ISIT'2013 1569717375

Separating Redundancy of Linear MDS Codes

Khaled A. S. Abdel-Ghaffar
University of California, Dept. ECE
Davis, CA 95616, USA
Email: ghaffar@ece.ucdavis.edu

Jos H. Weber
Delft University of Technology, IRCTR/CWPC
Mekelweg 4, 2628 CD Delft, The Netherlands
Email: j.h.weber@tudelft.nl

***Abstract*—Linear codes over channels causing erasures and errors can be decoded by deleting the erased symbols and decoding the resulting vector with respect to a punctured code. To facilitate decoding of MDS codes, parity-check matrices are proposed that contain, as submatrices, parity-check matrices of the punctured codes. Depending on the maximum number of erasures, the separating redundancy, which is the smallest number of rows in the proposed parity-check matrices, is determined.**

I. Introduction

Typically, noisy communication channels cause errors in unknown positions and erasures, denoted by "?", in known positions. For reliable communication, linear coding schemes are introduced to combat channel noise. In such schemes, decoding a received word suffering from erasures and errors can be accomplished by deleting the erased symbols and decoding the resulting vector with respect to a punctured code. The erased symbols can then be retrieved. To implement this decoding scheme, the decoder needs a characterization, e.g, a parity-check matrix, of the punctured code which depends on the positions of the erasures in the received word. For the decoder to compute a parity-check matrix for the punctured code after receiving the word may require an unacceptable time delay. On the other hand, storing precomputed parity-check matrices of all punctured codes corresponding to all erasure patterns may not be feasible either.

In [1], the authors introduced the concept of separating erasures from errors. A parity-check matrix *separates* a given set of erased positions if it contains a number of rows with zeros in all erased positions such that the matrix composed of these rows, after deleting these zeros, is a parity-check matrix for the code punctured at the erased positions. An ℓ-*separating parity-check matrix* for a given code separates all sets of size up to ℓ. In particular, an ℓ-separating parity-check matrix embodies parity-check matrices of all punctured codes corresponding to ℓ erasures or less. Having a separating parity-check matrix available makes it unnecessary to compute or store parity-check matrices of the punctured codes for the purpose of decoding. Typically, a separating parity-check matrix contains redundant rows. It is shown in [1] that the parity-check matrix consisting of all codewords in the dual code is ℓ-separating for every ℓ smaller than the Hamming distance of the code. However, for ease of decoder implementation, parity-check matrices with small number of redundant rows are preferred. For a given linear code, the minimum number of rows in an ℓ-separating parity-check matrix, denoted by s_ℓ, is called the ℓ-*separating redundancy* of the code. Constructions of separating parity-check matrices of linear codes and bounds on their separating redundancies are presented in [1], [2], [13]. These papers seem to indicate that determining the separating redundancies of linear codes, or even obtaining reasonably tight bounds, is generally a difficult problem. In this paper, we focus on MDS (which refers to maximum-distance-separable) codes and show that their rich structure allows us to determine exactly their ℓ-separating redundancy for some values of ℓ.

The rest of the paper is organized as follows. In Section II we provide basic definitions and notation and we state some relevant results from [1]. Section III contains our main contributions related to MDS codes. Section IV concludes the paper.

II. Preliminaries

Let $\mathcal{C}$ be an $[n, k, d]$ linear code over GF(q), where n, k, and d denote the code's length, dimension, and Hamming distance, respectively, and q is a prime power. Such a code is a k-dimensional subspace of the space of vectors of length n over GF(q), in which any two different vectors differ in at least d positions. The set of codewords of $\mathcal{C}$ can be defined as the null space of the row space of an $r \times n$ parity-check matrix $\mathbf{H} = (h_{i,j})$ of rank $n-k$. The row space of $\mathbf{H}$ is the $[n, n-k, d^\perp]$ dual code $\mathcal{C}^\perp$ of $\mathcal{C}$.

The support of a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ over GF(q) is the set $\{j : x_j \neq 0\}$ and the weight of $\mathbf{x}$ is the size of its support set. If $\mathbf{x}$ is a nonzero vector, i.e., its support is nonempty, then $\mathbf{x}$ is *normalized* if its leading nonzero term is equal to 1, i.e., if $x_i = 0$ for all $i < j$ and $x_j = 1$ for some $j = 1, 2, \ldots, n$.

Let $\mathcal{U}$ be a subset of $\{1, 2, \ldots, n\}$ and $\mathcal{T}$ be a subset of $\{1, 2, \ldots, r\}$. For any $\mathbf{H} = (h_{i,j})$ of size $r \times n$, let $\mathbf{H}_\mathcal{U}^\mathcal{T} = (h_{i,j})$ where $i \in \mathcal{T}$ and $j \in \mathcal{U}$. Then, $\mathbf{H}_\mathcal{U}^\mathcal{T}$ is a $|\mathcal{T}| \times |\mathcal{U}|$ submatrix of $\mathbf{H}$. For simplicity, we write $\mathbf{H}_\mathcal{U}$ and $\mathbf{H}^\mathcal{T}$ to denote $\mathbf{H}_\mathcal{U}^\mathcal{T}$ in case $\mathcal{T} = \{1, 2, \ldots, r\}$ and $\mathcal{U} = \{1, 2, \ldots, n\}$, respectively. We allow for empty matrices, i.e., with no rows or no columns, which is the case if either $\mathcal{U}$ or $\mathcal{T}$ or both are empty. The rank of an empty matrix is defined to be zero. If $\mathbf{x}$ is a vector of length n, then $\mathbf{x}_\mathcal{U}$ denotes the vector whose components are indexed by $\mathcal{U}$. Furthermore, for the code $\mathcal{C}$ of length n, we define $\mathcal{C}_\mathcal{U} = \{\mathbf{c}_\mathcal{U} : \mathbf{c} \in \mathcal{C}\}$.

