Towards Practical Minimum-Entropy Universal Decoding

Todd P. Coleman, Muriel Médard; Michelle Effros
{colemant,medard}@mit.edu; effros@caltech.edu
Massachusetts Institute of Technology; California Institute of Technology

Abstract

Minimum-entropy decoding is a universal decoding algorithm used in decoding block compression of discrete memoryless sources as well as block transmission of information across discrete memoryless channels. Extensions can also be applied for multiterminal decoding problems, such as the Slepian-Wolf source coding problem. The 'method of types' has been used to show that there exist linear codes for which minimum-entropy decoders achieve the same error exponent as maximum-likelihood decoders. Since minimum-entropy decoding is NP-hard in general, minimum-entropy decoders have existed primarily in the theory literature. We introduce practical approximation algorithms for minimum-entropy decoding. Our approach, which relies on ideas from linear programming, exploits two key observations. First, the 'method of types' shows that that the number of distinct types grows polynomially in n. Second, recent results in the optimization literature have illustrated polytope projection algorithms with complexity that is a function of the number of vertices of the projected polytope. Combining these two ideas, we leverage recent results on linear programming relaxations for error correcting codes to construct polynomial complexity algorithms for this setting. In the binary case, we explicitly demonstrate linear code constructions that admit provably good performance.

1 Introduction

Information theory has had a profound and significant impact on the design and understanding of digital communication systems. Since the inception of information theory, it has been known that there exist codes and decoding algorithms that permit block compression of memoryless sources and block transmission of information across uncertain, memoryless channels with an error probability that decays exponentially in the length of the block code [1]. Recent insights into iterative decoding techniques for linear codes based on graphs have sharply narrowed the gap between channel coding theory and practice [2, 3]. Applying these channel decoding techniques for near-lossless compression of one or many sources yields similar results [4, 5, 6, 7, 8]. All of these decoding techniques can be thought of as low-complexity approximations to maximum a posteriori (MAP) decoding and require a priori knowledge of the source or channel's statistics [9].

Since limited feedback and rate loss can make it difficult to estimate source and channel statistics in many problems of practical interest, a natural question to consider is whether or not there exist universal decoding algorithms that deliver the same performance (in terms of achievable rates and rate of decay of the error probability). For discrete memoryless systems,

Proceedings of the 2005 Data Compression Conference (DCC'05)
1068-0314/05 $ 20.00 IEEE

in the absence of complexity constraints, the answer is yes [10, 11]. Csiszár's *minimum-entropy* decoder [12] addresses the universal decoding problem for near-lossless data compression, which is where we will focus our attention for this setting. We note that our discussion and approach also directly apply to the universal channel coding setting by replacing 'minimum-entropy' with 'maximum-mutual information' (see [11]). While minimum-entropy decoders are well-established as a proof technique in the theory literature, we know of no prior literature that tackles practical minimum-entropy decoder design.

2 Model and Definitions

Throughout this discussion we consider a discrete memoryless source (DMS) U over alphabet $\mathcal{U} = \{0, 1, \ldots, Q-1\}$. We use the following definitions:

$$\begin{aligned}
CH(\mathcal{S}) &= \text{ the convex hull of all } s \in \mathcal{S} \\
\mathcal{V}(\mathcal{B}) &= \{v \in \mathcal{B} : v \text{ is a vertex of the polytope } \mathcal{B}\} \\
\mathcal{H}(\mathcal{B}) &\triangleq \text{the number of half-spaces representing } \mathcal{B} \\
\mathcal{P}(\mathcal{U}) &= \left\{ P = \left(\{P_a\}_{a \in \mathcal{U}} \right) : P \geq \underline{0}, \sum_{a \in \mathcal{U}} P_a = 1 \right\} \\
h_b(p) &= -p \log_2 p - (1-p) \log_2 (1-p) \text{ for } p \in [0,1] \\
P_{\underline{u}} &= \left(\left\{ \frac{1}{n} \sum_{i=1}^{n} 1_{u_i = a} \right\}_{a \in \mathcal{U}} \right) \text{ for } \underline{u} \in \mathcal{U}^n \qquad (1) \\
\mathcal{P}_n(\mathcal{U}) &= \{P \in \mathcal{P}(\mathcal{U}) : P = P_{\underline{u}} \text{ for some } \underline{u} \in \mathcal{U}^n\}
\end{aligned}$$

