The Density of MDS Codes With Subfield Linearity

Nadja Willenborg[1] and Anna-Lena Horlemann[1]

School of Computer Science, University of St. Gallen, 9000 St. Gallen, Switzerland

Abstract. By extending the Hamming metric from $\mathbb{F}_{q^m}^n$ to $\mathbb{F}_q^{m\times n}$ in a natural way, we show that the limiting proportion of $\mathbb{F}_q$-linear MDS codes in $\mathbb{F}_q^{m\times n}$ is 1 within the set of all matrix codes of the same dimension as $q \to \infty$. The same question of density of codes with subfield linearity has already been studied and answered in the rank metric, see, e.g., [2, 3]. Our results differ from the recent results in [1] where it is shown that $\mathbb{F}_q$-linear MRD codes in $\mathbb{F}_{q^m}^n$ are sparse for $q \to \infty$, unless the minimum rank distance d is $d = 1$ or $n = d = 2$.

Keywords: MDS codes · Hamming metric · density problems · additive codes

1 Introduction

Density questions have been well studied in coding theory, where the fraction of codes with certain properties among all codes within a certain space is studied. In particular, it has been long known that maximum distance separable (MDS) codes are dense within the set of linear codes of some fixed length n over some prescribed finite field $\mathbb{F}_q$, for $q \to \infty$. Similarly, it was shown in [2] that the same is true for linear maximum rank-distance (MRD) codes in $\mathbb{F}_{q^m}^n$, for fixed n and growing q or m. However, when the linearity is relaxed to linearity over the subfield $\mathbb{F}_q$, MRD codes are not dense anymore; on the contrary, it was shown in [1] that $\mathbb{F}_q$-linear MRD codes in $\mathbb{F}_{q^m}^n$ are sparse for $q \to \infty$, unless the minimum rank distance d is $d = 1$ or $n = d = 2$.

This peculiar behavior of MRD codes gives rise to the question of the density of $\mathbb{F}_q$-linear MDS codes in $\mathbb{F}_{q^m}^n$. In Section 3 we will see that $\mathbb{F}_q$-linear MDS codes behave similar to $\mathbb{F}_{q^m}$-linear MDS codes concerning the density. For the special case where q is a prime, these codes are known as *additive codes*, and have been studied in the literature, recently in particular with applications to quantum codes.

Density results like these have applications both in coding theory and cryptography. In particular, they represent the probability that a randomly chosen (linear) code has the required property (e.g., being MDS or MRD).

Our paper is structured as follows. In the following section we describe the preliminary definitions about MDS codes. In Section 3 we determine an upper and a lower bound for the fraction of MDS codes in the set of all $\mathbb{F}_q$-linear codes

in $\mathbb{F}_{q^m}^n$. For this we represent the codes as matrices in $\mathbb{F}_q^{m\times n}$. Furthermore, we determine the asymptotics (and hence the densities) of these fractions in q and in m. We conclude this paper in Section 4.

2 Preliminaries

We start by introducing the terminology used throughout the paper and recall some facts from classical coding theory. For more details and proofs we refer the interested reader to [5].

Let q be a prime power and k, n, m be non-negative integers. The finite field of cardinality q is denoted by $\mathbb{F}_q$ and an extension field of extension degree m is denoted by $\mathbb{F}_{q^m}$. Let $\Gamma = \{\gamma_1, \ldots, \gamma_m\}$ be a basis of $\mathbb{F}_{q^m}$ as an $\mathbb{F}_q$-vector space. Further we define the isomorphism

$$\phi_\Gamma \colon \mathbb{F}_{q^m} \to \mathbb{F}_q^m, \ \sum_{i=1}^{m} \alpha_i\gamma_i \mapsto \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix} =: [\alpha]_\Gamma,$$

which extends to the isomorphism

$$\Phi_\Gamma \colon \mathbb{F}_{q^m}^n \to \mathbb{F}_q^{m\times n}, \ (\alpha_1, \ldots, \alpha_n) \mapsto ([\alpha_1]_\Gamma, \ldots, [\alpha_n]_\Gamma).$$

The q-ary binomial coefficient, denoted by $\begin{bmatrix} n \\ k \end{bmatrix}_q$, counts the number of k-dimensional vector subspaces of $\mathbb{F}_q^n$. It is well-known that the identity

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \prod_{i=0}^{k-1} \frac{q^n - q^i}{q^k - q^i}$$

holds.