Let $\mathcal{S}$ be a subset of $\{1, 2 \ldots, n\}$ and $\bar{\mathcal{S}} = \{1, 2, \ldots, n\} \backslash \mathcal{S}$. Then, $\mathcal{C}_{\bar{\mathcal{S}}}$ is called the *punctured* code of $\mathcal{C}$ at $\mathcal{S}$, i.e., $\mathcal{C}_{\bar{\mathcal{S}}}$ consists of all codewords in $\mathcal{C}$ in which the components in the positions belonging to the set $\mathcal{S}$ are deleted. Clearly $\mathcal{C}_{\bar{\mathcal{S}}}$ is

a linear code over GF(q) of length $n' = n - |\mathcal{S}|$, dimension $k' \leq k$, and Hamming distance $d' \geq d - |\mathcal{S}|$. Furthermore, if $|\mathcal{S}| \leq d-1$, then $k' = k$ since the deletion of any number of components less than the Hamming distance of a code from two distinct codewords results in distinct vectors. It follows that if $|\mathcal{S}| \leq d-1$, then there is a one-to-one correspondence between the codes $\mathcal{C}$ and $\mathcal{C}_{\bar{\mathcal{S}}}$ such that $\mathbf{c}' \in \mathcal{C}_{\bar{\mathcal{S}}}$ if and only if there is a unique $\mathbf{c} \in \mathcal{C}$ such that $\mathbf{c}_{\bar{\mathcal{S}}} = \mathbf{c}'$. The dual code, $\mathcal{C}_{\bar{\mathcal{S}}}^{\perp}$, of $\mathcal{C}_{\bar{\mathcal{S}}}$ consists of all codewords in the dual code, $\mathcal{C}^{\perp}$, of $\mathcal{C}$ with zeros in all positions indexed by $\mathcal{S}$ after deleting these zeros, i.e., $\mathcal{C}_{\bar{\mathcal{S}}}^{\perp} = \{\mathbf{c}_{\bar{\mathcal{S}}} : \mathbf{c} \in \mathcal{C}^{\perp},\ \mathbf{c}_{\mathcal{S}} = \mathbf{0}\}$. We conclude, for $|\mathcal{S}| \leq d-1$, that a matrix is a parity-check matrix of the punctured code $\mathcal{C}_{\bar{\mathcal{S}}}$ if and only if it has rank $n-k-|\mathcal{S}|$ and all its rows are obtained from codewords in $\mathcal{C}^{\perp}$ with supports disjoint from $\mathcal{S}$ after deleting all symbols in positions indexed by $\mathcal{S}$, which are all zeros, from these codewords.

Let $\mathbf{r}$ be a received word corresponding to a transmitted codeword $\mathbf{c} \in \mathcal{C}$ and suppose that $\mathbf{r}$ suffers from $t_?$ erasures in the positions indexed by the set $\mathcal{S}$ in addition to $t_{\neq}$ errors, where $t_? + 2t_{\neq} \leq d-1$. Then, it is possible to decode $\mathbf{r}$ by deleting all its erasures to obtain the vector $\mathbf{r}_{\bar{\mathcal{S}}}$ and decode this vector with respect to the punctured code $\mathcal{C}_{\bar{\mathcal{S}}}$. If a parity-check matrix for the punctured code is available, then decoding the punctured code can be done by applying generic error-decoding techniques, ranging from syndrome decoding to iterative decoding methods based on belief propagation, to the parity-check matrices of the punctured codes. Once this is done, all errors in $\mathbf{r}_{\bar{\mathcal{S}}}$ are corrected and the word $\mathbf{r}$ is updated accordingly such that $\mathbf{r}_{\bar{\mathcal{S}}} = \mathbf{c}_{\bar{\mathcal{S}}}$. To retrieve the erased symbols, notice that $\mathbf{c}\mathbf{H}^{\mathrm{T}} = \mathbf{0}$, where the superscript T denotes transpose, gives r equations where each equation corresponds to a row in $\mathbf{H}$. By solving the simultaneous system of these equations, the $t_?$ erased symbols from $\mathbf{c}$ can be determined. This decoding algorithm is successful provided that $t_? + 2t_{\neq} \leq d-1$. However, to implement it, we need a parity-check matrix for the punctured code.

