EFFICIENT COMPUTATION OF THE SET OF CODEWORDS OF MINIMAL SUPPORT

IRENE MÁRQUEZ-CORBELLA AND EDGAR MARTÍNEZ-MORO

Abstract. In order to obtain the set of codewords of minimal support of codes defined over $\mathbb{Z}_q$ we must compute a Graver basis of the ideal associated to such codes, see [9]. The main aim of this article is to reduce the complexity of the previous algorithm taking advantage of the powerful decomposition theory for linear codes provided by the decomposition theory of representable matroids over finite fields. Following the works of Kashyap [6] we achieve our goal for every binary linear code and for the rest several improvement are presented.

Introduction

By $\mathbb{Z}$, $\mathbb{Z}_q$, $\mathbb{F}_q$ we mean the ring of integers, the ring of integers modulo q and the finite field with q elements. A *modular code* $\mathcal{C}$ over $\mathbb{Z}_q$ of length n and dimension k is a subspace of the abelian group $\left\langle \mathbb{Z}_q^n, + \right\rangle$, but for the shake of brevity we will just say an $[n,k]$ code. Note that elementary row operations can also be carried over any generator matrix of modular codes with the understanding that only multiplication of a row by a unit is allowed.

Let $\mathbf{x} \in \mathbb{Z}_q^n$ and J be a subset of $\{1,\ldots,n\}$, we denote by $\mathbf{x}_J$ the restriction of $\mathbf{x}$ to the coordinates indexed by J and by $\overline{J}$ the relative complement of J in $\{1,\ldots,n\}$. The process of deleting columns from a parity check matrix of an $[n,k,d]$ code $\mathcal{C}$ is known as *shortening* where d denotes the minimum distance between distinct codewords. The shortened code $\mathcal{C}._J$ is obtained by puncturing at J the set of codewords that have a zero in the J-locations, i.e. $\mathcal{C}._J = \left\{ \mathbf{c}_{\overline{J}} \mid \mathbf{c} \in \mathcal{C} \text{ and } \mathbf{c}_J = 0 \right\}$. If J consist of m elements then the shortened code $\mathcal{C}._J$ has parameters $[n-m, k', d']$ with $k - m \leq k' \leq k$ and $d \leq d'$.

From now on, for any positive integer i, let $\mathcal{C}_i$ be an $[n_i, k_i]$ modular code over $\mathbb{Z}_q$ defined on the index set I_i. We define the *star product* of two linear codes $\mathcal{C}_1$ and $\mathcal{C}_2$, denoted by $\mathcal{C}_1 * \mathcal{C}_2$ as the set of words of the form the form $\mathbf{c} = (c_i \mid i \in I_1 \cup I_2)$ where

$$c_i = \begin{cases} c_1^{(i)} & i \in I_1 \setminus I_2 \\ c_2^{(i)} & i \in I_2 \setminus I_1 \\ c_1^{(i)} + c_2^{(i)} & i \in I_1 \cap I_2 \end{cases} \quad \text{for some } \mathbf{c}_1 = (c_1^{(i)} \mid i \in I_1) \in \mathcal{C}_1 \text{ and } \mathbf{c}_2 = (c_2^{(i)} \mid i \in I_2) \in \mathcal{C}_2.$$

From $\mathcal{C}_1 * \mathcal{C}_2$ we can obtain a new code $\mathcal{S}(\mathcal{C}_1, \mathcal{C}_2)$ by shortening at the $m = |I_1 \cap I_2|$ positions where $\mathcal{C}_1$ and $\mathcal{C}_2$ overlap. The codewords of this code will be denoted by $\mathbf{c}_1 \ ||_m \ \mathbf{c}_2$ where $\mathbf{c}_1 \in \mathcal{C}_1$ and $\mathbf{c}_2 \in \mathcal{C}_2$. We are interested in just an specific case of the above construction called *r-sum* and denoted by $\mathcal{C}_1 \oplus_r \mathcal{C}_2$, for each positive integer r, in which the codes $\mathcal{C}_1$ and $\mathcal{C}_2$ must verify some specific conditions. For further details and terminology, we refer the reader to [6, 7]. $\mathcal{C}_1 \oplus_r \mathcal{C}_2$ is a linear code of lenght $n_1 + n_2 - 2|I_1 \cap I_2|$ and dimension $\dim(\mathcal{C}_1) + \dim(\mathcal{C}_2) - r$.

We will use the following *characteristic crossing functions*:

$$\blacktriangledown : \ \mathbb{Z}^s \ \longrightarrow \ \mathbb{Z}_q^s \qquad \text{and} \qquad \blacktriangle : \ \mathbb{Z}_q^s \ \longrightarrow \ \mathbb{Z}^s$$

The map $\blacktriangledown$ is reduction modulo q whereas the map $\blacktriangle$ replaces the class of $0, 1 \ldots, q-1$ by the same symbols regarded as integers. The integer s is determined by context and both maps act coordinate-wise.

