q, where q is a fixed prime-power. Since $AG(2,q)$ can be coordinatized over $GF(q)$, the finite field of q elements, its points can be identified with ordered pairs of elements of the underlying field. After this identification, non-vertical lines have equation $y = mx + b$, where m is the *slope* (*direction*) of the line in question; vertical lines have equation $x = c$, their direction is denoted by the symbol ∞.

Let q be a fixed prime-power and $\{U = (a_i, b_i) : i = 1, ..., q\}$ a set of q points in $AG(2,q)$. The elements of the set $D = \{\frac{b_i - b_j}{a_i - a_j} : i \neq j\}$ are called the *directions determined by* U, these are the slopes of lines joining at least two points of U. N will denote the cardinality of D.

In his book [1] Rédei proved deep theorems about fully reducible, lacunary polynomials. As an application of his algebraic results, he considered the problem of determining, what are the possible values N can take, that is, how many different directions a point set can determine. For an arbitrary prime-power q, he proved that for $N < \frac{q+1}{2}$, N cannot take all values, it has to be an element of the union of some intervals.

For $q = p$ prime, he proved much more, namely that unless U is a line (and hence $N = 1$), N is at least $\frac{p+1}{2}$. Later Megyesi realised that the case $\frac{p+1}{2}$ is impossible, so their result can be formulated as follows:

Theorem 1.1 (Rédei and Megyesi [1]) If a set of p points in $AG(2,p)$ is not a line, then it determines at least $\frac{p+3}{2}$ directions.

This bound is sharp, we give the (unique) example constructed by Megyesi at the end of this section.

After the publication of Rédei's book, it turned out that the problem of determining the possible values of N and looking for sets determining a small number of directions has an important application: a certain type of blocking set in $PG(2,q)$ (the projective

plane of order q) of size $q+N$ is equivalent to a set of q points in $AG(2,q)$ determining N directions. For more information, we refer to [2]. Later we will give another reformulation of the problem.

In the general case today almost everything is known for $N < \frac{q+3}{2}$: Ball, Blokhuis, Brouwer, Storme and Szőnyi determined all possible values, see [3].

For the prime case, the next step in the history of the problem was the characterization of equality in Theorem 1.1:

Theorem 1.2 (Lovász and Schrijver [4]) For every prime $p > 2$, up to affine transformation there is a unique set of p points in $AG(2,p)$ determining $\frac{p+3}{2}$ directions.

Recently in [5] and [6] we proved results suggesting that after $\frac{p+3}{2}$, there is a big gap in the possible values of N. In this paper we prove the following:

Theorem 1.3 For every prime p, besides lines and the example characterized by Lovász and Schrijver, any set of p points in $AG(2,p)$ determines at least $[2\frac{p-1}{3}]+1$ directions.

We close this introduction by giving an infinite series of examples constructed by Megyesi [1] for arbitrary prime-powers.

Example 1.4 For a $d|q-1$, $1 < d < q-1$, let G be the multiplicative subgroup of $GF(q)$ of order d. Set $U = \{(x,0) : x \in G\} \cup \{(0,x) : x \notin G\}$, that is put G and the complement of G on the x and y axes, respectively. An easy calculation shows that the number of determined directions is $q+1-d$.

Choosing $d = \frac{q-1}{2}$, we find an example with $N = \frac{q+3}{2}$. For $q = p$ prime, this is the unique example characterized in 1.2. The next value for N in this construction (whenever $3|q-1$) is $q+1-\frac{q-1}{3} = 2\frac{q-1}{3}+2$, so for primes, Theorem 1.3 is one less than a possible sharp result.

The examples of Megyesi are all contained in the union of two lines. This property was characterized by Szőnyi. For primes the result can be formulated as follows:

Theorem 1.5 (Szőnyi [7]) Suppose U is a set of p points in $AG(2,p)$, p prime, which is contained in the union of two lines and denote by N the number of determined directions. If $1 < N < p-1$, then $N = p+1-d$ for a $d|p-1$, and after linear transformation, $U = \{(x,0) : x \in K\} \cup \{(0,x) : x \notin K\}$, where K is the union of some cosets of the multiplicative subgroup of $GF(p)$, of order d.

2. Algebraic reformulation

In this section we give an algebraic reformulation of the problem, which in fact was the original terminology Rédei used.

