

Factorization patterns on nonlinear families of univariate polynomials over a finite field

Guillermo Matera

Universidad de Buenos Aires and CONICET

Joint work with [Mariana Pérez](#) and [Melina Privitelli](#)

Universidad Nacional de General Sarmiento and CONICET

Carleton Finite Fields eSeminar
July 22, 2020

Notations

\mathbb{F}_q finite field of q elements and characteristic p ,

$\overline{\mathbb{F}}_q$ algebraic closure of \mathbb{F}_q ,

$M(r)$ the set of monic polynomials of $\mathbb{F}_q[T]$ of degree r .

Let $\lambda := 1^{\lambda_1} \dots r^{\lambda_r}$ be such that $r = \lambda_1 + 2\lambda_2 + \dots + r\lambda_r$.

$f \in M(r)$ has **factorization pattern** λ if it has λ_i irreducible factors of degree i in $\mathbb{F}_q[T]$ for $1 \leq i \leq r$.

For $A \subset M(r)$, we denote

$$\mathcal{T}_\lambda(A) := |\{f \in A : f \text{ has factorization pattern } \lambda\}|.$$

Factorization patterns on linear families

[Cohen, Acta Arith. 17, 1970] For fixed r ,

$$\mathcal{T}_\lambda(M(r)) = \mathcal{T}(\lambda) q^r + \mathcal{O}(q^{r-\frac{1}{2}}),$$

where $\mathcal{T}(\lambda)$ is the proportion of elements in the r th symmetric group with cycle pattern λ .

Examples

- If $\lambda := (0, \dots, 0, 1)$ (irreducible polynomials), then

$$\mathcal{T}(\lambda) = \frac{1}{r} \quad \text{and} \quad \mathcal{T}_\lambda(M(r)) \approx \frac{q^r}{r} \quad (\text{Gauss}).$$

- For $\lambda := (r, 0, \dots, 0)$ (linear factors),

$$\mathcal{T}(\lambda) = \frac{1}{r!} \quad \text{and} \quad \mathcal{T}_\lambda(M(r)) \approx \frac{q^r}{r!}.$$

We call $A \subset M(r)$ **uniformly distributed** if

$$\mathcal{T}_\lambda(A) \sim \mathcal{T}(\lambda)|A|.$$

[Cohen, J. London Math. Soc. 6, 1972] Let $p > r$. Then the following **linear families** $A \subset M(r)$ are uniformly distributed:

- The elements of $M(r)$ with **s coefficients preassigned** (assuming $A \not\subset \mathbb{F}_q[T^l]$ for any $l > 1$);
- $C_r(f, g) := \{h \in M(r) : h \equiv f \pmod{g}\}$, where $f, g \in \mathbb{F}_q[T]$ are relatively prime with $\deg f < r$.

Factorization patterns on linear families

[Bank et al, Duke J Math 164, 2015] For any characteristic p , the following linear families $A \subset M(r)$ are uniformly distributed:

- The elements of $M(r)$ with the first $r - s \geq 3$ consecutive coefficients preassigned;
- $C_r(f, g) := \{h \in M(r) : h \equiv f \pmod{g}\}$, where $f, g \in \mathbb{F}_q[T]$ are relatively prime with $\deg f \leq r - 4$.

[Cesaratto, M, Pérez, Combinatorica 37, 2017] For characteristic $p > 2$, let $A_s \subset M(r)$ be set of $f \in M(r)$ with the first $r - s \geq 3$ consecutive coefficients preassigned. Then

$$|\mathcal{T}_\lambda(A_s) - \mathcal{T}(\lambda)q^s| \leq q^{s-1} \left(2\mathcal{T}(\lambda)rs \frac{(r-1)!}{(r-s)!} q^{\frac{1}{2}} + 20\mathcal{T}(\lambda)r^2s^2 \frac{(r-1)!^2}{(r-s)!^2} \right).$$