Let $\mathbf{x}$ denote the set of variables $x_1, \ldots, x_n$. Given an $[n, k]$ code $\mathcal{C}$ and letting $\mathbf{w}_1, \ldots, \mathbf{w}_k$ be the rows of a generator matrix of $\mathcal{C}$, then we define the *ideal associated to* $\mathcal{C}$, denoted by $I(\mathcal{C})$ as the subset: $\langle \{\mathbf{x}^{\blacktriangle \mathbf{w}_1} - 1, \ldots, \mathbf{x}^{\blacktriangle \mathbf{w}_k} - 1\} \cup \{x_i^q - 1 \mid i = 1, \ldots, n\}\rangle \subseteq \mathbb{F}_2[\mathbf{x}]$. Ikegami and Kaji [5] gives us a method for computing a test set for the code $\mathcal{C}$ which works only in the binary case. This complete decoding scheme is equivalent to the gradient descent decoding given by Barg [1] which has been proven to be equivalent to Liebler approach [8] in [4]. In [9] we consider the Graver basis associated to a modular integer programming problem that provides us a universal test set, which turns out to be the set of codewords of minimal support of codes defined on $\mathbb{Z}_q$. A codeword $\mathbf{c}$ is a *minimal support codeword* if it is non-zero and $\mathrm{supp}(\mathbf{c})$ is not contained in the support of any other codeword. We will denote by $\mathcal{M}_{\mathcal{C}}$ the set of codewords of minimal support of $\mathcal{C}$. This result give us a method to compute $\mathcal{M}_{\mathcal{C}}$ for any linear code defined on $\mathbb{Z}_q$. In particular for codes over $\mathbb{F}_p$ with p prime. But not for the case $q = p^r$ since $\mathbb{F}_{p^r} \neq \mathbb{Z}_{p^r}$.

Therefore, in order to obtain a Gröbner test-set, for the binary case, or the set of codewords of minimal support for any linear code $\mathcal{C}$ defined over $\mathbb{Z}_q$, we must compute a reduced Gröbner basis of an ideal from whom we known a generating set, thus we can use the FGLM-based trick in [3]. Note that the complexity of this algorithm was stated in [3, 2] and it is $\mathcal{O}\left(n^2 q^{n-k}\right)$ where k is the dimension of the code and n is the number of variables involved in our ideal. The main task of this paper is trying to reduce the complexity of the previous algorithms by using the decomposition of a code into smaller ones.

1. Direct sum

Definition 1.1. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be linear codes over $\mathbb{Z}_q$ defined on mutually disjoint index set I_1 and I_2, i.e. $I_1 \cap I_2 = \emptyset$. We can construct the code $\mathcal{C}_1 \oplus \mathcal{C}_2$ with $I_1 \cup I_2$ as its index set such that any codeword $\mathbf{c}$ of $\mathcal{C}$ is defined by $\mathbf{c} = (\mathbf{c}_i \mid i \in I_1 \cup I_2)$ where

$$c_i = \begin{cases} c_1^{(i)} & \text{for } i \in I_1 \quad \text{for some } \mathbf{c}_1 = (c_i \mid i \in I_1) \in \mathcal{C}_1 \\ c_2^{(i)} & \text{for } i \in I_2 \quad \text{for some } \mathbf{c}_2 = (c_i \mid i \in I_2) \in \mathcal{C}_2 \end{cases}$$

The following proposition states the connection between the direct sum of two codes and the sum of its associated ideals.

Proposition 1.2. *$\mathcal{C} = \mathcal{C}_1 \oplus \mathcal{C}_2$ if and only if $I(\mathcal{C}) = I(\mathcal{C}_1) + I(\mathcal{C}_2)$.*

Corollary 1.3. *Let $\mathcal{C} = \mathcal{C}_1 \oplus \mathcal{C}_2$ then $I(\mathcal{C})$ is generated by the disjoint union of the generators of each $I(\mathcal{C}_i)$ with $i = 1, 2$.*

The above result is also true for Gröbner basis and Graver basis. Consequently we can study the test-sets (or the set of codewords of minimal support) of $\mathcal{C}$ by using the test-set (or the set codewords of minimal support) of each $\mathcal{C}_i$ with $i = 1, 2$.