Suppose U is a set of q points which does not determine all directions. Then applying an affine transformation, we can achieve that the direction of vertical lines is not determined. But this means that our point set in question can be considered as the graph of

a function $f(x)$, over the underlying field. Since over a finite field every function can be represented by a polynomial, we may assume $f \in GF(q)[x]$. (Note that lines correspond to linear polynomials.) We will say that a polynomial f *determines a direction*, if its graph determines it, which means that the set of determined directions is $D = \{\frac{f(x)-f(y)}{x-y} : x \neq y\}$. This was the original question Rédei considered, he tried to determine the number of difference quotients a polynomial can have. It is easy to verify, that $D = \{c \in GF(q) : f(x) - cx$ is not bijective $\}$, so a third formulation of our problem is to look for polynomials, for which there are many c-s such that $f(x) - cx$ is a permutation, this was the terminology of [8].

Perhaps it is useful to give Example 1.4 in this form as well. In general, the polynomial giving the example on two lines with $N = q + 1 - d$ is $x\chi_G$, where χ_G is the characteristic function of G, the multiplicative subgroup of order d. Hence the polynomial having $N = \frac{q+3}{2}$ is $f(x) = x^{\frac{q+1}{2}}$.

We end this section by summarizing some properties of polynomials over finite fields. All proofs can be found in [6]. Let $f \in GF(q)[x]$. Then f, as a function, can be represented by a polynomial f_1 of degree at most $q-1$, this is called the *reduced form of* f. The degree of f_1 is the *reduced degree of* f.

Now let $f(x) = c_{q-1}x^{q-1} + ... + c_0$, that is, suppose f is reduced. Then $\sum_{x \in GF(q)} f(x) = -c_{q-1}$. If f is bijective, then $\sum_{x \in GF(q)} f(x)^k = 0$ for all $k = 1, ..., q-2$. The last two remarks together yield that if f is bijective, then the reduced degree of f^k is at most $q-2$ for $k = 1, ..., q-2$.

$GF(q)[x]$ can be considered as a vector space over $GF(q)$. For any subspace V, $\dim(V) = |\{\deg(f) : f \in V\}|$.

3. Another function on polynomials

In this section we introduce another parameter on polynomials over finite fields which has a close relationship with N and will enable us to gain from the algebraic terminology mentioned in Section 2.

From now on our underlying field will be $GF(p)$, where p is a prime, for brevity we will denote it by $\mathbf{F}$. For any polynomial f, $N(f)$ will denote the number of directions f determines. First we formulate a lemma about the function N.

Lemma 3.1 Suppose f and g are affine transforms of each other (that is $f(x) = ag(bx + c) + dx + e$, where $a, b, c, d, e \in \mathbf{F}$, $a, b \neq 0$), or f and g are bijective and $f^{-1} = g$. Then $N(f) = N(g)$.

Proof Straightforward.

Now we define another function on polynomials, which might seem at first glance a little bit strange, but Lemma 3.2 will connect it to our problem.

For any polynomial f, let $I(f) = \min\{k + l : \sum_{x \in \mathbf{F}} x^k f(x)^l \neq 0\}$. Here k and l are non-negative integers; for $k = 0$ or $l = 0$, x^0 or $f(x)^0$ is defined to be the polynomial 1. By the remarks at the end of the previous section, $I(f)$ is the smallest $k + l$, for which

$x^k f(x)^l$ has reduced degree $p-1$. The second and third statements of the next lemma can be implicitly found in [4].

Lemma 3.2 (i) For any of the situations mentioned in Lemma 3.1, we also have $I(f) = I(g)$;

(ii) For any polynomial f, $I(f) + N(f) \geq p+1$;

(iii) For any non-linear polynomial, $I(f) \leq \frac{p-1}{2}$ with equality if and only if $f(x)$ is affinely equivalent to x^2 or $x^{\frac{p+1}{2}}$;

(iv) For any polynomial of degree between 2 and $\frac{p+1}{2}$, $I(f) \leq \frac{p-1}{3}$, unless it is affinely equivalent to x^2 or $x^{\frac{p+1}{2}}$.

Proof (i) is straightforward. For the rest see [6].

A polynomial of the form $x^k f(x)^l$ will be called a *double power of* f, the *degree of* $x^k f(x)^l$ is defined to be $k+l$.

There is a result of A. Biró, which is the analogous statement of 1.5 for I.

Theorem 3.3 (A. Bíró [9]) Suppose the graph of f is contained in the union of two intersecting lines. Then $I(f) = \frac{p-1}{2}$ or $I(f) = \frac{p-1}{3}$ or $I(f) \leq \frac{p-1}{4}$. There is an example with $\frac{p-1}{4} > I(f) > \frac{p-1}{5}$.

