Improvement of Generic Attacks on the Rank Syndrome Decoding Problem

Nicolas Aragon* Philippe Gaborit† Adrien Hauteville‡
Jean-Pierre Tillich§

Abstract

Rank metric code-based cryptography exists for several years. The security of many cryptosystems is based on the difficulty of decoding a random code. Any improvement in the complexity of the best decoding algorithms can have a big impact on the security of these schemes.

In this article, we present an improvement on the recent GRS algorithm [1] and we obtain a complexity of $\mathcal{O}\big((n-k)^3m^3q^{w\frac{(k+1)m}{n}-m}\big)$ for decoding an error of weight w in an $[n,k]$ $\mathbb{F}_{2^m}$-linear code.

1 Presentation of rank metric codes

1.1 Definitions

Let us introduce the matrix codes.

Definition 1.1 (Linear matrix codes)**.** *A linear matrix code $\mathcal{C}$ of length $m \times n$ and dimension K over $\mathbb{F}_q$ is a subspace of dimension K of $\mathcal{M}_{m\times n}(\mathbb{F}_q)$. We say $\mathcal{C}$ is an $[m \times n, K]_q$ linear matrix cod, or simply an $[m \times n, K]$ code if there is no ambiguity.*

We define the rank metric over matrix codes as such:

- the distance between two words $\boldsymbol{A}$ and $\boldsymbol{B}$ is $d_R(\boldsymbol{A}, \boldsymbol{B}) \stackrel{\text{def}}{=} \text{Rank}(\boldsymbol{A} - \boldsymbol{B})$.

*XLIM, Université de Limoges, 123, Av. Albert Thomas, 87000 Limoges, France. nicolas.aragon@etu.unilim.fr

†XLIM, Université de Limoges, 123, Av. Albert Thomas, 87000 Limoges, France. gaborit@unilim.fr

‡XLIM, Université de Limoges, 123, Av. Albert Thomas, 87000 Limoges, France. adrien.hauteville@unilim.fr

§INRIA, 2, rue Simone Iff, 75012 Paris, France. jean-pierre.tillich@inria.fr

- the weight of a word $\boldsymbol{A}$ is $|\boldsymbol{A}|_R \stackrel{\text{def}}{=} \text{Rank}(\boldsymbol{A}) = d_R(\boldsymbol{A}, \boldsymbol{0})$.

An important family of matrix code are the $\mathbb{F}_{q^m}$-linear codes.

Definition 1.2 ($\mathbb{F}_{q^m}$-linear codes)**.** *A $\mathbb{F}_{q^m}$-linear code $\mathcal{C}$ of length n and dimension k is a subspace of dimension k of $\mathbb{F}_{q^m}^n$. We say $\mathcal{C}$ is an $[n,k]_{q^m}$ linear code, or simply an $[n,k]$ code if there is no ambiguity.*

A natural way to define the rank metric over $\mathbb{F}_{q^m}$-linear codes is to consider the matrix associated to a word of $\mathbb{F}_{q^m}^n$.

Definition 1.3 (Associated matrix)**.** *Let $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{F}_{q^m}^n$ and $\boldsymbol{\beta} = (\beta_1, dots, \beta_m)$ a basis of $\mathbb{F}_{q^m}/\mathbb{F}_q$.*
$\boldsymbol{x}$ can be represented by an $m \times n$ matrix $\boldsymbol{M_x}$ such that its i^{th} column represents the coordinates of x_i in the basis $\boldsymbol{\beta}$.

$$\boldsymbol{x} \leftrightarrow \begin{pmatrix} x_{11} & \ldots & x_{1n} \\ \vdots & & \vdots \\ x_{m1} & \ldots & x_{mn} \end{pmatrix}$$

with $x_j = \sum_{i=1}^{m} x_{ij}\beta_i$ for all $j \in [1..n]$.

By definition, $|\boldsymbol{x}|_R \stackrel{\text{def}}{=} |\boldsymbol{M_x}|_R$ and $d_R(\boldsymbol{x}, \boldsymbol{y}) = d_R(\boldsymbol{M_x}, \boldsymbol{M_y})$. These definition do not depend on the choice of the basis.