Factorization patterns on nonlinear families

Problem 2.2 of [Gao, Howell, Panario, Proc. Fq4, 1999] asks for estimates on the number of polynomials of a given degree with a given factorization pattern lying in **nonlinear families**:

for $m < r$, indeterminates $\mathbf{A} := (A_{r-1}, \dots, A_0)$ over $\overline{\mathbb{F}}_q$, and $G_1, \dots, G_m \in \mathbb{F}_q[\mathbf{A}]$, consider the **algebraic variety**

$$W = \{\mathbf{a} \in \overline{\mathbb{F}}_q^r : G_1(\mathbf{a}) = 0, \dots, G_m(\mathbf{a}) = 0\},$$

and the family

$$\mathcal{A} := \{T^r + a_{r-1}T^{r-1} + \dots + a_0 \in M(r) : (a_{r-1}, \dots, a_0) \in W\}.$$

[Chatzidakis et al., J. Reine Angew. Math. 427, 1992]

[Fried et al., Israel J. Math. 85, 1994] Let $n := \dim W$. There is a constant $d \geq 0$ such that, for large q ,

$$|\mathcal{T}_\lambda(\mathcal{A})| = d\mathcal{T}(\lambda)q^n + \mathcal{O}(q^{n-\frac{1}{2}}).$$

Factorization patterns on nonlinear families

Aims:

- provide a general criterion for a **nonlinear family** $\mathcal{A} \subset M(r)$ to be **uniformly distributed** (in the sense of Cohen);
- find **explicit estimates** on $|\mathcal{A}_\lambda|$ for any factorization pattern λ .

For a fixed k , let $\mathbb{F}_q[\mathbf{A}_k] := \mathbb{F}_q[A_{r-1}, \dots, A_{k+1}, A_{k-1}, \dots, A_0]$, let $G_1, \dots, G_m \in \mathbb{F}_q[\mathbf{A}_k]$ and $W := \{G_1 = 0, \dots, G_m = 0\}$. Let $\mathcal{A} := \{T^r + a_{r-1}T^{r-1} + \dots + a_0 \in M(r) : G_i(\mathbf{a}_k) = 0 \ (1 \leq i \leq m)\}$.

For the weight $\text{wt} : \mathbb{F}_q[\mathbf{A}_k] \rightarrow \mathbb{N}_0$, $\text{wt}(A_j) := r - j$ ($0 \leq j \leq r - 1$), denote by $G_1^{\text{wt}}, \dots, G_m^{\text{wt}}$ the **components of highest weight** of G_1, \dots, G_m . Let $(\partial \mathbf{G} / \partial \mathbf{A}_k)$ be the **Jacobian matrix** of G_1, \dots, G_m with respect to \mathbf{A}_k . Assume that G_1, \dots, G_m satisfy the conditions:

- (H₁) G_1, \dots, G_m form a **regular sequence** of $\mathbb{F}_q[\mathbf{A}_k]$.
- (H₂) $(\partial \mathbf{G} / \partial \mathbf{A}_k)$ has **full rank** on every point of W .
- (H₃) $G_1^{\text{wt}}, \dots, G_m^{\text{wt}}$ satisfy (H₁) and (H₂).

Factorization patterns on nonlinear families

Let $\overline{\mathbb{F}}_q[T]_r$ be the set of monic polynomials of $\overline{\mathbb{F}}_q[T]$ of degree r .
For $\mathcal{B} \subset \overline{\mathbb{F}}_q[T]_r$, the **discriminant locus** $\mathcal{D}(\mathcal{B})$ of \mathcal{B} is

$$\begin{aligned}\mathcal{D}(\mathcal{B}) &:= \{f \in \mathcal{B} : f \text{ not square-free}\} \\ &:= \{f \in \mathcal{B} : \text{Disc}(f) := \text{Res}(f, f') = 0\}.\end{aligned}$$

(see [Fried, Smith, Acta Arith 44, 1984] and [M, Pérez, Privitelli, Acta Arith 165, 2014] for the study of discriminant loci).

Our next conditions require that the **discriminant intersects well** W , and the same happens on the highest weight:

(H₄) $\mathcal{D}(W)$ has codimension ≥ 1 in W .

(H₅) $\mathcal{D}(V(G_1^{\text{wt}}, \dots, G_m^{\text{wt}}))$ has codim ≥ 1 in $V(G_1^{\text{wt}}, \dots, G_m^{\text{wt}})$.