We will not need this result (Theorem 1.5 will be good for our purposes), but it seems to show the limits of the method we are using.

4. Proof of Theorem 1.3

Let $f(x)$ be a non-linear polynomial over $\mathbf{F} = GF(p)$ and suppose $N(f) \leq [2\frac{p-1}{3}]$. By Lemma 3.2, supposing f is not affinely equivalent to $x^{\frac{p+1}{2}}$ (which is the unique example with $N(f) = \frac{p+3}{2}$), we have $[\frac{p+1}{3}] + 1 < I(f) < \frac{p-1}{2}$. (Note that $N(x^2) = p$.)

In Lemma 3.1 and 3.2 (i) we saw that I and N are invariant under affine transformations and inversion. Take the transitive closure of the corresponding relation on polynomials and let $E(f)$ denote the equivalence class of f. Note that the property of having a graph which is contained in the union of at most two lines is also invariant under the relation in question. This means that we will be allowed to change f to any element of $E(f)$ throughout the proof (since all we use from now on are the estimations on N, I and Theorem 1.5.)

Write $s = [\frac{p+1}{6}]$.

First we summarize what we can suppose by choosing an appropriate element of $E(f)$ (by f^i we mean the reduced form of the i-th power of f):

Lemma 4.1 We can suppose:
(i) $t := \deg(f) \in [\frac{p+3}{2}, \frac{p-3}{2} + s]$;
(ii) $\deg(f^i) \leq t + i - 1$ for $1 \leq i \leq 2s$;
(iii) $\deg(f^2) = t + 1$;
(iv) $x^k f(x)^l$ has reduced degree at most $p-2$ for $k+l \leq I-1$ (and we have $I-1 \geq 2s+1$);
(v) f has no root in $\mathbf{F}$.

Proof (iv) follows from the definition of I.

For (i) one should use Lemma 3.2 (iv) to see that $t \geq \frac{p+3}{2}$ (recall that $N(x^2) = p$) and the just proved (iv) with $l = 1$, $k = 0, 1, ..., 2s$ to find $t \leq p - 2 - 2s \leq \frac{p-3}{2} + s$.

For (ii) consider the following function on polynomials:

$$d(f) = \max\{\deg(f^i) - i : 1 \leq i \leq 2s\}.$$

Using (iv) with $l = i$, $k = 0, ..., 2s+1-i$, we can immediately see that $d(f) \leq p-3-2s$.

What we need is a $g \in E(f)$ with $\deg(g) = d(g) + 1$. First choose an element f_1 of $E(f)$ with $d(f_1)$ maximal. We will look for a c for which $f_1(x) - cx$ is bijective and such that for the inverse f_2 of it, $\deg(f_2) \geq d(f_1) + 1$ holds. By the maximality of $d(f_1)$, this will mean that $d(f_2) = d(f_1)$, hence f_2 is appropriate for (ii). Now write $d = d(f_1)$ and note that it suffices to show that $\sum_x f_2(x)x^{p-1-d-1} \neq 0$, that is that $F(c) := \sum_x (f_1(x) - cx)^{p-1-d-1}x \neq 0$. There are $p - N(f_1) \geq 2s$ c-s for which $f_1(x) - cx$ is bijective, suppose that $F(c) = 0$ for all of them. Consider the following polynomial: $G(c) = \frac{-1}{p-1-d} \sum_x (f_1(x) - cx)^{p-1-d}$. On the one hand, this is also zero for these c-s, on the other, $F = G'$, so we find that these c-s are roots of G of multiplicity at least two. Hence $2 \cdot 2s \leq \deg(G) \leq p - 2 - d$, so $d \leq p - 2 - 4s$. But this contradicts $\deg(f) \geq \frac{p+3}{2}$, so (ii) is satisfied. Note that this property is invariant under affine transformations. (Also note that $G = 0$ identically cannot hold, since it would imply $F = 0$ identically, but the coefficients of F are the double power sums $\sum_x f_2(x)^i x^{p-1-d-i}$; and by $2s \leq p-1-d$ and by the definition of d, at least one of them is not zero.)

For (iii) and (v) we need an affine transform $f_3 = f_2(x)+cx$ for which $\deg(f_3^2) \geq t+1$, and f_3 is not bijective. Then (iii) is automatically true by (ii); and choosing a c' not taken by f_3, $f_4(x) = f_3(x) - c'$ will be an element of $E(f)$ satisfying all the desired conditions.