Let $\mathbf{H} = (h_{i,j})$ be an $r \times n$ matrix over GF(q) and $\mathcal{S}$ be a subset of $\{1, 2, \ldots, n\}$. We define

$$\widehat{\mathcal{S}} = \{i : 1 \leq i \leq r, h_{i,j} = 0 \quad \forall j \in \mathcal{S}\},$$

i.e., $\widehat{\mathcal{S}}$ is the set of indices of rows of the matrix $\mathbf{H}$ whose supports are disjoint from $\mathcal{S}$. Let

$$\mathbf{H}(\mathcal{S}) = \mathbf{H}_{\bar{\mathcal{S}}}^{\widehat{\mathcal{S}}},$$

i.e., $\mathbf{H}(\mathcal{S})$ is the submatrix of $\mathbf{H}$ obtained by deleting the columns indexed by $\mathcal{S}$ and all rows except those with supports disjoint from $\mathcal{S}$.

Suppose that $\mathbf{H}$ is a parity-check matrix for $\mathcal{C}$, then $\mathbf{H}$ separates $\mathcal{S}$ if $\mathbf{H}(\mathcal{S})$ is a parity-check matrix for the punctured code $\mathcal{C}_{\bar{\mathcal{S}}}$. For $\ell = 0, 1, \ldots, d-1$, $\mathbf{H}$ is ℓ-separating for $\mathcal{C}$ if it separates every set $\mathcal{S}$ of size $|\mathcal{S}| = 0, 1, \ldots, \ell$. Clearly, any parity-check matrix of any linear code is 0-separating. If we have an ℓ-separating parity-check matrix, $\mathbf{H}$, for $\mathcal{C}$, then we can successfully decode every word suffering from $t_? \leq \ell$ erasures and $t_{\neq}$ errors provided that $t_? + 2t_{\neq} \leq d-1$ with all parity-check matrices of punctured codes at the erased positions readily available from $\mathbf{H}$ as submatrices. It was shown in [1] that any ℓ-separating parity-check matrix for an $[n, k, d]$ linear code, where $\ell \leq \min\{d, n-k\} - 1$, has no stopping set of size ℓ or less. Hence, after decoding $\mathbf{r}_{\bar{\mathcal{S}}}$ successfully with respect to the punctured code $\mathcal{C}_{\bar{\mathcal{S}}}$, the erased symbols can be retrieved one-by-one from $\mathbf{H}$ since it has a row with exactly one position, among all positions indexed by $\mathcal{S}$, occupied by a nonzero entry. This row yields a parity-check equation involving only one unknown which is the erased symbol in that position. Determining this symbol reduces the number of erasures by one. The process can be repeated until all erased symbols are retrieved. In fact, this work is related to work on stopping sets, especially to [3], [5]–[10], [12], [14], with its emphasis on constructing parity-check matrices with minimum number of rows satisfying certain conditions to facilitate decoding. However, whereas work on stopping sets assumes that the channel does not cause errors, our work deals with errors in addition to erasures.

The basic concept behind the proposed decoding technique is best illustrated by an example as follows.

Example 1. Let $\mathcal{C}$ be the linear code over GF(8) with full-rank parity-check matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \alpha & \alpha^2 & \alpha^3 & \alpha^4 & \alpha^5 \\ 1 & \alpha^2 & \alpha^4 & \alpha^6 & \alpha^8 & \alpha^{10} \\ 1 & \alpha^3 & \alpha^6 & \alpha^9 & \alpha^{12} & \alpha^{15} \end{pmatrix}, \tag{1}$$

where α is a primitive element over GF(8) satisfying $\alpha^3 = \alpha + 1$. Then, $\mathcal{C}$ is a $[6, 2, 5]$ code.

Now, suppose that a codeword in $\mathcal{C}$ is transmitted and $\mathbf{r} = (\alpha, 0, \alpha^2, ?, \alpha, ?)$ is the received word. In this instance, the channel caused two erasures and possibly errors also. Deleting the erased symbols from $\mathbf{r}$ yields the vector $\mathbf{r}_{\overline{\{4,6\}}} = (\alpha, 0, \alpha^2, \alpha)$. To decode $\mathbf{r}_{\overline{\{4,6\}}}$ with respect to the punctured code $\mathcal{C}_{\overline{\{4,6\}}}$, we need a parity-check matrix for $\mathcal{C}_{\overline{\{4,6\}}}$. Deriving such a parity-check matrix from the parity-check matrix of $\mathcal{C}$ in (1) requires computations and time. Alternatively, storing a precomputed parity-check matrix for the punctured code requires memory to store parity-check matrices of all punctured codes as we do not know in advance which symbols will be erased by the channel.