We also need the **first subdiscriminant locus** $\mathcal{S}_1(\mathcal{B})$ of $\mathcal{B} \subset \overline{\mathbb{F}}_q[T]_r$:

$$\begin{aligned}\mathcal{S}_1(\mathcal{B}) &:= \{f \in \mathcal{D}(\mathcal{B}) : \deg \gcd(f, f') > 1\} \\ &:= \{f \in \mathcal{D}(\mathcal{B}) : \text{Subdisc}(f) := \text{Subres}(f, f') = 0\}.\end{aligned}$$

We require that $\mathcal{D}(W)$ and $\mathcal{S}_1(W)$ intersect well W :

(H₆) $(A_0 \cdot \mathcal{S}_1)(W) := \{\mathbf{a}_0 \in W : a_0 = 0\} \cup \mathcal{S}_1(W)$ has codimension at least one in $\mathcal{D}(W)$.

Examples of linear and nonlinear families

Suppose that $\text{char}(\mathbb{F}_q) > 3$. Let $r, m \in \mathbb{Z}_{\geq 0}$ be such that $3 \leq r - m$ and $L_1, \dots, L_m \in \mathbb{F}_q[A_{r-1}, \dots, A_3]$ linear polynomials which are linearly independent. In [Cesaratto, M, Pérez, [Combinatorica 37, 2017](#)] the following linear family is considered:

$$\mathcal{A} := \{ T^r + a_{r-1}T^{r-1} + \dots + a_0 \in M(r) : L_j(a_{r-1}, \dots, a_3) = 0 \ \forall j \}.$$

We have:

Lemma: L_1, \dots, L_m satisfy hypotheses (H₁)–(H₆).

Examples of linear and nonlinear families

In [Gao, Howell, Panario, Proc. Fq4, 1999] there are experimental results on the number of irreducible polynomials on certain families over \mathbb{F}_q . In particular, the following family is considered.

Suppose that $\text{char}(\mathbb{F}_q) > 3$. For $s, r \in \mathbb{Z}_{\geq 0}$ with $3 \leq s \leq r - 2$, let

$$\mathcal{A} := \{T^r + g(T)T + 1 : g \in \mathbb{F}_q[T] \text{ and } \deg g \leq s - 1\}.$$

Observe that \mathcal{A} is isomorphic to the set of \mathbb{F}_q -rational points of the affine \mathbb{F}_q -subvariety of \mathbb{A}^r defined by

$$G_1 := A_0 - 1, \quad G_2 := A_{s+1}, \dots, \quad G_{r-s} := A_{r-1}.$$

Lemma: \mathcal{A} satisfies hypotheses (H₁)–(H₆) are fulfilled.

Examples of linear and nonlinear families

Let $r, t_1, \dots, t_r \in \mathbb{Z}_{\geq 0}$ with r even. Suppose that $\text{char}(\mathbb{F}_q) > 3$ does not divide $(r-1)(r+1)((r-1)^{r-1} + r^r)$. Consider the polynomial $G \in \mathbb{F}_q[A_1, \dots, A_r]$ defined in the following way:

$$G := \det \begin{pmatrix} A_r & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ A_1 & \dots & \dots & A_r & 1 \end{pmatrix}$$
$$:= \sum_{t_1+2t_2+\dots+rt_r=r} (-1)^{\Delta(t_1, \dots, t_r)} \frac{(t_1 + \dots + t_r)!}{t_1! \dots t_r!} A_r^{t_1} \dots A_1^{t_r},$$

where $\Delta(t_1, t_2, \dots, t_r) := r - \sum_{i=1}^r t_i$ (this is the well-known [Trudi formula](#)). $H_r := G(\Pi_r, \dots, \Pi_1)$ is critical in the study of deep holes of the standard Reed–Solomon codes (see [Cafure, M, Privitelli, Adv. Math. Commun. 6, 2012](#)).

We consider the following family of polynomials:

$$\mathcal{A}_{\mathcal{N}} := \{T^{r+1} + a_r T^r + \cdots + a_0 \in M(r+1) : G(a_r, \dots, a_1) = 0\}.$$

Observe that $\mathcal{A}_{\mathcal{N}}$ may be seen as the set of \mathbb{F}_q -rational points of the \mathbb{F}_q -variety $W := V(G) \subset \mathbb{A}^{r+1}$.

Lemma: $\mathcal{A}_{\mathcal{N}}$ satisfies hypotheses (H₁)–(H₆).

