On the Convex Geometry of Binary Linear Codes

Navin Kashyap
Dept. Mathematics and Statistics
Queen's University
Kingston, ON, K7L 3N6, Canada.
Email: nkashyap@mast.queensu.ca

***Abstract*—A code polytope is defined to be the convex hull in $\mathbb{R}^n$ of the points in $\{0,1\}^n$ corresponding to the codewords of a binary linear code. This paper contains a collection of results concerning the structure of such code polytopes. A survey of known results on the dimension and the minimal polyhedral representation of a code polytope is first presented. We show how these results can be extended to obtain the complete facial structure of the polytope determined by the $[n, n-1]$ even-weight code. We then give a result classifying the types of 3-faces a general code polytope can have, which shows that the faces of such a polytope cannot be completely arbitrary. Finally, we show how geometrical arguments lead to a simple lower bound on the number of minimal codewords of a code, and characterize the codes for which this bound is attained with equality. This also yields an interesting intermediate result that classifies simple code polytopes. The motivation for our study of code polytopes comes from the formulation by Feldman, Wainwright and Karger of maximum-likelihood decoding as a linear programming problem over the code polytope.**

I. Introduction

The recent work of Feldman, Wainwright and Karger [6] shows that maximum likelihood (ML) decoding of a binary linear code $\mathcal{C}$ over a discrete memoryless channel can be formulated as a linear programming problem. Recall that the ML decoding problem is: given a received word $\mathbf{y}$ at the channel output, find a codeword $\mathbf{x} \in \mathcal{C}$ that maximizes the probability, $\Pr[\mathbf{y}|\mathbf{x}]$, of receiving $\mathbf{y}$ conditioned on the event that $\mathbf{x}$ was transmitted. As observed by Feldman *et al.*, under the assumption of a discrete memoryless channel, given a received word $\mathbf{y} = y_1 y_2 \dots y_n$, the problem of determining $\arg\max_{\mathbf{x}\in\mathcal{C}} \Pr[\mathbf{y}|\mathbf{x}]$ is equivalent to the problem of finding $\arg\min_{\mathbf{x}\in\mathcal{C}} \langle \gamma, \mathbf{x}\rangle$, where $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)$ is given by

$$\gamma_i = \log\left(\frac{\Pr[y_i|x_i = 0]}{\Pr[y_i|x_i = 1]}\right)$$

and $\langle\cdot\rangle$ is the standard inner product on $\mathbb{R}^n$. Here, for the inner product $\langle \gamma, \mathbf{x}\rangle$ to make sense, a binary codeword $\mathbf{x} = x_1 x_2 \dots x_n \in \mathcal{C}$ is identified with the real vector $(x_1, x_2, \dots, x_n) \in \{0,1\}^n \subset \mathbb{R}^n$.

The above formulation shows ML decoding to be equivalent to the minimization of a linear function over a finite set $\mathcal{C} \subset \{0,1\}^n$. Let $P(\mathcal{C})$ be the *code polytope* of $\mathcal{C}$, *i.e.*, the convex hull in $\mathbb{R}^n$ of the finite set $\mathcal{C}$. It can be shown that the set of vertices of $P(\mathcal{C})$ coincides with $\mathcal{C}$. The key point now is that over a polytope P, a linear function ϕ attains its minimum value $\phi_{\min} = \min\{\phi(\mathbf{x}) : \mathbf{x} \in P\}$ at a vertex of P. In particular, $\arg\min_{\mathbf{x}\in P(\mathcal{C})} \langle \gamma, \mathbf{x}\rangle = \arg\min_{\mathbf{x}\in\mathcal{C}} \langle \gamma, \mathbf{x}\rangle$. Thus, ML decoding is in fact equivalent to the minimization of a linear function over a polytope. This, as noted in [6], is a classic linear program (LP) since a polytope can always be represented as a polyhedron, which by definition, is the solution set to a finite system of linear inequalities.

An LP is simply the minimization of a linear cost function over a polyhedron. The degree of difficulty involved in solving an LP is directly related to the number of variables in the problem and the number of inequalities defining the polyhedron. Since it is possible for two distinct systems of linear inequalities to have the same solution set, a given polyhedron may in general have many different representations in terms of linear inequalities. Thus, with a view towards reducing the complexity of an LP over a given polyhedron, it is clearly important to find a *minimal* representation of the polyhedron that involves the least number of inequalities among all its representations via linear inequalities. In particular, an efficient implementation of the ML decoding problem for a code $\mathcal{C}$ as an LP would require a minimal polyhedral representation of the code polytope $P(\mathcal{C})$.

The work of Barahona and Grötschel [1] is the obvious starting point in any study of code polytopes. The results in that work are couched in the language of binary matroids, and translations of those results into the language of coding theory can be found scattered throughout our paper. Section III of our paper describes one of the main theorems of [1], which gives a minimal polyhedral representation of $P(\mathcal{C})$ for nearly any code $\mathcal{C}$, and also precisely characterizes the exceptional codes not covered by this representation.

However, ML decoding of an arbitrary code is known to be NP-hard. So, in general, no polyhedral representation of the code polytope can make linear programming over the polytope an attractive, or even feasible, option. A strategy often followed in such a situation is to "relax" the problem. The idea is to look for a polytope that contains the code as a subset of its vertex set, but has a more manageable polyhedral representation; the cost function is then minimized over this "relaxed" polytope.