Therefore, we propose a different parity-check matrix for the same code $\mathcal{C}$, namely,

$$\mathbf{H} = \begin{pmatrix} 1 & \alpha^5 & \alpha^6 & 0 & 0 & 0 \\ 1 & \alpha^4 & 0 & \alpha^5 & 0 & 0 \\ 1 & 0 & \alpha^4 & 0 & 1 & 0 \\ 0 & 1 & 0 & \alpha^2 & 1 & 0 \\ 0 & 0 & 1 & \alpha^4 & \alpha & 0 \\ 1 & 0 & 0 & \alpha^4 & 0 & \alpha^2 \\ 1 & 0 & 0 & 0 & \alpha^6 & \alpha \\ 0 & 1 & 1 & 0 & 0 & \alpha \\ 0 & 1 & 0 & 0 & \alpha^5 & \alpha^2 \\ 0 & 0 & 1 & \alpha^3 & 0 & \alpha^6 \end{pmatrix}. \tag{2}$$

Notice that each of the first and third rows in $\mathbf{H}$ has zeros as their fourth and sixth symbols. Hence, by deleting these symbols from these rows, we obtain the matrix

$$\mathbf{H}(\{4,6\}) = \begin{pmatrix} 1 & \alpha^5 & \alpha^6 & 0 \\ 1 & 0 & \alpha^4 & 1 \end{pmatrix},$$

of rank two and, therefore, is a parity-check matrix for the punctured code $\mathcal{C}_{\overline{\{4,6\}}}$. Hence, $\mathbf{H}$ separates the set $\{4,6\}$. The punctured code $\mathcal{C}_{\overline{\{4,6\}}}$ is a $[4,2,3]$ code capable of correcting one error. We can then decode $\mathbf{r}_{\overline{\{4,6\}}} = (\alpha, 0, \alpha^2, \alpha)$ based on $\mathbf{H}(\{4,6\})$ to yield the codeword $(\alpha, 0, \alpha^2, \alpha^5)$ in $\mathcal{C}_{\overline{\{4,6\}}}$. Updating $\mathbf{r}$ to be $(\alpha, 0, \alpha^2, ?, \alpha^5, ?)$, the fourth and sixth erased symbols can be retrieved from the parity-check equations arising from the second and seventh rows of $\mathbf{H}$ as α^3 and α, respectively. The estimated transmitted codeword corresponding to $\mathbf{r}$ is $(\alpha, 0, \alpha^2, \alpha^3, \alpha^5, \alpha)$, which is correct provided that the channel caused at most one error in addition to the two erasures.

It can be checked that $\mathbf{H}$ not only separates the set $\{4,6\}$ but also every set of size at most two. Hence, $\mathbf{H}$ is a 2-separating parity-check matrix for the code $\mathcal{C}$. Notice that $\mathbf{H}$ has ten rows. It will be shown in Section III that no 2-separating parity-check matrix for the code $\mathcal{C}$ has fewer than ten rows. In particular, for this code $s_2 = 10$. However, $\mathbf{H}$ is not a 3-separating parity-check matrix since it does not separate the set $\{1,2,3\}$ since $\mathbf{H}(\{1,2,3\})$ is the null matrix, as there is no row in $\mathbf{H}$ with zeros as its first three symbols while the rank of any parity-check matrix of $\mathcal{C}_{\overline{\{1,2,3\}}}$ is one. It will be shown in Theorem 1 that the minimum number of rows in a 3-separating parity-check matrix for $\mathcal{C}$ is 20.

In the next section, we will make use of the following four results which appear, along with their proofs, as Lemma 1, Lemma 2, Lemma 3, and Theorem 4, respectively, in [1].

Lemma 1. *Let $\mathbf{H}$ be a parity-check matrix of an $[n,k,d]$ linear code $\mathcal{C}$ over GF(q) and $\mathcal{S} \subset \{1,2,\ldots,n\}$ of size $|\mathcal{S}| \leq d-1$. Then, $\mathcal{C}_{\bar{\mathcal{S}}}$ is in the null space of $\mathbf{H}(\mathcal{S})$ which has rank at most equal to $n-k-|\mathcal{S}|$.*

Lemma 2. *A parity-check matrix $\mathbf{H}$ of an $[n,k,d]$ linear code $\mathcal{C}$ over GF(q) separates a set $\mathcal{S}$ of size $|\mathcal{S}| \leq d-1$ if and only if $\mathbf{H}(\mathcal{S})$ has rank $n-k-|\mathcal{S}|$.*

Lemma 3. *Let $\mathbf{H}$ be a parity-check matrix of an $[n,k,d]$ linear code over GF(q). If $\mathbf{H}$ separates all sets of size ℓ for a fixed $\ell \leq \min\{d, n-k\} - 1$, then it is ℓ-separating.*

Lemma 4. *Let $\mathcal{C}$ be an $[n,k,d]$ linear code over GF(q). Then, for each nonnegative integer $\ell \leq \min\{d, n-k\} - 1$,*

$$s_\ell \geq \frac{\binom{n}{\ell}(n-k-\ell)}{\binom{n-d^\perp}{\ell}},$$

where $d^\perp$ is the Hamming distance of $\mathcal{C}^\perp$.

III. Results on the Separating Redundancy of MDS Codes

An $[n,k,d]$ linear code is MDS if its Hamming distance, d, satisfies the Singleton bound $d \leq n-k+1$ with equality [11]. It is well known that the dual code, $\mathcal{C}^\perp$, of an $[n,k,n-k+1]$ linear MDS code $\mathcal{C}$ is an $[n,n-k,k+1]$ MDS code. Furthermore, for any subset $\mathcal{U} \subseteq \{1,2,\ldots,n\}$ of size $k+1$, there is exactly one normalized codeword in $\mathcal{C}^\perp$ whose support is $\mathcal{U}$ [11]. Therefore, there are exactly

$$\binom{n}{k+1}$$

normalized codewords in $\mathcal{C}^\perp$ of weight $k+1$.

