

Practical Universal Decoding for Combined Routing and Compression in Network Coding

Todd P. Coleman

Department of EECS

Massachusetts Institute of Technology
Cambridge, MA 02139

colemant@mit.edu

Muriel Médard

Department of EECS

Massachusetts Institute of Technology
Cambridge, MA 02139

medard@mit.edu

Michelle Effros

Department of Electrical Engineering

California Institute of Technology
Pasadena, CA 91125

effros@caltech.edu

Abstract—Minimum-entropy decoding is a universal decoding algorithm used in decoding block compression of discrete memoryless sources as their multiterminal counterparts, such as the Slepian-Wolf problem. It has recently been shown that such an algorithm can be applied for combined distributed compression and distributed routing in a randomized distributed network coding framework. The ‘method of types’ shows that there exist linear codes that when applied with such an algorithm, can attain the same error exponent as that of a maximum-likelihood decoder. Owing to the algorithm being generally NP-hard, the traditional rationale for discussing this technique has been mostly theoretical pursuit. Here we discuss practical approximation algorithms to minimum-entropy decoding by using ideas from linear programming. We exploit two main facts. First, the ‘method of types’ shows that the number of distinct types is polynomial in the block length n . Second, recent results in the optimization literature have illustrated polytope projection algorithms whose complexity is a function of the number of vertices of the polytope projection. Combining these two ideas, we leverage recent results on linear programming as a relaxation for error correcting codes to construct polynomial complexity algorithms for this setting.

I. INTRODUCTION

Distributed compression of correlated sources has become of interest in the research community recently because of its possible promise in efficient transmission of information where energy, computation, and communication constraints prohibit nodes from significantly cooperating with one another. The Slepian-Wolf framework [1] addresses near-lossless distributed compression and has served as a substructure in a number of distributed data dissemination strategies. For a set of M correlated discrete memoryless sources (DMS)

$(U^1, \dots, U^M) \sim P(u^1, \dots, u^M)$, the achievable rate region $\mathcal{R}[P(u^1, \dots, u^M)]$ is given by

$$\sum_{i \in S} R_i > H(U(S)|U(S^c)) \quad \forall S \subseteq \{1, 2, \dots, M\} \quad (1)$$

where $U(S) = \{U^j\}_{j \in S}$. For a set of rates $R = (R_1, \dots, R_M)$ and a block code of length n , the probability of error for a set of encoders \mathcal{C} is given by $P_{e,n}^{\mathcal{C}}(R)$. The random coding error exponent, given by

$$E_r(R) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} E_{\mathcal{C}}[-\log P_{e,n}(R)],$$

quantifies the ensemble average exponential rate of decay in error probability for all achievable rates. In [2], Csiszár showed that random linear block codes asymptotically achieve optimal performance - in terms of $\mathcal{R}[P(u^1, \dots, u^M)]$ and $E_r(R)$. Practically speaking, this problem is difficult because of the complexity in jointly decoding the M sources. On top of this, Csiszár showed [2] that the same performance can be attained universally using linear encoding: without the a priori knowledge of $P(u^1, \dots, u^M)$ at the encoder nor decoder. The universal ‘minimum-entropy’ decoder, however, is nonlinear and computationally infeasible with practical limitations. Consequently, discussions of universal decoding have been mostly confined to the realm of proofs of existence.

Recently Ho. et al [3] brought Csiszár’s framework to the realm of distributed randomized network coding by considering a situation where arbitrarily correlated sources must traverse through a network. By performing linear block operations for compression as well as linear operations for network coding, the authors showed that the aggregate linear transformation attains all achievable rates, and the same ‘minimum-entropy’ decoder provides universality.

Recently, consideration of separating distributed source coding from network coding illustrated that this is in general a suboptimal strategy [4]. Thus we can conclude that when considering combined routing and distributed data compression, the framework discussed in [3] is crucial.

After bringing forth definitions and preliminaries in section II, we construct in section III a low-complexity universal relaxation to the proposed universal decoder that has a certificate property. The basis for our approach centrally relies on tying recent results in the optimization and polyhedral theory literature on polytope projection to well-known information-theoretic results from the ‘method of types’.