The advantage of $\mathbb{F}_{q^m}$-linear codes with respect to matrix codes is that they have a much compact representation. Indeed, we can associate an $[m \times n, km]_q$ matrix code to an $[n,k]_{q^m}$ linear code. The matrix code can be represented by $(n-k)km^2$ coefficients in $\mathbb{F}_q$ that is $(n-k)km^2 \lceil \log_2 q \rceil$ bits whereas the $\mathbb{F}_{q^m}$-linear can be represented by $(n-k)k$ coefficients in $\mathbb{F}_{q^m}$ that is $(n-k)km \lceil \log_2 q \rceil$ bits.

Definition 1.4 (Support of a word)**.** *Let $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$. The support of $\boldsymbol{x}$ is the $\mathbb{F}_q$-subspace of $\mathbb{F}_{q^m}$ generated by the coordinates of $\boldsymbol{x}$. It is denoted* $\text{Supp}(\boldsymbol{x})$.

$$\text{Supp}(\boldsymbol{x}) = \langle x_1, \ldots, x_n \rangle_{\mathbb{F}_q}$$

The weight of a word is equal to the dimension of its support.

In the case of matrix codes, we can consider the subspace of $\mathbb{F}_q^m$ generated by the columns of the matrix, called the columns support, or the subspace of $\mathbb{F}_q^n$ generated by its rows, called the rows support.

1.2 Hard problem in rank metric

Rank-based cryptography relies on difficult problems in rank metric. These problem are the same as in the Hamming metric, but with the rank metric. In this subsection, we only consider $\mathbb{F}_{q^m}$-linear codes but all the problems are defined exactly the same way for linear matrix codes.

The first one corresponds to the decoding problem by syndromes.

Definition 1.5 (Rank Syndrome Decoding (RSD) problem)**.** *Let $\boldsymbol{H}$ be the parity-check matrix of an $[n,k]$ code, $\boldsymbol{s} \in \mathbb{F}_{q^m} n-k$ and w an integer. The RSD problem consists to find $\boldsymbol{e} \in \mathbb{F}_{q^m}^n$ such that:*

- $\boldsymbol{H}\boldsymbol{e}^T = \boldsymbol{s}$
- $|\boldsymbol{e}|_R = w$

The second problem is the search for small-weight codewords.

Definition 1.6 (Small-weight codeword problem)**.** *Let $\mathcal{C}$ be an $[n,k]$ code and w an integer. The problem consists to find a codeword $\boldsymbol{c} \in \mathcal{C}$ such that $|\boldsymbol{c}|_R = w$.*

Remark: If $\boldsymbol{H}$ is an parity-check matrix of $\mathcal{C}$, the small-weight codeword problem corresponds to an instance of the RSD problem with $\boldsymbol{s} = \boldsymbol{0}$.

1.3 Generic algorithms

1.4 Basic algorithm

In the article [1], the general idea to solve the RSD problem is to find a subspace F such that $\mathrm{Supp}(\boldsymbol{e}) \subset F$. Then we can express the coordinates of $\boldsymbol{x}$ in a basis of F and solve the linear system given by the parity-check equations. There are two possible cases we will describe more precisely.

- **first case:** $n \geqslant m$**.**
  Let F be a subspace of $\mathbb{F}_{q^m}$ of dimension r and $(F_1, \ldots, F_r)$ a basis of F. Let us assume that $\mathrm{Supp}(\boldsymbol{e}) \subset F$.

$$\Rightarrow \forall i \in [1..n], e_i = \sum_{j=1}^{r} \lambda_{ij} F_j$$

  This gives us nr unknowns over $\mathbb{F}_q$.
  We can now rewrite the parity-check equations in these unknowns:

$$\boldsymbol{H}\boldsymbol{e}^T = \boldsymbol{s} \tag{1}$$

$$\Leftrightarrow \begin{cases} H_{11}e_1 & +\cdots+ & H_{1n}e_n & = & s_1 \\ \vdots & & \vdots & & \vdots \\ H_{n-k,1}e_1 & +\cdots+ & H_{n-k,n}e_n & = & s_{n-k} \end{cases}$$

$$\Leftrightarrow \begin{cases} \sum_{j=1}^{r} (\lambda_{1j} H_{11} F_j & +\cdots+ & \lambda_{nj} H_{1n} F_j) & = & s_1 \\ \vdots & & \vdots & & \vdots \\ \sum_{j=1}^{r} (\lambda_{1j} H_{n-k,1} F_j & +\cdots+ & \lambda_{nj} H_{n-k,n} F_j) & = & s_{n-k} \end{cases} \tag{2}$$