Definition 1. *Let $X \in \mathbb{F}_q^{m\times n}$. We define the* column Hamming weight *of X by*

$$\tilde{w}_H(X) := |\{j \in [n] \mid Xe_j \neq 0\}|,$$

where e_j denotes the jth unit vector in $\mathbb{F}_q^n$.

Note that the column Hamming weight corresponds to the usual Hamming weight of the preimage under Φ_Γ, for any $\mathbb{F}_q$-basis Γ of $\mathbb{F}_{q^m}$, i.e.,

$$w_H(v_1, \ldots, v_n) = \tilde{w}_H(\Phi_\Gamma(v_1, \ldots, v_n))$$

for $(v_1, \ldots, v_n) \in \mathbb{F}_{q^m}^n$. Since the (column) Hamming weight satisfies the triangular inequality we can define a distance on $\mathbb{F}_q^{m\times n}$ as follows:

Definition 2. *Let $X, Y \in \mathbb{F}_q^{m\times n}$. Then the* column Hamming distance *between X and Y is defined as*

$$\tilde{d}_H(X,Y) := \tilde{w}_H(X-Y).$$

We call a $\mathbb{F}_q$-linear subspace $\mathcal{C}$ of the metric space $(\mathbb{F}_q^{m\times n}, \tilde{d}_H)$ a linear Hamming-metric matrix code. *The* minimum column Hamming distance *of the code $\mathcal{C}$ is*

$$\tilde{d}_H(\mathcal{C}) := \min\{\tilde{w}_H(X) \mid X \in \mathcal{C} \setminus \{0\}\}.$$

A $[m \times n, k, d]_q$-code is a linear Hamming-metric matrix code in $\mathbb{F}_q^{m\times n}$ of $\mathbb{F}_q$-dimension k and minimum column Hamming distance d.

Let $2 \leq d \leq n$ and $\mathcal{C} \subseteq \mathbb{F}_q^{m\times n}$ be a $[m \times n, k, d]_q$-code. For $X \in \mathcal{C}$ with columns $\underline{x_1}, \ldots, \underline{x_n}$ we write $X = (\underline{x_1}, \ldots, \underline{x_n})$. Then the projective map $\rho\colon \mathcal{C} \to \mathbb{F}_q^{m\times(n-d+1)}$ given by $(\underline{x_1}, \ldots, \underline{x_n}) \mapsto (\underline{x_d}, \ldots, \underline{x_n})$ must be injective. Otherwise $\rho(X) = \rho(X')$ for some $X, X' \in \mathcal{C}, X \neq X'$. But then $\tilde{d}_H(X, X') \leq d-1$ which is a contradiction to $\tilde{d}_H(\mathcal{C}) = d$. Hence $|\mathcal{C}| \leq q^{m(n-d+1)}$ and since the dimension of the code $\mathcal{C}$ is given by $k = \log_q(|\mathcal{C}|)$ we can state the Singleton bound in the context of Hamming-metric matrix codes:

Theorem 1. *Let $\mathcal{C} \subseteq \mathbb{F}_q^{m\times n}$ be a nonzero $[m \times n, k, d]_q$-code. Then*

$$\tilde{d}_H(\mathcal{C}) \leq n - \frac{k}{m} + 1.$$

Definition 3. *A $[m \times n, k, d]_q$ code $\mathcal{C} \subseteq \mathbb{F}_q^{m\times n}$ with $2 \leq d \leq n$ and dimension $k = m(n-d+1)$ is called a* MDS-code (maximum distance separable code) *and denoted as $[m \times n, k]_q$-MDS code.*

Recalling the weight preservation of Φ_Γ, we get that, if $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ is a $[n, k, d]_{q^m}$-code, then $\Phi_\Gamma(\mathcal{C})$ is a Hamming-metric matrix code of dimension mk over $\mathbb{F}_q$, with the same minimum distance as $\mathcal{C}$. Thus, a $[m \times n, k, d]_q$ Hamming-metric matrix code $\mathcal{C}$ is MDS, if and only if $\Phi_\Gamma^{-1}(\mathcal{C})$ is MDS for an arbitrary $\mathbb{F}_q$-basis Γ of $\mathbb{F}_{q^m}$.

