Linear-programming Decoding of Non-binary Linear Codes

Mark F. Flanagan[1], Vitaly Skachek[2], Eimear Byrne[3], Marcus Greferath[4]

[1] Institute for Digital Communications, The University of Edinburgh, Edinburgh EH9 3JL, Scotland
mark.flanagan@ieee.org

[2,3,4] Claude Shannon Institute, University College Dublin, Belfield, Dublin 4, Ireland
[2] vitaly.skachek@ucd.ie
[3] ebyrne@ucd.ie
[4] marcus.greferath@ucd.ie

Abstract

We develop a framework for linear-programming (LP) decoding of non-binary linear codes over rings. We prove that the resulting LP decoder has the 'maximum likelihood certificate' property, and we show that the decoder output is the lowest cost pseudocodeword. Equivalence between pseudocodewords of the linear program and pseudocodewords of graph covers is proved. LP decoding performance is illustrated for the $(11, 6, 5)$ ternary Golay code with ternary PSK modulation over AWGN, and in this case it is shown that the LP decoder performance is comparable to codeword-error-rate-optimum hard-decision based decoding.

1 Introduction

For high-data-rate communication systems, bandwidth-efficient signalling schemes are required which necessitate the use of higher-order modulation. This may be achieved in conjunction with coding by the use of non-binary codes whose symbols map directly to modulation signals. A study of such codes over rings, particularly over the integers modulo 8, for use with PSK modulation was performed in [7].

Of course, within such a framework it is desirable to use state-of-the-art error-correcting codes. *Low-density parity-check* (LDPC) codes have become very popular in recent years due to their practical effectiveness under message-passing decoding. However, the analysis of LDPC codes is a difficult task. One approach was proposed in [8], and it is based on the consideration of so-called *pseudocodewords* and their *pseudoweights*. The approach was further explored in [3], [6]. In [1] and [2], the decoding of *binary* LDPC codes using linear-programming decoding was proposed, and the connections between linear-programming decoding and classical belief propagation decoding were established. Recently, pseudocodewords of non-binary codes were defined and some bounds on the pseudoweights were derived in [4].

In this work, we extend the approach in [2] towards coded modulation, in particular to codes over rings mapped to non-binary modulation signals. As was done in [2], we show that the problem of decoding may be formulated as a linear-programming (LP) problem for the non-binary case. We also show that an appropriate relaxation of the LP leads to a solution which has the 'maximum likelihood (ML) certificate' property, i.e. if the LP outputs a codeword, then it must be the ML codeword. Moreover, we show that if the LP output is integral, then it must correspond to the ML codeword. We define the *graph-cover pseudocodewords* of the code, and the *LP pseudocodewords* of the code, and prove the equivalence of these two concepts. This shows that the links between LP decoding on the relaxed polytope and message-passing decoding on the Tanner graph generalize to the non-binary case.

To demonstrate performance, LP decoding of the ternary Golay code is simulated, and the LP decoder is seen to perform approximately as well as codeword-error-rate optimum hard-decision decoding, and approximately 1.5 dB from the union bound for codeword-error-rate optimum soft-decision decoding.

2 General Settings

We consider codes over finite rings (this includes codes over finite fields, but may be more general). Denote by $\mathfrak{R}$ a ring with q elements, by 0 its additive identity, and let $\mathfrak{R}^- = \mathfrak{R}\backslash\{0\}$. Let $\mathcal{C}$ be a linear $[n, k]$ code with parity-check matrix $\mathcal{H}$ over $\mathfrak{R}$. The parity check matrix $\mathcal{H}$ has $m \geq n - k$ rows.

Denote the set of column indices and the set of row indices of $\mathcal{H}$ by $\mathcal{I} = \{1, 2, \cdots, n\}$ and $\mathcal{J} = \{1, 2, \cdots, m\}$, respectively. We use notation $\mathcal{H}_j$ for the j-th row of $\mathcal{H}$. Let the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the Tanner graph of $\mathcal{C}$ associated with the matrix $\mathcal{H}$, namely $\mathcal{V} = \{u_1, u_2, \cdots, u_n\} \cup \{v_1, v_2, \cdots, v_m\}$, and there is an edge between u_i and v_j if and only if $\mathcal{H}_{j,i} \neq 0$. We denote by $\mathcal{N}(v_j)$ the set of neighbors of the vertex v_j, and by $\text{supp}(\boldsymbol{c})$ the support of a vector $\boldsymbol{c}$. Let $d = \max_{j \in \mathcal{J}}\{|\text{supp}(\mathcal{H}_j)|\}$.

For a word $\boldsymbol{c} = (c_1, c_2, \cdots, c_n) \in \mathfrak{R}^n$, we associate the value c_i with variable vertex u_i for each $i \in \mathcal{I}$. Parity-check $j \in \mathcal{J}$ is said to be *satisfied* if and only if $\sum_{i \in \mathcal{I}} \mathcal{H}_{j,i} \cdot c_i = 0$. We say that the vector $\boldsymbol{c}$ is a codeword of the single parity-check code $\mathcal{C}_j$ if and only if parity check $j \in \mathcal{J}$ is satisfied. Also, we say that the vector $\boldsymbol{c}$ is a codeword of $\mathcal{C}$ if and only if all parity checks $j \in \mathcal{J}$ are satisfied.