Note that 4.1. (i) implies in particular, that $p \geq 19$.

Just like in the previous lemma, throughout this text f^i will mean the reduced form of the i-th power of f. This implies that in most of the cases when we consider a polynomial which is the linear combination of double powers of f, it will automatically be reduced, so its degree will coincide with its reduced degree. We will remark whenever this is not the case.

During the proof we will often look for elements of the following two sets of polynomials, which both will be considered as vector spaces over $\mathbf{F}$:

$$\Phi = \{F_1 f + F_2 f^2 + ... + F_{2s} f^{2s} : \deg(F_i) + i \leq 2s\},$$

$$\Psi = \{F_1 f + F_2 f^2 + ... + F_s f^s : \deg(F_i) + i \leq s\}.$$

Note that they are both spaces generated by double powers of f mentioned in Lemma 4.1 (iv). First we summarize some properties of these spaces:

Lemma 4.2 (i) The elements of Φ have degree at most $p - 2$;

(ii) For $F, G \in \Psi$, $\deg(F) + \deg(G) \neq p - 1$;

(iii) In Ψ all elements have degree from $[0, s] \cup [p - t, t] \cup [t + 1, t + s - 1]$;

(iv) $|\{\deg(F) : F \in \Psi, \deg(F) \in [p - t, t]\}| \leq t - \frac{p-1}{2}$;

(v) There is no polynomial in Φ of degree $p-1-t$;

(vi) Whenever $F = F_1 f + ... + F_{2s-1} f^{2s-1} \in \Phi$ has degree at most $p-t-1$, then in fact $\deg(F) \leq \max\{\deg(F_i) + i : i = 1, ..., 2s-1\}$.

Proof All elements in Φ (and hence in $\Psi \subset \Phi$) are linear combinations of polynomials occurring in Lemma 4.1 (iv), so (i) is true. For $F, G \in \Psi$, we have $FG \in \Phi$, so (ii) follows from (i).

According to Lemma 4.1 (ii), all degrees of Ψ are at most $t+s-1$, so all we need for (iii) is that an $F \in \Psi$ cannot have $\deg(F) \in [s+1, p-t-1]$, this follows from (vi).

For (vi) suppose to the contrary and let $F = F_1 f + ... + F_{2s-1} f^{2s-1} \in \Phi$ be of degree bigger than $d = \max\{\deg(F_i)+i : i\}$. Multiplying by f, we find $F_1 f^2 + ... + F_{2s-1} f^{2s} = Ff$. By Lemma 4.1 this is a contradiction, since the left hand side has degree at most $t+d$, while the right has degree between $t+d+1$ and $p-1$.

For (iv) note that by (ii) at most half of the interval $[p-t, t]$ can occur as a degree and $\frac{p-1}{2}$ cannot occur at all.

For (v) suppose we have an $F \in \Phi$ with $\deg(F) = p-1-t$. Then Ff is a polynomial of degree $p-1$ and is the linear combination of double powers of f of degree at most $2s+1$, a contradiction by Lemma 4.1 (iv).

From now on, the proof will consist of five steps. First we find polynomials F, G and H with $Ff + Gf^2 = H$, where F, G and H are of small degree.

In the second step by induction on i we investigate equations of the form $A_i f + B_i f^i = C_i$, where A_i, B_i, C_i are still considerably of small degree (and in fact $B_i = G^{i-1}$).

In the next step we prove that any equation $Af + Bf^i = C$ with A, B, C of small degree has to be a multiple of the equation found in Step 2.

In the fourth step we prove that G is a constant by finding another equation $Af + Bf^i = C$ for a sufficiently chosen i, where B has degree less than G^{i-1}, unless G is a constant.

Finally in Step 5 we prove that if G is a constant, then the graph of f is contained in the union of two lines contradicting Theorem 1.5.

Step 1 (Maybe after transformation) there are polynomials F, G and H with the following properties:
(i) $Ff + Gf^2 = H$;
(ii) $s \geq \deg(H) = \deg(F) + 1 = \deg(G) + 2$;
(iii) $(F, G) = 1$.