3 Proportions of MDS Matrix Codes

First we state the results and notations taken from [1] which we use later for estimating the proportion of MDS matrix codes.

Notation 1 *[1, Notation 2.1] Let q be a prime power and m, n, k, l be non-negative integers with $mn \geq \max\{3, k\}$ and $nm - 2k \leq l \leq nm - k$. Define*

$$\upsilon_q(nm, k, l) := \begin{bmatrix} nm \\ k \end{bmatrix}_q - 2q^{k(nm-k)} + q^{(2k-nm+l)(nm-k)} \prod_{i=l}^{nm-k-1} (q^{nm-k} - q^i).$$

Theorem 2. *[1, Theorem 3.6] Let $mn \geq 3$ and let $1 \leq k \leq mn-1$. Let $\mathcal{A}$ be a collection of subspaces $A \subseteq \mathbb{F}_q^{m\times n}$ with at least two elements that all have codimension k. Let $\mathcal{F}$ be a collection of k-dimensional spaces $W \subseteq \mathbb{F}_q^{m\times n}$ that intersect in at least one subspace $A \in \mathcal{A}$ and define*

$$l_{max} := \max\{\dim(A \cap A') | A, A' \in \mathcal{A}, A \neq A'\},$$

then

$$\frac{\upsilon_q(nm,k,nm-k)^2|\mathcal{A}|}{\upsilon_q(nm,k,nm-k)+(|\mathcal{A}|-1)\upsilon_q(nm,k,l_{max})} \leq |\mathcal{F}| \leq |\mathcal{A}|\upsilon_q(nm,k,nm-k).$$

We now turn to the question of estimating the proportion of MDS-matrix codes within the family of codes having the same dimension. We adopt the notation of [4].

Definition 4. *Suppose that $1 \leq n, m$ are integers and let $k \in \{m, 2m, \ldots, nm\}$, as for other choices of the dimension $[m \times n, k]_q$-MDS codes do not exist. We define*

$$\begin{aligned} T_q^{m,n,k} &:= \{\mathcal{C} \in \mathbb{F}_q^{m\times n} |\ \dim(\mathcal{C}) = k\} \\ \hat{T}_q^{m,n,k} &:= \{\mathcal{C} \in T_q^{m,n,k} |\ d_H(\mathcal{C}) = n - \frac{k}{m} + 1\}, \end{aligned} \tag{1}$$

thus $\hat{T}_q^{m,n,k}$ is the set of $[m \times n, k]_q$-MDS codes and $T_q^{m,n,k}$ counts the number of linear codes $\mathcal{C} \subseteq \mathbb{F}_q^{m\times n}$ of dimension k. The fraction $|\hat{T}_q^{m,n,k}|/|T_q^{m,n,k}|$ describes the proportion of MDS-matrix codes among the space of all k-dimensional matrix-codes of $\mathbb{F}_q^{m\times n}$.

We study the asymptotic proportion of MDS-matrix codes as the field size q tends to infinity and as their column length m tends to infinity.

1. If $|\hat{T}_q^{m,n,k}|/|T_q^{m,n,k}| = 0$ as $q \to \infty$, respectively $m \to \infty$, then the family of $[m \times n, k]_q$-MDS codes is called sparse.
2. If $|\hat{T}_q^{m,n,k}|/|T_q^{m,n,k}| = 1$ as $q \to \infty$, respectively $m \to \infty$, then the family of $[m \times n, k]_q$-MDS codes is called dense.