Definition 2.1: ([5]) A graph $\tilde{\mathcal{G}} = (\tilde{\mathcal{V}}, \tilde{\mathcal{E}})$ is a *finite cover* of the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if there exists a mapping $\Pi : \tilde{\mathcal{V}} \rightarrow \mathcal{V}$ which is a graph homomorphism (Π takes adjacent vertices of $\tilde{\mathcal{G}}$ to adjacent vertices of $\mathcal{G}$), such that for every vertex $v \in \mathcal{G}$ and every $\tilde{v} \in \Pi^{-1}(v)$, the neighborhood $\mathcal{N}(\tilde{v})$ of $\tilde{v}$ is mapped bijectively to $\mathcal{N}(v)$.

Definition 2.2: ([5]) A cover of the graph $\mathcal{G}$ is called an M-cover, where M is a positive integer, if $|\Pi^{-1}(v)| = M$ for every vertex $v \in \mathcal{V}$.

Fix some positive integer M. Let $\tilde{\mathcal{G}} = (\tilde{\mathcal{V}}, \tilde{\mathcal{E}})$ be an M-cover of the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ representing the code $\mathcal{C}$ with parity-check matrix $\mathcal{H}$. Denote the vertices in the sets $\Pi^{-1}(u_i)$ and $\Pi^{-1}(v_j)$ by $\{u_{i,1}, u_{i,2}, \cdots, u_{i,M}\}$ and $\{v_{j,1}, v_{j,2}, \cdots, v_{j,M}\}$, respectively, where $i \in \mathcal{I}$ and $j \in \mathcal{J}$.

Consider the linear code $\tilde{\mathcal{C}}$ of length Mn over $\mathfrak{R}$, defined by the $Mm \times Mn$ parity-check matrix $\tilde{\mathcal{H}}$. For $1 \leq i^*, j^* \leq M$ and $i \in \mathcal{I}$, $j \in \mathcal{J}$, we let $i' = (i-1)M + i^*, j' = (j-1)M + j^*$, and

$$\tilde{\mathcal{H}}_{j',i'} = \begin{cases} \mathcal{H}_{j,i} & \text{if } u_{i,i^*} \in \mathcal{N}(v_{j,j^*}) \\ 0 & \text{otherwise} \end{cases} .$$

Then, any vector $\boldsymbol{p} \in \tilde{\mathcal{C}}$ has the form

$$\begin{aligned} \boldsymbol{p} \;\; = \;\; & (p_{1,1}, p_{1,2}, \cdots, p_{1,M}, p_{2,1}, p_{2,2}, \\ & \cdots, p_{2,M}, \cdots, p_{n,1}, p_{n,2}, \cdots, p_{n,M}) . \end{aligned}$$

We associate the value $p_{i,\ell} \in \mathfrak{R}$ with the vertex $u_{i,\ell}$ in $\tilde{\mathcal{G}}$ ($i \in \mathcal{I}$, $\ell = 1, 2, \cdots, M$).

The word $\boldsymbol{p} \in \tilde{\mathcal{C}}$ as above is called a *graph-cover pseudocodeword* of the code $\mathcal{C}$. Sometimes, we consider the following $n \times q$ matrix representation, denoted $\mathcal{P}$, of the pseudocodeword $\boldsymbol{p}$:

$$\Big(m_i(\alpha)\Big)_{i \in \mathcal{I};\, \alpha \in \mathfrak{R}} ,$$

where

$$m_i(\alpha) = |\{\ell \in \{1, 2, \cdots, M\} \;:\; p_{i,\ell} = \alpha\}| \geq 0 ,$$

for $i \in \mathcal{I}$, $\alpha \in \mathfrak{R}$.

3 Decoding as a Linear-Programming Problem

Assume throughout that the codeword $\bar{\boldsymbol{c}} = (\bar{c}_1, \bar{c}_2, \cdots, \bar{c}_n) \in \mathcal{C}$ has been transmitted over a q-ary input memoryless channel, and a corrupted word $\boldsymbol{y} = (y_1, y_2, \cdots, y_n) \in \Sigma^n$ has been received. Here Σ denotes the set of channel output symbols; we assume that this set either has finite cardinality, or is equal to $\mathbb{R}^l$ or $\mathbb{C}^l$ for some integer $l \geq 1$. In practice, this channel may represent the combination of modulator and physical channel. We assume hereafter that all information words are equally probable, and so all codewords are transmitted with equal probability.

It was suggested in [1] to represent each symbol as a binary vector of length $|\mathfrak{R}^-|$, where the entries in the vector are indicators of a symbol taking on a particular value. Below, we elaborate on this approach. It should be mentioned that by using such a representation, the non-binary code is converted into a binary code. However, this binary code is not linear, and therefore the analysis in [1], [2] is not directly applicable.

For use in the following derivation, we shall define the mapping

$$\xi \;:\; \mathfrak{R} \longrightarrow \{0,1\}^{q-1} \subset \mathbb{R}^{q-1} ,$$

defined by

$$\xi(b) = \boldsymbol{x} = (x^{(\alpha)})_{\alpha \in \mathfrak{R}^-} ,$$

such that, for all $\alpha \in \mathfrak{R}^-$,

$$x^{(\alpha)} = \begin{cases} 1 & \text{if } b = \alpha \\ 0 & \text{otherwise} \end{cases} .$$

We note that the mapping $\xi(\cdot)$ is one-to-one, and its image is the set of binary vectors of length $q-1$ with Hamming weight 0 or 1.