From [10, 11] we note the following:

$$\begin{aligned}
|\mathcal{P}_n(\mathcal{U})| &= \binom{n + |\mathcal{U}| - 1}{|\mathcal{U}| - 1} \\
&\leq (n+1)^{|\mathcal{U}|} \qquad (2)
\end{aligned}$$

Thus *the number of types is polynomial in n.* We also note that

$$\binom{n}{k} = \binom{n}{n-k} = O(n^k). \qquad (3)$$

3 The General Problem

Consider a DMS U with probability distribution $\Pr(\cdot) \in \mathcal{P}(\mathcal{U})$. Our goal is to design a low-complexity, fixed-rate universal code. For any fixed coding dimension n and rate R, a fixed-rate source code's encoder maps length-n input sequences to binary strings of length nR while the code's decoder maps length-nR binary sequences back to length-n sequences from the input alphabet $\mathcal{U}$. The encoder and decoder operate without any knowledge of the source distribution $\Pr(\cdot)$. A code is universal if its error probability can be made arbitrarily small on *all* sources $\Pr(\cdot)$ with $H(U) < R$.

Without loss of generality, we assume that $\mathcal{U} = \{0, 1, \ldots, Q-1\}$ where $Q = 2^m$ for some integer $m > 1$. Thus we may assume that U takes on values in field $\mathbb{F}_{2^m}$. We use a linear block source encoder, which maps source vector $\underline{u} \in \mathcal{U}^n$ to syndrome $\underline{s} \in \mathcal{U}^{n-k}$ using linear code

$$H = \begin{bmatrix} -H'_1- \\ -H'_2- \\ \vdots \\ -H'_{n-k}- \end{bmatrix} \in \mathcal{U}^{(n-k)\times n}$$

according to $\underline{s} = H\underline{u}$. The source coding rate is $R = \frac{n-k}{n} m$.

A universal decoder must select the 'best' source vector consistent with the observation $\underline{s}$. The set of source vectors consistent with $\underline{s}$ is the coset

$$\mathrm{Co}\,(H, \underline{s}) = \{\underline{u} \bigm| H\underline{u} = \underline{s}\}.$$

Csiszár's 'minimum-entropy' decoder selects as the source reconstruction the coset's entropy minimizer

$$\hat{\underline{u}} = \arg \min_{\underline{u} \in \mathrm{Co}(H,\underline{s})} H\left(P_{\underline{u}}\right). \tag{4}$$

In [12], Csiszár shows that not only do there exist linear codes such whose rates can be arbitrarily close to $H(U)$ when such a decoder is applied, but also that minimum entropy decoding achieves the same error exponent as the optimal maximum-likelihood (ML) decoder.

3.1 From Discrete to Continuous Optimization

Note that (4) is a discrete optimization problem with an exponential number of feasible solutions. Our first step is to replace (4) by a continuous optimization problem. We first construct indicator variables $I_i^k \in \{0, 1\}$, for $k \in \mathcal{U}, i \in \{1, \ldots, n\}$, such that $I_i^k = 1$ if $u_i = k$ and $I_i^k = 0$ otherwise. Thus I_i^k specifies $\underline{u} \in \mathcal{U}^n$ as

$$\underline{u} \;\; = \;\; \mu(I), \text{ where } u_i = \mu_i(I) = \sum_{k \in \mathcal{U}} k I_i^k. \tag{5}$$

Note that any $\underline{u} \in \mathrm{Co}\,(H, \underline{s})$ must satisfy the constraints of the linear code. We impose these code constraints on I by defining

$$\mathcal{I}(H, \underline{s}) = \{I \; s.t. \; \mu(I) \in \mathrm{Co}\,(H, \underline{s})\}. \tag{6}$$

For any $I \in \mathcal{I}(H, \underline{s})$ and the corresponding $\underline{u} = \mu(I)$, we can construct $P_{\underline{u}}$ as a linear mapping

$$P = \tau(I), \text{ where } P(k) = \tau_k(I) = \frac{1}{n} \sum_{i=1}^{n} I_i^k, \;\; k \in \mathcal{U}.$$