A simple way of relaxing a code polytope $P(\mathcal{C})$ is to cast out some subset of the inequalities forming a polyhedral representation of $P(\mathcal{C})$, while restricting the resulting polytope

This work was supported in part by a research grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.
It was presented in part at the Inaugural UC-San Diego Workshop on Information Theory and its Applications, La Jolla, CA, USA, Feb. 6–10, 2006.

to remain within the n-cube $[0,1]^n$. This relaxation procedure ensures that the codewords in $\mathcal{C}$ continue to be vertices of the relaxed polytope. The "projected polytope" $\overline{Q}$ of Feldman *et al*. [6, p. 958] and the equivalent "fundamental polytope" of Vontobel and Koetter [11], [13], as well as the "rth relaxation" polytope defined in [13, Definition 11], are examples of such relaxations of the code polytope. Now, the polytopes so obtained generally have more vertices than the original code polytope. These additional vertices are also potential minimizers of the cost function $\langle \gamma, \mathbf{x} \rangle$ over the relaxed polytope, but as they are not codewords, they are spurious solutions to the decoding algorithm and represent decoding failure. This was noted in [6], where the term "pseudocodeword" is used[1] to denote a vertex (or more accurately, a vertex scaled so as to lie in the integer lattice $\mathbb{Z}^n$) of the polytope $\overline{Q}$.

Thus, relaxing the code polytope to reduce the complexity of the LP creates extra vertices that causes the LP over the relaxed polytope to perform worse as a decoding algorithm than ML decoding. There is thus a trade-off between complexity of the LP and its performance as a decoding algorithm, which is poorly understood at present. It is our thesis that in order to understand this trade-off, it is important to gain a thorough understanding of the structure of a code polytope. This should help in finding sharp estimates of the number of extra vertices in relaxations of the polytope, and may also help in finding relaxations that introduce relatively few extra vertices. One of the aims of this paper is to initiate a study of the complete facial structure of a code polytope, and we present some preliminary results in Section IV.

It turns out that the geometry of code polytopes implies a simple lower bound on the number of minimal codewords in a code. In Section V, we present this bound along with a necessary and sufficient condition for the bound to be met with equality. This derivation yields an interesting intermediate result that characterizates code polytopes that are simple, *i.e.*, polytopes in which each vertex is incident with exactly n edges, n being the dimension of the polytope.

II. Formal Definitions and Notation

We establish in this section the language used to present our results in later sections. We follow for the most part the standard terminology of coding theory [12] and discrete convex geometry [2]. As the average reader is expected to be familiar with the former, we elaborate on the latter.

Given a binary linear code $\mathcal{C}$, we let $d(\mathcal{C})$ denote its minimum distance. We define the *support* of a codeword $\mathbf{c} = c_1 c_2 \dots c_n \in \mathcal{C}$ to be $\mathsf{supp}(\mathbf{c}) = \{i : c_i = 1\}$, and also define $\mathsf{supp}(\mathcal{C}) = \bigcup_{\mathbf{c} \in \mathcal{C}} \mathsf{supp}(\mathbf{c})$ to be the support of $\mathcal{C}$. A codeword $\mathbf{c} \in \mathcal{C}$, $\mathbf{c} \neq \mathbf{0}$, is called *minimal* if its support $\mathsf{supp}(\mathbf{c})$ does not contain as a subset the support of any other nonzero codeword in $\mathcal{C}$. The set of all minimal codewords of $\mathcal{C}$ is denoted by $\mathcal{M}(\mathcal{C})$.

[1]The term "pseudocodeword" is used rather differently in [11], [13] where it denotes any rational point within the fundamental polytope. However, the authors of [14] use the term "minimal pseudocodewords" to describe the vertices adjacent to $\mathbf{0}$ in the fundamental polytope of [11], [13].

The dual code of $\mathcal{C}$ will be denoted by $\mathcal{C}^\perp$. A code $\mathcal{D}$ is said to be a *minor* of $\mathcal{C}$ if $\mathcal{D}$ can be obtained from $\mathcal{C}$ by a series of puncturing and shortening operations.

A *polytope* in $\mathbb{R}^n$ is the convex hull of a finite set of points in $\mathbb{R}^n$. In this paper, we will be interested in polytopes associated with binary linear codes. Let $\mathcal{C}$ be an $[n,k]$ binary linear code. Each codeword $c_1 c_2 \dots c_n \in \mathcal{C}$, $c_i \in \{0,1\}$, can be identified with the point $(c_1, c_2, \dots, c_n) \in \mathbb{R}^n$. Thus, $\mathcal{C}$ is identified with a set of 2^k points in $\{0,1\}^n \subset \mathbb{R}^n$, which we will continue to denote by $\mathcal{C}$. The *code polytope* of $\mathcal{C}$ is defined to be the convex hull of $\mathcal{C}$ in $\mathbb{R}^n$, and will be denoted by $P(\mathcal{C})$. A code polytope is thus a *0/1-polytope*, meaning that it is the convex hull of some subset of $\{0,1\}^n$ (cf. [16]).

The *dimension* of a polytope $P \subset \mathbb{R}^n$, denoted by $\dim(P)$, is defined to be the dimension of its affine hull (which is the smallest affine subspace in $\mathbb{R}^n$ containing P). Thus, $\dim(P) \leq n$, with strict inequality being possible. The following result explicitly determines $\dim(P(\mathcal{C}))$ for a binary linear code $\mathcal{C}$.

Theorem 2.1 ([1], Theorem 4.1): Given a binary linear code $\mathcal{C}$, let K be a largest subset of $\mathsf{supp}(\mathcal{C})$ with the property that for all $i, j \in K$, $i \neq j$, there exists $c_1 c_2 \dots c_n \in \mathcal{C}$ with $c_i \neq c_j$. Then, $\dim(P(\mathcal{C})) = |K|$.

A polytope $P \subset \mathbb{R}^n$ with $\dim(P) = n$ is said to be *full-dimensional*. The following corollary to the above theorem succinctly characterizes full-dimensional code polytopes.

Corollary 2.2: For any binary linear code $\mathcal{C}$ of length n, $\dim(P(\mathcal{C})) = n$ iff $d(\mathcal{C}^\perp) \notin \{1,2\}$.

For the most part, in this paper, we will consider full-dimensional code polytopes only, as codes $\mathcal{C}$ with $d(\mathcal{C}^\perp) \in \{1,2\}$ are not very interesting, and can always be punctured to obtain a code whose polytope is full-dimensional.