The proportion above relates to the probability that k randomly and independently choosen matrices $A_1, \ldots, A_k \in \mathbb{F}_q^{m\times n}$ generate a $[m \times n, k]_q$-MDS code which means that for all $(\lambda_1, \ldots, \lambda_k) \in \mathbb{F}_q^k \setminus \{0\}$ the matrix $\sum_{i=1}^k \lambda_i A_i$ has at least $n - \frac{k}{m} + 1$ nonzero columns. Intuitively, this probability approaches 1, as the field size q grows. The following two theorems show that this intuition is indeed correct.

Theorem 3. *Suppose that $1 \leq n, m$ are integers and let $k \in \{m, 2m, \ldots, (n-1)m\}$. Then the proportion of $[m \times n, k]_q$-MDS codes is bounded by*

$$\max(0, L_q^k(m,n)) \leq \frac{|\hat{T}_q^{m,n,k}|}{|T_q^{m,n,k}|} \leq \min(1, U_q^k(m,n)),$$

where

$$L_q^k(m,n) := 1 - \frac{\binom{n}{k/m} \upsilon_q(nm,k,mn-k)}{\begin{bmatrix} mn \\ k \end{bmatrix}_q}$$

$$U_q^k(m,n) := 1 - \frac{\binom{n}{k/m} \upsilon_q(mn,k,mn-k)^2}{\begin{bmatrix} mn \\ k \end{bmatrix}_q \left(\upsilon_q(nm,k,nm-k) + (\binom{n}{k/m} - 1)\upsilon_q(nm,k,m(n-1)-k) \right)}.$$

Proof. Define the family of sets $\mathcal{S}_{n-k/m} := \{S \subseteq [n] \mid |S| = n - \frac{k}{m}\}$ *and the* $(mn-k)$*-dimensional linear spaces* $\mathbb{F}_q^{m\times n}(S) := \{A \in \mathbb{F}_q^{m\times n} \mid Ae_i = 0 \, \forall i \notin S\}$. *By this definition every* $A \in \mathbb{F}_q^{m\times n}(S)$ *has Hamming weight at most* $n - \frac{k}{m}$. *Let* $\mathcal{A} = \{\mathbb{F}_q^{m\times n}(S) \mid S \in \mathcal{S}_{n-k/m}\}$ *be the collection of those linear spaces. Then* $|\mathcal{A}| = \binom{n}{k/m}$ *and any code* $\mathcal{C} \in T_q^{m,n,k}$ *that intersects one of the spaces in* $\mathcal{A}$ *non-trivially has minimum Hamming distance at most* $n - \frac{k}{m}$. *Furthermore let* $\mathcal{C} \in T_q^{m,n,k}$ *be a code with* $\tilde{d}_H(\mathcal{C}) \leq n - \frac{k}{m}$, *i.e. there exists a nonzero matrix* $X \in \mathcal{C}$ *with potentially nonzero columns* $\underline{x}_{i_1}, \ldots, \underline{x}_{i_{n-k/m}}$ *(and else zero). Then* $S := \{i_1, \ldots, i_{n-k/m}\} \in \mathcal{S}_{n-k/m}$ *and thus the* $[m\times n, k]_q$*-MDS codes are precisely those subspaces that intersect none of the spaces in* $\mathcal{A}$. *Furthermore we have*

$$\max_{\substack{S,S' \in \mathcal{S}_{n-k/m} \\ S \neq S'}} \{\dim(\mathbb{F}_q^{m\times n}(S) \cap \mathbb{F}_q^{m\times n}(S'))\} = m(n-1) - k.$$

Now the desired bounds follow from Theorem 2.

Table 1. The table below shows the lower and upper bound, calculated with SageMath 9.4, for the proportion of $[m \times n, k]_q$-MDS codes as stated in Theorem 3.

n	q	m	k	Lower Bound	Upper Bound
5	2	2	2	0	0.107
5	2	2	6	0	0.076
5	2	8	8	0	0.075
5	2	8	16	0	0.039
5	337	2	2	0.985	0.985
5	337	2	6	0.97	0.971
5	337	8	8	0.985	0.985
5	337	8	16	0.97	0.971

We can now state the asymptotics of the proportion as the field size q tends to infinity.