We also define a function $\boldsymbol{\lambda} : \Sigma \longrightarrow \mathbb{R} \cup \{\pm\infty\}$ by

$$\boldsymbol{\lambda} = (\lambda^{(\alpha)})_{\alpha \in \mathfrak{R}^-} ,$$

where, for each $y \in \Sigma$, $\alpha \in \mathfrak{R}^-$,

$$\lambda^{(\alpha)}(y) = \log\left(\frac{p(y|0)}{p(y|\alpha)}\right) ,$$

and $p(y|c)$ denotes the channel output probability (density) conditioned on the channel input. Extend $\boldsymbol{\lambda}$ to a map on Σ^n by $\boldsymbol{\lambda}(\boldsymbol{y}) = (\boldsymbol{\lambda}(y_1) \mid \boldsymbol{\lambda}(y_2) \mid \ldots \mid \boldsymbol{\lambda}(y_n))$.

The codeword-error-rate-optimum receiver operates according to the *maximum a posteriori* (MAP) decision rule:

$$\begin{aligned} \hat{\boldsymbol{c}} \;\; &= \;\; \arg\max_{\boldsymbol{c} \in \mathcal{C}} p(\, \boldsymbol{c} \mid \boldsymbol{y} \,) \\ &= \;\; \arg\max_{\boldsymbol{c} \in \mathcal{C}} \frac{p(\, \boldsymbol{y} \mid \boldsymbol{c} \,) p(\, \boldsymbol{c} \,)}{p(\, \boldsymbol{y} \,)} . \end{aligned}$$

Here $p\,(\cdot)$ denotes probability if Σ has finite cardinality, and probability density if Σ has infinite cardinality.

By assumption, the *a priori* probability $p(\boldsymbol{c})$ is uniform over codewords, and $p(\boldsymbol{y})$ is independent of $\boldsymbol{c}$. Therefore, the decision rule reduces to maximum likelihood (ML) decoding:

$$\begin{aligned} \hat{\boldsymbol{c}} \;\; &= \;\; \arg\max_{\boldsymbol{c} \in \mathcal{C}} p(\, \boldsymbol{y} \mid \boldsymbol{c} \,) \\ &= \;\; \arg\max_{\boldsymbol{c} \in \mathcal{C}} \prod_{i=1}^{n} p(y_i | c_i) \\ &= \;\; \arg\max_{\boldsymbol{c} \in \mathcal{C}} \sum_{i=1}^{n} \log(p(y_i | c_i)) \end{aligned}$$

$$= \arg\min_{\boldsymbol{c}\in\mathcal{C}} \sum_{i=1}^{n} \log\left(\frac{p(y_i|0)}{p(y_i|c_i)}\right)$$
$$= \arg\min_{\boldsymbol{c}\in\mathcal{C}} \sum_{i=1}^{n} \boldsymbol{\lambda}(y_i)\xi(c_i)^T \ ,$$

where we have made use of the memoryless property of the channel, and of the fact that if $c_i = \alpha \in \mathfrak{R}^-$, then $\boldsymbol{\lambda}(y_i)\xi(c_i)^T = \lambda^{(\alpha)}(y_i)$. This is then equivalent to

$$\begin{aligned}(\xi(\hat{c}_1) \mid \xi(\hat{c}_2) \mid \ \ldots \mid \xi(\hat{c}_n)) \\ &= \arg\min_{\boldsymbol{f}\in\mathcal{K}(\mathcal{C})} \sum_{i=1}^{n} \boldsymbol{\lambda}(y_i)\boldsymbol{f}_i^T \\ &= \arg\min_{\boldsymbol{f}\in\mathcal{K}(\mathcal{C})} \boldsymbol{\lambda}(\boldsymbol{y})\boldsymbol{f}^T \ ,\end{aligned} \tag{1}$$

where

$$\boldsymbol{f} = (\boldsymbol{f}_1 \mid \boldsymbol{f}_2 \mid \ \cdots \mid \boldsymbol{f}_n)$$

and

$$\boldsymbol{f}_i = (f_i^{(\alpha)})_{\alpha\in\mathfrak{R}^-} \text{ for all } i \in \mathcal{I} \ ,$$

and where $\mathcal{K}(\mathcal{C})$ represents the convex hull of all points $\boldsymbol{f} \in \mathbb{R}^{(q-1)n}$ which correspond to codewords, i.e.

$$\mathcal{K}(\mathcal{C}) = H_{\text{conv}}\big\{(\xi(c_1) \mid \xi(c_2) \mid \ \ldots \mid \xi(c_n)) \ : \ \boldsymbol{c} \in \mathcal{C}\big\} \ .$$

Therefore it is seen that the ML decoding problem reduces to the minimization of a linear objective function (or cost function) over a polytope in $\mathbb{R}^{(q-1)n}$. The number of variables and constraints for this linear program is exponential in n, and it is therefore too complex for practical implementation. To circumvent this problem, we formulate a relaxed LP problem, as shown next.