Proof Let $\Psi_1 = \{Ff + Gf^2 : \deg(F) \leq s-1, \deg(G) \leq s-2\} \subseteq \Psi$. First we prove that there is a $H \in \Psi_1$ with $\deg(H) \leq s$. Note that if $Ff + Gf^2 = 0$ identically, then $f(F + Gf) = 0$ as a function, but since f has no root and $\deg(F + Gf) < p$, this is only possible, if $F = G = 0$. This means that $\dim(\Psi_1) = s + s - 1 = 2s - 1$. Now according to Lemma 4.2 (iii) and (iv), there are at most $s - 1 + t - \frac{p-1}{2}$ degrees in Ψ_1 bigger than s. Hence the dimension of the subspace of Ψ_1 spanned by polynomials of degree more than s is at most $s - 1 + t - \frac{p-1}{2}$ and by Lemma 4.1 (i), this is less than $2s - 1 = \dim(\Psi_1)$.

Now $\deg(F) = \deg(G) + 1$ follows from the fact that $\deg(f^2) = \deg(f) + 1$ (note that we have $\deg(F) + \deg(f) = \deg(G) + \deg(f^2)$). If $\deg(H) \leq \deg(F)$, then we change f

to $f_1(x) = f(x) + cx + d$ still satisfying all properties proved in Lemma 4.1. It is easy to see that for f_1 we have $F_1 f_1 + G_1 f_1^2 = H_1$, where $F_1 = F - 2(cx+d)G$, $G_1 = G$, $H_1 = H - (cx+d)^2 G + (cx+d)F$, so here $\deg(H_1) \geq \deg(F_1) + 1$ (maybe except for the case when $\deg(F) \leq 1$, but in this case we can jump to Step 5).

Now by Lemma 4.2 (vi), we automatically have $\deg(H) = \deg(F) + 1$.

Finally $(F, G) = 1$ is automatic if we look for an equation with $\deg(F)$ minimal.

Let $r = \deg(G)$, from now on our main purpose is to prove that $r = 0$.

Step 2 For $2 \leq i \leq s$ there are polynomials A_i and B_i satisfying the following conditions:
(i) $A_i f + G^{i-1} f^i = B_i$;
(ii) $\deg(A_i) \leq (i-1)r + i - 1$, $\deg(B_i) \leq (i-1)r + i$;
(iii) $(A_i, G) = 1$.

Proof We use induction on i. For $i = 2$, we have $A_2 = F$ and $B_2 = H$ from Step 1.

For the general case, suppose we have already found the equation

$$A_i f + G^{i-1} f^i = B_i$$

(with all properties we need). Multiplying by Gf and using $Gf^2 = H - Ff$, after a little manipulation we find

$$-(A_i F + B_i G)f + G^i f^{i+1} = -HA_i.$$

It is easy to verify that $A_{i+1} = -(A_i F + B_i G)$ and $B_{i+1} = -HA_i$ satisfy all the wished conditions.

Step 3 Suppose $\deg(A_i) \leq 2s-1$, $\deg(G^{i-1}) \leq 2s - i$, $\deg(B_i) \leq 2s$ holds for a fixed i and we also have $r > 0$. Then all elements of the set $\Phi_1 = \{C : C = Af + Bf^i, \deg(A) \leq 2s-1, \deg(B) \leq 2s - i, \deg(C) \leq 2s\} \subseteq \Phi$ are multiples of B_i in the sense that for any equation $Af + Bf^i = C$ from Φ_1, we have a polynomial K such that $A = A_i K$, $B = G^{i-1} K$, $C = B_i K$.

Proof Let $\Phi_2 = \{Af + Bf^i : \deg(A) \leq 2s - 1, \deg(B) \leq 2s - i\}$.

First we show that the only solution for $Af + Bf^i = 0$ with $\deg(A) \leq 2s-1$, $\deg(B) \leq 2s - i$, is $A = B = 0$. This means on the one hand that $\dim(\Phi_2) = 4s - i + 1$, on the other that for any $C \in \Phi_1$, the corresponding A and B are unique. Suppose that $Af + Bf^i = 0$ in Φ_1 with $(A, B) \neq (0, 0)$. $Af + Bf^i = f(A + Bf^{i-1})$, so by Lemma 4.1 (v), we have $A + Bf^{i-1} = 0$. On the other hand an easy calculation shows that the degree of this polynomial is smaller than p, so it has to be identically zero. (To see this, one should use $i \leq s$ and apply Lemma 4.1 (ii): $\deg(B) + \deg(f^{i-1}) \leq 2s - i + t + i - 2 < p$.) Now if $A + Bf^{i-1}$ is the zero polynomial, then $\deg(f^{i-1}) \leq \deg(A) \leq 2s - 1 \leq p - t - 1$, so by Lemma 4.2 (vi), $\deg(f^{i-1}) \leq i - 1$. For $i = 2$ this is impossible, so we can suppose $i > 2$ and use Step 2. with $i - 1$ to find a contradiction by considering degrees in the equation $A_{i-1} f + G^{i-2} f^{i-1} = B_{i-1}$: $\deg(G^{i-2} f^{i-1}) \leq (i-2)r + i - 1 = \deg(G^{i-1}) - r + i - 1 \leq 2s - i - r + i - 1 = 2s - r - 1$; $\deg(B_{i-1}) \leq (i-2)r + i - 1 \leq 2s - r - 1$ similarly; while $\deg(A_{i-1} f) \geq t > 2s - r - 1$

