FINDING POINTS ON CURVES OVER FINITE FIELDS

JOACHIM VON ZUR GATHEN[*], IGOR SHPARLINSKI[†], AND ALISTAIR SINCLAIR[‡]

September 17, 2003

Abstract. We solve two computational problems concerning plane algebraic curves over finite fields: generating a uniformly random point, and finding all points deterministically in amortized polynomial time (over a prime field, for non-exceptional curves).

1. Introduction. Let q be a prime power, $\mathbb{F}_q$ a finite field with q elements, $f \in \mathbb{F}_q[x,y]$ of total degree n, and $\mathcal{C} = \{(a,b) \in \mathbb{F}_q^2 : f(a,b) = 0\} = \{f = 0\}$ the plane curve defined by f. We consider two problems of finding points on this curve: probabilistically finding a uniformly distributed random point, and deterministically computing all its points.

Curves over finite fields play a role in several applications: factoring integers with elliptic curves, testing primality with elliptic curves (or more general algebraic varieties), algebro-geometric Goppa codes, and fast multiplication over finite fields. For these applications, special methods for finding points (if needed) are used. This paper presents the first general and systematic approach to the problem, to the authors' knowledge.

Throughout this paper, we will assume that f is squarefree, and denote by σ the number of absolutely irreducible components of $\mathcal{C}$ which are defined over $\mathbb{F}_q$. The famous theorem of Weil says that the number of points $\#\mathcal{C}$ on $\mathcal{C}$ satisfies

$$| \, \#\mathcal{C} - \sigma q \, | \leq n^2 q^{1/2}. \tag{1.1}$$

The case of an *exceptional curve*, corresponding to $\sigma = 0$, needs special treatment and is dealt with in Section 5. So for now we assume that $\sigma \geq 1$.

In Section 2 we provide a polynomial-time solution for the probabilistic variant of our question: generating a uniform random point on $\mathcal{C}$. The algorithm is elementary and is based on the idea of rejection sampling. We also use this algorithm to obtain arbitrarily good probabilistic estimates of $\#\mathcal{C}$.

With deterministic methods, the "brute force" approach to computing all points on $\mathcal{C}$ via finding, for each $a \in \mathbb{F}_q$, all $b \in \mathbb{F}_q$ with $f(a,b) = 0$, takes $O\tilde{~}(n^2 q^{3/2})$ operations in $\mathbb{F}_q$, using the fastest known deterministic algorithms to factor the univariate polynomial $f(a,y)$, for all $a \in \mathbb{F}_q$ (Shoup 1990; Section 1.1 of Shparlinski 1999, von zur Gathen & Shoup 1992). We present in Section 3 a deterministic method that uses $O\tilde{~}(n^5 q)$ operations, i.e., polynomial time per point. The central tool for our estimates is a bound of Perel'muter's (1969) on a certain exponential sum. In order to use this, we have to study in Section 4 some geometric and arithmetic properties of the fibre square $\mathcal{C} \times_\pi \mathcal{C}$. Our approach only works in the case of a prime field $\mathbb{F}_q$, with $q = p$ prime, and does not work for exceptional curves.

Shoup (1990) has exhibited a deterministic univariate factoring algorithm which for almost all polynomials runs in polynomial time. Our deterministic result has two interpretations: the first is that the members of a "small" parametrized family

[*] FB Mathematik-Informatik, Universität Paderborn, 33095 Paderborn, Germany, gathen@upb.de

[†] Department of Computing, Macquarie University, Sydney, NSW 2109, Australia, igor@comp.mq.edu.au

[‡] Computer Science Division, University of California, Berkeley, CA 94720-1776, USA, sinclair@cs.berkeley.edu

$f(a,y)$ of univariate polynomials, for all $a \in \mathbb{F}_p$, can be factored deterministically in (amortized) polynomial time. The second is that all points on a plane algebraic curve over $\mathbb{F}_p$ can be found deterministically in (amortized) polynomial time.

Finally, Section 5 presents a discussion of the case of exceptional curves which has been excluded in the other sections.

A different set of results on our problem (and higher-dimensional varieties) was obtained by Adleman & Huang (2001), Huang & Wong (1999), Huang & Ierardi (1998), and Huang & Wong (1998).

2. Generating uniform random points. In order to generate random points on a plane curve, it is natural to take random points on a coordinate axis and compute points "above" them. So let $\pi\colon \mathcal{C} \to \mathbb{F}_q$ be the projection onto the first coordinate. For $0 \leq i \leq n$ let

$$R_i = \{a \in \mathbb{F}_q : \#\pi^{-1}(\{a\}) = i\}$$

be the set of points with exactly i preimages, and $r_i = \#R_i$. We assume that $\mathcal{C}$ contains no vertical lines, so that no $x-a$ with $a \in \mathbb{F}_q$ divides f. Then $\mathbb{F}_q = \bigcup_{0\leq i\leq n} R_i$ is a partition, and

$$q = \sum_{0\leq j\leq n} r_j, \quad \#\mathcal{C} = \sum_{1\leq j\leq n} jr_j.$$

Algorithm 2.1. Random point.
Input: $f \in \mathbb{F}_q[x,y]$ of degree n.
Output: Either a uniform random point (a,b) on $\mathcal{C} = \{f=0\} \subseteq \mathbb{F}_q^2$, or "failure".

```
1. Choose $a \in \mathbb{F}_q$ uniformly at random.
2. Compute $f_a = \gcd(y^q - y, f(a,y)) \in \mathbb{F}_q[y]$.
3. Choose a random root $b \in \mathbb{F}_q$ of $f_a$. [Then $(a,b) \in \mathcal{C}$.]
4. Set $i = \deg f_a$. [Then $a \in R_i$.]
5. Choose YES with probability $i/n$, and NO with probability $1 - i/n$. If YES was
   chosen, return $(a,b)$, and otherwise return "failure".
```

Theorem 2.2. *Suppose that $\mathcal{C}$ is a nonexceptional curve without vertical lines. Then the algorithm returns a uniform random point on $\mathcal{C}$ with probability*

$$\frac{\#\mathcal{C}}{nq} \geq \frac{1}{n}\left(1 - n^2 q^{-1/2}\right),$$

and "failure" with probability $1 - \#\mathcal{C}/nq$. For every $P \in \mathcal{C}$, P is returned with probability $1/nq$. The algorithm can be performed with an expected number of $O(n \log n \log(nq) \operatorname{loglog} n)$ operations in $\mathbb{F}_q$.