The solution we seek for $\boldsymbol{f}$ (i.e. the desired LP output) is

$$\boldsymbol{f} = (\xi(\bar{c}_1) \mid \xi(\bar{c}_2) \mid \ \ldots \mid \xi(\bar{c}_n)) \ .$$

We introduce auxiliary variables whose constraints, along with those of the elements of $\boldsymbol{f}$, will form the relaxed LP problem. First, for each $j \in \mathcal{J}$, we define the mapping $\boldsymbol{X}_j(\boldsymbol{c})$ of the words $\boldsymbol{c} \in \mathfrak{R}^n$, $\boldsymbol{X}_j(\boldsymbol{c}) = (X_{j,\alpha}(\boldsymbol{c}))_{\alpha\in\mathfrak{R}^-}$, where

$$X_{j,\alpha}(\boldsymbol{c}) = \{i \in \text{supp}(\mathcal{H}_j) \ : \ c_i = \alpha\} \ ,$$

for $\alpha \in \mathfrak{R}^-$. For each word $\boldsymbol{c} \in \mathfrak{R}^n$, $X_{j,\alpha}(\boldsymbol{c})$ is the set of word indices where symbol α appears in parity check j, for $j \in \mathcal{J}$, $\alpha \in \mathfrak{R}^-$. We define the set E_j as

$$E_j \ = \ \big\{\boldsymbol{S} = (S_\alpha)_{\alpha\in\mathfrak{R}^-} = \boldsymbol{X}_j(\boldsymbol{c}) \ : \ \boldsymbol{c} \in \mathcal{C}_j\big\} \ .$$

In other words, $\boldsymbol{X}_j(\boldsymbol{c}) \in E_j$ if and only if parity check j is satisfied by the word $\boldsymbol{c} \in \mathfrak{R}^n$.

We now introduce the auxiliary variables

$$w_{j,\boldsymbol{S}} \ \text{ for } \ j \in \mathcal{J}, \boldsymbol{S} \in E_j \ ,$$

and denote the vector containing these variables as

$$\boldsymbol{w} = \big(\ w_{j,\boldsymbol{S}}\ \big)_{j\in\mathcal{J},\boldsymbol{S}\in E_j} \ ,$$

with respect to some ordering on the elements of E_j. The solution we seek for these variables is

$$\forall j \in \mathcal{J} \ \ : \ \ w_{j,\boldsymbol{S}} = \begin{cases} 1 & \text{if } \boldsymbol{S} = \boldsymbol{X}_j(\bar{\boldsymbol{c}}) \\ 0 & \text{otherwise} \end{cases} \ .$$

To this end, we impose the constraints

$$\forall j \in \mathcal{J}, \ \forall \boldsymbol{S} \in E_j, \quad 0 \le w_{j,\boldsymbol{S}} \le 1 \ , \tag{2}$$

and

$$\forall j \in \mathcal{J}, \quad \sum_{\boldsymbol{S}\in E_j} w_{j,\boldsymbol{S}} = 1 \ . \tag{3}$$

Finally, we note that the solution we seek satisfies the further constraints

$$\begin{aligned}\forall j \in \mathcal{J}, \ \forall i \in \text{supp}(\mathcal{H}_j), \ \forall \alpha \in \mathfrak{R}^-, \\ f_i^{(\alpha)} = \textstyle\sum_{\boldsymbol{S}\in E_j, \ i\in S_\alpha} w_{j,\boldsymbol{S}} \ .\end{aligned} \tag{4}$$

Constraints (2)-(4) form a polytope which we denote $\mathcal{Q}$. The minimization of the objective function (1) over $\mathcal{Q}$ forms the relaxed LP decoding problem. This LP is defined by $O(qn + q^d m)$ variables and $O(qn + q^d m)$ constraints. We note that the further constraints

$$\forall i \in \mathcal{I}, \ \forall \alpha \in \mathfrak{R}^-, \quad 0 \le f_i^{(\alpha)} \le 1 \ , \tag{5}$$

and

$$\forall i \in \mathcal{I}, \quad \sum_{\alpha\in\mathfrak{R}^-} f_i^{(\alpha)} \le 1 \ , \tag{6}$$

follow from the constraints (2)-(4), for any $(\boldsymbol{f}, \boldsymbol{w}) \in \mathcal{Q}$.

Now we may define the decoding algorithm, which works as follows. The decoder solves the LP problem of minimizing the objective function (1) subject to the constraints (2)-(4). If $\boldsymbol{f} \in \{0,1\}^{(q-1)n}$, the output is the codeword $(\xi^{-1}(\boldsymbol{f}_1), \xi^{-1}(\boldsymbol{f}_2), \cdots, \xi^{-1}(\boldsymbol{f}_n))$ (we shall prove in the next section that this output is indeed a codeword). Otherwise, the decoder outputs an 'error'.

4 Polytope Properties

The analysis in this section is a direct generalization of the results in [2].

Definition 4.1: An *integral point* in a polytope is a point with all *integer* coordinates.

Proposition 4.1:

1) Let $(\boldsymbol{f}, \boldsymbol{w}) \in \mathcal{Q}$, and $f_i^{(\alpha)} \in \{0,1\}$ for every $i \in \mathcal{I}$, $\alpha \in \mathfrak{R}^-$. Then,
$$(\xi^{-1}(\boldsymbol{f}_1) \ , \xi^{-1}(\boldsymbol{f}_2) \ , \cdots \ , \xi^{-1}(\boldsymbol{f}_n)) \in \mathcal{C} \ .$$
2) Conversely, for every codeword $\boldsymbol{c} = (c_1, c_2, \cdots, c_n) \in \mathcal{C}$, there exists $\boldsymbol{w}$ such that $(\boldsymbol{f}, \boldsymbol{w})$ is an integral point in $\mathcal{Q}$ with $\boldsymbol{f}_i = \xi(c_i)$ for all $i \in \mathcal{I}$.

Proof.

1) Suppose $(\boldsymbol{f}, \boldsymbol{w}) \in \mathcal{Q}$, and $f_i^{(\alpha)} \in \{0,1\}$ for every $i \in \mathcal{I}$, $\alpha \in \mathfrak{R}^-$.