Factorization patterns on nonlinear families

A simple example: consider $\lambda := (r, 0, \dots, 0)$ and the family

$$A_s := \{ T^r + a_{r-1} T^{r-1} + \dots + a_0 : a_{r-s-1}, \dots, a_0 \in \mathbb{F}_q \}.$$

Let X_1, \dots, X_r be indeterminates, $\mathbf{X} := (X_1, \dots, X_r)$ and

$$G(\mathbf{X}, T) := (T + X_1) \cdots (T + X_r) = T^r + \Pi_1 T^{r-1} + \dots + \Pi_r,$$

where $\Pi_1, \dots, \Pi_r \in \mathbb{F}_q[\mathbf{X}]$ are the elementary symmetric polynomials.

- $f \in M(r)$ has pattern $\lambda \Leftrightarrow \exists \mathbf{x} \in \mathbb{F}_q^r$ with $f = G(\mathbf{x}, T)$.
- $G(\mathbf{x}, T) \in A_s \Leftrightarrow \Pi_j(\mathbf{x}) = a_{r-j}$ for $1 \leq j \leq s$.

We conclude that

$$\mathcal{T}_\lambda(A_s) \sim \frac{1}{r!} \cdot |\{\Pi_1 = a_{r-1}, \dots, \Pi_s = a_{r-s}\} \cap \mathbb{F}_q^r|.$$

Factorization patterns on nonlinear families

Fix $a_{r-1}, \dots, a_{r-s} \in \mathbb{F}_q$ and consider the \mathbb{F}_q -variety

$$V := \{\mathbf{x} \in \overline{\mathbb{F}}_q^r : \Pi_1(\mathbf{x}) = a_{r-1}, \dots, \Pi_s(\mathbf{x}) = a_{r-s}\}.$$

Fact: V is a **complete intersection**. In particular,

- all the irreducible components of V have **dimension** $r - s$;
- the **degree** of V is $\leq \deg \Pi_1 \cdots \deg \Pi_s = s!$.

To estimate $|V(\mathbb{F}_q)|$, we need to prove that V is **absolutely irreducible** (=irreducible as an $\overline{\mathbb{F}}_q$ -variety). For this purpose, we study its **singular locus**.

Factorization patterns on nonlinear families

In Cesaratto, M, Pérez, Privitelli [J. Combin. Theory A 124, 2014], M, Pérez, Privitelli [Acta Arith. 165, 2014], Cesaratto, M, Pérez [Combinatorica 37, 2017] we study the **singular locus** of complete intersections defined by **symmetric polynomials**.

Theorem: **Singular points** $\mathbf{x} := (x_1, \dots, x_r) \in V$ correspond to polynomials $f \in A_s$ which are **not square-free**.

This leads us to consider the **discriminant locus** of A_s . Let $\mathbf{a} := (a_{r-s-1}, \dots, a_0) \in \overline{\mathbb{F}}_q^{r-s}$ and let

$$f_{\mathbf{a}} := T^r + a_{r-1}T^{r-1} + \dots + a_{r-s}T^{r-s} + a_{r-s-1}T^{r-s-1} + \dots + a_0 \in A_s.$$

Then the discriminant locus of A_s is

$$\mathcal{D}(A_s) := \{\mathbf{a} \in \overline{\mathbb{F}}_q^{r-s} : f_{\mathbf{a}} \text{ is not square-free}\}.$$

Factorization patterns on nonlinear families

Theorem (Fried, Smith [Acta Arith. 44, 1984]): Let $A(i_1, \dots, i_s) \subset M(r)$ be the family of monic polynomials with fixed coefficients a_{i_1}, \dots, a_{i_s} . There exists $n(i_1, \dots, i_s) \in \mathbb{N}$ such that $\mathcal{D}(A(i_1, \dots, i_s))$ is **absolutely irreducible** if $\gcd(n(i_1, \dots, i_s), p) = 1$.

In M, Pérez, Privitelli [Acta Arith. 165, 2014] we prove:

Theorem: For $p > 2$ and $r - s \geq 3$, the discriminant locus $\mathcal{D}(A_s)$ is **absolutely irreducible**.

Corollary: $\text{Sing}(V)$ has dimension $\leq \dim(V) - 2$.