The definition of direct sum can be easily extends to a finite family of linear codes. Indeed let $\{\mathcal{C}_\alpha\}_{\alpha\in A}$ be linear codes over $\mathbb{Z}_q$ of parameters $[n_\alpha, k_\alpha]$ and defined on mutually disjoint index sets I_α with $\alpha \in A$ then we can define $\oplus_{\alpha\in A}\mathcal{C}_\alpha$ and a similar study of this code can be done. Therefore if we achieve to decompose a code $\mathcal{C}$ as a direct sum of several smaller codes $\mathcal{C}_i$, then the cost of computing a Gröbner test set (or the set of codewords of minimal support) for $\mathcal{C}$ is reduced to computing a Gröbner test set for every C_i that appears on its decomposition. Furthermore, since this procedure can be parallelize, then we can reduce the time required for computing a Gröbner test set of $\mathcal{C}$ to the time needed to compute a Gröbner test set of the largest $\mathcal{C}_i$ that appears on its decomposition.

2. 2-sum

Definition 2.1. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be modular codes over $\mathbb{Z}_q$ of length at least 3 such that $I_1 \cap I_2 = j$. Moreover, if $\mathbf{c} = \big(c_i \in \mathbb{Z}_q^*$ for $i = j$ and 0 otherwise$\big)$ is not a codeword of $\mathcal{C}_i$ and the j-th coordinate of $\mathcal{C}_i$ is neither identically zero nor a zero divisor, for $i = 1$ or 2, then the 2-sum code $\mathcal{C}_1 \oplus_2 \mathcal{C}_2$ can be defined.

The following lemma allows us to characterize the set of codewords of the code $\mathcal{C}_1 \oplus_2 \mathcal{C}_2$.

Lemma 2.2. $\mathbf{c} \in \mathcal{C}_1 \oplus_2 \mathcal{C}_2$ *if and only if there exists two codewords* $\mathbf{c}_1 \in \mathcal{C}_1$ *and* $\mathbf{c}_2 \in \mathcal{C}_2$ *such that* $c_1^{(n_1)} + c_2^{(1)} = 0$ *and* $\mathbf{c} = \mathbf{c}_1 \mid\mid_2 \mathbf{c}_2$.

The following proposition states the connection between the 2-sum of two codes and the sum of its associated ideals.

Proposition 2.3. $\mathcal{C} = \mathcal{C}_1 \oplus_2 \mathcal{C}_2$ *if and only if* $I(\mathcal{C}) = I(\mathcal{C}_{1\cdot\{n_1\}}) + I(\mathcal{C}_{2\cdot\{1\}}) + \langle \mathbf{x}^{\blacktriangle\gamma_x} - \mathbf{y}^{\blacktriangle\gamma_y}\rangle$ *for some special vectors* $\gamma_x \in \mathbb{Z}_q^{n_1-1}$ *and* $\gamma_y \in \mathbb{Z}_q^{n_2-1}$.

Next proposition is related to the set of codewords of minimal support. In its proof we use some notions from $\mathbb{F}_q$-representable matroid theory, thus our results are restricted to linear codes defined over $\mathbb{F}_p$ with p prime, or similarly for any code defined on $\mathbb{Z}_p$ with p prime.

Proposition 2.4. *Let* $\mathcal{C} = \mathcal{C}_1 \oplus_2 \mathcal{C}_2$ *be a linear code defined over* $\mathbb{F}_p$ *with* p *prime, then the following statements are equivalent:*

(1) $\mathbf{c} \in \mathcal{M}_\mathcal{C}$.
(2) $\mathbf{c}$ *belongs to one of the following sets:*

$$\mathcal{M}_{\mathcal{C}_{1\cdot\{n_1\}}} \mid\mid_2 \mathbf{0},\quad \mathbf{0} \mid\mid_2 \mathcal{M}_{\mathcal{C}_{2\cdot\{1\}}},\quad \textit{or}\quad \left\{\mathbf{c}_1 \mid\mid_2 \mathbf{c}_2 \;:\mathbf{c}_1 \in \mathcal{M}_{\mathcal{C}_1\setminus\mathcal{C}_{1\cdot\{n_1\}}}, \mathbf{c}_2 \in \mathcal{M}_{\mathcal{C}_2\setminus\mathcal{C}_{2\cdot\{1\}}}\right\}.$$

3. 3-sum

Definition 3.1. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be modular codes over $\mathbb{Z}_q$ of length at least 7 such that $|I_1 \cap I_2| = 3$. If $\mathbf{c} = (c_j = 1$ for $j \in I_1 \cap I_2$ and 0 otherwise$) \in \mathcal{M}_{\mathcal{C}_i}$ and all possible 3-bit words appear in the $|I_1 \cap I_2|$ coordinates of $\mathcal{C}_i$ for $i = 1$ or 2, then the 3-sum code $\mathcal{C}_1 \oplus_3 \mathcal{C}_2$ can be defined.

Analogous to the 2-sum case we have a characterization Lemma for the set of codewords of the code $\mathcal{C}_1 \oplus_3 \mathcal{C}_2$ and a result which states the connection between the 3-sum of two codes and the sum of its associated ideals, but for lack of space we will not introduce them here.

Next proposition is related to the set of codewords of minimal support.