Proof. Let $P = (a,b) \in \mathcal{C}$ with $a \in R_i$. Then

$$\operatorname{prob}\{P \text{ is returned}\} = \frac{1}{q} \cdot \frac{1}{i} \cdot \frac{i}{n} = \frac{1}{nq}.$$

We denote by $\mathsf{M}(n)$ a *multiplication time*, so that the product of two polynomials in $\mathbb{F}_q[x]$ of degree at most n can be computed with $O(\mathsf{M}(n))$ operations in $\mathbb{F}_q$. Then we can take $\mathsf{M}(n) = n \log n \operatorname{loglog} n$, and a gcd can be computed with $O(\mathsf{M}(n) \log n)$ operations. Using repeated squaring to calculate $y^q \bmod f(a, y)$ with $O(\mathsf{M}(n) \log q)$ operations, the cost of step 2 is $O(\mathsf{M}(n) \log(nq))$. The polynomial f_a is a product of $i = \deg f_a$ many linear factors in $\mathbb{F}_q[x]$. If we find a root using the randomized algorithms of Cantor & Zassenhaus (1981), it will be uniformly randomly distributed among these i roots. The algorithm splits the polynomial recursively into two factors, one of which is $\gcd(y^{(q-1)/2} - 1), f_a(y + b)$ for a random $b \in \mathbb{F}_q$, and continues with the smaller factor. (For even q, a different formula is used.) We expect $O(\log i)$ splits to suffice, and each costs $O(M(i) \log(qi))$ operations in $\mathbb{F}_q$. □

We think of q as being much larger than n, say $q \geq c^2 n^4$ for some constant c. Then the success probability of Algorithm 2.1 is at least $\frac{1}{n}(1 - c^{-1})$. Of course, we can increase the success probability by repeated runs of the algorithm.

We can adapt Algorithm 2.1 to obtain an arbitrarily good approximation for $\#\mathcal{C}$, the number of points on $\mathcal{C}$. An (ϵ, δ)-*approximation* ρ to $\#\mathcal{C}$ satisfies

$$\text{prob } \{ \mid \rho - \#\mathcal{C} \mid \leq \epsilon \#\mathcal{C} \} \geq 1 - \delta.$$

To achieve this, we simply run Algorithm 2.1 k times, count the number t of times that YES was chosen in step 5, and return the value $\rho = tnq/k$. Since YES is output with probability $\#\mathcal{C}/nq$, the expected value of ρ is exactly $\#C$, so it is an unbiased estimator. The unbiased estimator theorem of Karp *et al.* (1989) tells us how large k, the number of samples, should be to guarantee an (ϵ, δ)-approximation. This value is

$$k = \lceil 4\beta \log_e(2/\delta)\epsilon^{-2} \rceil, \tag{2.3}$$

where β is an upper bound on $nq/\#\mathcal{C}$. But $nq/\#\mathcal{C} \leq n(1 - n^2 q^{-1/2})^{-1}$, so β is not very large. In fact, assuming as before that $q \gg n^4$, the number of samples required is only about $4n \log_e(2/\delta)\epsilon^{-2}$.

It is even easier in principle to estimate the individual r_i's. We choose k random values $a \in \mathbb{F}_q$, determine for each the j with $a \in R_j$, count the number t of times that $j = i$ occurred, and return the value $\rho_i = tq/k$. This is obviously an unbiased estimator of r_i, and the number of samples required for an (ϵ, δ)-approximation is as in (2.3), where now $\beta = \beta_i$ is an upper bound on q/r_i. With a parameter α, this implies that, by taking

$$k = \lceil 4\alpha n \log_e(2/\delta)\epsilon^{-2} \rceil,$$

we get an (ϵ, δ)-approximation for any r_i satisfying $r_i \geq q/\alpha$. Since

$$n \sum_{1 \leq i \leq n} r_i \geq \sum_{1 \leq i \leq n} i r_i = \#\mathcal{C},$$

the r_i's are on average at least $\#\mathcal{C}/n^2 \geq q(n^{-2} - q^{-1/2})$. Thus "on average" k will only be about $4n^2 \log_e(2/\delta)\epsilon^{-2}$, assuming as before that $q \gg n^4$. Such a value will enable us to estimate the "large" r_i's, though not of course the small ones. In fact, when q is large compared to n^{6n}, then the r_i separate into two classes: Lemma 2.3 of von zur Gathen & Shparlinski (1998) implies that either $r_i \geq \frac{q}{i!(n-i)!} - 2n^{2n} q^{1/2}$ is reasonably large, or $r_i \leq 2n^{2n} q^{1/2}$ is very small. Of course, the "reasonably large" may still be very small, and about q/r_i samples are required. Thus if we use $\beta_i = n!$, then in the first case we obtain an (ϵ, δ)-approximation scheme for r_i, and in the second we expect to find no $a \in R_i$.

Since

$$\frac{\#\mathcal{C}}{n} \leq \sum_{1 \leq i \leq n} r_i \leq \sum_{1 \leq i \leq n} i r_i = \#\mathcal{C},$$

the r_i's are on average at least $\#\mathcal{C}/n^2$. To find approximations only to the "large" r_i's, we might use $\beta_i = \lambda n^2$, with some small number λ.

3. Deterministic construction of all points. In this section, we present a deterministic algorithm for finding all points on $\mathcal{C} = \{f = 0\}$ over a prime field $\mathbb{F}_p$. It employs a deterministic polynomial–time algorithm for finding all roots of the univariate polynomials $f(a, y)$, with $a \in \mathbb{F}_p$. This algorithm does not factor $f(a, y)$ completely for all a, but we show that there are only about $\sqrt{p}$ exceptional a, and for these we use an always successful deterministic algorithm with time about $\sqrt{p}$; thus the total time is proportional to p, which is about the size of $\mathcal{C}$. Everything is polynomial in the degree n.

As a first step, we factor f into irreducible factors in $\mathbb{F}_p[x, y]$. The bivariate factoring algorithms (Lenstra 1985; von zur Gathen 1984; von zur Gathen & Kaltofen 1985) can actually be made into deterministic reductions from bivariate to univariate factorization over finite fields. Thus f can be factored with $n^{O(1)} p^{1/2}$ operations in $\mathbb{F}_p$. From now on, we assume that f is irreducible.