Let $\langle \cdot \rangle$ denote the standard inner product in $\mathbb{R}^n$. Given $\mathbf{a} \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$, the set $H = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x} \rangle = \beta\}$ is called a(n affine) *hyperplane*, and the set $H^{\leq} = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x} \rangle \leq \beta\}$ is called a *closed halfspace*. Given a polytope $P \subset \mathbb{R}^n$, a (possibly empty) subset $F \subset P$ is called a *face* of P if there exists a hyperplane $H \subset \mathbb{R}^n$ such that P is contained in the closed halfspace $H^{\leq}$, and $F = P \cap H$. A face of a polytope is itself a polytope. A 0-dimensional face of a polytope P is called a *vertex*, a 1-dimensional face is called an *edge*, and a face of dimension $\dim(P) - 1$ is called a *facet*. The only face of dimension $\dim(P)$ is P itself. In general, a k-dimensional face is referred to as a *k-face*. An inequality $\langle \mathbf{a}, \mathbf{x} \rangle \leq \beta$ is called *k-face-defining* (resp. *vertex-defining*, *edge-defining*, *facet-defining*) for P if $\langle \mathbf{a}, \mathbf{x} \rangle \leq \beta$ holds for all $\mathbf{x} \in P$, and $\{\mathbf{x} \in P : \langle \mathbf{a}, \mathbf{x} \rangle = \beta\}$ is a k-face (resp. vertex, edge, facet) of P.

The two basic n-dimensional $0/1$-polytopes that we will often encounter are the *n-cube*, $[0,1]^n$, which we denote by $\square_n$, and the *n-simplex* Δ_n, which is the convex hull of $\{\mathbf{0}, \mathbf{e}_1, \dots, \mathbf{e}_n\} \subset \mathbb{R}^n$, where $\mathbf{e}_1, \dots, \mathbf{e}_n$ is the standard basis of $\mathbb{R}^n$. Thus, for example, Δ_1 is a line segment, Δ_2 a triangle, and Δ_3 a tetrahedron.

We will need the notion of affine equivalence of polytopes: two polytopes $P \subset \mathbb{R}^m$ and $Q \subset \mathbb{R}^n$ are *affinely equivalent*

if there exists an affine map $f : \mathbb{R}^m \to \mathbb{R}^n$ that is a bijection between P and Q. Any n-dimensional polytope with $n+1$ vertices is affinely equivalent to the n-simplex Δ_n, and so is also usually referred to as a *simplex*.

A *polyhedron* is the intersection of finitely many closed halfspaces in $\mathbb{R}^n$. By the Weyl-Minkowski Theorem [2, p. 9], a set $P \subset \mathbb{R}^n$ is a polytope iff it is a bounded polyhedron. Thus, a polytope has a dual representation as a polyhedron. However, a polytope does not in general have a unique polyhedral representation, *i.e.*, it could be represented as an intersection of closed halfspaces in a variety of ways. But, there is a unique *minimal* polyhedral representation of a polytope P which requires the minimum number of halfspaces (or equivalently, linear inequalities) among all polyhedral representations of P. For a full-dimensional polytope P, the minimal polyhedral representation consists of precisely the facet-defining inequalities for P.

III. Minimal Polyhedral Representation of $P(\mathcal{C})$

It is a somewhat surprising fact that the minimal polyhedral representation can be explicitly given for most code polytopes, and we quickly summarize this result in this section. For simplicity, we state the result only in the case when $d(\mathcal{C}^\perp) \notin \{1, 2\}$, so that $P(\mathcal{C})$ is a full-dimensional code polytope. The most general statements can be found in [1], and are given there in the language of matroid theory.

The result given below is valid for a class of codes with a certain set of excluded minors. To be precise, it holds for binary linear codes $\mathcal{C}$ with the property that $\mathcal{C}$ does not have as a minor any code equivalent to one of the following:

1) the $[7, 3, 4]$ simplex code, which we denote by F_7^*;
2) the $[10, 5, 4]$ code, denoted by R_{10}, the columns of whose parity check matrix are the ten words of length 5 and Hamming weight 3;
3) the $[10, 4, 4]$ code, denoted by $M^*(K_5)$, the columns of whose generator matrix are the ten nonzero words of length 4 and Hamming weight at most 2.

We need one last piece of notation to state the theorem on polyhedral representations. Recall that $\mathcal{M}(\mathcal{C})$ denotes the set of minimal codewords of $\mathcal{C}$. The *boundary*, $\partial\mathcal{M}(\mathcal{C})$, of $\mathcal{M}(\mathcal{C})$ is defined to be the set of all $\mathbf{c} \in \mathcal{M}(\mathcal{C})$ such that $\mathbf{c} = \mathbf{p} \oplus \mathbf{q}$ for some $\mathbf{p}, \mathbf{q} \in \mathcal{M}(\mathcal{C})$ with $|\mathsf{supp}(\mathbf{p}) \cap \mathsf{supp}(\mathbf{q})| = 1$. Here and hereafter, $\oplus$ denotes modulo-2 addition. Let $\mathcal{M}(\mathcal{C})^\circ = \mathcal{M}(\mathcal{C}) \setminus \partial\mathcal{M}(\mathcal{C})$. We are now in a position to state the rather elaborate result on minimal polyhedral representations of code polytopes.

Theorem 3.1 ([1], Theorem 4.22): Let $\mathcal{C}$ be a binary linear code of length n, with $d(\mathcal{C}^\perp) \notin \{1, 2\}$, and let I denote the union of the supports of weight-3 codewords in $\mathcal{C}^\perp$. Then the system of linear inequalities in the variables $x_1, x_2, \ldots, x_n$ given by

(a) $0 \le x_j \le 1$ for each $j \notin I$, and
(b) for each $\mathbf{u} \in \mathcal{M}(\mathcal{C}^\perp)^\circ$ and each $K \subset \mathsf{supp}(\mathbf{u})$, $|K|$ odd,

$$\sum_{i \in K} x_i - \sum_{j \in \mathsf{supp}(\mathbf{u}) \setminus K} x_j \le |K| - 1,$$

is a system of facet-defining inequalities comprising the minimal polyhedral representation of $P(\mathcal{C})$ iff $\mathcal{C}$ does not have as a minor any code equivalent to F_7^*, R_{10} or $M^*(K_5)$.