II. PRELIMINARIES

Throughout this discussion we consider a discrete memoryless source (DMS) pair $(U^1, U^2) \in \mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2$ with joint probability distribution $\Pr(u)$ where $u = (u^1, u^2)$. We adhere

to the following definitions:

$$\begin{aligned}
CH(\mathcal{S}) &= \text{the convex hull of all } s \in \mathcal{S} \\
\mathcal{V}(\mathcal{B}) &= \{v \in \mathcal{B} \mid v \text{ is a vertex of the polytope } \mathcal{B}\} \\
\mathcal{H}(\mathcal{B}) &= \text{the number of half-spaces representing } \mathcal{B} \\
\mathcal{P}(\mathcal{U}) &= \left\{ P = (\{P_a\}_{a \in \mathcal{U}}) : P \geq \underline{0}, \sum_{a \in \mathcal{U}} P_a = 1 \right\} \\
P_{\underline{u}} &= \left(\left\{ \frac{1}{n} \sum_{i=1}^n 1_{u_i=a} \right\}_{a \in \mathcal{U}} \right) \text{ for } \underline{u} \in \mathcal{U}^n \\
\mathcal{P}_n(\mathcal{U}) &= \{P \in \mathcal{P}(\mathcal{U}) : P = P_{\underline{u}} \text{ for some } \underline{u} \in \mathcal{U}^n\}
\end{aligned} \tag{2}$$

We exploit the following property repeatedly:

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

The ‘method of types’ [2] exploits this property

$$|\mathcal{P}_n(\mathcal{U})| = \binom{n + |\mathcal{U}| - 1}{|\mathcal{U}| - 1} \tag{4a}$$

$$= O(n^{|\mathcal{U}|-1}) \tag{4b}$$

to show that **the number of types is polynomial in n** .

Here we consider the case where $r \in \{1, 2\}$, $|\mathcal{U}_r| = 2^t$ and block compression transforms $\underline{u}^r \in \mathcal{U}_r^n$ to $\underline{s}^r \in \mathcal{U}_r^{m_r}$ via

a linear code $H^r = \begin{bmatrix} -H_1^{r'} & - \\ \vdots & \\ -H_{m_r}^{r'} & - \end{bmatrix} \in \mathcal{U}_r^{m_r \times n}$ according to

$\underline{s}^r = H^r \underline{u}^r$ where algebraic operations are performed over \mathbb{F}_{2^t} . For $j \in \{1, \dots, m_r\}$ we define $N(j) \triangleq \{i \mid H_{j,i} = 1\}$ and $\delta_j = |N(j)|$. Throughout this discussion we consider achievable rates, as given by (1).

III. UNIVERSAL MINIMUM-ENTROPY DECODING

The minimum-entropy (ME) decoder uses $\{\underline{s}^1, \underline{s}^2\}$ along with $\{H^1, H^2\}$ to select $\hat{\underline{u}} = \hat{\underline{u}}^1, \hat{\underline{u}}^2$ according to:

$$\hat{\underline{u}} = \arg \min_{\{\underline{u}^r \in \text{Co}(H^r, \underline{s}^r)\}_{r=1,2}} H(P_{\underline{u}^1, \underline{u}^2}) \tag{5}$$

$$\text{where } \text{Co}(H^r, \underline{s}^r) = \{\underline{u} \mid H^r \underline{u} = \underline{s}^r\}.$$