Thus we can define the polytope $\mathcal{B}^{i,p}(H, \underline{s})$ as

$$\mathcal{B}^{i,p}(H, \underline{s}) = \left\{(I, P) \; s.t. \; I \in CH(\mathcal{I}(H, \underline{s})), \; P = \tau(I)\right\}.$$

Note that for every $(I, P) \in \mathcal{V}\left(\mathcal{B}^{i,p}(H, \underline{s})\right)$:

- I corresponds to a coset member $\underline{u} = \mu(I) \in \mathrm{Co}\,(H, \underline{s})$.
- The empirical type $P_{\underline{u}}$ associated with $\underline{u} = \mu(I)$ satisfies $P_{\underline{u}} = P = \tau(I)$.

Since the entropy function is *strictly concave*, and since minimizing a strictly concave function over a polytope $\mathcal{B}$ has the property that an optimal solution lies in $\mathcal{V}\,(\mathcal{B})$, we can perform (4) in the continuous domain as

$$\begin{aligned} \min \quad & H(P) && (7a) \\ s.t. \quad & (I, P) \in \mathcal{B}^{i,p}(H, \underline{s}) && (7b) \end{aligned}$$

and take the minimum-entropy solution as $\underline{u}^* = \mu(I^*)$ where (I^*, P^*) is an optimal solution to (7). At first glance, there are two difficulties that arise in trying to perform (7):

1) Since ML-decoding for linear codes is generally NP-hard, the best bound on $\mathcal{H}\,(\mathcal{B})$ (and thus $\mathcal{H}\,(\mathcal{B}^{i,p})$) is $O(2^n)$. As a result, it is not obvious how to efficiently represent $\mathcal{B}^{i,p}$.

2) In (7), $|\mathcal{V}\,(\mathcal{B}^{i,p})| = O(2^n)$ and concave minimization over a polytope is NP-hard [13] - generally requiring a visit to every $v \in \mathcal{V}\,(\mathcal{B}^{i,p})$.

However, even though $|\mathrm{Co}\,(H, \underline{s})| = O(2^n)$, from (2) it follows that the number of distinct *types* associated with $\mathrm{Co}\,(H, \underline{s})$ is polynomial in n. This observation suggests the following strategy.

a) Project $\mathcal{B}^{i,p}(H, \underline{s})$ onto $\mathcal{B}^{p}(H, \underline{s}) = \{P \mid (I, P) \in \mathcal{B}^{i,p}(H, \underline{s}) \text{ for some } I\}$.

b) Perform the minimization

$$\begin{aligned} \min \quad & H(P) && (8a) \\ s.t. \quad & P \in \mathcal{B}^{p}(H, \underline{s}). && (8b) \end{aligned}$$

In worst case any 'concave minimization over a polytope' algorithm might have to iterate through every $v \in \mathcal{V}\,(\mathcal{B}^{p}(H, \underline{s}))$, but here $|\mathcal{V}\,(\mathcal{B}^{p}(H, \underline{s}))|$ is polynomial in n and thus this is not too severe of a problem. Denote the vertex P^* as the minimizer in (8).

c) Find an I^* such that (I^*, P^*) is a vertex of $\mathcal{B}^{i,p}(H, \underline{s})$ and let $\underline{u}^* = \mu(I^*)$ be the estimated codeword.

Performing the the projection of a polytope, as in a), was originally addressed with Fourier-Motzkin elimination [14, section 2.8] and is usually a computationally difficult task. However, we can here leverage (2) to use polytope projection algorithms; these algorithms have low complexity under conditions that our problem satisfies. More explicitly, recent developments [15, section 3],[16] in the optimization literature have illustrated polytope projection algorithms that are *linear in the number of vertices or halfspaces of the polytope projection.* Since the polytope projection $\mathcal{B}^{p}(H, \underline{s})$ has a polynomial number of vertices, these algorithms directly apply to give polynomial-complexity solutions. Finally, we can perform c) using a single linear program [14]. It thus follows that part 2) of the difficulties in performing (7) can be alleviated.