Thus, for a code $\mathcal{C}$ of length n, with $d(\mathcal{C}^\perp) \notin \{1, 2\}$, and no F_7^*, R_{10} or $M^*(K_5)$ minor, the number of facets of $P(\mathcal{C})$ is

$$2(n - |I|) + \sum_{\mathbf{u} \in \mathcal{M}(\mathcal{C}^\perp)^\circ} 2^{w_H(\mathbf{u})-1}$$

where $w_H(\mathbf{u})$ is the Hamming weight of $\mathbf{u}$. Thus, implementing ML decoding as an LP is truly an attractive option only when $\mathcal{M}(\mathcal{C}^\perp)^\circ$ is of relatively small size, and consists of words of small Hamming weight relative to the codelength.

IV. Other Faces of $P(\mathcal{C})$

In the previous section, we saw that it is possible (with not insignificant effort) to characterize the facets of code polytopes, at least when the codes have no F_7^*, R_{10} or $M^*(K_5)$ minor. But what information do we have about the other faces of $P(\mathcal{C})$? It is of course obvious that the vertices of $P(\mathcal{C})$ are precisely the codewords of $\mathcal{C}$. The edges of $P(\mathcal{C})$ also have a simple characterization.

Theorem 4.1 ([1], Theorem 5.1): Let $\mathcal{C}$ be a binary linear code. There is an edge connecting two distinct vertices $\mathbf{c}_1$ and $\mathbf{c}_2$ in $P(\mathcal{C})$ iff $\mathbf{c}_1 \oplus \mathbf{c}_2 \in \mathcal{M}(\mathcal{C})$.

There is unfortunately (or fortunately for this author) very little known about faces of $P(\mathcal{C})$ other than vertices, edges and facets. One might naturally ask why it is necessary to investigate faces in general. A reason we give is that it could help in understanding what happens when the code polytope is relaxed by removing some facets. A beautiful result of Figiel, Lindenstrauss and Milman [2, p. 275] states that there is a constant $\delta > 0$ such that any centrally symmetric polytope P satisfies $\ln|V| \cdot \ln|F| \ge \delta \dim(P)$, where $|V|$ and $|F|$ denote the number of vertices and facets, respectively, of P. In particular, it applies to the code polytope $P(\mathcal{C})$ of a code $\mathcal{C}$ containing the all-ones word $\mathbf{1}$, and any of its relaxations obtained by throwing away facets while still maintaining central symmetry. Thus, for instance, the number of vertices and the number of facets in any of these polytopes cannot both be polynomial in the codelength. A more complete understanding of the facial structure of a code polytope and its relaxations can only help in improving such estimates of the number of vertices in terms of the number of facets. It may also be useful in the search for "good" relaxations that introduce relatively few extra vertices.

A code polytope is highly symmetric; it "looks the same" from any of its vertices. This symmetry is formally captured by the following statement, which essentially rephrases Theorem 3.1 from [1].

Lemma 4.2: For any binary linear code $\mathcal{C}$, $\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m\}$ is the set of vertices of a k-face of $P(\mathcal{C})$ iff $\{\mathbf{0}, \mathbf{c}_2 \oplus \mathbf{c}_1, \ldots, \mathbf{c}_m \oplus \mathbf{c}_1\}$ is the set of vertices of a k-face of $P(\mathcal{C})$.

Thus, it is enough to study the structure of faces containing $\mathbf{0}$, as an arbitrary face F of $P(\mathcal{C})$ can be transformed, via

a symmetry of the n-cube $\square_n$, to a face F' containing $\mathbf{0}$. Faces of $P(\mathcal{C})$ containing $\mathbf{0}$ actually have an important structural property, which is best described in the terminology of partially-ordered sets (posets) [5].

Any code $\mathcal{C}$ can be made a poset by endowing it with the partial ordering $\preceq$ as follows: $\mathbf{c}_1 \preceq \mathbf{c}_2$ iff $\mathsf{supp}(\mathbf{c}_i) \subset \mathsf{supp}(\mathbf{c}_2)$. A *downset* in $\mathcal{C}$ is a subset $\mathcal{D} \subset \mathcal{C}$ with the property that, for each $\mathbf{b} \in \mathcal{D}$ and $\mathbf{c} \in \mathcal{C}$, $\mathbf{c} \preceq \mathbf{b}$ implies that $\mathbf{c} \in \mathcal{D}$. Clearly, any non-empty downset in $\mathcal{C}$ contains $\mathbf{0}$.

Lemma 4.3: Let $\mathcal{C}$ be a binary linear code, and let F be a face of $P(\mathcal{C})$ containing $\mathbf{0}$. Then, the set of vertices of F is a downset of $\mathcal{C}$.

Proof: Let $\mathbf{b}$ be a vertex of F, and let $\mathbf{c} \in \mathcal{C}$ be such that $\mathbf{c} \preceq \mathbf{b}$. Note that since $\mathsf{supp}(\mathbf{c}) \subset \mathsf{supp}(\mathbf{b})$, we have $\mathbf{b} \oplus \mathbf{c} = \mathbf{b} - \mathbf{c}$, and thus, $\mathbf{b} - \mathbf{c} \in \mathcal{C}$.