Since the Hamming distance of $\mathcal{C}$ is $d = n-k+1$, we are interested in the ℓ-separating redundancies for $\ell \leq d-1 = n-k$. As the Hamming distance of $\mathcal{C}^\perp$ is $d^\perp = k+1$, it follows from Lemma 4 that

$$s_\ell \geq \frac{\binom{n}{\ell}(n-k-\ell)}{\binom{n-k-1}{\ell}} \tag{3}$$

for $\ell \leq n-k-1$. Next we determine the exact values of the separating redundancies s_{n-k} and s_{n-k-1}.

Theorem 1. *For an $[n,k,n-k+1]$ linear MDS code $\mathcal{C}$ over GF(q), where $n-k \geq 2$,*

$$s_{n-k} = s_{n-k-1} = \binom{n}{k+1}.$$

Proof. From Lemmas 1 and 2, it follows that any parity-check matrix for $\mathcal{C}$ separates every subset $\{1,2,\ldots,n\}$ of size $n-k$. Hence, $s_{n-k} = s_{n-k-1}$. From (3), s_{n-k-1} is lower bounded by $\binom{n}{k+1}$. In the following, we construct a parity-check matrix for $\mathcal{C}$ that has that many rows and which is $(n-k-1)$-separating. Let $\mathbf{H}$ be a matrix whose rows are all the normalized codewords in $\mathcal{C}^\perp$ of weight $k+1$. Since $\mathcal{C}^\perp$ is an $[n,n-k,k+1]$ MDS code, for every subset of $\{1,2,\ldots,n\}$ of size $k+1$, there is a unique normalized codeword in $\mathcal{C}^\perp$ whose support is that subset. Hence, $\mathbf{H}$ has exactly $\binom{n}{k+1}$ rows. The $n-k$ codewords with supports $\{1,2,\ldots,k\} \cup \{j\}$, for $j = k+1, k+2, \ldots, n$, are linearly independent. Hence, $\mathbf{H}$ is a parity-check matrix for $\mathcal{C}$. To show that $\mathbf{H}$ is $(n-k-1)$-separating, consider an arbitrary subset $\mathcal{S} \in \{1,2,\ldots,n\}$ of size $n-k-1$. As $|\bar{\mathcal{S}}| = k+1$, there is exactly one row in $\mathbf{H}$ whose support is $\bar{\mathcal{S}}$. We conclude that $\mathbf{H}(\mathcal{S})$ contains exactly one row of weight $k+1$ and, therefore, its rank is 1. From Lemma 2, $\mathbf{H}$ separates $\mathcal{S}$. Since $\mathcal{S}$ is arbitrary, $\mathbf{H}$ separates all sets of size $n-k-1$. From Lemma 3, it follows that $\mathbf{H}$ is $(n-k-1)$-separating. □

Next, we consider the separating redundancy s_{n-k-2} of the MDS code $\mathcal{C}$. For this purpose, recall that a (v,u,t) Turán design, where $1 \leq t \leq u \leq v$, is a collection of subsets of $\{1,2,\ldots,v\}$ of size t, called blocks, such that every subset of $\{1,2,\ldots,v\}$ of size u contains at least one block [4]. We use a variation of this concept and define a $(v,u,t)_p$ Turán design to be a collection of distinct subsets of $\{1,2,\ldots,v\}$ of size t, also called blocks, such that every subset of $\{1,2,\ldots,v\}$ of

size u contains at least p blocks. Notice that in this definition we require that the blocks of the design be distinct. Such a design exists if and only if $p \leq \binom{u}{t}$. Provided that this is the case, let $T_p(v,u,t)$ be the minimum size of a $(v,u,t)_p$ Turán design, where the size of a Turán design is the number of its blocks.

Theorem 2. *For an $[n,k,n-k+1]$ linear MDS code $\mathcal{C}$ over GF(q), where $n-k \geq 3$,*