Note that (5) is a discrete optimization problem with an exponential number of candidates. We now discuss formulating the ME decoding problem in terms of a continuous concave minimization problem. For $a = (a_1, a_2) \in \mathcal{U}_1 \times \mathcal{U}_2$, we define I_i^a to be the indicator variable for the event $(u_i^1, u_i^2) = (a_1, a_2)$. Define

$$\iota^r(I) = \sum_{a_{\bar{r}} \in \mathcal{U}_{\bar{r}}} I^{a_1, a_2} \text{ where } \bar{r} = \{1, 2\} \setminus r$$

$$\mu^r(I) = \sum_{a_r \in \mathcal{U}_r} a_r \iota^r(I)$$

$$\mathcal{I}(H^r, \underline{s}^r) = \{I \mid \mu^r(I) \in \text{Co}(H^r, \underline{s}^r)\} \tag{6}$$

$$\mathcal{B} = \{I \mid \iota^r(I) \in CH(\mathcal{I}(H^r, \underline{s}^r)), r = 1, 2\} \tag{7}$$

Note that \mathcal{B} represents the convex hull of all indicators I that are consistent with members of $\text{Co}(H^1, \underline{s}^1) \times \text{Co}(H^2, \underline{s}^2)$.

To associate the joint type of $(u^1, u^2) \in \text{Co}(H^1, \underline{s}^1) \times \text{Co}(H^2, \underline{s}^2)$ through the indicator variable I we construct the linear mapping

$$P = \tau(I), \text{ where } P(a) = \tau_a(I) = \frac{1}{n} \sum_{i=1}^n I_i^a$$

and note that for each $I \in \mathcal{V}(\mathcal{B})$, $P = \tau(I)$ is the joint type given by (2) associated with $(\underline{u}^1, \underline{u}^2) = (\mu^1(I), \mu^2(I))$.

Since $H(P)$ is *strictly concave* in P , and since for concave minimization over a polytope an optimal solution lies in $\mathcal{V}(\mathcal{B})$ [5], we can perform (5) in the continuous domain:

$$\min H(P) \tag{8a}$$

$$s.t. (I, P) \in \mathcal{B}^{i,p} \tag{8b}$$

$$\text{where } \mathcal{B}^{i,p} = \{(I, P) \mid I \in \mathcal{B}, P = \tau(I)\} \tag{8c}$$

and take $(\underline{u}^{1*}, \underline{u}^{2*}) = (\mu^1(I^*), \mu^2(I^*))$ where (I^*, P^*) is an optimal solution to (8).

This formulation presents two major problems:

- 1) By virtue of ML-decoding for linear codes generally being NP-hard, the best bound on $\mathcal{H}(\mathcal{B})$ (and thus $\mathcal{H}(\mathcal{B}^{i,p})$) is $O(2^n)$.
- 2) Along with not knowing how to efficiently represent $\mathcal{B}^{i,p}$, another problem manifests itself in (8): $|\mathcal{V}(\mathcal{B}^{i,p})| = O(2^n)$ and ‘concave minimization over a polytope’ is NP-hard [5], generally requiring to visit every $v \in \mathcal{V}(\mathcal{B}^{i,p})$.

To avoid problem 2), we note that although $|\mathcal{V}(\mathcal{B}^{i,p})| = O(2^n)$, from (4) we have that $|\mathcal{P}_n(\mathcal{U})| = O(n^{|\mathcal{U}|-1})$. We thus consider the following strategy:

- a) Project $\mathcal{B}^{i,p}$ onto $\mathcal{B}^p = \{P \mid (I, P) \in \mathcal{B}^{i,p} \text{ for some } I\}$.
- b) Perform the minimization

$$\min H(P) \tag{9a}$$

$$s.t. P \in \mathcal{B}^p(H, \underline{s}). \tag{9b}$$

Note that from (4) the worst-case scenario of enumerating through each $v \in \mathcal{V}(\mathcal{B}^p)$ involves a polynomial number of visits. Denote the vertex P^* as the minimizer in (9).

- 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 projection of a polytope, as in a), was originally addressed with Fourier-Motzkin elimination [6, section 2.8] and is in general extremely computationally complex. However, in this situation, \mathcal{B}^p is fixed dimension, invariant of n , and (4) suggests using special-purpose polytope projection algorithms that are low-complexity in this case. Recent developments [7, Sec. 3], [8] in the optimization literature have illustrated polytope projection algorithms whose complexity is only a *linear* function of $|\mathcal{V}(\mathcal{B}^p)|$ or $\mathcal{H}(\mathcal{B}^p)$. Instantiation of a single LP [6] addresses c).