Let $\langle \mathbf{a}, \mathbf{x} \rangle \leq \beta$ be a face-defining inequality for F. Note that β must be 0 since $\mathbf{0} \in F$. We thus have $\langle \mathbf{a}, \mathbf{c} \rangle \leq 0$ and $\langle \mathbf{a}, \mathbf{b} \rangle = 0$, and hence, $\langle \mathbf{a}, \mathbf{b} - \mathbf{c} \rangle \geq 0$. On the other hand, since $\mathbf{b} - \mathbf{c} \in \mathcal{C}$, we also have $\langle \mathbf{a}, \mathbf{b} - \mathbf{c} \rangle \leq 0$. Therefore, $\langle \mathbf{a}, \mathbf{b} - \mathbf{c} \rangle = 0$, and hence, $\langle \mathbf{a}, \mathbf{c} \rangle = 0$, implying that $\mathbf{c} \in F$. Hence, F is a downset of $\mathcal{C}$. ∎

While the converse to the above lemma is, in general, not true, a partial converse does hold. Given a subset $\mathcal{D}$ of $\mathcal{C}$, a codeword $\mathbf{b} \in \mathcal{D}$ is called a *maximum element* of $\mathcal{D}$ if $\mathbf{c} \preceq \mathbf{b}$ for all $\mathbf{c} \in \mathcal{D}$. The maximum element of $\mathcal{D}$, if it exists, is unique.

Lemma 4.4: Let $\mathcal{C}$ be a binary linear code, and let $\mathcal{D}$ be a subset of $\mathcal{C}$ with a maximum element. Then, $\mathsf{conv}(\mathcal{D})$ is a face of $P(\mathcal{C})$ containing $\mathbf{0}$ iff $\mathcal{D}$ is a downset.

Proof: We only have to prove the "if" part of the lemma. So, let $\mathcal{D}$ be a downset in $\mathcal{C}$, and let $\mathbf{b}$ be the maximum element of $\mathcal{D}$. Note that for any $\mathbf{c} \in \mathcal{C}$, we have $\mathbf{c} \preceq \mathbf{b}$ iff $\mathbf{c} \in \mathcal{D}$.

Let $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ be the vector with $a_i = -1$ for all $i \notin \mathsf{supp}(\mathbf{b})$ and $a_i = 0$ otherwise. Note that $\langle \mathbf{a}, \mathbf{c} \rangle \leq 0$ for all $\mathbf{c} \in \mathcal{C}$, with equality iff $\mathbf{c} \preceq \mathbf{b}$, or equivalently, iff $\mathbf{c} \in \mathcal{D}$. Hence, $\langle \mathbf{a}, \mathbf{x} \rangle \leq 0$ for all $\mathbf{x} \in P(\mathcal{C})$, with equality iff $\mathbf{x} \in \mathsf{conv}(\mathcal{D})$. Thus, $\mathsf{conv}(\mathcal{D})$ is a face of $P(\mathcal{C})$. Furthermore, $\mathbf{0} \in \mathsf{conv}(\mathcal{D})$ since $\mathbf{0} \in \mathcal{D}$. ∎

Lemmas 4.3 and 4.4 contain information sufficient to derive some interesting properties of k-faces of code polytopes. We consider first the problem of characterizing the 2-faces of an arbitrary code polytope. A 2-face of any $0/1$-polytope is a 2-dimensional $0/1$-polytope. Now, any 2-dimensional $0/1$-polytope is affinely equivalent to $\mathsf{conv}(V) \subset [0,1]^2$, where V is either $\{00, 01, 10\}$ or $\{0,1\}^2$. Hence, 2-faces of $0/1$-polytopes are either $0/1$-triangles or $0/1$-rectangles. In fact, if $Q \subset \mathbb{R}^n$ is a $0/1$-rectangle containing $\mathbf{0}$, the affine map taking Q bijectively to $\{0,1\}^2$ may be taken to be linear, and so the vertices of Q must be $\{\mathbf{0}, \mathbf{a}, \mathbf{b}, \mathbf{a}+\mathbf{b}\}$ for some $\mathbf{a}, \mathbf{b} \in \{0,1\}^n$. Furthermore, $\mathbf{a}, \mathbf{b}, \mathbf{a}+\mathbf{b}$ can all be in $\{0,1\}^n$ only if $\mathsf{supp}(\mathbf{a}) \cap \mathsf{supp}(\mathbf{b}) = \emptyset$. Thus, to summarize, if Q is a $0/1$-rectangle in $\mathbb{R}^n$ containing $\mathbf{0}$, then $Q = \mathsf{conv}(\{\mathbf{0}, \mathbf{a}, \mathbf{b}, \mathbf{a}+\mathbf{b}\})$ for some $\mathbf{a}, \mathbf{b} \in \{0,1\}^n$ such that $\mathsf{supp}(\mathbf{a}) \cap \mathsf{supp}(\mathbf{b}) = \emptyset$. The converse is also easily seen to be true. We can now prove the following simple characterization of the rectangular 2-faces in a code polytope.

Proposition 4.5: Let $\mathcal{C}$ be a binary linear code, and let $\mathcal{D} = \{\mathbf{0}, \mathbf{a}, \mathbf{b}, \mathbf{c}\} \subset \mathcal{C}$ be the vertex set of a $0/1$-rectangle. Then, $\mathsf{conv}(\mathcal{D})$ is a 2-face of $P(\mathcal{C})$ iff $\mathcal{D}$ is a downset in $\mathcal{C}$.

Proof: $\mathcal{D} = \{0, \mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is the vertex set of a $0/1$-rectangle only if $\mathsf{supp}(\mathbf{a}) \cap \mathsf{supp}(\mathbf{b}) = \emptyset$ and $\mathbf{c} = \mathbf{a} + \mathbf{b}$. Therefore, we have $\mathbf{a} \preceq \mathbf{c}$ and $\mathbf{b} \preceq \mathbf{c}$, and hence, $\mathbf{c}$ is the maximum element of $\mathcal{D}$. The result now follows from Lemma 4.4. ∎

The following corollary is essentially a re-statement of the above proposition in more coding-theoretic terms.

Corollary 4.6: Let $\mathcal{C}$ be a binary linear code. F is a rectangular 2-face of $P(\mathcal{C})$ containing $\mathbf{0}$ iff $F = \mathsf{conv}(\{\mathbf{0}, \mathbf{a}, \mathbf{b}, \mathbf{c}\})$, where $\mathbf{a}$ and $\mathbf{b}$ are minimal codewords with disjoint support, $\mathbf{c} = \mathbf{a} + \mathbf{b}$, and $\{\mathbf{x} \in \mathcal{C} : \mathbf{x} \preceq \mathbf{c}\} = \{0, \mathbf{a}, \mathbf{b}, \mathbf{c}\}$.