Theorem 4. *Suppose that* $1 \leq n, m$ *are integers and let* $k \in \{m, 2m, \ldots, nm\}$. *Then*

$$\lim_{q\to\infty} \frac{|\hat{T}_q^{m,n,k}|}{|T_q^{m,n,k}|} = 1,$$

that is $[m \times n, k]_q$-MDS codes are dense within all linear codes of a fixed dimension.

Proof. If $k = mn$ the statement of the theorem directly follows, as the minimum column Hamming distance of the code $\mathbb{F}_q^{m\times n}$ is 1. Hence this unique code of dimension mn is MDS. Now let $k < mn$, substituting $l = mn - k$ into the formula given in Notation 1, the lower bound given in Theorem 3 can be written as

$$L_q^k(m,n) = 1 - \binom{n}{k/m} + \binom{n}{k/m} \cdot \frac{q^{k(nm-k)}}{\begin{bmatrix} mn \\ k \end{bmatrix}_q}.$$

Using the asymptotic estimate

$$\begin{bmatrix} mn \\ k \end{bmatrix}_q \in \Theta(q^{k(mn-k)}) \ \text{ as } q \to \infty,$$

the statement of the theorem follows.

We now consider the asymptotic proportion of MDS-codes as their column length tends to ∞. First we state the asymptotics for the number given in Notation 1. Before giving the explicit results for this quantity, we introduce the Euler function and give an asymptotic estimate of the q-binomial coefficient involving this function.

Definition 5. *[6, Section 14] The Euler function $\phi\colon (-1,1) \to \mathbb{R}$ is defined by*

$$x \mapsto \prod_{i=1}^{\infty}(1 - x^i).$$

Lemma 1. *[1, Section 6] Suppose that $m \geq 1$ and $a > b > 0$ are integers. Then*

$$\begin{bmatrix} ma \\ mb \end{bmatrix}_q \sim \frac{q^{m^2b(a-b)}}{\phi(q^{-1})}.$$

Proof.

$$\begin{aligned}
\begin{bmatrix} ma \\ mb \end{bmatrix}_q &= \frac{q^{mamb}\prod_{i=0}^{mb-1}(1-q^{i-ma})}{q^{mbmb}\prod_{i=0}^{mb-1}(1-q^{i-mb})} \\
&= q^{m^2b(a-b)}\frac{\prod_{i=m(a-b)+1}^{ma}\left(1-q^{-i}\right)}{\prod_{i=1}^{mb}\left(1-q^{-i}\right)} \\
&= q^{m^2b(a-b)}\frac{\prod_{i=1}^{ma}\left(1-q^{-i}\right)}{\prod_{i=1}^{mb}\left(1-q^{-i}\right)\prod_{i=1}^{m(a-b)}\left(1-q^{-i}\right)} \\
&\sim \frac{q^{m^2b(a-b)}}{\phi(q^{-1})} \ \text{ as } m \to \infty,
\end{aligned}$$

by the definition of the Euler function.

Lemma 2. *[1, Lemma 6.5] Suppose that $1 \leq n, m$, let i be an integer and let $k \in \{m, 2m, \ldots, (n-1)m\}$. For $m \to \infty$ the following asymptotic estimates hold:*

$$\upsilon_q(mn, k, mi) \sim \begin{cases} q^{k(mn-k)} \cdot \frac{(1-\phi(q^{-1}))^2}{\phi(q^{-1})} & \text{if } 0 \leq i \leq n - \frac{k}{m} - 1 \\ q^{k(mn-k)} \cdot \frac{(1-\phi(q^{-1}))}{\phi(q^{-1})} & \text{if } i = n - \frac{k}{m} \end{cases}$$

Using similar techniques as in the proof of Theorem 6.6 in [1], we can now prove an asymptotic bound for the proportion of MDS-matrix codes as $m \to \infty$.