Define $\boldsymbol{c}$ by $c_i = \xi^{-1}(\boldsymbol{f}_i)$ for all $i \in \mathcal{I}$. By (6), this is well defined. Define $\boldsymbol{T} = (T_\alpha)_{\alpha\in\mathfrak{R}^-} = \boldsymbol{X}_j(\boldsymbol{c})$, i.e.

$$T_\alpha = \{i \in \text{supp}(\mathcal{H}_j) \ : \ f_i^{(\alpha)} = 1\} \ , \qquad (7)$$

for $\alpha \in \mathfrak{R}^-$. Now, fix some $j \in \mathcal{J}$ and let $\boldsymbol{P} = (P_\alpha)_{\alpha\in\mathfrak{R}^-} \in E_j$, $\boldsymbol{P} \neq \boldsymbol{T}$. There must exist $\alpha \in \mathfrak{R}^-$ and $i_0 \in \mathcal{I}$ such that either $i_0 \in P_\alpha \backslash T_\alpha$ or $i_0 \in T_\alpha \backslash P_\alpha$.
If $i_0 \in P_\alpha \backslash T_\alpha$, then by (4) and (7)

$$f_{i_0}^{(\alpha)} = 0 = \sum_{\boldsymbol{S}\in E_j, \ i_0\in S_\alpha} w_{j,\boldsymbol{S}} \ .$$

Therefore $w_{j,\boldsymbol{S}} = 0$ for all $\boldsymbol{S} \in E_j$ with $i_0 \in S_\alpha$, and in particular $w_{j,\boldsymbol{P}} = 0$.
If $i_0 \in T_\alpha \backslash P_\alpha$, then by (3), (4), and (7)

$$\begin{aligned} 0 &= 1 - f_{i_0}^{(\alpha)} \\ &= \sum_{\boldsymbol{S}\in E_j} w_{j,\boldsymbol{S}} - \sum_{\boldsymbol{S}\in E_j, \ i_0\in S_\alpha} w_{j,\boldsymbol{S}} \\ &= \sum_{\boldsymbol{S}\in E_j, \ i_0\notin S_\alpha} w_{j,\boldsymbol{S}} \ . \end{aligned}$$

Therefore $w_{j,\boldsymbol{S}} = 0$ for all $\boldsymbol{S} \in E_j$ with $i_0 \notin S_\alpha$, and in particular $w_{j,\boldsymbol{P}} = 0$.
It follows that $w_{j,\boldsymbol{S}} = 0$ for all $\boldsymbol{S} \in E_j$, $\boldsymbol{S} \neq \boldsymbol{T}$. But by (3) this implies that $\boldsymbol{T} \in E_j$ (and that $w_{j,\boldsymbol{T}} = 1$). Applying this argument for every $j \in \mathcal{J}$ implies $\boldsymbol{c} \in \mathcal{C}$.

2) For $\boldsymbol{c} \in \mathcal{C}$, we let $\boldsymbol{f}_i = \xi(c_i)$ for $i \in \mathcal{I}$. For each parity check $j \in \mathcal{J}$, we let $\boldsymbol{T} = (T_\alpha)_{\alpha\in\mathfrak{R}^-} = \boldsymbol{X}_j(\boldsymbol{c}) \in E_j$ and then set

$$\forall j \in \mathcal{J}: \qquad w_{j,\boldsymbol{S}} = \begin{cases} 1 & \text{if } \boldsymbol{S} = \boldsymbol{T} \\ 0 & \text{otherwise.} \end{cases}$$

It is easily checked that the resulting point $(\boldsymbol{f}, \boldsymbol{w})$ is integral and satisfies constraints (2)-(4). □

The following proposition assures the so-called *ML certificate* property.

Proposition 4.2: Suppose that the decoder outputs a codeword $\boldsymbol{c} \in \mathcal{C}$. Then, $\boldsymbol{c}$ is the maximum-likelihood codeword.

The proof of this proposition is straightforward. The reader can refer to a similar proof for the binary case in [2].

5 Transmission-Independent Decoder Performance

In this section, we state a theorem on decoder performance, namely, that under a certain symmetry condition, the probability of decoder failure is independent of the transmitted codeword. Decoder failure is defined as the event where the decoder output is not equal to the transmitted codeword (this could correspond to a non-integral value of $\boldsymbol{f}$, or to an erroneous output codeword).

Symmetry Condition.
For each $\alpha \in \mathfrak{R}$, there exists a bijection

$$\tau_\alpha \ : \ \Sigma \longrightarrow \Sigma \ ,$$

such that the channel output probability (density) conditioned on the channel input satisfies

$$p(y|\beta) = p(\tau_\alpha(y)|\beta - \alpha) \ ,$$

For all $y \in \Sigma$, $\beta \in \mathfrak{R}$. When Σ is equal to $\mathbb{R}^l$ or $\mathbb{C}^l$ for $l \geq 1$, the mapping τ_α is assumed to be isometric with respect to Euclidean distance in Σ, for every $\alpha \in \mathfrak{R}$.

Theorem 5.1: Under the stated symmetry condition, the probability of decoder failure is independent of the transmitted codeword.
The proof of this theorem is omitted due to space limitations. Examples of modulator-channel combinations for which this assumption holds are: q-ary PSK modulation over AWGN (where the additive group of $\mathfrak{R}$ is cyclic); orthogonal modulation over AWGN; and the discrete memoryless q-ary symmetric channel.