$$s_{n-k-2} = T_2(n,k+2,k+1).$$

Proof. From Lemmas 2 and 3, a necessary and sufficient condition for a parity-check matrix $\mathbf{H}$ to be $(n-k-2)$-separating is that for any subset $\mathcal{S} \subseteq \{1,2,\ldots,n\}$ of size $n-k-2$, the set $\{1,2,\ldots,n\}\backslash\mathcal{S}$ of size $k+2$ contains the supports of two linearly independent rows in $\mathbf{H}$. Let $\mathbf{H}$ be an $(n-k-2)$-separating parity-check matrix for $\mathcal{C}$ that has s_{n-k-2} rows. Suppose that $\mathbf{H}$ has a row, $\mathbf{r}$, of weight $k+2$, whose support is the set $\mathcal{U}$. We will argue that we can replace $\mathbf{r}$ by another vector of weight $k+1$ such that the resulting matrix is still an $(n-k-2)$-separating parity-check matrix of $\mathcal{C}$. Notice that this replacement does not affect the capability to separate any set of size $n-k-2$ other than $\{1,2,\ldots,n\}\backslash\mathcal{U}$. The row $\mathbf{r}$ is not a linear combination of other rows in $\mathbf{H}$ whose supports are contained in $\mathcal{U}$; otherwise by deleting $\mathbf{r}$ from $\mathbf{H}$, we obtain a parity-check matrix which is $(n-k-2)$-separating but has fewer rows than $\mathbf{H}$. Hence, there is another row, $\mathbf{r}'$, in $\mathbf{H}$ whose support is contained in $\mathcal{U}$ such that $\mathbf{r}$ and $\mathbf{r}'$ are linearly independent. There is a nonzero element $\alpha \in$ GF(q) such that $\mathbf{r}+\alpha\mathbf{r}'$ is a nonzero vector whose support is of size at most $k+1$. Since $\mathbf{r}+\alpha\mathbf{r}'$ is a codeword in $\mathcal{C}^\perp$, the size of its support is exactly $k+1$. As $\mathbf{r}$ and $\mathbf{r}'$ are linearly independent, $\mathbf{r}'$ and $\mathbf{r}+\alpha\mathbf{r}'$ are also linearly independent. By replacing $\mathbf{r}$ in $\mathbf{H}$ by $\mathbf{r}+\alpha\mathbf{r}'$, the resulting matrix is still an $(n-k-2)$-separating parity-check matrix for $\mathcal{C}$ where a row of weight $k+2$ is replaced by a row of weight $k+1$. By repeating this process, we obtain an $(n-k-2)$-separating parity-check matrix for $\mathcal{C}$ that has no rows of weight $k+2$. Let $\mathcal{Y}$ be the set of supports of all rows in $\mathbf{H}$ of weight $k+1$. Every subset $\mathcal{U} \subseteq \{1,2,\ldots,n\}$ of size $k+2$ contains the supports of at least two linearly independent rows of weight $k+1$. The supports of these two rows are distinct otherwise a linear combination of them gives a nonzero codeword in $\mathcal{C}^\perp$ of weight less than $k+1$. We conclude that every subset $\mathcal{U} \subseteq \{1,2,\ldots,n\}$ of size $k+2$ contains two subsets in $\mathcal{Y}$. This gives an $(n,k+2,k+1)_2$ Turán design. The size, $|\mathcal{Y}|$, of this design is at least equal to $T_2(n,k+2,k+1)$, which is the minimum size of an $(n,k+2,k+1)_2$ Turán design. Since $\mathbf{H}$ has at least $|\mathcal{Y}|$ rows, $s_{n-k-2} \geq T_2(n,k+2,k+1)$. Next, we show that if we are given an $(n,k+2,k+1)_2$ Turán design, then we can construct an $(n-k-2)$-separating parity-check matrix for $\mathcal{C}$ whose number of rows equals the number of blocks. The key point is to notice that every subset of $\{1,2,\ldots,n\}$ of size $k+1$ is the support of a unique normalized codeword in $\mathcal{C}^\perp$. This gives a one-to-one correspondence between subsets of size $k+1$ and normalized codewords in $\mathcal{C}^\perp$ of weight $k+1$. Let $\mathbf{H}$ be the matrix whose rows are the codewords corresponding to the blocks of the $(n,k+2,k+1)_2$ Turán design. If $\mathcal{S}$ is a subset of $\{1,2,\ldots,n\}$ of size $n-k-2$, then its complement $\mathcal{U} = \{1,2,\ldots,n\}\backslash\mathcal{S}$ of size $k+2$ contains two blocks, i.e., $\mathbf{H}$ has two rows with distinct supports that are disjoint from $\mathcal{S}$. Hence, these two rows are linearly independent and the rank of $\mathbf{H}(\mathcal{S})$ is at least two. From Lemma 1, it follows that the rank of $\mathbf{H}(\mathcal{S})$ is two. It remains to show that $\mathbf{H}$ is a parity-check matrix of $\mathcal{C}$. Let $\mathcal{V}$ be a block in the $(n,k+2,k+1)_2$ Turán design. For every $j \in \{1,2,\ldots,n\}\backslash\mathcal{V}$, the subset $\mathcal{V} \cup \{j\} \subseteq \{1,2,\ldots,n\}$ is of size $k+2$ and hence should contain a block $\mathcal{V}_j$ other than $\mathcal{V}$. Notice that the $n-k$ blocks $\mathcal{V}$ and $\mathcal{V}_j$, for $j \in \{1,2,\ldots,n\}\backslash\mathcal{V}$, are distinct and, therefore, they correspond to linearly independent rows in $\mathbf{H}$. This proves that $\mathbf{H}$ has rank $n-k$. □

As an example, let $n=6$ and $k=2$. Then, $s_2 = T_2(6,4,3)$, which is the minimum size of a $(6,4,3)_2$ Turán design. Notice that in such a design, every block, being of size three, is contained in exactly $6-3=3$ subsets of $\{1,2,\ldots,6\}$ of size four and there are exactly $\binom{6}{4}=15$ such subsets. Since each one of these subsets contains at least two blocks, it follows that the number of blocks is at least $T_2(6,4,3) \geq 2\times 15/3 = 10$. The 2-separating parity-check matrix, $\mathbf{H}$ in (2), for the $[6,2,5]$ MDS code presented in Example 1 is obtained from a $(6,4,3)_2$ Turán design of size ten. The blocks of the Turán design are the supports of the rows of $\mathbf{H}$.