Theorem 5. *Suppose that $1 \leq n, m$ are integers and let $k = mi$ with $i \in [n-1]$. Then we have*

$$\limsup_{m\to\infty} \frac{\left|\hat{T}_q^{m,n,k}\right|}{\left|T_q^{m,n,k}\right|} \leq \frac{\phi(q^{-1})}{\phi(q^{-1}) + \binom{n}{i}(1-\phi(q^{-1}))} < 1,$$

i.e., $[m \times n, k]_q$-MDS codes are not dense with respect to the parameter m.

Proof. Define the sequences $(a_m)_m, (b_m)_m, (c_m)_m$ by

$$\begin{aligned} a_m &:= \upsilon_q(mn, k, mn-k), \\ b_m &:= \left(\binom{n}{k/m} - 1\right) \upsilon_q(mn, k, m(n-1)-k), \\ c_m &:= \frac{\binom{n}{k/m} \upsilon_q(mn, k, mn-k)^2}{\begin{bmatrix} mn \\ k \end{bmatrix}_q}. \end{aligned}$$

Then the upper bound given in Theorem 3 can be written as

$$\frac{\left|\hat{T}_q^{m,n,d}\right|}{\left|T_q^{m,n,d}\right|} \leq \frac{a_m + b_m - c_m}{a_m + b_m}.$$

Using Lemma 1 and Lemma 2 the asymptotic behaviour of the three sequences above can be described by

$$\begin{aligned} a_m &\sim q^{k(mn-k)} \cdot \frac{(1-\phi(q^{-1}))}{\phi(q^{-1})}, \\ b_m &\sim q^{k(mn-k)} \cdot \left(\binom{n}{i} - 1\right) \cdot \frac{(1-\phi(q^{-1}))^2}{\phi(q^{-1})}, \\ c_m &\sim q^{k(mn-k)} \cdot \binom{n}{i} \cdot \frac{(1-\phi(q^{-1}))^2}{\phi(q^{-1})}. \end{aligned}$$

The three estimates are of the form $a_m \sim a f_m, b_m \sim b f_m$ and $c_m \sim c f_m$, where $f_m = q^{k(mn-k)}$ and $a, b, c \in \mathbb{R}$ are positive constants such that

$$\frac{a+b-c}{a+b} = \frac{\phi(q^{-1})}{\phi(q^{-1}) + \binom{n}{i}(1-\phi(q^{-1}))} < 1.$$

Taking the limit superior as $m \to \infty$ we obtain

$$\limsup_{m\to\infty} \frac{\left|\hat{T}_q^{m,n,k}\right|}{\left|T_q^{m,n,k}\right|} \leq \frac{\phi(q^{-1})}{\phi(q^{-1}) + \binom{n}{i}(1-\phi(q^{-1}))}.$$

4 Conclusion

We derived an upper and a lower bound on the fraction of MDS codes within the set of all $\mathbb{F}_q$-linear codes of fixed dimension in $\mathbb{F}_{q^m}^n$. We showed that for $q \to \infty$ this fraction approaches 1, whereas for $m \to \infty$ this fraction is strictly smaller than 1. This behavior differs from the one of $\mathbb{F}_{q^m}$-linear MDS codes that are dense for both q and m, as well as from the one of $\mathbb{F}_q$-linear MRD codes that are sparse with respect to q.

References

1. A. Gruica and A. Ravagnani: Common complements of linear subspaces and the sparseness of MRD Codes. https://arxiv.org/abs/2011.02993, 2021.
2. A. Neri, A.-L. Horlemann-Trautmann, T. Randrianarisoa, and J. Rosenthal: On the genericity of maximum rank distance and Gabidulin codes. Designs, Codes and Cryptography 86 (2018), no. 2, 341-363.
3. E. Byrne and A. Ravagnani: Partition-balanced families of codes and asymptotic enumeration in coding theory, Journal of Combinatorial Theory, Series A 171 (2020).
4. H. Gluesing-Luerssen: On the sparseness of certain linear MRD codes, Linear Algebra and its Applications 596 (2020), 145-168.
5. J. MacWilliams and N. Sloane: The Theory of Error-Correcting Codes, Elsevier, 1977.
6. T.M. Apostol: Introduction to Analytic Number Theory, Springer Science and Business Media, 2013.