6 Linear-Programming Pseudocodewords

Definition 6.1: A *linear-programming pseudocodeword* (LP pseudocodeword) of the code $\mathcal{C}$ is a vector $(\boldsymbol{h}, \boldsymbol{z})$ where

$$\boldsymbol{h} = (\boldsymbol{h}_1 \mid \boldsymbol{h}_2 \mid \cdots \mid \boldsymbol{h}_n) \ ,$$

$$\forall i \in \mathcal{I}, \ \boldsymbol{h}_i = (h_i(\alpha))_{\alpha\in\mathfrak{R}^-} \ ,$$

$$\boldsymbol{z} = \left(z_{j,\boldsymbol{S}} \right)_{j\in\mathcal{J}, \boldsymbol{S}\in E_j} \ ,$$

where the elements of $\boldsymbol{z}$ are nonnegative integers, and the following two conditions hold for all $j \in \mathcal{J}$:

$$\forall i \in \text{supp}(\mathcal{H}_j), \ \forall \alpha \in \mathfrak{R}^-, \quad h_i(\alpha) = \textstyle\sum_{\boldsymbol{S}\in E_j, \ i\in S_\alpha} z_{j,\boldsymbol{S}} \ , \qquad (8)$$

$$\forall i \in \text{supp}(\mathcal{H}_j), \ h_i(0) = \sum_{\substack{\boldsymbol{S}\in E_j \\ \forall \alpha\in\mathfrak{R}^- \, : \, i\notin S_\alpha}} z_{j,\boldsymbol{S}} \ . \qquad (9)$$

From (8) and (9) it follows that the elements of $\boldsymbol{h}$ are nonnegative integers, and that for each $i \in \text{supp}(\mathcal{H}_j) \cap \text{supp}(\mathcal{H}_{j'})$, we have

$$\sum_{\alpha\in\mathfrak{R}} h_i(\alpha) = \sum_{\boldsymbol{S}\in E_j} z_{j,\boldsymbol{S}} = \sum_{\boldsymbol{S}\in E_{j'}} z_{j',\boldsymbol{S}} \ . \qquad (10)$$

We assume that the Tanner graph of $\mathcal{H}$ is connected; it then follows from (10) that

$$\forall i \in \mathcal{I} \ : \ \sum_{\alpha\in\mathfrak{R}} h_i(\alpha) = M \ ,$$

for some fixed nonnegative integer M.

We note that the LP pseudocodeword $(\boldsymbol{h}, \boldsymbol{z})$ defined above can be represented by the $n \times q$ matrix

$$\mathsf{H} = \Big(h_i(\alpha) \Big)_{i\in\mathcal{I}; \, \alpha\in\mathfrak{R}} \ .$$

In the following, we say that the decoder *fails* if the decoder output is not equal to the transmitted codeword.

Theorem 6.1: Assume that the all-zero codeword was transmitted.

1) If the LP decoder fails, then there exists some LP pseudocodeword $(\boldsymbol{h}, \boldsymbol{z})$, $\boldsymbol{h} \neq \mathbf{0}$, such that
$$\sum_{i=1}^{n} \left(\sum_{\alpha \in \mathfrak{R}^-} \lambda^{(\alpha)}(y_i) h_i(\alpha) \right) \leq 0 \ . \qquad (11)$$
2) If there exists some LP pseudocodeword $(\boldsymbol{h}, \boldsymbol{z})$, $\boldsymbol{h} \neq \mathbf{0}$, such that
$$\sum_{i=1}^{n} \left(\sum_{\alpha \in \mathfrak{R}^-} \lambda^{(\alpha)}(y_i) h_i(\alpha) \right) < 0 \ , \qquad (12)$$
then the LP decoder fails.

Proof. The proof follows the lines of its counterpart in [2].

1) Let $(\boldsymbol{f}, \boldsymbol{w})$ be the point in $\mathcal{Q}$ which minimizes $\boldsymbol{\lambda}(\boldsymbol{y})\boldsymbol{f}^T$. Suppose the decoder fails; then $\boldsymbol{f} \neq \mathbf{0}$, and we must have $\boldsymbol{\lambda}(\boldsymbol{y})\boldsymbol{f}^T \leq 0$.
Next, we construct the LP pseudocodeword $(\boldsymbol{h}, \boldsymbol{z})$ as follows. Since the LP has rational coefficients, all elements of the vectors $\boldsymbol{f}$ and $\boldsymbol{w}$ must be rational. Let M denote their lowest common denominator; since $\boldsymbol{f} \neq \mathbf{0}$ we may have $M > 0$. Now set $h_i(\alpha) = M \cdot f_i^{(\alpha)}$ for all $i \in \mathcal{I}$, $\alpha \in \mathfrak{R}^-$, set $z_{j,\boldsymbol{S}} = M \cdot w_{j,\boldsymbol{S}}$ for all $j \in \mathcal{J}$ and $\boldsymbol{S} \in E_j$, and then define $h_i(0)$ as in (9) for all $i \in \mathcal{I}$. By (2) and (4), $(\boldsymbol{h}, \boldsymbol{z})$ is an LP pseudocodeword and $\boldsymbol{h} \neq \mathbf{0}$ since $\boldsymbol{f} \neq \mathbf{0}$. Also $\boldsymbol{\lambda}(\boldsymbol{y})\boldsymbol{f}^T \leq 0$ implies (11).
2) Now, suppose that an LP pseudocodeword $(\boldsymbol{h}, \boldsymbol{z})$ with $\boldsymbol{h} \neq \mathbf{0}$ satisfies (12). Let
$$M = \sum_{\alpha \in \mathfrak{R}} h_i(\alpha) \ .$$
Since $\boldsymbol{h} \neq \mathbf{0}$ we have $M > 0$. Now:
   - Set $f_i^{(\alpha)} = h_i(\alpha)/M$ for all $i \in \mathcal{I}$, $\alpha \in \mathfrak{R}^-$;
   - Set $w_{j,\boldsymbol{S}} = z_{j,\boldsymbol{S}}/M$ for all $j \in \mathcal{J}$ and $\boldsymbol{S} \in E_j$.