We next consider part 1) of the difficulties in performing (7). This problem remains since ML decoding for linear codes is in general NP-hard. We next introduce a relaxed polytope $\tilde{\mathcal{B}}^{i,p}(H, \underline{s})$ that can be efficiently represented. For low-density parity check codes (LDPCs), $\tilde{\mathcal{B}}^{i,p}(H, \underline{s})$ has a projection vertex count $\left|\mathcal{V}\left(\tilde{\mathcal{B}}^{p}(H, \underline{s})\right)\right|$ that is polynomial in n. This implies our aforementioned approach has polynomial complexity.

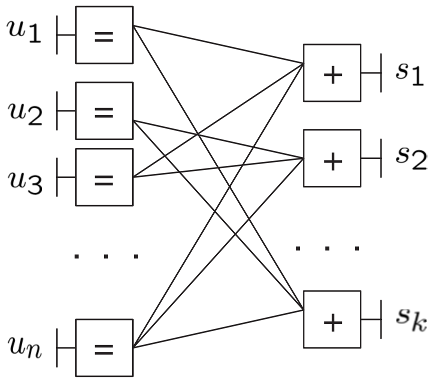

Figure 1: Normal Syndrome-Former Encoding Graph

3.2 A Polynomial Complexity Continuous Relaxation

We now restrict our attention to low density linear codes over $\mathbb{F}_{2^m}$. A linear code H has low density if there exists a constant d independent of n such that the number δ_j of nonzero elements in H_j satisfies $\delta_j \leq d$ for all j.

For a linear code H, each local constraint is a smaller linear code. Figure 1 illustrates a normal graph representation [9], where codeword symbols are associated with edges and constraint codes are associated with nodes. The jth node with a '+' sign is a single parity check code connected to one syndrome symbol s_j and a set $N(j)$ of δ_j adjacent variable nodes. This parity check enforces the constraint that s_j and the symbols in $N(j)$ must sum to 0 (over $\mathbb{F}_{2^m}$). Each node with an '=' sign is a repetition code enforcing the constraint that all symbols lying on its adjacent edges must be equal. It thus follows that

$$\begin{aligned} \mathrm{Co}(H, \underline{s}) &= \bigcap_{j=1}^{n-k} \left\{ \underline{u} \mid \underline{u}_{|N(j)} \in \mathrm{Co}(H_j, s_j) \right\} \\ \Rightarrow \mathcal{I}(H, \underline{s}) &= \bigcap_{j=1}^{n-k} \mathcal{I}(H_j, s_j), \qquad (9) \\ \text{where } \mathcal{I}(H_j, s_j) &\triangleq \left\{ I \mid \mu^r(I)_{|N(j)} \in \mathrm{Co}(H_j, s_j) \right\}. \end{aligned}$$

Since $\mathcal{I}(H, \underline{s})$ can be represented as (9), it is natural to consider the relaxed polytope, analogous to [8, section 4] (which originated from the LP relaxations of Feldman et al. for channel coding [17, 18, 19]),

$$\begin{aligned} \tilde{\mathcal{B}}_j^{i,p}(H_j, s_j) &= \left\{ (I, P) \ s.t. \ I \in CH(\mathcal{I}(H_j, \underline{s}_j)), \ P = \tau(I) \right\}, j = 1, \ldots, n-k \\ \tilde{\mathcal{B}}^{i,p}(H, \underline{s}) &= \bigcap_{j=1}^{n-k} \tilde{\mathcal{B}}_j^{i,p}(H_j, s_j). \end{aligned}$$

Since the degrees of the checks are bounded, $CH(\mathrm{Co}(H_j, s_j))$ can be compactly represented in terms of the $Q^{\delta_j - 1} \leq Q^{d-1}$ configurations consistent with check j. Thus $\tilde{\mathcal{B}}_j^{i,p}(H, \underline{s})$ can be represented with fixed complexity in terms of n. It therefore follows that

$$\mathcal{H}\left(\tilde{\mathcal{B}}^{i,p}(H, \underline{s})\right) = O(n). \qquad (10)$$