The projection $\pi : \mathcal{C} = \{f = 0\} \to \mathbb{F}_p$ onto the first coordinate is called *separable* if and only if $h_y = \partial h / \partial y \neq 0$ for each irreducible factor $h \in \mathbb{F}_p[x, y]$ of f. A simple example of an inseparable projection is given by $f = x - y^p \in \mathbb{F}_p[x, y]$. The curve $\mathcal{C} = \{x = y^p\}$ is smooth, and all tangents to $\mathcal{C}$ are vertical.

Let $\varphi \colon \mathbb{F}_p \to \mathbb{F}_p$ denote the absolute Frobenius map, with $\varphi(a) = a^p$. For our algorithms, it is convenient to have π separable, and the next lemma describes a simple procedure for achieving this by factoring out φ. (It actually works over any finite field of characteristic p.)

LEMMA 3.1. *Let $f \in \mathbb{F}_p[x, y]$ be irreducible. We can compute in polynomial time $g \in \mathbb{F}_p[x, y]$ and an integer $k \leq \log_p(\deg_y f)$ such that*

$$id \times \varphi^k \colon \mathbb{F}_p^2 \longrightarrow \mathbb{F}_p^2$$

gives a bijection between $\{f = 0\}$ and $\{g = 0\}$, $\deg_x g = \deg_x f, \deg_y g \leq \deg_y f$, and $\pi \colon \{g = 0\} \to \mathbb{F}_p$ is separable.

Proof. We write $f = \sum_{i,j} f_{ij} x^i y^j$, with each $f_{ij} \in \mathbb{F}_p$. Then

$$f_y = 0 \iff \forall i, j \ (f_{ij} \neq 0 \Rightarrow p \mid j).$$

If $f_y = 0$ and

$$h = \sum_{\substack{i,j \\ p \mid j}} f_{ij} x^i y^{j/p} \in \mathbb{F}_p[x, y],$$

then $f(a, b) = h(a, b^p)$ for all $(a, b) \in \mathbb{F}_p^2$, and thus $\mathrm{id} \times \varphi \colon \mathbb{F}_p^2 \to \mathbb{F}_p^2$ gives a bijection between $\{f = 0\}$ and $\{h = 0\}$. Furthermore, h is irreducible. We repeat this process until we obtain a polynomial $g \in \mathbb{F}_p[x, y]$ and $k \in \mathbb{N}$ with $g_y \neq 0$ and $\mathrm{id} \times \varphi^k$ a bijection between $\{f = 0\}$ and $\{g = 0\}$. □

ALGORITHM 3.2. Finding all points.
Input: $f \in \mathbb{F}_p[x,y]$ of degree n, where p is a prime.
Output: A list of all points $(a,b) \in \mathbb{F}_p^2$ with $f(a,b)=0$.

```
1. Set $h = 288n^4\lceil \log_2 p\rceil^2$.
2. For all $a \in \mathbb{F}_p$ do 3–7
3.     Compute $f_a = f(a,y) \in \mathbb{F}_p[y]$.
4.     Compute $f_a^* = \gcd(y^p - y, f_a) \in \mathbb{F}_p[y]$.
5.     For $0 \le t < h$ compute the two factors
           $g_{a,t} = \gcd\big((y-t)^{(p-1)/2} - 1, f_a^*\big),\ g_{a,t}^* = \gcd(y-t, f_a^*) \in \mathbb{F}_p[y]$
       of $f_a^*$.
6.     Compute the common refinement of the partial factorizations from Step 5.
7.     If Step 6 returns only linear factors $y - b$, then add all these $(a,b)$ to the list.
       Otherwise completely factor $f_a^*$ with the deterministic algorithm of von zur
       Gathen & Shoup (1992), and add all resulting $(a,b)$ to the list.
```

THEOREM 3.3. *Let p be a prime, $f \in \mathbb{F}_p[x,y]$ squarefree and non-exceptional, and $\pi\colon \mathcal{C} = \{f=0\} \to \mathbb{F}_p$ separable. Then the algorithm correctly computes all points on $\mathcal{C}$. It uses*

$$O(n^5 p \log n \operatorname{loglog} n \log(np)\log^2 p)$$

or $O\tilde{}(n^5p)$ operations in $\mathbb{F}_p$.

Proof. For all $a,b \in \mathbb{F}_p$ we have

$$f(a,b) = 0 \iff f_a^*(b) = 0 \iff y - b \mid f_a^*.$$

Since Step 7 returns all linear factors of f_a^*, the final list correctly contains all points of $\mathcal{C} = \{f=0\}$.

It remains to analyze the running time. The crucial point is to understand when Step 6 succeeds in completely factoring f_a^*. Denote by $S \subseteq \mathbb{F}_p$ the set of all a for which this is not the case, and $s = \#S$. Furthermore, $\mathcal{C}_a = \pi_2(\mathcal{C} \cap (\{a\} \times \mathbb{F}_p))$ consists of all $b \in \mathbb{F}_p$ with $(a,b) \in \mathcal{C}$. Thus

$$S = \{a \in \mathbb{F}_p : \exists b,c \in \mathcal{C}_a\ b \neq c; b,c \geq h, \text{ and } \forall t < h\ (y-b \mid g_{a,t} \iff y-c \mid g_{a,t})\}.$$

The refinement cost in Step 6, if done along a binary tree, is $O(\mathsf{M}(n)\log n)$ for each t, or $O(h\,\mathsf{M}(n)\log n)$ in total. For $a \in S$, an application of the algorithm from von zur Gathen & Shoup (1992) costs $O(\mathsf{M}(n)\,p^{1/2}\log(np))$ operations in $\mathbb{F}_p$. The gcds in Steps 4 and 5 are computed by repeated squaring for the required power of y and $y-t$, reducing after each multiplication modulo f_a and f_a^*, respectively.

For each a in Step 2, we find the following number of operations in $\mathbb{F}_p$:

- Step 3: $O(n^2)$,
- Step 4: $O(\mathsf{M}(n)\log(np))$,
- Step 5: $O(h\,\mathsf{M}(n)\log(np))$,
- Step 6: $O(h\,\mathsf{M}(n)\log n)$,
- Step 7: 0 if $a \in \mathbb{F}_p \setminus S$, and $O(\mathsf{M}(n)\,p^{1/2}\log(np))$ if $a \in S$.