Next, we consider the case $\ell=1$ for an $[n,k,n-k+1]$ MDS code $\mathcal{C}$ over GF(q). From (3), we have the lower bound $s_1 \geq n$. The following result shows that equality holds if $n-k \geq 1$.

Theorem 3. *For an $[n,k,n-k+1]$ linear MDS code $\mathcal{C}$ over GF(q), where $n-k \geq 1$,*

$$s_1 = n.$$

Proof. For $i=1,2,\ldots,n$, let $\mathcal{U}_i$ be the set of $k+1$ consecutive integers starting with i and reduced modulo n to be positive integers less than or equal to n. Then, there is a normalized codeword in $\mathcal{C}^\perp$ whose support is $\mathcal{U}_i$. Let $\mathbf{H}$ be the $n\times n$ matrix whose ith row is this normalized codeword. Since all the rows of $\mathbf{H}$ are codewords in $\mathcal{C}^\perp$ which has dimension $n-k$, the rank of $\mathbf{H}$ is at most $n-k$. Notice that for $i=1,2,\ldots,n$, the ith row in $\mathbf{H}$ has a run of $n-k-1$ cyclically consecutive zeros and this run is cyclically followed by a nonzero element in the ith column. Hence, for $j=1,2,\ldots,n$, the jth column in $\mathbf{H}$ has a run of $n-k-1$ cyclically consecutive zeros and this run cyclically follows a nonzero element in the jth row. It follows that any $n-k$ cyclically consecutive rows are linearly independent. We conclude that $\mathbf{H}$ is a parity-check matrix for $\mathcal{C}$ and for every $j=1,2,\ldots,n$, the set $\widehat{\{j\}}$ contains the indices of $n-k-1$ cyclically consecutive rows in $\mathbf{H}$. These rows are linearly independent and, therefore, $\mathbf{H}(\{j\})$, which is obtained by deleting the jth component, which is zero, from each of these rows, has rank $n-k-1$. This proves that $\mathbf{H}$ separates $\{j\}$ for every $j=1,2,\ldots,n$. □

IV. Discussion and Conclusion

In this paper we determined the ℓ-separating redundancy of MDS codes for some values of ℓ. In all cases considered so far in which s_ℓ is determined for an $[n, k, n-k+1]$ MDS code $\mathcal{C}$, an ℓ-separating parity-check matrix is constructed with s_ℓ rows of weight $k+1$. However, this is not true in general, i.e., a code may have an ℓ-separating parity-check matrix whose number of rows is less than the number of rows in any ℓ-separating parity-check matrix all of whose rows are of weight $k+1$. Indeed, consider the $[7, 1, 7]$ binary repetition code, $\mathcal{C}$, which is an MDS code. Suppose that $\mathbf{H}$ is a 2-separating parity-check matrix for $\mathcal{C}$ with $r \leq 10$ rows, each having weight two. Then the number of ones in $\mathbf{H}$ is at most $2r \leq 20$. In particular, there is a column with no more than $\lfloor 2r/7 \rfloor \leq \lfloor 20/7 \rfloor = 2$ ones, i.e., of weight 0, 1, or 2. Without loss of generality, assume that it is the first column. Choose j_1 and j_2, where $1 < j_1 < j_2 \leq 7$, such that each row in $\mathbf{H}$ with a one in the first column, if any, has a one in either column j_1 or column j_2. Then the matrix $\mathbf{H}(\{j_1, j_2\})$ has five columns, the first of which is all zeros and the sum of the columns is the all-zeros vector as each row in $\mathbf{H}$ has two ones. We conclude that the rank of $\mathbf{H}(\{j_1, j_2\})$ is at most three and, therefore, $\mathbf{H}$ does not separate $\{j_1, j_2\}$ and hence is not a 2-separating parity-check matrix. On the other hand, it is possible to check that