(note that A_{i-1} cannot be the zero polynomial by $(A_{i-1}, G) = 1$, and $A_{i-1}f$ cannot have degree more than $p-1$, since $\deg(A_{i-1}f) \leq (i-2)r + i - 2 + t \leq 2s - r - 2 + t < p$).

Let $A_0 f + B_0 f^i = C_0$ be an element of Φ_1 with $a_0 = \deg(A_0)$ minimal. Then writing $b_0 = \deg(B_0)$, we have $a_0 - b_0 = \deg(f^i) - \deg(f) \leq i - 1$ by Lemma 4.1 (ii).

If we can prove that Φ_1 consists of the multiples of C_0, then by Step 2 (iii) we will be done.

We prove that Φ_2 is the direct sum of the following spaces:
$\Phi_3 = \{KC_0 : \deg(K) \leq 2s - i - b_0\}$;
$\Phi_4 = \{Af + Bf^i : \deg(A) < a_0, \deg(B) < b_0\}$;
$\Phi_5 =< x^{a_0} f, x^{a_0+1} f, ..., x^{2s-1} f >$.

It is easy to see that only Φ_3 has elements of degree at most $2s$, so this will imply $\Phi_1 = \Phi_3$.

The sum of the three dimensions is exactly the dimension of Φ_2, so what we need is that these spaces are disjoint and contained in Φ_2.

$\Phi_4 \subseteq \Phi_2$ and $\Phi_5 \subseteq \Phi_2$ automatically hold. For $\Phi_3 \subseteq \Phi_2$ we need that $a_0 + 2s - i - b_0 \leq 2s - 1$, this is equivalent to $a_0 - b_0 \leq i - 1$, this is true.

To show that the spaces are disjoint, we consider the degrees of their non-zero elements.

All elements of Φ_3 have degree at most $2s$, since by Lemma 4.2 (vi) $\deg(C_0) \leq \max(a_0 + 1, b_0 + i) = b_0 + i$.

Elements of Φ_5 have degree $t + a_0, t + a_0 + 1, ..., 2s - 1 + t$, while in Φ_4 the degrees are bigger than $2s$ (by the choice of A_0) but below $a_0 + t$. This completes the proof of Step 3.

Note that Step 3 implies that in Step 1 all elements of Ψ_1 are multiples of H.

Step 4 G is a constant.

Proof Suppose to the contrary that $r \geq 1$ and let i be such that $(i-2)r+1 \leq s < (i-1)r+1$ (note that $r \leq s - 2$ implies $i \geq 3$). For $r = 1$ this gives $i = s + 1$, in this case redefine i to be s.

We will find an equation of the form $Af + Bf^j = C$ for a $j \leq i$, where the degrees are small enough to use Step 3 and $\deg(B) < \deg(G^{j-1})$ will give the contradiction.

First we look for a polynomial

$$U = F_1 f + F_2 f^2 + ... + F_i f^i$$

with the following properties:
(i) $\deg(U) \leq s$, $\deg(F_1) \leq s - 1$, $\deg(F_2) \leq s - 2$;
(ii) $\deg(F_3), ..., \deg(F_i) \leq r - 1$;
(iii) F_j is not zero for at least one $j \geq 3$.

Consider the following space:

$$\Psi_2 = \{F_1 f + F_2 f^2 + ... + F_i f^i : \deg(F_1) \leq s-1, \deg(F_2) \leq s-2, \deg(F_3), ..., \deg(F_i) \leq r-1\}.$$

If we can achieve $F_1 f + ... + F_i f^i = 0$ with not all F_j-s zero, then $U = 0$ is appropriate, otherwise we have $\dim(\Psi_2) = 2s - 1 + (i-2)r$.