The total cost is

$$O\Big(p \cdot (n^2 + n^4 \mathsf{M}(n) \log(np) \log^2 p) + s \, \mathsf{M}(n) p^{1/2} \log(np)\Big) \tag{3.4}$$

operations, and we now show that s is $O(n^2(n^2 + \log p)p^{1/2})$. This will imply the claim about the running time. We let

$$Q = \{u \in \mathbb{F}_p^\times : \exists v \in \mathbb{F}_p^\times \; u = v^2\} = \{u \in \mathbb{F}_p^\times : u^{(p-1)/2} = 1\}$$

be the set of nonzero squares in $\mathbb{F}_p$, and χ the quadratic character on $\mathbb{F}_p$, with

$$\chi(b) = \begin{cases} 1, & \text{if } b \in Q, \\ -1, & \text{if } b \notin Q, \; b \neq 0, \\ 0, & \text{if } b = 0. \end{cases}$$

For the time being, we work with an arbitrary integer parameter h; only at the end will we substitute the value from Step 1. Set $H = \{0, \ldots, h-1\} \subseteq \mathbb{F}_p$, where we identify $\mathbb{F}_p$ with $\{0, \ldots, p-1\}$. Two distinct elements $b, c \in \mathbb{F}_p$ are *h-separated* if and only if $\chi(b-t) \neq \chi(c-t)$ for some $t \in H$. A set $B \subseteq \mathbb{F}_p$ is *h-separated* if any two distinct elements of B are. With this notation, we have for $a \in \mathbb{F}_p$

$$a \in S \Longrightarrow \mathcal{C}_a \text{ is not } h\text{-separated.}$$

The reverse implication is true if the non-h-separated $b, c \in \mathcal{C}_a$ are both at least h. If $a \in S$, then for at least one pair of distinct elements $b, c \in \mathcal{C}_a$,

$$h = \sum_{0 \le t < h} \chi\big((t-b)(t-c)\big).$$

Now we let $k \in \mathbb{N}$ and

$$\begin{aligned} w &= \sum_{a \in \mathbb{F}_p} \sum_{\substack{b,c \in \mathcal{C}_a \\ b \neq c}} \Big| \sum_{0 \le t < h} \chi\big((t-b)(t-c)\big) \Big|^{2k} \\ &= \sum_{0 \le t_1, \ldots, t_{2k} < h} \; \sum_{a \in \mathbb{F}_p} \sum_{\substack{b,c \in \mathcal{C}_a \\ b \neq c}} \chi\big((t_1 - b)(t_1 - c) \cdots (t_{2k} - b)(t_{2k} - c)\big). \end{aligned}$$

Then, by the above, $sh^{2k} \le w$. We consider the set

$$\mathcal{D}_0 = \{(a, b, c) \in \mathbb{F}_p^3 : f(a, b) = f(a, c) = 0, b \neq c\} \subseteq \mathbb{F}_p^3.$$

The fibre product $\mathcal{D} = \mathcal{C} \times_\pi \mathcal{C}$ is the closure of $\mathcal{D}_0$ in $\mathbb{F}_p^3$; it has degree at most $n(n-1) < n^2$ and is discussed in detail in Section 4. Then

$$w = \sum_{t \in H^{2k}} \sum_{P \in \mathcal{D}} \chi\big(\psi_t(P)\big),$$

where the inner sum is over all $\mathbb{F}_p$-rational points $P = (a, b, c) \in \mathcal{D}$ with $b \neq c$, ψ_t is the polynomial

$$\psi_t = (y - t_1) \cdots (y - t_{2k})(z - t_1) \cdots (z - t_{2k}) \in \mathbb{F}_p[y, z]$$

in indeterminates y and z, and $\psi_t\big((a,b,c)\big)$ is obtained by substituting b and c for y and z, respectively.

Theorem 4.6 below says that there are at most $(12kn^2h^{1/2})^{2k}$ values of $t \in H^{2k}$ for which $\rho(\psi_t)$ is a square in the global ring $\mathcal{O}_{\mathcal{A}}$ of some irreducible component $\mathcal{A} \subseteq \mathbb{F}^3$ of $\mathcal{D}$, where $\rho \colon \mathbb{F}[x,y,z] \longrightarrow \mathcal{O}_{\mathcal{A}}$ is the restriction map.

For other vectors $t \in \mathbb{F}^{2k}$, we may apply the bound on character sums along a curve from Perel'muter (1969) that gives

$$\sum_{P \in \mathcal{D}} \chi\big(\psi_t(P)\big) \le d \cdot \big(n^2(n^2+2k)p^{1/2}\big) \tag{3.5}$$

for some constant d. Perel'muter's bound holds for each irreducible component of $\mathcal{D}$; we also use the fact that no such component is vertical (Lemma 3.1 of von zur Gathen *et al.* 1996). Since their degrees sum to $\deg \mathcal{D} < n^2$, (3.5) follows. Therefore

$$w \le (12kn^2h^{1/2})^{2k}p + d \cdot n^2(n^2+2k)h^{2k}p^{1/2},$$

$$s \le (12kn^2h^{-1/2})^{2k}p + d \cdot n^2(n^2+2k)p^{1/2}.$$

Now, using $k = \lceil \log_2 p \rceil$ and h as in Step 1 of Algorithm Algorithm 3.2, we find

$$(12kn^2h^{-1/2})^{2k} \le 2^{-k} \le 2^{-\log_2 p} = p^{-1}.$$

Hence

$$s = O\big(n^2(n^2+\log p)p^{1/2}\big).$$

Together with (3.4), this proves the estimate of the total cost. □

4. Squares on the fibre product. The goal of this section is to bound the number of products Ψ_t which are squares on some irreducible component of $\mathcal{D}$; this was used in the previous proof.

Let $\mathbb{F}$ be an algebraically closed field, $f \in \mathbb{F}[x,y]$ squarefree of degree $n \ge 1$, $\mathcal{C} = \{f = 0\} \subseteq \mathbb{F}^2$ the associated plane curve, and $\pi \colon \mathcal{C} \longrightarrow \mathbb{F}$ the first projection. We assume that π is separable. Then $\mathcal{D} = \mathcal{C} \times_\pi \mathcal{C} \subseteq \mathbb{F}^3$, the fibre square over π, can be defined as the closure in $\mathbb{F}^3$ of

$$\mathcal{D}_0 = \{(a,b,c) \in \mathbb{F}^3 : f(a,b) = f(a,c) = 0, b \ne c\}.$$

Furthermore, let $g = (f(x,y) - f(x,z))/(y-z) \in \mathbb{F}[x,y,z]$.