Now, we need embed these $n-k$ equations over $\mathbb{F}_{q^m}$ into $(n-k)m$ equations over $\mathbb{F}_q$.
Let φ_i the i^{th} canonical projection from $\mathbb{F}_{q^m}$ on $\mathbb{F}_q$:

$$\begin{aligned}\varphi_i : \quad \mathbb{F}_{q^m} &\to \mathbb{F}_q \\ \sum_{i=1}^{m} x_i\beta_i &\mapsto x_i\end{aligned}$$

We have:

$$\begin{array}{l}\boldsymbol{H}\boldsymbol{e}^T = \boldsymbol{s} \\ \Leftrightarrow \forall i \in [1..m], \\ \left\{\begin{array}{ccccc} \sum_{j=1}^{r}\left(\lambda_{1j}\varphi_i(H_{11}F_j)\right. & +\cdots+ & \left.\lambda_{nj}\varphi_i(H_{1n}F_j)\right) & = & \varphi_i(s_1) \\ \vdots & & \vdots & & \vdots \\ \sum_{j=1}^{r}\left(\lambda_{1j}\varphi_i(H_{n-k,1}F_j)\right. & +\cdots+ & \left.\lambda_{nj}\varphi_i(H_{n-k,n}F_j)\right) & = & \varphi_i(s_{n-k}) \end{array}\right.\end{array}$$

Since we assume $\text{Supp}(\boldsymbol{e}) \subset F$, this system has at least one solution. We want more equations than unknowns to "eliminate" the false solutions. So we have:

$$(n-k)m \geqslant nr \Leftrightarrow r \leqslant m - \left\lceil \frac{km}{n} \right\rceil$$

The complexity of the algorithm is $\mathcal{O}\big(\frac{(n-k)^3m^3}{p}\big)$ where p is the probability that $\text{Supp}(\boldsymbol{e}) \subset F$.
p is equal to the number of subspaces of dimension w in a subspace of dimension r divided by the total number of subspaces of dimension w in $\mathbb{F}_{q^m}$. By definition, these numbers are expressed by the Gaussian coefficients.

$$p = \frac{\begin{bmatrix} r \\ w \end{bmatrix}_q}{\begin{bmatrix} m \\ w \end{bmatrix}_q} \approx q^{-w(m-r)}$$

By taking $r = m - \left\lceil \frac{km}{n} \right\rceil$ we obtain a complexity of $\mathcal{O}\big((n-k)^3m^3q^{w\left\lceil \frac{km}{n} \right\rceil}\big)$

- **second case:** $m > n$. In this case we consider the matrix code associated to the $\mathbb{F}_{q^m}$-linear code. Instead of searching for a subspace which contains the columns support of the matrix of the error, we search for a subspace F of $\mathbb{F}_q^n$ which contains the rows support of the error. The remainder of the algorithm is the same, the only differences are:
  - the number of unknowns is mr, which implies $r \leqslant n-k$

– the probability to choose a "good" F is $\dfrac{\begin{bmatrix} r \\ w \end{bmatrix}_q}{\begin{bmatrix} n \\ w \end{bmatrix}_q} \approx q^{-w(n-r)}$.

Thus, the complexity in this case is $\mathcal{O}\big((n-k)^3 m^3 q^{wk}\big)$.

1.5 Improvements of the algorithm

In the last algorithm, we do not use the $\mathbb{F}_{q^m}$-linearity of the code. In this subsection, we will see two ways to use the linearity to improve the complexity of our algorithm. The main idea is to consider the code $\mathcal{C}' = \mathcal{C} + \mathbb{F}_{q^m}\boldsymbol{e}$, where $\mathcal{C}$ is the code with parity-check matrix $\boldsymbol{H}$. The problem is reduced to the search for a codeword of weight w in $\mathcal{C}'$. If $\boldsymbol{e}$ is the only solution of the equation $\boldsymbol{H}\boldsymbol{x}^T = \boldsymbol{s}, |\boldsymbol{x}|_R = w$, then the only codewords of $\mathcal{C}'$ of weight w are of the form $\alpha\boldsymbol{e}, \alpha \in \mathbb{F}_{q^m}^*$. Instead of looking for the support E of $\boldsymbol{e}$, we can look for any multiple αE of the support. There is two strategies to exploit this idea:

- in the first case, we specialize one element of the support by searching for small weight codeword $\boldsymbol{c}$ such that $1 \in \mathrm{Supp}(\boldsymbol{c})$ (or any other scalar). We need to compute the probability that $F \supset \mathrm{Supp}(\boldsymbol{c})$, knowing that $1 \in F$. This probability is given by the formula $\dfrac{\begin{bmatrix} w-1 \\ r-1 \end{bmatrix}_q}{\begin{bmatrix} w-1 \\ m-1 \end{bmatrix}_q} \approx q^{(w-1)(m-r)}$. Since $\mathcal{C}'$ is of dimension $k+1$, we take $r = \left\lfloor \frac{m(n-k-1)}{\max(m,n)} \right\rfloor$ which gives us a complexity of $\mathcal{O}\big((n-k)^3 m^3 q^{(w-1)\min(k+1, \frac{(k+1)m}{n})}\big)$.
- the idea is to choose F at random like in the first algorithm, if F contains a multiple αE of E, we can compute the codeword $\alpha\boldsymbol{e}$ of $\mathcal{C}'$. There is at most $\frac{q^m-1}{q-1}$ subspace of this form, because $\alpha E = \beta E$ if $\alpha/\beta \in \mathbb{F}_q^*$. In the following we will suppose that all this subspaces are different, which is always true if m is prime (see appendix A). We need to compute the probability that F of dimension $\left\lfloor \frac{m(n-k-1)}{n} \right\rfloor = m - \left\lceil \frac{m(k+1)}{n} \right\rceil$ contains any subspaces of the form $\alpha E, \alpha \in \mathbb{F}_{q^m}^*$. We can approximate it by the product of the number of of these subspaces by the probability that F contains a fixed subspace of dimension w. This approximation is correct if

$$\frac{q^m-1}{q-1}\begin{bmatrix} r \\ w \end{bmatrix}_q \ll \begin{bmatrix} m \\ w \end{bmatrix}_q \Leftrightarrow q^{m+w(r-w)} \ll q^{w(m-w)}.$$

We obtain a complexity of $\mathcal{O}\big((n-k)^3 m^3 q^{w\frac{(k+1)m}{n} - m}\big)$.

In the case $m \leqslant n$, the second strategies is always better. The gain in the exponent is equal to $m - \left\lceil \frac{(k+1)m}{n} \right\rceil = m - \lceil mR' \rceil \approx m(1-R')$ where R' is the rate of $\mathcal{C}'$. In the case $m > n$ the second strategy may also be faster for some parameters.

References

[1] P. Gaborit, O. Ruatta and J. Schrek, "On the complexity of the rank syndrome decoding problem," in IEEE Trans. Information Theory 62(2): pp. 1006-1019 (2016).

A On the multiples of subspace of $\mathbb{F}_{q^m}$

In this section, we study the probability that $E = \alpha E$, where E is a subspace of $\mathbb{F}_{q^m}$ and $\alpha \in \mathbb{F}_{q^m}$.

Proposition A.1. *Let E a subspace of $\mathbb{F}_{q^m}$ of dimension d and $\alpha \in \mathbb{F}_{q^m}^*$ such that $E = \alpha E$. Then $[\mathbb{F}_q(\alpha) : \mathbb{F}_q]$ divides d. In particular, if $d \wedge m = 1$, $\alpha \in Fq^*$.*

Proof. Let $m' = \mathbb{F}_q(\alpha) : \mathbb{F}_q]$. Since $\alpha E = E$, we have $\mathbb{F}_q(\alpha)E = E$. So E is an $\mathbb{F}_q(\alpha)$-subspace of $\mathbb{F}_{q^m}$. Thus

$$d = \dim_{\mathbb{F}_{q^m}} E = m' \dim_{\mathbb{F}_q(\alpha)} E$$

which proves the first point. Moreover m' divides $m \Rightarrow m'$ divides $m \wedge d = 1$. So $\mathbb{F}_q(\alpha) = \mathbb{F}_q$ which proves the second point. □

Now let $m = am'$ with $a > 1$ and E an $\mathbb{F}_q$-subspace of $\mathbb{F}_{q^m}$ of dimension dm'. The probability that E is an $\mathbb{F}_{q^{m'}}$-subspace of dimension d is

$$\frac{\begin{bmatrix} a \\ d \end{bmatrix}_{q^{m'}}}{\begin{bmatrix} m \\ m'd \end{bmatrix}_q} \approx q^{\frac{m'd(a-d)}{m'd(m-m'd)}} = q^{-m'd(a-d)(m'-1)}$$