Triangular 2-faces of code polytopes do not appear to have a similar easy characterization. Clearly, by Theorem 4.1, $\mathsf{conv}(\{\mathbf{0}, \mathbf{a}, \mathbf{b}\})$ is a 2-face of some $P(\mathcal{C})$ only if $\mathbf{a}, \mathbf{b}, \mathbf{a} \oplus \mathbf{b} \in \mathcal{M}(\mathcal{C})$. We believe that the converse is not true, although we do not have a counterexample either.

Despite this high degree of symmetry, determining the complete facial structure of an arbitrary code polytope is a very difficult problem, for which there is no known general solution. However, using Theorem 3.1 as a starting point, it is fairly easy to derive the facial structure in the important special case of the polytope of the $[n, n-1]$ even-weight code $\mathcal{E}_n$, as we now show.

The cases $n = 1, 2, 3$ are trivial: $\mathcal{E}_1 = P(\mathcal{E}_1) = \{0\}$; $\mathcal{E}_2 = \{00, 11\}$, hence $P(\mathcal{E}_2) = \Delta_1$; and $\mathcal{E}_3 = \{000, 011, 101, 110\}$, hence $P(\mathcal{E}_3) = \Delta_3$. The code polytopes for $\mathcal{E}_n$, $n \geq 4$, are much more interesting, and Theorem 4.7 below, in conjunction with Lemma 4.2, gives their complete facial structure.

We introduce some simplifying pieces of notation. Given a positive integer n, we let $[n]$ denote the set $\{1, 2, \ldots, n\}$. If S is some set in $\mathbb{R}^n$, we let $\mathsf{conv}(S)$ denote its convex hull. For $j = 1, 2, \ldots, n$, we define the hyperplanes $H_j = \{\mathbf{x} \in \mathbb{R}^n : x_j = 0\}$, and $L_j = \{\mathbf{x} \in \mathbb{R}^n : x_j - \sum_{s \in [n], s \neq j} x_s = 0\}$.

Theorem 4.7: For $n \geq 4$, let $\mathcal{E}_n$ denote the $[n, n-1]$ binary even-weight code, so that $\mathcal{M}(\mathcal{E}_n)$ is the set of binary words of length n and Hamming weight 2.

(a) $\mathsf{conv}(\mathbf{0}, \mathbf{c})$ is an edge of $P(\mathcal{E}_n)$ iff $\mathbf{c} \in \mathcal{M}(\mathcal{E}_n)$.
(b) F is a 2-face of $P(\mathcal{E}_n)$ containing $\mathbf{0}$ iff $F = \mathsf{conv}(\mathbf{0}, \mathbf{a}, \mathbf{b})$ for some $\mathbf{a}, \mathbf{b} \in \mathcal{M}(\mathcal{E}_n)$ such that $\mathbf{a} \oplus \mathbf{b}$ is also in $\mathcal{M}(\mathcal{E}_n)$.
(c) For $3 \leq k \leq n-1$, F is a k-face of $P(\mathcal{E}_n)$ containing $\mathbf{0}$ iff $F = \mathsf{conv}(\mathcal{E}_n \cap V)$, where V is either
  (i) $\bigcap_{j \in J} H_j$ for some $J \subset [n]$ with $|J| = n - k$, or
  (ii) $\bigcap_{j \in J} H_j \cap L_i$ for some $J \subset [n]$ with $|J| = n - k - 1$ and $i \in [n] \setminus J$.

Sketch of Proof: Part (a) is a consequence of Theorem 4.1. For part (b), we first observe that 2-dimensional 0/1-polytopes must either be triangles or rectangles [16, Section 1.1]. We then show that no rectangle can be a 2-face of $P(\mathcal{E}_n)$, and a triangle with vertices $\mathbf{0}, \mathbf{a}, \mathbf{b}$ can be a 2-face of $P(\mathcal{E}_n)$ iff $\mathbf{a}, \mathbf{b}, \mathbf{a} \oplus \mathbf{b} \in \mathcal{M}(\mathcal{E}_n)$.

For part (c), we note that the result for $k = n-1$ follows from Theorem 3.1.[2] For the general result, we observe first that a k-face incident with $\mathbf{0}$ must be contained in at least $n-k$ facets incident with $\mathbf{0}$. We then make use of the result for $k = n-1$ to decide which intersections of $n-k$ (or more) facets yields a k-face. ∎

For $n \geq 4$ and $3 \leq k \leq n-1$, define a k-face of $P(\mathcal{E}_n)$ to be of *Type A* (resp. *Type B*) if it is of the form $\mathsf{conv}(\mathcal{E}_n \cap V)$ for some V as in Theorem 4.7(c)(i) (resp. Theorem 4.7(c)(ii)). It is easy to see that k-faces of Type A are affinely equivalent (via projection onto the coordinates x_i, $i \notin J$) to $P(\mathcal{E}_k)$. Note that since $P(\mathcal{E}_3) = \Delta_3$, 3-faces of Type A are in fact affinely equivalent to Δ_k. On the other hand, k-faces of Type B are always affinely equivalent to the k-simplex Δ_k, due to the following corollary to the theorem.

Corollary 4.8: F is a k-face of Type B iff $F = \mathsf{conv}(\mathbf{0}, \mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_k)$ for $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_k \in \mathcal{M}(\mathcal{E}_n)$ such that $\left|\bigcap_{i=1}^{k} \mathsf{supp}(\mathbf{c}_i)\right| = 1$.

Corollary 4.9: For $k \leq 3$, all k-faces of $P(\mathcal{E}_n)$ are affinely equivalent to Δ_k. For $k > 3$, a k-face of $P(\mathcal{E}_n)$ is affinely equivalent to either $P(\mathcal{E}_k)$ or Δ_k.

The results presented so far in this section are enough to patch together a proof of the following useful result.