   It is straightforward to check that $(\boldsymbol{f}, \boldsymbol{w})$ satisfies all the constraints of the polytope $\mathcal{Q}$. Also, $\boldsymbol{h} \neq \mathbf{0}$ implies $\boldsymbol{f} \neq \mathbf{0}$. Finally, (12) implies $\boldsymbol{\lambda}(\boldsymbol{y})\boldsymbol{f}^T < 0$. Therefore, the LP decoder will produce an output other than the all-zero codeword, resulting in decoder failure.

□

7 Equivalence Between Pseudocodeword Sets

In this section, we show the equivalence between the set of LP pseudocodewords and the set of graph-cover pseudocodewords. The result is summarized in the following theorem.

Theorem 7.1: There exists an LP pseudocodeword $(\boldsymbol{h}, \boldsymbol{z})$ for the code $\mathcal{C}$ with matrix representation H if and only if there exists a graph-cover pseudocodeword $\boldsymbol{p}$ with the same matrix representation.

Proof.

1) Let $(\boldsymbol{h}, \boldsymbol{z})$ be an LP pseudocodeword, and let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the Tanner graph $\mathcal{C}$ associated with the parity-check matrix $\mathcal{H}$. We define
$$M = \sum_{\alpha \in \mathfrak{R}} h_i(\alpha) \ .$$
(Recall that under our assumption that the Tanner graph is connected, the value of M is independent of i.) Below, we construct a corresponding M-cover graph $\tilde{\mathcal{G}} = (\tilde{\mathcal{V}}, \tilde{\mathcal{E}})$.
   - For every $i \in \mathcal{I}$, and for every $\alpha \in \mathfrak{R}$, the graph $\tilde{\mathcal{G}}$ will contain $h_i(\alpha)$ copies of the vertex u_i associated with the value α.
   - For every $j \in \mathcal{J}$, $\boldsymbol{S} \in E_j$, the graph $\tilde{\mathcal{G}}$ will contain $z_{j,\boldsymbol{S}}$ copies of the check vertex v_j, associated with the $(q-1)$-tuple $\boldsymbol{S}$.
   - The edges in the graph are connected according to the membership in the sets S_α, for $\alpha \in \mathfrak{R}^-$. Namely, each copy of check vertex v_j will be connected to one copy of u_i for every $u_i \in \mathcal{N}(v_j)$. A copy of a check vertex v_j associated with the $(q-1)$-tuple $\boldsymbol{S}$ will be connected to a copy of u_i associated with the value $\alpha \in \mathfrak{R}^-$ if and only if $i \in S_\alpha$. A copy of v_j associated with the $(q-1)$-tuple $\boldsymbol{S}$ will be connected to a copy of u_i associated with the value 0 if and only if $i \notin \cup_{\alpha \in \mathfrak{R}^-} S_\alpha$.

   By using (8), we see that for every $j \in \mathcal{J}$, $i \in \text{supp}(\mathcal{H}_j)$, $\alpha \in \mathfrak{R}$, there are exactly $h_i(\alpha)$ edges connecting the copies of the vertex v_j with the copies of u_i associated with the value α. Therefore, the graph $\tilde{\mathcal{G}}$ is well-defined, and the neighborhood of a copy of v_j contains exactly one copy of u_i for every $u_i \in \mathcal{N}(v_j)$. Furthermore, it can be seen that the neighborhood of a copy of u_i contains exactly one copy of v_j for every $v_j \in \mathcal{N}(u_i)$. In addition, all copies of all check vertices v_j represent satisfied checks, and therefore $\boldsymbol{p}$, induced by the graph $\tilde{\mathcal{G}}$, is a graph cover pseudocodeword of $\mathcal{C}$, as claimed.
2) Now let $\boldsymbol{p}$ be a graph-cover pseudocodeword corresponding to some M-cover of the Tanner graph of $\mathcal{C}$. Then,
   - for every $i \in \mathcal{I}$, and for every $\alpha \in \mathfrak{R}$, we define $h_i(\alpha)$ to be the number of copies of the vertex u_i associated with value α.
   - for every $j \in \mathcal{J}$, and for every $\boldsymbol{S} \in E_j$, we define $z_{j,\boldsymbol{S}}$ to be the number of copies of the check vertex v_j connected to copies of u_i, associated with $\alpha \in \mathfrak{R}^-$ for $i \in S_\alpha$, and associated with 0 for $i \notin \cup_{\alpha \in \mathfrak{R}^-} S_\alpha$.