Next we prove that $\Psi_2 \subseteq \Psi$. What we need is $i+r-1 \leq s$. For $r=1$, this is obvious, so suppose $r>1$. For $i=3$ this is true, since $r \leq s-2$; for $i>3$, using $s \geq (i-2)r+1$, it suffices to show that $i+r-1 \leq (i-2)r+1$ and this is equivalent to $(r-1)(i-3) \geq 1$, this is true. We will use the just found condition $i+r-1 \leq s$ several times later.

Suppose to the contrary that we do not have an appropriate U. This means that whenever we have an element of Ψ_2 of degree at most s, it is automatically in Ψ_1 (see Step 1), so by the remark after the proof of Step 3, it is a multiple of H.

We count the number of different degrees in Ψ_2. There are exactly $s-1-r$ degrees from $[0,s]$ (the degrees of multiples of H), at most $t-\frac{p-1}{2}$ from $[p-t,t]$ (by Lemma 4.2 (iv)), and at most $s-1$ in $[t+1,t+s-1]$, so all in all we have at most $s-1-r+t-\frac{p-1}{2}+s-1$, and using $t \leq \frac{p-1}{2}+s-1$, at most $3s-3-r$ degrees. By $(i-1)r+1>s$, this is smaller than the dimension, a contradiction (or $r=1$, but this case is again easy).

Now redefine i to be maximal with $F_i \neq 0$. We multiply the equation $U = F_1 f + F_2 f^2 + ... + F_i f^i$ by G^{i-2} and for $1 \leq j \leq i-2$ we change $G^j f^{j+1}$ by $B_{j+1} - A_{j+1} f$ (by Step 2). After a little manipulation, we find the equation:

$$Af + Bf^i = C,$$

where

$$A = F_1 G^{i-2} - F_2 A_2 G^{i-3} - .. - F_j A_j G^{i-1-j} - ... - F_{i-1} A_{i-1}; \tag{1}$$

$$B = G^{i-2} F_i; \tag{2}$$

$$C = G^{i-2} U - F_2 B_2 G^{i-3} - ... - F_j B_j G^{i-j-1} - ... - F_{i-1} B_{i-1}. \tag{3}$$

Now if we can show that the conditions from Step 3 are satisfied, then we will be done, since $G^{i-1} | B$ cannot be true by $\deg(F_i) \leq r-1$.

First we need $\deg(A_i) \leq 2s-1$. By Step 2 (ii) it suffices to show that $(i-1)r+i-1 \leq 2s-1$, this is the sum of $r+i-1 \leq s$ and $(i-2)r \leq s-1$, both of which are true. The estimations of $\deg(B_i)$ and $\deg(G^{i-1})$ are similar.

We prove $\deg(A) \leq 2s-1$ by proving that all terms in (1) are of degree at most $2s-1$:
$\deg(F_1 G^{i-2}) \leq s-1+(i-2)r \leq s-1+s-1 = 2s-2$;
$\deg(F_2 A_2 G^{i-3}) \leq s-2+r+1+(i-3)r = s+(i-2)r-1 \leq 2s-2$;
$\deg(F_j A_j G^{i-1-j}) \leq (r-1)+(j-1)r+j-1+(i-1-j)r = (i-1)r-2+j \leq (i-1)r-2+i = (i-2)r+1+r+i-3 \leq s+s-2 = 2s-2$ (for $3 \leq j \leq i-1$).

The estimation of $\deg(B)$ is easy: $\deg(B) \leq r-1+(i-2)r \leq s+r-2 < 2s-i$ (by $i+r-1 \leq s$).

For $\deg(C)$ the estimation is similar to that of $\deg(A)$:
$\deg(G^{i-2}U) \leq (i-2)r+s \leq s-1+s = 2s-1$;
$\deg(F_2 B_2 G^{i-3}) \leq s-2+r+2+(i-3)r = s+(i-2)r \leq 2s-1$;
$\deg(F_j B_j G^{i-1-j}) \leq (r-1)+(j-1)r+j+(i-1-j)r = (i-1)r-1+j \leq (i-1)r-1+i = (i-2)r+1+r+i-2 \leq s+s-1 = 2s-1$.

Step 5 The graph of f is contained in the union of two lines, which is a contradiction by Theorem 1.5.

Proof We prove that for an affine transform g of f, $g^2(x) = (ax+b)^2$ (as a function), this means that the graph of g is contained in the lines $y = ax + b$ and $y = -ax - b$.