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix},$$

composed of nine rows of weight two and one row of weight four, is a 2-separating parity-check matrix for the $[7, 1, 7]$ binary repetition code. Actually, for this code, this matrix has the smallest number of rows in any 2-separating parity-check matrix, i.e., $s_2 = 10$. Indeed, if $\mathbf{H}$ is a 2-separating parity-check matrix with $r \leq 9$ rows, then as we just argued, not all its rows are of weight two. Let $r_2 \leq r - 1 \leq 8$ be the number of rows in $\mathbf{H}$ of weight two. Then, $\mathbf{H}$ has $r - r_2 \geq 1$ rows of weight at least four. Consider the collection of submatrices $\mathbf{H}^{\widehat{\mathcal{S}}}$ for all subsets $\mathcal{S} \subseteq \{1, 2, \ldots, 7\}$ of size two. There are $\binom{7}{2} = 21$ such submatrices. As $\mathbf{H}(\mathcal{S})$ has rank four for each subset $\mathcal{S}$ of size two, $\mathbf{H}^{\widehat{\mathcal{S}}}$ also has rank four and, therefore, has at least four rows. We conclude that the sum of the number of rows in the submatrices in the collection is at least $4 \times 21 = 84$. Each row in $\mathbf{H}^{\widehat{\mathcal{S}}}$ has zeros in the two positions indexed by $\mathcal{S}$. Hence, a row in $\mathbf{H}$ of weight two appears in at most $\binom{5}{2} = 10$ such submatrices and a row of weight four or more appears in at most $\binom{3}{2} = 3$ such submatrices. Therefore the total number of appearances of the rows of $\mathbf{H}$ in the collection of submatrices $\mathbf{H}^{\widehat{\mathcal{S}}}$ for all subsets $\mathcal{S} \subseteq \{1, 2, \ldots, 7\}$ of size two is at most $10r_2 + 3(r - r_2) = 3r + 7r_2 \leq 3 \times 9 + 7 \times 8 = 83$. This number is less than 84, which is a lower bound on the sum of the number of rows in these submatrices. This contradiction proves that $\mathbf{H}$ is not a 2-separating parity-check matrix for the $[7, 1, 7]$ linear MDS code. We conclude that $s_2 = 10$ for this code. Notice that this is strictly larger than the ceiling of the lower bound on s_2 as given by (3). Hence, although in all cases considered in Section III, s_ℓ equals the lower bound (3), this is not generally true.

Our work also indicates some similarity in the difficulties faced in determining the separating redundancies and the stopping redundancies of MDS codes. For example, we have shown that the $(n - k - 2)$-separating redundancy of MDS codes equals the minimum size of some variation of Turán designs. It is interesting to note that the stopping redundancy for an MDS code is lower bounded by the minimum size of Turán designs and upper bounded by the minimum size of some special Turán designs called single-exclusion systems [5], [7]. Except for some special cases, the minimum sizes of these Turán designs or their variations are not known.

References

[1] K. A. S. Abdel-Ghaffar and J. H. Weber, "Separating erasures from errors for decoding," in *Proc. Int. Symp. Inform. Theory*, Toronto, Canada, pp. 215–219, July 6–11, 2008.

[2] K. A. S. Abdel-Ghaffar and J. H. Weber, "An upper bound on the separating redundancy of linear block codes," in *Proc. Int. Symp. Inform. Theory*, Austin, TX, pp. 1173–1177, July 13–18, 2010.

[3] R. Ahlswede and H. Aydinian, "On generic erasure correcting sets and related problems," *IEEE Trans. Inform. Theory*, vol. 58, no. 2, pp. 501–508, February 2012.

[4] C. J. Colbourn and J. H. Dinitz, Eds., *The CRC Handbook of Combinatorial Designs*. Boca Raton, FL, USA: CRC Press, 1996.

[5] J. Han and P. H. Siegel, "Improved upper bounds on stopping redundancy," *IEEE Trans. Inform. Theory*, vol. 53, no. 1, pp. 90–104, January 2007.

[6] J. Han and P. H. Siegel, "On ML redundancy of codes," in *Proc. Int. Symp. Inform. Theory*, Toronto, Canada, July 6–11, 2008, pp. 280–284.

[7] J. Han, P. H. Siegel, and R. M. Roth, "Single-exclusion number and the stopping redundancy of MDS codes," *IEEE Trans. Inform. Theory*, vol. 55, no. 9, pp. 4155–4166, September 2009.

[8] J. Han, P. H. Siegel, and A. Vardy, "Improved probabilistic bounds on stopping redundancy," *IEEE Trans. Inform. Theory*, vol. 54, no. 4, pp. 1749–1753, April 2008.

[9] H. D. L. Hollmann and L. M. G. M. Tolhuizen, "Generic erasure correcting sets: bounds and constructions," *J. Combin. Theory, Ser. A*, vol. 113, pp. 1746–1759, 2006.

[10] H. D. L. Hollmann and L. M. G. M. Tolhuizen, "On parity check collections for iterative erasure decoding that correct all correctable erasure patterns of a given size," *IEEE Trans. Inform. Theory*, vol. 53, no. 2, pp. 823–828, February 2007.

[11] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1977.

[12] M. Schwartz and A. Vardy, "On the stopping distance and the stopping redundancy of codes," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 922–932, March 2006.

[13] N. M. Tri, J. H. Weber, and K. A. S. Abdel-Ghaffar, "New upper bounds on the separating redundancy of linear block codes," in *Proc. Thirtieth Symposium on Information Theory in the Benelux*, Eindhoven, The Netherlands, pp. 209-216, May 28-29, 2009.

[14] J. H. Weber and K. A. S. Abdel-Ghaffar, "Results on parity-check matrices with optimal stopping and/or dead-end set enumerators," *IEEE Trans. Inform. Theory*, vol. 54, no. 3, pp. 1368–1374, March 2008.