To avoid problem 1) we consider a relaxed polytope $\tilde{\mathcal{B}}$, in the spirit of LP relaxations for channel coding [9], to replace \mathcal{B} . For a linear code H^r , each local constraint is a smaller

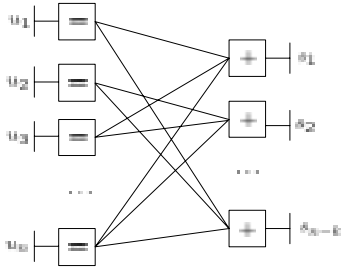


Fig. 1. Normal Syndrome-Former Encoding Graph

linear code and

$$\begin{aligned} \text{Co}(H^r, \underline{s}^r) &= \bigcap_{j=1}^{m_r} \left\{ \underline{u} \mid \underline{u}_{|N(j)} \in \text{Co}(H_j^r, s_j^r) \right\} \\ \Rightarrow \mathcal{I}(H^r, \underline{s}^r) &= \bigcap_{j=1}^{m_r} \mathcal{I}(H_j^r, s_j^r), \end{aligned} \quad (10)$$

$$\text{where } \mathcal{I}(H_j^r, s_j^r) = \left\{ I \mid \mu^r(I)_{|N(j)} \in \text{Co}(H_j^r, s_j^r) \right\}$$

Figure 1 illustrates a normal graph representation [10], where codeword symbols are associated with edges and constraint codes are associated with nodes. The j th node with a ‘+’ sign is a single parity check code connected to one syndrome symbol s_j and a set of δ_j adjacent variable nodes, given by $N(j)$. It enforces the constraint that the symbols in $N(j)$ along with s_j must sum (over \mathbb{F}_{2^m}) to 0. Each node with an ‘=’ sign is a repetition code enforcing the constraint that all symbols lying on its adjacent edges must be equal. Since $\mathcal{I}(H^r, \underline{s}^r)$ can be represented as (10) we consider

$$\begin{aligned} \tilde{\mathcal{B}}_j^r(H_j^r, s_j^r) &= \left\{ I \mid \iota^r(I)_{|N(j)} \in \text{CH}(\mathcal{I}(H_j^r, \underline{s}_j^r)) \right\} \quad (11) \\ \tilde{\mathcal{B}} &= \bigcap_{r=1}^2 \bigcap_{j=1}^{m_r} \tilde{\mathcal{B}}_j^r(H_j^r, s_j^r). \end{aligned}$$

Analogous to Feldman’s illustration [9, sec. 5.5] for channel decoding of binary linear codes, it can be shown that $\tilde{\mathcal{B}}_j^r(H_j^r, s_j^r)$ can be compactly represented:

Lemma 3.1: The polytope $\tilde{\mathcal{B}}_j^r(H_j^r, s_j^r)$ can be represented as

$$\tilde{\mathcal{B}}_j^r(H_j^r, s_j^r) = \left\{ I : (I, \alpha, z) \in \tilde{\mathcal{W}}_j^r(H_j^r, s_j^r) \right\}$$

where $|\{\alpha_k, z_l\}| = O(n^{|\mathcal{U}|})$ and $\mathcal{H}(\tilde{\mathcal{W}}_j^r(H_j^r, s_j^r)) = O(n^{|\mathcal{U}|})$.

The proof details follow in the appendix. For any graphical representation as denoted in Figure 1 other than a tree, however, $\mathcal{V}(\tilde{\mathcal{B}})$ includes fractional entries, termed ‘pseudocodewords’ [11]. Nonetheless it can be shown that

$$I \in \mathcal{V}(\tilde{\mathcal{B}}) \text{ is integral} \Rightarrow \{\mu^r(I) \in \text{Co}(H^r, \underline{s}^r)\}_{r=1,2}. \quad (12)$$