Proposition 3.2. *Let $\mathcal{C} = \mathcal{C}_1 \oplus_3 \mathcal{C}_2$ be a linear code defined over $\mathbb{F}_p$ with p prime. If $\mathbf{c} \in \mathcal{M}_\mathcal{C}$ then $\mathbf{c}$ belongs to one of the following sets:*

$$\mathcal{M}_{\mathcal{C}_{1\cdot\{n_1-2,n_1-1,n_1\}}} \,||_3\, \mathbf{0},\ \mathbf{0} \,||_3\, \mathcal{M}_{\mathcal{C}_{2\cdot\{1,2,3\}}},$$
$$\text{or } \left\{\mathbf{c}_1 \,||_3\, \mathbf{c}_2 \ :\ \mathbf{c}_1 \in \mathcal{M}_{\mathcal{C}_1\setminus\mathcal{C}_{1\cdot\{n_1-2,n_1-1,n_1\}}},\ \mathbf{c}_2 \in \mathcal{M}_{\mathcal{C}_2\setminus\mathcal{C}_{2\cdot\{1,2,3\}}}\right\}.$$

4. Concluding remarks

If we know the decomposition of a modular code $\mathcal{C}$ defined over $\mathbb{F}_p$ with p prime as a m-sum of smaller codes with $m \leq 3$, then we have found an effective method to reduce the complexity of computing the set of codewords of minimal support of $\mathcal{C}$, since the computation can be carried out in parallel for each component. In particular we have reduced the complexity for any binary linear code, since it follows from [6] that for any binary linear code $\mathcal{C}$ there exists linear codes $\mathcal{C}_1$ and $\mathcal{C}_2$, both obtained from $\mathcal{C}$ via a sequence of shortening and puncturing operations, and a permutation π of the coordinate set of $\mathcal{C}$ such that

$$\mathcal{C} = \pi(\mathcal{C}_1 \oplus \mathcal{C}_2), \quad \mathcal{C} = \pi(\mathcal{C}_1 \oplus_2 \mathcal{C}_2) \quad \text{or} \quad \mathcal{C} = \pi(\mathcal{C}_1 \oplus_3 \mathcal{C}_2).$$

Such decomposition of any given binary linear code can be obtained in polynomial time in the length of the code. Furthermore, by definition the codes $\mathcal{C}_1$ and $\mathcal{C}_2$ have smaller length than $\mathcal{C}$. Therefore the complexity of the search problem is reduced from $\mathcal{O}(n^2 2^{n-k})$ to $\mathcal{O}(m^2 2^{m-k'})$ where $m = \max\{n_1, n_2\} < n$ and $k' \leq k$.

References

[1] Alexander Barg. Complexity issues in coding theory. In *Handbook of coding theory, Vol. I, II*, pages 649–754. North-Holland, Amsterdam, 1998.

[2] M. Borges-Quintana, M. A. Borges-Trenard, P. Fitzpatrick, and E. Martínez-Moro. Gröbner bases and combinatorics for binary codes. *Appl. Algebra Engrg. Comm. Comput.*, 19(5):393–411, 2008.

[3] M. Borges-Quintana, M. A. Borges-Trenard, and E. Martínez-Moro. On a Gröbner bases structure associated to linear codes. *J. Discrete Math. Sci. Cryptogr.*, 10(2):151–191, 2007.

[4] M. Borges-Quintana, M.A. Borges-Trenard, I. Marquez-Corbella, and E. Martinez-Moro. An algebraic view to gradient descent decoding. In *Information Theory Workshop (ITW), 2010 IEEE*, pages 1–4, 30 2010-sept. 3 2010.

[5] Daisuke Ikegami and Yuichi Kaji. Maximum Likelihood Decoding for Linear Block Codes using Grobner Bases. *IEICE Trans. Fund. Electron. Commun. Comput. Sci.*, E86-A(3):643–651, 2003.

[6] Navin Kashyap. A decomposition theory for binary linear codes. *IEEE Trans. Inform. Theory*, 54(7):3035–3058, 2008.

[7] Navin Kashyap. On minimal tree realizations of linear codes. *IEEE Trans. Inform. Theory*, 55(8):3501–3519, 2009.

[8] Robert A. Liebler. Implementing gradient descent decoding. *Michigan Math. J.*, 58(1):285–291, 2009.

[9] Irene Márquez-Corbella and Edgar Martínez-Moro. Algebraic structure of the minimal support codewords set of some linear codes. *Adv. Math. Commun.*, 5(2):233–244, 2011.

University of Valladolid. IMUVa (Instituto de Matemáticas de la Universidad de Valladolid).
E-mail address: imarquez@agt.uva.es

University of Valladolid. IMUVa (Instituto de Matemáticas de la Universidad de Valladolid).
E-mail address: edgar@maf.uva.es