Combining this result with **explicit** estimates for **singular projective complete intersections** we obtain:

Theorem: For $p > 2$ y $r - s \geq 3$, we have

$$\left| \mathcal{T}_\lambda(A_s) - \frac{q^{r-s}}{r!} \right| \leq \frac{(r+2)!}{r!} q^{r-s-\frac{1}{2}} + 6 \frac{((s+2)!)^2}{r!} q^{r-s-1}.$$

(precise for $s \lesssim r/2$)

Our main result shows that:

- any family \mathcal{A} satisfying (H_1) – (H_6) is uniformly distributed (in the sense of Cohen),
- provides explicit estimates on $|\mathcal{A}_\lambda|$.

More precisely, we have the following result:

Theorem: For $m < r$ and a factorization pattern λ , we have

$$\left| |\mathcal{A}_\lambda| - \mathcal{T}(\lambda) q^{r-m} \right| \leq q^{r-m-1} (\mathcal{T}(\lambda) (D\delta q^{\frac{1}{2}} + 14D^2\delta^2 + r^2\delta) + r^2\delta),$$

where $\delta := \prod_{i=1}^m \text{wt}(G_i)$ and $D := \sum_{i=1}^m (\text{wt}(G_i) - 1)$.

Average-case analysis of factorization

As an application of our theorem, we determine the **average-case analysis** of the **classical factorization algorithm** applied to any family \mathcal{A} satisfying (H_1) – (H_6) .

Problem: given $f \in M(r)$, **find the factorization of f** as $f = f_1^{e_1} \cdots f_r^{e_r}$, where the $f_i \in \mathbb{F}_q[T]$ are **irreducible**, monic, pairwise distinct and $e_i > 0$.

The **classical factorization algorithm** roughly proceeds by the following steps:

- 1 Elimination of repeated factors (ERF).
- 2 Distinct-degree factorization (DDF).
- 3 Equal-degree factorization (EDF).

Average-case analysis of factorization

Let $\mathcal{M}(r) := r \log r \log \log r$, $\mathcal{U}(r) := \mathcal{M}(r) \log r$.

There exist $\tau_1, \tau_2 > 0$ such that:

- multiplication of $f, g \in M(r)$: $\tau_1 \mathcal{M}(r)$ operations in \mathbb{F}_q ,
- division with remainder of $f, g \in M(r)$: $\tau_1 \mathcal{M}(r)$ ops in \mathbb{F}_q ,
- gcd of $f, g \in M(r)$: $\tau_2 \mathcal{U}(r)$ operations in \mathbb{F}_q .

Von zur Gathen, Gerhard [Modern computer algebra, CUP, 1999]:

On input $f \in M(r)$, in **worst-case**, the classical factorization algorithm performs $\mathcal{O}(r \mathcal{M}(r) \log(rq))$ operations in \mathbb{F}_q :

ERF: $\mathcal{O}(\mathcal{U}(r) + r \log(\frac{q}{p}))$ operations in \mathbb{F}_q .

DDF: $\mathcal{O}(s \mathcal{M}(r) \log(rq))$ operations in \mathbb{F}_q ,

where $s =$ highest degree of the irreducible factors of f .

EDF: $\mathcal{O}((k \log q + \log r) \mathcal{M}(r) \log s)$ operations in \mathbb{F}_q ,

where $s =$ number of irreducible factors of degree k of f .

Average-case analysis of factorization

Flajolet, Gourdon, Panario [J. Algorithms 40, 2001]: **average-case analysis** (based on the **distribution of factorization patterns** in $M(r)$). Assuming that **classical polynomial multiplication** is used:

- **ERF**: $\mathcal{O}(r^2)$ operations in \mathbb{F}_q .
- **DDF**: $\mathcal{O}(r^3 \log q)$ operations in \mathbb{F}_q .
- **EDF**: $\mathcal{O}(r^2 \log q)$ operations in \mathbb{F}_q .

We consider the **uniform probability on \mathcal{A}** and the random variable $\mathcal{X} : \mathcal{A} \rightarrow \mathbb{Z}_{\geq 0}$, $\mathcal{X}(f)$ = **number of operations in \mathbb{F}_q** performed by the classical factorization algorithm on input f .