Since this is a relaxation, for any graph other than a tree, $\mathcal{V}\left(\tilde{\mathcal{B}}^{i,p}(H,\underline{s})\right)$ includes fractional vectors. As in the binary case [17, 18, 8], the polytope has the property that

$$v = (I,P) \in \mathcal{V}\left(\tilde{\mathcal{B}}^{i,p}(H,\underline{s})\right) \quad \text{is integral } \Rightarrow \underline{u} = \mu(I) \in \mathrm{Co}\,(H,s)\,. \tag{11}$$

As a consequence of (11) along with the application of the vertex enumeration algorithm in a),b), and c) to the relaxed polytope $\tilde{\mathcal{B}}^{i,p}(H,\underline{s})$, we have a relaxed minimum-entropy decoder with an analogous property to the ***ML-certificate property*** in [17, 18, 8]. In particular, we have the ***ME-certificate property****: if an integral solution is found, it is guaranteed to be the minimum-entropy solution.*

Since $\tilde{\mathcal{B}}^{i,p}(H,\underline{s})$ has vertices corresponding to both true elements of $\mathrm{Co}\,(H,s)$ and 'pseudocodewords' [18, 17], it is important to understand the impact these pseudocodewords have on $\mathcal{V}\left(\tilde{\mathcal{B}}^{p}(H,\underline{s})\right)$, and whether or not $\mathcal{H}\left(\tilde{\mathcal{B}}^{p}(H,\underline{s})\right)$ and $\left|\mathcal{V}\left(\tilde{\mathcal{B}}^{p}(H,\underline{s})\right)\right|$ are polynomial in n. The following Lemma addresses this question.

Lemma 3.1. *Both* $\left|\mathcal{V}\left(\tilde{\mathcal{B}}^{p}(H,\underline{s})\right)\right|$ *and* $\mathcal{H}\left(\tilde{\mathcal{B}}^{p}(H,\underline{s})\right)$ *are polynomial in* n.

Proof details appear in the appendix.

4 The Binary Case

In the binary case ($\mathcal{U} = \{0,1\}$), the concave minimization problem for minimum-entropy decoding corresponds to

$$\begin{array}{ll} \min & h_b(p) \\ s.t. & \underline{u} \in CH\,(\mathrm{Co}\,(H,\underline{s})) \\ & p = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} u_i. \end{array}$$

Since $|\mathcal{V}\,(\mathcal{B}^p(H,\underline{s}))| \leq 2$, we may instead perform the following pair of optimization problems to obtain the two vertices:

LP-MIN		**LP-MAX**	
min	$\frac{1}{n}\sum_{i=1}^{n} u_i$	max	$\frac{1}{n}\sum_{i=1}^{n} u_i$
s.t.	$\underline{u} \in CH\,(\mathrm{Co}\,(H,\underline{s}))$	*s.t.*	$\underline{u} \in CH\,(\mathrm{Co}\,(H,\underline{s}))$

and let $\underline{u}^{\min,*}, \underline{u}^{\max,*}$ be the optimal solutions to **LP-MIN** and **LP-MAX**, respectively. From there we may take

$$\underline{u}^* = \arg_{\underline{u} \in \{\underline{u}^{\min,*},\underline{u}^{\max,*}\}} \min H(P_{\underline{u}})$$

and arrive at the same optimal solution.

4.1 A Figure of Merit for Binary Linear Codes under Minimum-Entropy Decoding

Notice that **LP-MIN** is equivalent to ML-decoding where $\Pr(0) < \Pr(1)$ and is thus NP-hard. Likewise, **LP-MAX** corresponds to ML-decoding where $\Pr(1) < \Pr(0)$ and is also NP-hard.

The difficulty in these two problems manifests itself in the inability to efficiently represent the polytope $CH\left(\mathrm{Co}\left(H,\underline{s}\right)\right)$. However, this can be relaxed while maintaining provably good performance by using the relaxed polytope $\tilde{\mathcal{B}}^{i,p}(H,\underline{s})$ in **LP-MIN** and **LP-MAX**.