Theorem 4.10: Fix an $n \geq 1$, and let f_k, $0 \leq k \leq n$, denote the number of k-faces of $P(\mathcal{E}_n)$. We have $f_0 = 2^{n-1}$, $f_1 = \binom{n}{2} 2^{n-2}$, $f_2 = \binom{n}{3} 2^{n-1}$, and

$$f_k = \binom{n}{k} 2^{n-k} + \binom{n}{k+1} 2^{n-1} \quad \text{for } 3 \leq k \leq n.$$

The vector $(f_0, f_1, \ldots, f_n)$ is called the *f-vector* of the polytope $P(\mathcal{E}_n)$. It should be noted that f-vectors of polytopes are in general notoriously difficult to compute.

It is interesting to note that the polytope $P(\mathcal{E}_n)$ only has triangular 2-faces. Code polytopes in general can have quadrilateral 2-faces as evidenced by the 2-faces of the n-cube, which is the code polytope of $\{0,1\}^n$. Examples of code polytopes with both kinds of 2-faces abound, one such being the $[4,3]$ code obtained as the direct sum of $\mathcal{E}_3$ and $\{0,1\}$.

This raises the question of whether *any* $0/1$-polytope can arise as a face of some code polytope. This would make code polytopes a hopelessly complex class of polytopes. Indeed, there are classes of polytopes with this property, a celebrated example being the class of travelling salesman polytopes [3]. Luckily, code polytopes are not as complex. Up to affine equivalence, there are eight types of 3-dimensional 0/1-polytopes [16, Section 1.1], of which only four are allowed to be 3-faces of code polytopes. We omit the proof.

Theorem 4.11: If F is a 3-face of some code polytope, then up to affine equivalence, F is one of the following: (i) the 3-cube $\square_3$, (ii) the prism $\Delta_2 \times \Delta_1$, (iii) the square pyramid $\mathsf{conv}(\{000, 001, 010, 011, 111\})$, and (iv) the 3-simplex Δ_3.

[2]Actually, this representation of the facets of $\mathcal{E}_n$ can be found in the earlier work of Jeroslow [7].

Remark: As examples, the prism $\Delta_2 \times \Delta_1$ occurs as a 3-face of the [4,3] code that is the direct sum of $\mathcal{E}_3$ and $\{0,1\}$, while the square pyramid occurs as a 3-face of the $[5,3]$ code with generator matrix

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

V. Minimal Codewords and Simple Code Polytopes

Curiously, Theorem 4.1 leads to a simple lower bound on the number of minimal codewords in a binary linear code $\mathcal{C}$, which appears to be new.

Lemma 5.1: For any binary linear code $\mathcal{C}$, $|\mathcal{M}(\mathcal{C})| \geq \dim(P(\mathcal{C}))$. In particular, if $\mathcal{C}$ is a code of length n with $d(\mathcal{C}^\perp) \notin \{1,2\}$, then $|\mathcal{M}(\mathcal{C})| \geq n$.

Proof: The proof is simply the geometric fact that in any polytope P of dimension d, each vertex has at least d edges incident with it. By Theorem 4.1, the number of edges incident with any vertex of $P(\mathcal{C})$ is exactly $|\mathcal{M}(\mathcal{C})|$. ∎

It is natural to ask the question of when this bound is met with equality, and the following result provides the answer.

Theorem 5.2: For any binary linear code $\mathcal{C}$ of length n with $d(\mathcal{C}^\perp) \notin \{1,2\}$, we have $|\mathcal{M}(\mathcal{C})| \geq n$, with equality iff $\mathcal{C}$ is the direct sum of simplex codes.

We give a proof of Theorem 5.2 that is essentially geometric, although it would be interesting to find a purely coding-theoretic proof. In what follows, we let $\mathcal{C}$ be a code of length n with $d(\mathcal{C}^\perp) \notin \{1,2\}$, so that $\dim(P(\mathcal{C})) = n$. By Theorem 4.1, $|\mathcal{M}(\mathcal{C})| = n$ iff each vertex of $P(\mathcal{C})$ belongs to exactly n edges.

A polytope of dimension n is called *simple* if each vertex of the polytope is incident with exactly n edges.[3] Thus, $|M(\mathcal{C})| = n$ iff $P(\mathcal{C})$ is simple. So, Theorem 5.2 is proved once we show the following lemma to be true.

Lemma 5.3: $P(\mathcal{C})$ is simple iff $\mathcal{C}$ is the direct sum of simplex codes.

A result of Kaibel and Wolff [9, Theorem 1] states that a $0/1$-polytope is simple iff it is the (Cartesian) product of $0/1$-simplices (*i.e.*, simplices whose vertices are in $\{0,1\}^l$ for some $l > 0$). Therefore, the proof of Lemma 5.3, and hence that of Theorem 5.2, relies strongly on the next result.

Lemma 5.4: $P(\mathcal{C})$ is a simplex iff $\mathcal{C}$ is a simplex code.

Proof: The result is trivial when $\mathcal{C}^\perp = \{\mathbf{0}\}$, as the code $\mathcal{C} = \{0,1\}^n$ is the direct sum of n copies of $\{0,1\}$, which is degenerately a simplex code. So, assume that $d(\mathcal{C}^\perp) \geq 3$.

If $\mathcal{C}$ is a $[2^k - 1, k]$ simplex code, then $P(\mathcal{C})$ is a $(2^k - 1)$-dimensional polytope with 2^k vertices, which is a simplex.

Conversely, if $P(\mathcal{C})$ is a simplex of dimension n, it has $n+1$ vertices. Hence, $n + 1 = |\mathcal{C}| = 2^k$ for some k, and

[3]A simple polytope is usually defined as an n-dimensional polytope in which each vertex is in exactly n facets. The two definitions are equivalent [15, p. 66].

so, $n = 2^k - 1$. Let G be a $k \times (2^k - 1)$ generator matrix for $\mathcal{C}$. Since $d(C^\perp) \geq 3$, the columns of G are nonzero and distinct. Therefore, the $2^k - 1$ columns of G are the $2^k - 1$ distinct nonzero binary words of length k, which means that G generates the $[2^k - 1, k]$ simplex code. ∎

We can now provide a proof of Lemma 5.3, which completes the proof of Theorem 5.2.