By the previous step we have $(ax+b)f(x)+f^2(x) = dx^2+ex+h$, where $a \neq 0$, $d \neq 0$ ($G = 1$ can be achieved my dividing the equation with the constant G). It is easy to see that for $g(x) = f(x) + (ax + b)/2$ we have $g^2(x) = d'x^2 + e'x + h'$. Now all values of the left hand side is a square, so the right hand side has to be $(a'x + b')^2$.

5. Final remarks

There are two natural ways to generalize our result. One is to consider the problem in $AG(2, q)$ for any prime-power q. After the classification result in [3] for the case $N < \frac{q+3}{2}$, one should ask, whether Theorem 1.2 is true in general and is there a gap in the possible values of N after $\frac{q+3}{2}$. Both seem to be true, though we have a proof for only the case $q = p^2$:

Proposition 5.1 (G-Lovász-Szőnyi) For any prime p, in $AG(2, p^2)$ the only point set determining $\frac{p^2+3}{2}$ directions is the one coming from Example 1.4. There are no examples for $\frac{p^2+3}{2} < N < \frac{p^2+p}{2} + 1$.

This result is sharp, there is a construction by Polverino, Szőnyi and Weiner [10] for a point set determining $\frac{q+\sqrt{q}}{2} + 1$ directions, whenever q is a square.

The other possible generalization is to consider the direction problem in higher dimensions. For this we refer to Sziklai [11] and Storme-Sziklai [12]. We only mention one result from [11] which can be considered as the 3-dimensional analogue of Theorem 1.3.

Theorem 5.2 (Sziklai) Let $U \subset AG(3, p)$ be a point set of size p^2, $p > 11$. Then for the number $N = |D|$ of determined directions one of the following possibilities holds:
(i) U is a plane and $N = p + 1$;
(ii) U is a cylinder with the projective triangle as a base, and $N = 1 + p\frac{p+3}{2}$;
(iii) $N \geq \frac{2}{3}(p-1)p + 2p$.

I would like to thank P. Sziklai and T. Szőnyi for useful conversations.

I am very grateful to one of the referees for lots of corrections and suggestions which made the text much more easy to understand.

Acknowledgements The author acknowledges the financial support of COST Grant 3314 and OTKA Grants F-030737, T-019367, T-029255. The preparation of the final version was supported by FKFP Grant 0063/2001 and the Magyary Scholarship of the "Alapítvány a Magyar Felsőoktatásért és Kutatásért".

References

[1] L. Rédei: Lückenhafte Polynome über endlichen Körpern, *Birkhäuser Verlag, Basel* (1970) (English translation: Lacunary polynomials over finite fields, *North Holland, Amsterdam* (1973)).
[2] A. Blokhuis, A. E. Brouwer, T. Szőnyi: The number of directions determined by a function f on a finite field, *J. Comb. Theory, Ser. A.* **70** (1995), 349-353.
[3] S. Ball, A. Blokhuis, A. Brouwer, L. Storme, T. Szőnyi: On the number of directions determined by a polynomial, *J. Combin. Theory, Ser. A.* **86** (1999), 187-196.
[4] L. Lovász, A. Schrijver: Remarks on a theorem of Rédei, *Studia Scient. Math. Hungar.* **16** (1981), 449-454.
[5] A. Gács: On the number of directions determined by a point set in $AG(2,p)$, *Disc. Math.* **208/209** (1999), 299-309.
[6] A. Gács: On the size of the smallest non-classical blocking set of Rédei type in $PG(2,p)$, p prime, *J. of Combinatorial Theory Ser. A.* **89** (2000), 43-54.
[7] T. Szőnyi: Combinatorial problems for Abelian groups arising from geometry, *Periodica Polytechnica*, **19** (1991), 91-100.
[8] P. Niederreiter, K. H. Robinson: Complete mappings over finite fields, *J. of Australian Math. Soc. Ser. A.* **33** (1982), 197-212.
[9] A. Biró: On polynomials over prime fields taking only two values on the multiplicative group, *Finite Fields and Their Appl.* **6** (2000), 302-308.
[10] O. Polverino, T. Szőnyi and Zs. Weiner: Blocking sets in Galois planes of square order, *Acta Sci. Math. (Szeged)* **65** (1999), 773-784.
[11] P. Sziklai: Directions in $AG(3,p)$ and their applications, submitted to *Note di Matematica, Lecce.*
[12] L. Storme and P. Sziklai: Linear pointsets and Rédei type k-blocking sets in $PG(n,q)$, *J. Alg. Comb.* **14** (2001), 221 -228.