A smooth point $P = (a,b) \in \mathcal{C}$ is *critical* for π if and only if the tangent line $T_{P,\mathcal{C}}$ in $\mathbb{F}^2$ is vertical. If f is irreducible, this is equivalent to $f_y(a,b) = 0$, where $f_y = \partial f/\partial y \in \mathbb{F}[x,y]$; in general, we have to replace f by its (unique) irreducible factor on whose component P lies. Since π is separable, $\mathcal{C}$ has only finitely many critical points.

THEOREM 4.1. *Let $f \in \mathbb{F}[x,y]$ be squarefree and π separable.*

(i) $\mathcal{D} = \{f(x,y) = g(x,y,z) = 0\}$.

(ii) $\mathcal{D} = \mathcal{D}_0 \cup \{(a,b,b) \colon (a,b) \in \mathcal{C}$ *is singular or critical*$\}$.

(iii) $(a,b,c) \in \mathcal{D}$ *with* $b \ne c$ *is singular on* $\mathcal{D}$ *if and only if either* (a,b) *or* (a,c) *is singular on* $\mathcal{C}$, *or both* (a,b) *and* (a,c) *are critical on* $\mathcal{C}$. *All points of* $\mathcal{D} \setminus \mathcal{D}_0$ *are singular on* $\mathcal{D}$.

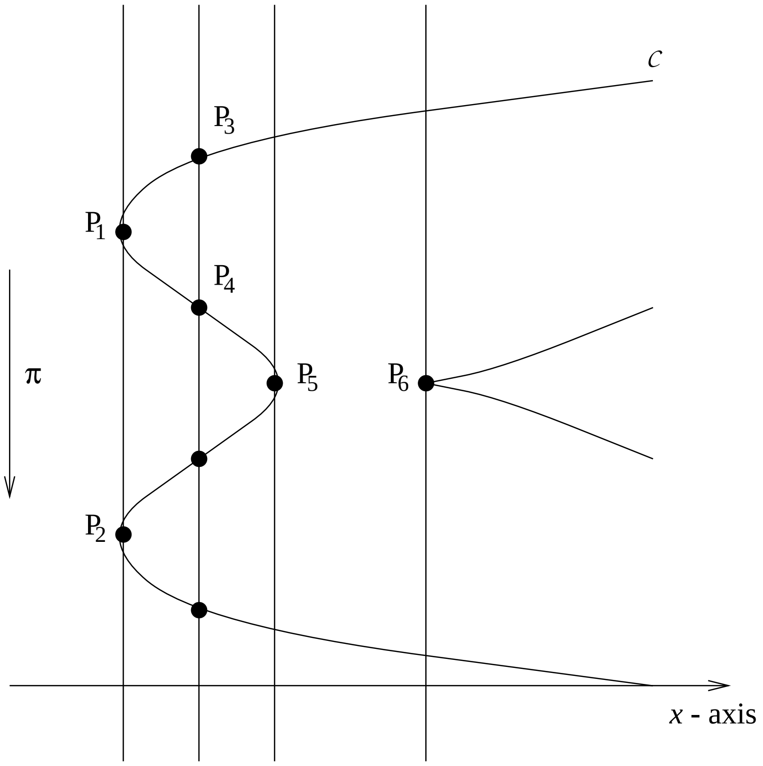


FIG. 4.1. *P_1, P_2, P_5 are critical for π, and P_6 is singular on $\mathcal{C}$. If $P_i = (a_i, b_i)$, then $(a_i, b_i, b_i) \in \mathcal{D} \cap \Delta$ for $i = 1, 2, 5, 6$. These four points are singular on $\mathcal{D}$. Furthermore, $(a_1, b_1, b_2) \in \mathcal{S} \subseteq \mathcal{D}$, and $(a_3, b_3, b_4) \in \mathcal{D} \setminus \mathcal{S}$.*

(iv) $\deg \mathcal{D} \le n(n-1) < n^2$.

Proof. Let

$$\Delta = \{(a, b, b) \in \mathbb{F}^3 : a, b \in \mathbb{F}\}, \quad \mathcal{D}_1 = \{f(x, y) = g(x, y, z) = 0\},$$

so that Δ is the diagonal. Clearly $\mathcal{D} \setminus \Delta = \mathcal{D}_0$, and $\mathcal{D}_0 \subseteq \mathcal{D}_1$. By definition, $\mathcal{D}$ is the closure of $\mathcal{D}_0$, and thus $\mathcal{D} \subseteq \mathcal{D}_1$. We prove in the following that (ii) is valid with $\mathcal{D}_1$ instead of $\mathcal{D}$. Thus $\mathcal{D}_1 \cap \Delta$ is finite, and $\mathcal{D} = \mathcal{D}_1$ follows, hence (i), (ii), and (iv).

So let u, v be indeterminates over $\mathbb{F}[x, y]$. Then the Taylor expansion of f around (u, v) of order 1 is

$$f(x, y) = f(u, v) + f_x(u, v)(x - u) + f_y(u, v)(y - v) + h$$

in $\mathbb{F}[x, y, u, v]$, with some $h \in (x - u, y - v)^2$. Therefore

$$\begin{aligned} & g(x, y, z) \\ & = \frac{1}{y - z} \cdot (f_y(u, v)(y - v) - f_y(u, v)(z - w) + h(x, y, u, v) - h(x, z, u, v)) \\ & = f_y(u, v) + H, \end{aligned}$$

with some $H \in (x - u, y - v, z - v)$. Thus for $(a, b) \in \mathcal{C}$

$$(a, b, b) \in \mathcal{D}_1 \iff f_y(a, b) = 0 \iff (a, b) \text{ is singular or critical on } \mathcal{C}.$$

For (iii), let $(a,b,c) \in \mathcal{D}$ with $b \neq c$. The Jacobian of $\mathcal{D}$ at (a,b,c) is

$$J(a,b,c) = \begin{pmatrix} f_x(a,b) & \dfrac{f_x(a,b) - f_x(a,c)}{b-c} \\ f_y(a,b) & \dfrac{f_y(a,b)}{b-c} \\ 0 & \dfrac{-f_y(a,c)}{b-c} \end{pmatrix}.$$