Proof of Lemma 5.3: Let $\mathcal{C}$ be the direct sum of some simplex codes $\mathcal{C}_i$, $i = 1, 2, \ldots, r$. Trivially, $P(\mathcal{C}) = P(\mathcal{C}_1) \times \cdots \times P(\mathcal{C}_r)$, and each $P(\mathcal{C}_i)$ is a simplex by Lemma 5.4. Hence, by the Kaibel-Wolff result [9], $P(\mathcal{C})$ is simple.

Conversely, if $P(\mathcal{C})$ is simple, then again by the Kaibel-Wolff result, $P(\mathcal{C}) = \Delta^{(1)} \times \cdots \times \Delta^{(r)}$ for some simplices $\Delta^{(1)}, \ldots, \Delta^{(r)}$. Letting $\mathcal{C}_i$ denote the vertex set of $\Delta^{(i)}$, $i = 1, \ldots, r$, we observe that $\mathcal{C} = \mathcal{C}_1 \times \cdots \times \mathcal{C}_r$. Since $\mathcal{C}$ is a linear code, it readily follows that each $\mathcal{C}_i$ is a linear code. Thus, $\mathcal{C}$ is actually the direct sum of the $\mathcal{C}_i$'s. It remains to show that each $\mathcal{C}_i$ is a simplex code.

Now, $\mathcal{C}^\perp$ is the direct sum of the codes $\mathcal{C}_i^\perp$, and so, the condition $d(\mathcal{C}^\perp) \notin \{1, 2\}$ implies that $d(\mathcal{C}_i^\perp) \notin \{1, 2\}$ for each i. By definition, $P(\mathcal{C}_i) = \mathsf{conv}(\mathcal{C}_i)$. But, $\mathsf{conv}(\mathcal{C}_i) = \Delta^{(i)}$, since $\Delta^{(i)}$ is the convex hull of its vertex set. So, Lemma 5.4 shows that each $\mathcal{C}_i$ is a simplex code. ∎

From a geometer's point of view, Lemma 5.3 is interesting by itself, as it provides a complete characterization of full-dimensional code polytopes that are simple. Another geometrically important class of polytopes is the class of *simplicial* polytopes, which are polytopes in which every facet is a simplex. The question of which code polytopes are simplicial remains open. A non-trivial example of such a polytope is the polytope of the $[6, 3]$ code generated by the matrix

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

As a final remark, we would like to mention that simple polytopes play a crucial role in the performance analysis of simplex algorithms in linear programming [10]. A simplex algorithm is a method to solve an LP over a polyhedron P by repeatedly moving from one vertex of P to an adjacent vertex so that in each step the value of the objective function is decreased. Such an algorithm is not just easily described, but is also an efficient general tool for solving LP's, and is the most promising candidate for a strongly polynomial-time linear programming algorithm [8]. Our results on simple code polytopes could be useful in analyzing the performance of simplex algorithms for decoding by linear programming.

REFERENCES

[1] F. Barahona and M. Grötschel, "On the cycle polytope of a binary matroid," *J. Comb. Theory Ser. B*, vol. 40, pp. 40–62, 1986.

[2] A. Barvinok, *A Course in Convexity*, Graduate Studies in Mathematics vol. 54, Providence, RI: AMS Publications, 2002.

[3] L. Billera and A. Sarangarajan, "All 0-1 polytopes are traveling salesman polytopes," *Combinatorica*, vol. 16, no. 2, pp. 175–188, 1996.

[4] A. Bonisoli, "Every equidistant linear code is a sequence of dual Hamming codes," *Ars Combinatoria*, vol. 18, pp. 181–186, 1984.

[5] B.A. Davey and H.A. Priestley, *Introduction to Lattices and Order*, Cambridge, UK: Cambridge Univ. Press, 1990.

[6] J. Feldman, M.J. Wainwright and D.R. Karger, "Using linear programming to decode binary linear codes," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 954–972, 2005.

[7] R.G. Jeroslow, "On defining sets of vertices of the hypercube by linear inequalities," *Discr. Math.*, vol. 11, pp. 119–124, 1975.

[8] V. Kaibel, R. Mechtel, M. Sharir and G. Ziegler, "The simplex algorithm in dimension three," *SIAM J. Comput.*, vol. 34, no. 2, pp. 475–497, 2005.

[9] V. Kaibel and M. Wolff, "Simple 0/1-polytopes," *Europ. J. Comb.*, vol. 21, no. 326, pp. 139–144, 2000.

[10] G. Kalai, "Linear programming, the simplex algorithm and simple polytopes," *Math. Programming Ser. B*, vol. 79, pp. 217–233, 1997.

[11] R. Koetter, W.-C.W. Li, P.O. Vontobel and J.L. Walker "Characterizations of pseudo-codewords of LDPC codes," ArXiv e-print cs.IT/0508049.

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

[13] P.O. Vontobel and R. Koetter, "Graph-cover decoding and finite-length analysis of message-passing iterative decoding of LDPC codes," preprint submitted to IEEE Trans. Inform. Theory, Dec. 2005. ArXiv e-print cs.IT/0512078.

[14] P.O. Vontobel, R. Smarandache, N. Kiyavash, J. Teutsch and D. Vukobratovic, "On the minimal pseudo-codewords of codes from finite geometries," *Proc. 2005 IEEE Int. Symp. Inform. Theory, Adelaide, Australia*, pp. 980–984, September 4–9, 2005.

[15] G. Ziegler, *Lectures on Polytopes*, New York: Springer, 1994.

[16] G. Ziegler, "Lectures on 0/1-Polytopes," in *Polytopes — Combinatorics and Computation*, (G. Kalai and G.M. Ziegler, eds.), DMV Seminars Series, Basel: Birkhäuser, 2000. ArXiv e-print math.CO/9909177.