For good performance under such a universal decoder, we must construct codes that simultaneously work well for **LP-MIN** and **LP-MAX** or any relaxations thereof. This leads to a new performance metric for a binary linear code, the 'minimum-maximum distance', given by

$$\min_{\underline{u}\in\mathrm{Co}(H,\underline{0}),\underline{u}\neq 0} \min\left(w_H(\underline{u}), w_H(\underline{1}\oplus\underline{u})\right),$$

where $w_H(\cdot)$ is the Hamming weight. We illustrate the motivation for using the minimum-maximum distance in code design using the following example. Consider any linear code H for which the all one's vector, $\underline{1}$ is a member of $\mathrm{Co}\left(H,\underline{0}\right)$. Then the minimum-entropy decoder has probability of error equal to $\frac{1}{2}$ by the following argument. For any $\underline{u}\in\mathrm{Co}\left(H,\underline{s}\right)$, $\underline{u}\oplus\underline{1}\in\mathrm{Co}\left(H,\underline{s}\right)$. Further, $H\left(P_{\underline{u}}\right)=H\left(P_{\underline{u}\oplus\underline{1}}\right)$. Thus $\underline{u}$ and $\underline{u}\oplus\underline{1}$ are indistinguishable to a minimum-entropy decoder. Note that the 'min-max' distance of any such linear code H is 0, which captures this undesirable effect.

Using properly constructed expander codes as discussed in [19], we may select component codes of the expander code to have good min-max distance so that the aggregate code has good min-max distance. We can then show that the performance of the above mentioned algorithm is provably good (has a positive error exponent). In fact, from [1], it follows that if we take as our component code of the expander code to be a random binary linear code chosen uniformly, then with *exponentially* high probability the distance spectrum of the code will satisfy our 'minimum-maximum distance' criterion.

We also note that for the binary case, we need not perform a Feldman-style LP. By constructing the same expander codes as mentioned above, iterative bit-flipping algorithms with provably good performance due to [20] can naturally be extended so that one algorithm searches in a manner analogous to **LP-MIN** and the other to **LP-MAX**. Thus in this setting we also get provably good performance.

5 Extensions and Conclusions

In this paper we bring minimum-entropy universal decoding from the realm of proof technique to the realm of practice. We do this by exploiting the fact that the number of types is polynomial in the block length and applying recent results in the optimization literature on polytope projection algorithms. We consider polynomial complexity relaxations using LDPC codes. In the binary case, we show that that by using proper codes with this algorithm, the performance is provably good. Further directions that we plan to pursue include

- As in the case of LP decoding for relaxations to ML decoding,

  - the quantification of the error exponent loss under this sub-optimal decoding

  - proof of attainability of all achievable rates under this suboptimal decoder.

- understanding how iterative decoding relates to this optimization problem. Considering how the min-sum algorithm operates on the polytope discussed here (see [21]), it would be interesting to see if there is an iterative equivalent of performing the concave minimization solver.

We note that although this discussion was limited to block source compression, these techniques directly apply to distributed source coding problems, such as Slepian-Wolf [22]. The discussion of this methodology for Slepian-Wolf is discussed in [23]. By noting that 'maximum-mutual information' decoding in the universal channel decoding domain [11] - as a replacement to 'minimum-entropy' decoding in this domain - also exploits the method of types, the techniques discussed here directly apply to universal channel decoding for discrete memoryless channels. In particular, section 4 directly applies to decoding on a binary symmetric channel with unknown crossover probability by application of the coset-leader approach to channel decoding.

Acknowledgement

The authors thank Jon Feldman and Ralf Koetter for their helpful discussions and comments on LP decoding.

References

[1] A. Barg and G. D. Forney, "Random codes: Minimum distances and error exponents," *IEEE Transactions on Information Theory*, pp. 2568–2573, 2002.

[2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting codes and decoding: Turbo codes," *Proc. IEEE International Communications Conference*, 1993.

[3] S. Chung, Jr. G.D. Forney, T.J. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 db of the Shannon limit," *IEEE Communications Letters*, vol. 5, no. 2, pp. 58–60, February 2001.

[4] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *IEEE Data Compression Conference*, April 2002, pp. 252–261.