After multiplying the second column by $c-b$ and then adding the first column to the second, we obtain the matrix

$$A = \begin{pmatrix} f_x(a,b) & f_x(a,c) \\ f_y(a,b) & 0 \\ 0 & f_y(a,c) \end{pmatrix}.$$

Thus
(a,b,c) is singular on $\mathcal{D} \iff$ rank $(J(a,b,c)) \leq 1$
$\iff$ rank $(A) \leq 1$
$\iff$ (a,b) or (a,c) is singular on $\mathcal{C}$, or both are critical on $\mathcal{C}$. □

The condition that π be separable is necessary, since otherwise all points on $\mathcal{C}$ are critical. Recall the example $\mathcal{C} = \{x = y^p\}$, where $p = \operatorname{char} \mathbb{F}$, from Section 3. Then $f_y = 0$, $\mathcal{C}$ is smooth, and all tangent lines to $\mathcal{C}$ are vertical. Furthermore, $\mathcal{D}_0 = \emptyset$, $g = (y^p - z^p)/(y-z) = (y-z)^{p-1}$, and $\mathcal{C} \times_\pi \mathcal{C}$ equals $\{(a,b,b) \in \mathbb{F}^3 : a = b^p\}$, counted $p-1$ times. On the other hand, when $\mathcal{C} = \{y = g(x)\}$ is the graph of a polynomial $g \in \mathbb{F}_q[x]$, then π is separable, and $\mathcal{D} = \emptyset$.

We define

$$\mathcal{S} = \{(a,b,c) \in \mathcal{D} \colon (a,b) \text{ or } (a,c) \text{ is singular or critical on } \mathcal{C}\}.$$

We now let $\mathcal{A}$ be an irreducible component of $\mathcal{D}$, and want to estimate the number of t such that

$$\psi_t = \prod_{1 \leq i \leq 2k} (t_i - y)(t_i - z)$$

is a square in $\mathcal{O}_{\mathcal{A}}$. We let $\rho \colon \mathbb{F}[x,y,z] \longrightarrow \mathcal{O}_{\mathcal{A}}$ be the restriction map.

Let $t \in \mathbb{F}^{2k}$, and $T = \{1, \ldots, 2k\}$. The overall goal of this section is to show in Theorem 4.6 that only few $\rho(\psi_t)$ are squares, when t is chosen from a finite subset H of $\mathbb{F}^{2k}$. For a simple example of a square, we take the parabola $f = x - y^2$, so that $\mathcal{C} = \{x = y^2\}$, and $\mathcal{D} = \{x - y^2 = y + z = 0\}$ is irreducible. If $k = 1$ and $t_2 = -t_1$, then

$$\rho(\psi_t) = \rho((t_1 - y)(t_1 - z)(t_2 - y)(t_2 - z)) = \rho((t_1 - y)^2 (t_1 + y)^2) \tag{4.2}$$

is a square on $\mathcal{D}$.

The condition that $\rho(\psi_t)$ not be a square for (3.5) to hold is not an artifact of Perel'muter's proof, but without it (3.5) may actually fail to be true.

In the sequel, we define several combinatorial objects on the index set T. We first collect pairs of equal values of t_i in a systematic way. Namely, we take the lexicographically first maximal matching on the directed graph with vertex set T, and where (i,j) are connected if and only if $i < j$ and $t_i = t_j$. Then $T_1 \subseteq T$ is defined as the set of these first coordinates i, and $\tau_1 : T_1 \to T$ is defined by $\tau_1(i) = j$ if (i, j) occurs in that matching. As an example, if $t_3 = t_5 = t_8 = t_{11} = t_{13}$ and no other t_i equals these, then $T_1 = \{3, 5\}$, $\tau_1(3) = 8$, and $\tau_1(5) = 11$.

Next, we set

$$T_2 = \{i \in T \setminus (T_1 \cup \tau_1(T_1)) \colon \mathcal{A} \cap \{y = t_i\} \subseteq \mathcal{S} \text{ or } \mathcal{A} \cap \{z = t_i\} \subseteq \mathcal{S}\}.$$

Then the t_i for

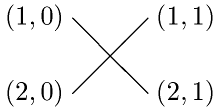
$$i \in T_3 = T \setminus (T_1 \cup \tau_1(T_1) \cup T_2)$$

are pairwise distinct, and $(T_1, \tau_1(T_1), T_2, T_3)$ is a partition of T. Next, we let

$$S_0 = T_3 \times \{0\}, \quad S_1 = T_3 \times \{1\}$$

be two disjoint copies of T_3, and define a bipartite undirected graph $G = (S_0 \cup S_1, E)$ as follows. For $i, j \in T_3$, $(i, 0)$ and $(j, 1)$ are connected in G if and only if there is some $(a, b, c) \in \mathcal{A} \setminus \mathcal{S}$ such that $b = t_i$ and $c = t_j$.

In the example (4.2) of a parabola, we have $T_1 = T_2 = \emptyset$, and

$G =$  .

Lemma 4.3. *If $t \in \mathbb{F}^{2k}$ is such that $\rho(\psi_t) \in \mathcal{O}_{\mathcal{A}}$ is a square, then each vertex in G has degree at least one.*

Proof. By symmetry, it is sufficient to show the claim for a vertex $(i, 0) \in S_0$.

Since $i \notin T_2$, we can choose some $P = (a, t_i, c) \in \mathcal{A} \setminus \mathcal{S}$; then $c \neq t_i$. Let

$$U_0 = \{j \in T \colon t_j = t_i\}, \quad U_1 = \{j \in T \colon t_j = c\},$$

$\rho \colon \mathbb{F}[x, y, z] \to \mathcal{O}_{\mathcal{A}}$ the restriction to $\mathcal{A}$, $\mathcal{R} = \mathcal{O}_{P,\mathcal{A}}$ the local ring at P, which is a Unique Factorization Domain (see e. g. Shafarevich 1974, Theorem II.3.2), and $\lambda = (\mathcal{O}_{\mathcal{A}} \to \mathcal{O}_{P,\mathcal{A}}) \circ \rho$ the composition of ρ with the localization at P. Then $i \in U_0$ and $U_0, U_1 \subseteq T \setminus T_2$.