Aim: To obtain an upper bound on

$$E[\mathcal{X}] := \frac{1}{|\mathcal{A}|} \sum_{f \in \mathcal{A}} \mathcal{X}(f).$$

Average-case analysis of factorization

Recall that $\text{ERF}(f_1^{e_1} \cdots f_r^{e_r}) = f_1 \cdots f_r$. Let $\mathcal{X}_1 : \mathcal{A} \rightarrow \mathbb{Z}_{\geq 0}$, $\mathcal{X}_1(f)$ = number of operations in \mathbb{F}_q of $\text{ERF}(f)$, and let

$$E[\mathcal{X}_1] := \frac{1}{|\mathcal{A}|} \sum_{f \in \mathcal{A}} \mathcal{X}_1(f).$$

Let $\mathcal{A}^{sq} = \{f \in \mathcal{A} : f \text{ is square-free}\}$ and $\mathcal{A}^{nsq} := \mathcal{A} \setminus \mathcal{A}^{sq}$.

- $f \in \mathcal{A}^{nsq} \Leftrightarrow \text{Disc}(f) = 0 \Rightarrow |\mathcal{A}^{nsq}| = \mathcal{O}(q^{r-m-1})$.
- For $q \gg 0$, $|\mathcal{A}| \geq \frac{1}{2}q^{r-m} \Rightarrow \text{Prob}[\mathcal{A}^{sq}] > 1/2$.

Theorem: For $q > 15\delta_G^{13/3}$, $\delta_G = \deg(G_1) \cdots \deg(G_m)$,

$$E[\mathcal{X}_1] \leq c_2 \mathcal{U}(r) + c_3 \log\left(\frac{q}{p}\right) \delta_G \frac{r^3}{q},$$

where c_2, c_3 are constants independent of r and q .

Average-case analysis of factorization

Next we consider **DDF**: $\text{DDF}(\text{ERF}(f)) := (b(1), \dots, b(s))$, where $b(k)$ = product of all irreducible factors of degree k of $\text{ERF}(f)$.

Let $\mathcal{X}_2 : \mathcal{A} \rightarrow \mathbb{Z}_{\geq 0}$, $\mathcal{X}_2(f)$ = number of operations in \mathbb{F}_q of $\text{DDF}(\text{ERF}(f))$, and

$$E[\mathcal{X}_2] := \frac{1}{|\mathcal{A}|} \sum_{f \in \mathcal{A}} \mathcal{X}_2(f).$$

Theorem: For $q > 15\delta_G^{13/3}$,

$$E[\mathcal{X}_2] \leq \xi (2\tau_1 \lambda(q) + \tau_1 + \tau_2 \log r) \mathcal{M}(r) (r+1) (1+o(1)),$$

where $\xi \sim 0.62432945\dots$ is the Golomb constant.

Theorem: The probability that **DDF** outputs the complete factorization of a random $f \in \mathcal{A}$ is

$$(e^{-\gamma} + \frac{e^{-\gamma}}{r} + \mathcal{O}(\frac{\log r}{r^2})) (1+o(1)), \quad e^{-\gamma} \sim 0.5614\dots, \quad \gamma \text{ Euler's constant.}$$

Average-case analysis of factorization

Finally we consider **EDF**: if $\text{DDF}(f) = (b(1), \dots, b(s))$, then **EDF**(f) factorizes each $b(k)$. Let $\mathcal{X}_3 : \mathcal{A} \rightarrow \mathbb{Z}_{\geq 0}$, $\mathcal{X}_3(f) =$ number of operations in \mathbb{F}_q of $\text{EDF}(\text{DDF}(\text{ERF}(f)))$, and

$$E[\mathcal{X}_3] := \frac{1}{|\mathcal{A}|} \sum_{f \in \mathcal{A}} \mathcal{X}_3(f) = \sum_{k=1}^{\lceil r/2 \rceil} \underbrace{\frac{1}{|\mathcal{A}|} \sum_{f \in \mathcal{A}} \mathcal{X}_{3,k}(f)}_{E[\mathcal{X}_{3,k}]},$$

$$\mathcal{X}_{3,k}(f) := \text{Cost}(\text{EDF}(b(k))).$$

Theorem: For $q > 15\delta_G^{13/3}$,

$$E[\mathcal{X}_3] = \tau \mathcal{M}(r) \log q (1 + o(1)),$$

where τ is a constant independent of q and r .

Thanks!!!