[5] D. Schonberg, S. S. Pradhan, and K. Ramchandran, "LDPC codes can approach the Slepian-Wolf bound for general binary sources," in *Proceedings of the 40th Allerton Conference on Communication, Control and Computing*, October 2002.

[6] A. Liveris, Z. Xiong, and C. Georghiades, "Distributed compression of binary sources using conventional parallel and serial concatenated convolutional codes," in *Proc. IEEE DCC*, Brest, France, March 2003, pp. 193–202.

[7] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes," *IEEE Communications Letters*, vol. 5, pp. 417–419, October 2001.

[8] T. P. Coleman, A. H. Lee, M. Médard, and M. Effros, "On some new approaches to practical Slepian-Wolf compression inspired by channel coding," in *IEEE Data Compression Conference*, Snowbird, Utah, March 23 – March 25 2004.

[9] G. D. Forney, "Codes on graphs: Normal realizations," *IEEE Transactions on Information Theory*, pp. 101–112, 2001.

[10] I. Csiszár and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Elsevier Science and Technology, 1982.

[11] I. Csiszár, "The method of types," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2205–2523, 1998.

[12] I. Csiszár, "Linear codes for sources and source networks: Error exponents, universal coding," *IEEE Transactions on Information Theory*, vol. 28, no. 4, pp. 585–592, 1982.

[13] R. Horst and H. Tuy, *Global Optimization: Deterministic Approaches*, Springer Verlag, Berlin, Germany, third revised and enlarged edition edition, 1996.

[14] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*, Athena Scientific, Belmont, MA, 1997.

[15] J. Ponce, S. Sullivan, A. Sudsang, J. Boissonnat, and J. Merlet, "On computing four-finger equilibrium and force-closure grasps of polyhedral objects," *International Journal of Robotics Research*, vol. 16, no. 1, pp. 11–35, 1997.

[16] C. N. Jones, E. C. Kerrigan, and J. M. Maciejowski, "Equality set projection: A new algorithm for the projection of polytopes in halfspace representation," Tech. Rep. CUED/F-INFENG/TR.463, Cambridge University Engineering Department, March 2004.

[17] J. Feldman, M. Wainwright, and D. R. Karger, "Using linear programming to decode linear codes," *Proceedings of Conference on Information Sciences and Systems, The John Hopkins University*, March 2003.

[18] J. Feldman, *Decoding Error-Correcting Codes via Linear Programming*, PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, September 2003.

[19] J. Feldman, T. Malkin, C. Stein, R. A. Servedio, and M. J. Wainwright, "LP decoding corrects a constant fraction of errors," in *IEEE International Symposium on Information Theory*, Chicago, Ill, June 27 – July 2 2004.

[20] A. Barg and G. Zémor, "Error exponents of expander codes," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1725–1729, 2002.

[21] R. Koetter and P. O. Vontobel, "Graph-covers and iterative decoding of finite length codes," *Proceedings of Turbo conference, Brest*, 2003.

[22] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[23] T. P. Coleman, M. Medard, and M. Effros, "Towards bridging the gap between theory and practice for the slepian-wolf problem," in *Special Session on 'Distributed Source and Joint Source-Channel Coding', 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, March 2005.

[24] G. M. Ziegler, *Lectures on Polytopes*, Springer-Verlag, New York, NY, 1994.

A Proof of Lemma 3.1

Proof. Consider the polytope

$$\tilde{\mathcal{B}}^{i,p} = \{x \in \mathbb{R}^N \mid Ax \leq b\}, \tag{12}$$

where $M = \mathcal{H}\left(\tilde{\mathcal{B}}^{i,p}\right)$ is the number of rows in A. Any subspace L and an affine set M parallel to L can be expressed as

$$\begin{aligned} L &= \{x \in \mathbb{R}^N \mid Cx = 0\}, \quad \text{where } \dim L \triangleq N - rk(C) \\ M &= L + a = \{x \in \mathbb{R}^N \mid Cx = d\}, \quad \text{where } Ca = d, \;\; \dim M \triangleq \dim L = N - rk(C). \end{aligned}$$