Then, $z_{j,\boldsymbol{S}}$ are all nonnegative integers for all $j \in \mathcal{J}$ and $\boldsymbol{S} \in E_j$. Moreover, (8) and (9) hold for all $j \in \mathcal{J}$ by construction of the graph. Therefore, $(\boldsymbol{h}, \boldsymbol{z})$ is an LP pseudocodeword of the code $\mathcal{C}$. □

8 Simulation Study

In this section we compare performance of the linear-programming decoder with hard-decision and soft-decision based ML decoding. For such a comparison, a code and modulation scheme are needed which possess sufficient symmetry properties to enable derivation of analytical ML performance results. We consider encoding of 6-symbol blocks according to the $(11, 6, 5)$ ternary Golay code, and modulation of the resulting ternary symbols with 3-PSK modulation prior to transmission over the AWGN channel. The symbol error rate (SER) and codeword error rate (WER) are shown in Figure 1. To quantify performance, we define the signal-to-noise ratio (SNR) per information symbol $\gamma_s = E_s/N_0$ as the ratio of receive signal energy per information symbol to the noise power spectral density. Also shown in the figure are two other performance curves for WER. The first is the exact result for ML hard-decision decoding of the ternary Golay code; since the Golay code is perfect, this is obtained from

$$\mathrm{WER}(\gamma_s) = \sum_{\ell=3}^{11} \binom{11}{\ell} (p(\gamma_s))^{\ell} (1 - p(\gamma_s))^{11-\ell} \ ,$$

where $p(\gamma_s)$ represents the probability of incorrect hard decision at the demodulator and was evaluated for each value of γ_s using numerical integration. The second WER curve represents the union bound for ML soft-decision decoding. Using the symmetry of the 3-PSK constellation, this may be obtained from

$$\mathrm{WER}(\gamma_s) < \frac{1}{2} \sum_{\boldsymbol{c} \in \mathcal{C}} \mathrm{erfc} \left(\sqrt{\frac{3}{4} w_H(\boldsymbol{c})\, r \gamma_s} \right) \ ,$$

where r denotes the code rate, and the Hamming weight of the codeword $\boldsymbol{c} \in \mathcal{C}$, $w_H(\boldsymbol{c})$, is given by the weight enumerating polynomial

$$W(x) = 1 + 132x^5 + 132x^6 + 330x^8 + 110x^9 + 24x^{11} \ .$$

The performance of LP decoding is approximately the same as that of codeword-error-rate optimum hard-decision decoding. The performance lies 0.1 dB from the result for ML hard-decision decoding and 1.53 dB from the union bound for codeword-error-rate optimum soft-decision decoding at a WER of 10^{-4}.

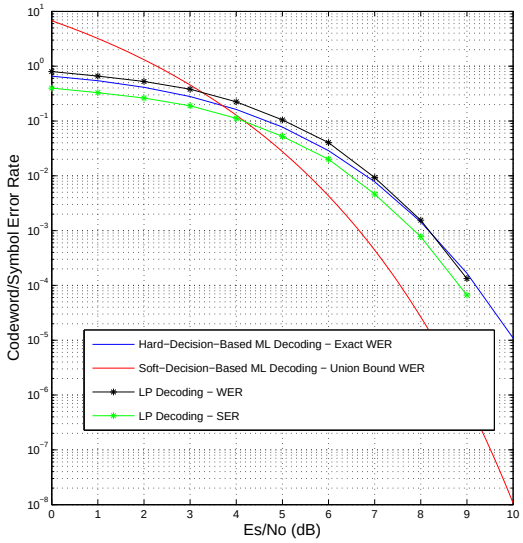

Fig. 1. Codeword error rate (WER) and symbol error rate (SER) for the $(11, 6, 5)$ ternary Golay code under 3-PSK modulation. The figure shows performance under LP decoding, as well as the exact result for hard-decision decoding and the union bound for soft-decision decoding.

Acknowledgements

This work was supported by the Claude Shannon Institute for Discrete Mathematics, Coding and Cryptography (Science Foundation Ireland Grant 06/MI/006). The authors would like to thank J. Feldman and R. Koetter for helpful discussions.

References

[1] J. FELDMAN, *Decoding Error-Correcting Codes via Linear Programming*, Ph.D. Thesis, Massachusetts Institute of Technology, Sep. 2003.

[2] J. FELDMAN, M.J. WAINWRIGHT, D.R. KARGER, *Using linear programming to decode binary linear codes, IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 954–972, Mar. 2005.

[3] G.D. FORNEY, R. KOETTER, F.R. KSCHISCHANG, A. REZNIK, *On the effective weights of pseudocodewords for codes defined on graphs with cycles, vol. 123 of Codes, systems, and graphical models, IMA Vol. Math. Appl., ch. 5, pp. 101-112*, Springer, 2001.

[4] C.A. KELLEY, D. SRIDHARA, J. ROSENTHAL, *Pseudocodeword weights for non-binary LDPC codes, Proc. IEEE International Symposium on Information Theory (ISIT), July 2006, p.1379-1383*, Seattle, USA.

[5] R. KOETTER, W.-C. W. LI, P.O. VONTOBEL, J.L. WALKER, *Characterizations of Pseudo-Codewords of LDPC Codes*, Arxiv report arXiv:cs.IT/0508049, Aug. 2005.

[6] R. KOETTER, P. VONTOBEL, *Graph-cover decoding and finite-length analysis of message-passing iterative decoding of LDPC codes*, submitted to *IEEE Trans. Inform. Theory*, Arxiv report arXiv:cs.IT/0512078, Dec. 2005.

[7] D. SRIDHARA, T. E. FUJA, *LDPC codes over rings for PSK modulation, IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3209–3220, Sep. 2005.

[8] N. WIBERG, *Codes and Decoding on General Graphs*, Ph.D. Thesis, Linköping University, Sweden, 1996.