Our aggregate strategy using $\tilde{\mathcal{B}}$ considers performing a relaxed universal decoder by performing steps a)-c) replacing

$$\begin{aligned} \mathcal{B}^{i,p} &\text{ with } \tilde{\mathcal{W}}^{i,p} = \{(I, \alpha, z, P) \mid (I, \alpha, z) \in \tilde{\mathcal{W}}, P = \tau(I)\}, \text{ and} \\ \mathcal{B}^p &\text{ with } \tilde{\mathcal{B}}^p = \{P \mid (I, \alpha, z, P) \in \tilde{\mathcal{W}}^{i,p} \text{ for some } (I, \alpha, z)\} \end{aligned}$$

where $\tilde{\mathcal{W}} = \bigcap_{r=1}^2 \bigcap_{j=1}^{m_r} \tilde{\mathcal{W}}_j^r(H_j^r, s_j^r)$. Because of the fractional ‘pseudocodewords’ in $\mathcal{V}(\tilde{\mathcal{B}})$, we must check that $|\mathcal{V}(\tilde{\mathcal{B}}^p)|$ is polynomial in n in order to guarantee a polynomial complexity decoder.

Because $\mathcal{H}(\tilde{\mathcal{W}}) = O(n^{|\mathcal{U}|+1})$, it follows that

$$\mathcal{H}(\tilde{\mathcal{W}}^{i,p}) = \mathcal{H}(\tilde{\mathcal{W}}) + 2|\mathcal{U}| = O(n^{|\mathcal{U}|+1}).$$

We now note the following lemma:

Lemma 3.2:

$$\mathcal{H}(\tilde{\mathcal{B}}^p) = O\left(\mathcal{H}(\tilde{\mathcal{W}}^{i,p})^{|\mathcal{U}|}\right).$$

Proof: Note that we are projecting the d -dimensional polytope $\tilde{\mathcal{B}}^{i,p} \subseteq \mathbb{R}^d$ (where $d = O(n^{|\mathcal{U}|})$) onto $\tilde{\mathcal{B}}^p \subseteq \mathbb{R}^{|\mathcal{U}|}$. Since $\mathcal{H}(\tilde{\mathcal{W}}^{i,p}) = O(n^{|\mathcal{U}|+1})$, we can construct an auxiliary polyhedron $\hat{\mathcal{W}}^{i,p} \subseteq \mathbb{R}^{\mathcal{H}(\tilde{\mathcal{W}}^{i,p})}$ by introducing auxiliary unconstrained variables

$$\{f_j\}, k = 1, \dots, T = \mathcal{H}(\tilde{\mathcal{B}}^{i,p}) - d$$

and adding f_j to the k th halfspace inequality representing $\tilde{\mathcal{W}}^{i,p}$. We now consider projecting $\hat{\mathcal{W}}^{i,p}$ onto $\tilde{\mathcal{B}}^p$. We note that in general, projecting a M -dimensional polyhedron \mathcal{B} with $\mathcal{H}(\mathcal{B}) = N$ onto a $|\mathcal{U}|$ dimensional polyhedron $\tilde{\mathcal{B}}^p$ requires at most

$$\binom{N}{M - |\mathcal{U}| + 1}$$

halfspaces to represent $\tilde{\mathcal{B}}^p$. For our case we have $N = M = \mathcal{H}(\hat{\mathcal{W}}^{i,p}) = \mathcal{H}(\tilde{\mathcal{W}}^{i,p})$ and thus we have from 3 that

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

Since any d -dimensional polytope \mathcal{B} has at most $\binom{\mathcal{H}(\mathcal{B})}{d}$ vertices [6], we have from (3) that

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

which is polynomial in n .

Thus both $|\mathcal{V}(\tilde{\mathcal{B}}^p)|$ and $\mathcal{H}(\tilde{\mathcal{B}}^p)$ are polynomial in n , and thus either of the approaches in [7, Sec. 3] or [8] leads to a polynomial complexity solution. By virtue of (12) we have therefore constructed a polynomial complexity universal decoder that has the **ME-certificate property**: if an integral solution is found, it is the minimum-entropy solution.