For every $j \in T \setminus (U_0 \cup \tau_1(U_0) \cup \{i\})$, we have $t_j \neq t_i$, and thus $\lambda(y - t_j)$ is a unit in $\mathcal{R}$. Similarly, each $\lambda(z - t_j)$ with $t_j \neq c$ is a unit in $\mathcal{R}$. Since $(a, t_i) \in \mathcal{C}$ is not critical for π, we have $f_y(a, t_i) \neq 0$, and therefore $\lambda(y - t_i) \in \mathcal{R}$ is a local parameter in $\mathcal{R}$. Similarly, each $\lambda(z - t_j)$ with $t_j = c$ is a local parameter in $\mathcal{R}$.

By the above, there is a unit $u \in \mathcal{R}$ such that

$$\begin{aligned} \lambda(\psi_t) &= \prod_{j \in T} \lambda(y - t_j) \cdot \prod_{j \in T} \lambda(z - t_j) \\ &= u \cdot \prod_{j \in U_0 \cup \tau_1(U_0) \cup \{i\}} \lambda(y - t_j) \cdot \prod_{j \in U_1} \lambda(z - t_j) \end{aligned}$$

is a square in $\mathcal{R}$. Thus the total number of local parameters in the product is even. We have $\#U_0 = \#\tau_1(U_0)$ and $i \notin U_0 \cup \tau_1(U_0)$. It follows that in the left hand product, the number of local parameters is odd, and therefore also in the right hand product. Thus there exists some $j \in T_3$ with $t_j = c$; then $\{(i, 0), (j, 1)\} \in E$. □

We now take a maximal "disjoint" matching (V_0, V_1) in G of the following type. The sets $V_0, V_1 \subseteq T_3$ are disjoint, G induces a perfect matching on $(V_0 \times \{0\}) \cup (V_1 \times \{1\})$, and this matching is maximal. Furthermore, let $\mu : V_0 \longrightarrow V_1$ be the corresponding bijection, with $\mu(i) = j$ if and only if $\{(i,0),(j,1)\}$ occurs in the matching.

For every $i \in V_2 = T_3 \setminus (V_0 \cup V_1)$, $(i,0)$ is connected to some $(j,1) \in T_3 \times \{1\}$, and by the maximality of the matching, we have $j \in V_0 \cup V_1$. We take $\mu\colon V_2 \longrightarrow V_0 \cup V_1$ such that $\mu(i) = j$ for some such j, and note that (V_0, V_1, V_2) is a partition of T_3.

Finally, we indicate how to describe t_i for $i \in V_0$ succinctly if $\{(i,0),(j,1)\} \in E$ and t_j is known. For this, we take an arbitrary total order $\prec$ on $\mathbb{F}$. For each $t \in \mathbb{F}$, $\mathcal{C} \cap \{y = t\}$ has at most n points, say $(a_1, t), \ldots, (a_l, t)$ with $l \leq n$ and $a_1 \prec \cdots \prec a_l$. If $j = \mu(i)$ and $t = t_j$, then $(a_r, t_i, t_j) \in \mathcal{D} \setminus \mathcal{S}$ for one of those points, with $1 \leq r \leq l$. We choose the smallest such r; then $\mathcal{C} \cap \{x = a_r\}$ consists again of at most n points. We let v be the position of (a_r, t_i) in this list, ordered according to $\prec$, and set $\tau_3(i) = (r, v)$. Then t_i is determined by $j = \mu(i)$, t_j, and $\tau_3(i)$.

Similarly, we define $\tau_3\colon V_2 \longrightarrow \{1, \ldots, n\}^2$ so that for $i \in V_2$, t_i is determined by $j = \mu(i)$, t_j, and $\tau_3(i)$.

We have thus associated to any $t \in \mathbb{F}^{2k}$ with $\rho(\psi_t)$ a square the following data:

$$T_1, \tau_1, T_2, V_0, \mu, \tau_3, \text{ and } t_i \text{ for } i \in T_1 \cup T_2 \cup V_1. \tag{4.4}$$

LEMMA 4.5. *If $\rho(\psi_t)$ is a square in $\mathcal{O}_{\mathcal{A}}$, then t is determined by the data in (4.4).*

Proof. $(T_1, \tau_1(T_1), T_2, V_0, V_1, V_2)$ is a partition of T, and $t_i = t_{\tau_2(i)}$ for each $i \in T_1$. Thus it remains to show that each t_i with $i \in V_0 \cup V_2$ is determined by (4.4). But that is precisely what the construction of μ and τ_3 achieves. □

We are now ready for the main result of this section, an upper bound on the number of ψ_t which are squares. The bound is rather coarse, but sufficient for our purposes.

THEOREM 4.6. *Let $\mathbb{F}$ be an algebraically closed field, $f \in \mathbb{F}[x,y]$ squarefree, $\mathcal{C} = \{f = 0\}$ with $\pi\colon \mathcal{C} \to \mathbb{F}$ separable, $H \subseteq \mathbb{F}$ be a finite set with h elements, and $k \in \mathbb{N}$ positive. The number of $t \in H^{2k}$ such that $\rho(\psi_t)$ is a square in $\mathcal{O}_{\mathcal{A}}$ for some irreducible component $\mathcal{A}$ of $\mathcal{C} \times_\pi \mathcal{C}$ is at most $(12kn^2h^{1/2})^{2k}$.*

Proof. We first fix a component $\mathcal{A}$ of $\mathcal{D}$, and show the corresponding bound. By Lemma 4.5, it is sufficient to give an upper bound on the number of choices for the data in (4.4).

The six sets $T_1, \tau_1(T_1), T_2, V_0, V_1, V_2$ form a partition of T, and there are at most 6^{2k} choices for this partition.

Suppose that these sets are chosen, with cardinalities $c_1, c_2, c_3, c_4, c_5, c_6$, respectively. Then $c_1 = c_2$, $c_3 < n^2$, and $c_4 = c_5$. The number of choices for τ_1 is at most $(2k)^{c_1}$, for μ at most $(2k)^{c_4+c_6}$, for τ_3 at most $(n^2)^{c_4+c_6}$, and for all t_i's required in (Theorem 4.1) at most $h^{c_1+c_5} \cdot (n^2)^{c_3}$. Since $c_1 + c_5 \leq 2k/2 = k$, the total comes to

$$m = 6^{2k} \cdot (2k)^{c_1+c_4+c_6} \cdot (n^2)^{c_3+c_4+c_6} \cdot h^{c_1+c_5}. \tag{4.7}$$

Since $\deg \mathcal{D} \leq n(n-1)$ by Theorem 4.1 (i), $\mathcal{D}$ has at most $n(n-1) < n^2$ irreducible components. So the total number of t considered is at most $n^2 m$, and

$$n^2 m \leq 6^{2k} \cdot (2k)^{2k} \cdot (n^2 h^{1/2})^{2k}.$$

Here we use that either $c_1 + c_2 + c_5 > 0$ and then $n^2 \cdot (n^2)^{c_3+c_4+c_6} \leq (n^2)^{2k}$, or $c_2 + c_3 + c_4 + c_6 > 0$ and then $n^2 (h)^{c_1+c_5} \leq h^k$. □

5. Exceptional polynomials. In this section, we deal with the somewhat troublesome case excluded so far: exceptional polynomials, for which $\sigma = 0$. No analogue of the deterministic result of Theorem 3.3 is known for them, while the probabilistic results of Section 2 carry over easily.