For an arbitrary set $\mathcal{S} \subseteq \mathbb{R}^N$, its affine hull, given by

$$\text{aff}\,(\mathcal{S}) = \left\{ x \in \mathbb{R}^N \mid x = \sum_{i=1}^{N} \lambda_i v_i, v_i \in \mathcal{S}, \sum_{i=1}^{N} \lambda_i = 1 \right\}, \tag{13}$$

is an affine set and thus can be expressed as $\text{aff}\,(\mathcal{S}) = \{x \in \mathbb{R}^N \mid Cx = d\}$. The dimension of a polytope $\tilde{\mathcal{B}}^{i,p}$ is given by $\dim \tilde{\mathcal{B}}^{i,p} \triangleq \dim \text{aff}\left(\tilde{\mathcal{B}}^{i,p}\right) = N - rk(C)$. Let

$$\begin{aligned} \hat{\tilde{\mathcal{B}}}^{i,p} &= \left\{ (z, w) \in \mathbb{R}^N \times \mathbb{R}^{M-N+rk(C)} \mid \begin{bmatrix} A & \underbrace{0}_{M-N+rk(C)} \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix} \leq b \right\} \\ \Rightarrow \text{aff}\left(\hat{\tilde{\mathcal{B}}}^{i,p}\right) &= \left\{ (z, w) \in \mathbb{R}^N \times \mathbb{R}^{M-N+rk(C)} \mid \begin{bmatrix} C & \underbrace{0}_{M-N+rk(C)} \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix} = d \right\} \\ \Rightarrow \dim \hat{\tilde{\mathcal{B}}}^{i,p} &= \dim \text{aff}\left(\hat{\tilde{\mathcal{B}}}^{i,p}\right) = N + M - N + rk(C) - rk\left(\begin{bmatrix} C & \underbrace{0}_{M-N+rk(C)} \end{bmatrix} \right) \\ &= M \end{aligned}$$

Note that $\dim \hat{\tilde{\mathcal{B}}}^{i,p} = \mathcal{H}\left(\hat{\tilde{\mathcal{B}}}^{i,p}\right) = M$ and the projection of $\hat{\tilde{\mathcal{B}}}^{i,p}$ onto any set of indices $\mathcal{S} \subseteq \{1, \ldots, N\}$ equals the projection of $\tilde{\mathcal{B}}^{i,p}$ onto $\mathcal{S}$.

In general, the P-dimensional projection $\tilde{\mathcal{B}}^{p}$ of a D-dimensional polyhedron $\tilde{\mathcal{B}}^{i,p}$ where $\mathcal{H}\left(\tilde{\mathcal{B}}^{i,p}\right) = H$ satisfies [24]

$$\mathcal{H}\left(\tilde{\mathcal{B}}^{p}\right) \leq \binom{H}{D - P + 1}.$$

For $\hat{\tilde{\mathcal{B}}}^{i,p}$ we have $D = H = \mathcal{H}\left(\hat{\tilde{\mathcal{B}}}^{i,p}\right) = \mathcal{H}\left(\tilde{\mathcal{B}}^{i,p}\right)$ and so from (3) and (10) it follows that

$$\mathcal{H}\left(\tilde{\mathcal{B}}^{p}\right) \leq \binom{\mathcal{H}\left(\tilde{\mathcal{B}}^{i,p}\right)}{\mathcal{H}\left(\tilde{\mathcal{B}}^{i,p}\right) - |\mathcal{U}| + 1} = O\left(\mathcal{H}\left(\tilde{\mathcal{B}}^{i,p}\right)^{|\mathcal{U}|}\right) = O\left(n^{|\mathcal{U}|}\right). \tag{14}$$

Since any $|\mathcal{U}|$-dimensional polytope $\tilde{\mathcal{B}}^{p}$ has at most $\binom{\mathcal{H}(\tilde{\mathcal{B}}^{p})}{|\mathcal{U}|}$ vertices [14], from (3) and (14) it follows that

$$\left|\mathcal{V}\left(\tilde{\mathcal{B}}^{p}\right)\right| = O\left(\mathcal{H}\left(\tilde{\mathcal{B}}^{p}\right)^{|\mathcal{U}|}\right) = O\left(n^{|\mathcal{U}|^2}\right).$$

□