We first note that it is not surprising that they are difficult to deal with, since any subset of $\mathbb{F}_q^2$ is an exceptional curve. If $c \in \mathbb{F}_q$ is a nonsquare and $f = x^2 + cy^2$, then f is exceptional and

$$\{f = 0\} = \{(0,0)\}, \tag{5.1}$$

and by translation and finite unions the claim follows. If $\operatorname{char} \mathbb{F}_q \geq 3$, then (5.1) also holds for $f = x^{q-1} + y^{q-1}$. If $b \in \mathbb{F}_{q^2} \backslash \mathbb{F}_q$ with $b^2 \in \mathbb{F}_q$, then $b^{q-1} = (b^2)^{(q-1)/2} = -1$. Thus f is the product of all $x - by$ with these b, and thus f is exceptional, too.

Now given an arbitrary $f \in \mathbb{F}_q[x,y]$ of degree n, there are well-known probabilistic algorithms with time polynomial in $n \log q$ that factor f into its irreducible factors over $\mathbb{F}_q$ (von zur Gathen & Kaltofen 1985) and test each such factor for absolute irreducibility (Kaltofen 1985). For simplicity, assume now that f is irreducible over $\mathbb{F}_q$, and not absolutely irreducible. Then Kaltofen's algorithm can be used to find a field extension K of $\mathbb{F}_q$ with $[K : \mathbb{F}_q] \leq n$ and a proper factorization of f over K. If g and h are two distinct factors, then the first coordinate of any common root is a root of

$$\operatorname{res}_y(g,h) \in K[x].$$

Thus it is easy to calculate all common roots of g and h, to check which ones are in $\mathbb{F}_q^2$, and to determine whether they are indeed roots of f. All roots of f are found in this way; there are at most $n^2/4$ of them (von zur Gathen *et al.* 1996).

Theorem 5.2. *Let $f \in \mathbb{F}_q[x,y]$ have degree n. There is a probabilistic algorithm using $(n \log q)^{O(1)}$ operations in $\mathbb{F}_q$ that determines whether f is exceptional and, if it is, finds all points of $\{f = 0\}$.*

Acknowledgements. Parts of this work were done on visits of the first author to Macquarie University and to ICSI, Berkeley, and of the second author to Universität Paderborn. The hospitality of these institutions is gratefully acknowledged. The third author was supported in part by NSF grant CCR-9505448 and by the International Computer Science Institute. A preliminary version of this paper appeared in the Proceedings 36th Annual IEEE Symp. Foundations of Computer Science, Milwaukee WI, 1995, 284-292.

References.

Leonard M. Adleman & M. D. Huang (2001). Counting points on curves and abelian varieties over finite fields. *Journal of Symbolic Computation* **32**, 171–189.

E. R. Berlekamp (1970). Factoring Polynomials Over Large Finite Fields. *Mathematics of Computation* **24**(11), 713–735.

David G. Cantor & Hans Zassenhaus (1981). A New Algorithm for Factoring Polynomials Over Finite Fields. *Mathematics of Computation* **36**(154), 587–592.

Joachim von zur Gathen (1984). Hensel and Newton methods in valuation rings. *Mathematics of Computation* **42**(166), 637–661.

J. von zur Gathen & E. Kaltofen (1985). Factorization of Multivariate Polynomials Over Finite Fields. *Mathematics of Computation* **45**, 251–261.

Joachim von zur Gathen, Marek Karpinski & Igor E. Shparlinski (1996). Counting curves and their projections. *computational complexity* **6**, 64–99.

Joachim von zur Gathen & Victor Shoup (1992). Computing Frobenius maps and factoring polynomials. *computational complexity* **2**, 187–224.

Joachim von zur Gathen & Igor E. Shparlinski (1998). Computing components and projections of curves over finite fields. *SIAM Journal on Computing* **28**(3), 822–840. URL `http://epubs.siam.org/sam-bin/dbq/article/27741`.

M.-D. Huang & D. Ierardi (1998). Counting Points on Curves over Finite Fields. *Journal of Symbolic Computation* **25**, 1–21.

M. D. Huang & Y. C. Wong (1999). Solvability of systems of polynomial congruences modulo a large prime. *computational complexity* **8**, 227–257.

Ming-Deh Huang & Yiu-Chung Wong (1998). An algorithm for approximate counting of points on algebraic sets over finite fields. In *Algorithmic Number Theory, Proceedings ANTS-III,* Portland OR, J. P. Buhler, editor, number 1423 in Lecture Notes in Computer Science, 514–527. Springer-Verlag. ISSN 0302-9743.

E. Kaltofen (1985). Fast parallel absolute irreducibility testing. *Journal of Symbolic Computation* **1**, 57–67.

R. M. Karp, M. Luby & N. Madras (1989). Monte-Carlo approximation algorithms for enumeration problems. *Journal of Algorithms* **10**(3), 429–448.

Arjen K. Lenstra (1985). Factoring Multivariate Polynomials over Finite Fields. *Journal of Computer and System Sciences* **30**, 235–248.

G. I. Perel'muter (1969). Оценка суммы вдоль алгебраической кривой (Bounds on sums along algebraic curves). *Mat. Zametki* **5**, 373–380.

I. R. Shafarevich (1974). *Basic algebraic geometry.* Number 213 in Grundlehren. Springer-Verlag.

Victor Shoup (1990). On the deterministic complexity of factoring polynomials over finite fields. *Information Processing Letters* **33**, 261–267.

Igor E. Shparlinski (1999). *Finite Fields: Theory and Computation.* Mathematics and Its Applications. Kluwer Academic Publishers, Dordrecht/Boston/London, 528+xiv pp.