IV. CONCLUSION

In this discussion we have considered low-complexity approximations to minimum-entropy universal decoding that have a certificate property. These results rely on exploiting well-known results from the ‘method of types’ and noting how this relates to low-complexity polytope projection algorithms. By using appropriately constructed graphical codes, such as

low-density parity check codes [10], and because of the linear programming decoder's strong connection with the min-sum iterative decoder [12], we conjecture this algorithm will have good decoding properties for appropriately constructed irregular codes. Because of the use in the network coding context, future work should consider successfully implementing this algorithm on general linear codes, and/or considering ways to construct randomized distributed network coding strategies that allow for good decoding with the aforementioned low-complexity decoder.

REFERENCES

- [1] 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.
- [2] 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.
- [3] T. Ho, M. Médard, M. Effros, and R. Koetter, "Network coding for correlated sources," in *Proceedings of CISS*, 2004.
- [4] A. Ramamoorthy, K. Jain, P. A. Chou, and M. Effros, "Separating distributed source coding from network coding," in *42nd Allerton Conference on Communication, Control, and Computing*, 2004.
- [5] R. Horst and H. Tuy, *Global Optimization: Deterministic Approaches*, Springer Verlag, Berlin, Germany, third revised and enlarged edition, 1996.
- [6] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*, Athena Scientific, Belmont, MA, 1997.
- [7] 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.
- [8] 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.
- [9] J. Feldman, *Decoding Error-Correcting Codes via Linear Programming*, PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, September 2003.
- [10] G. D. Forney, "Codes on graphs: Normal realizations," *IEEE Transactions on Information Theory*, pp. 101–112, 2001.
- [11] G. Forney, R. Koetter, J. Kschischang, and A. Reznik, "On the effective weights of pseudocodewords for codes defined on graphs with cycles," *Codes, Systems and graphical models*, pp. 101–112, 2001.
- [12] R. Koetter and P. O. Vontobel, "Graph-covers and iterative decoding of finite length codes," *Proceedings of Turbo conference, Brest*, 2003.

APPENDIX

Proof of Lemma 3.1:

For a check j with syndrome symbol s_j we let T_j be the set of all possible weighted types of local codewords consistent with check j and syndrome s_j . More explicitly, define:

$$T_j = \{T = |N(j)|P_{\underline{u}} : \underline{u} \in \text{Co}(H_j^r, s_j^r)\}.$$

Our new polytope has three sets of variables:

- For all $i \in \{1, \dots, n\}$, $a \in \mathcal{U}_r$ we have indicators $I_i^a \in [0, 1]$ that represent whether or not $u_i^r = a$.
- For all $T \in T_j$, we have a variable $\alpha_{j,T} \in [0, 1]$ that represents the contribution of local codewords of weighted type T .
- For all $T \in T_j$, $a \in \mathcal{U}$ and $i \in N(j)$, we have a variable $z_{i,j,T,a} \in [0, \alpha_{j,T}]$ that indicates the portion of $\mu(I_i)$ assigned to local codewords of weighted type T .

Using these variables, we have the following constraint set:

$$\forall i \in N(j), a, \quad I_i^a = \sum_{T \in T_j} z_{i,j,T,a} \quad (13)$$

$$\sum_{T \in T_j} \alpha_{j,T} = 1 \quad (14)$$

$$\forall a, T \in T_j, \quad \sum_{i \in N(j)} z_{i,j,T,a} = T(a)\alpha_{j,T} \quad (15)$$

$$\forall i \in N(j), a, \quad 0 \leq I_i^a \leq 1 \quad (16)$$

$$\forall T \in T_j, \quad 0 \leq \alpha_{j,T} \leq 1 \quad (17)$$

$$\forall a, i \in N(j), P \in T_j, \quad 0 \leq z_{i,j,T,a} \leq \alpha_{j,T} \quad (18)$$

Let $\tilde{\mathcal{W}}_j^r(H_j^r, s_j^r)$ be the set of points (I, α, z) such that all the above constraints hold. Since there are $|N(j)||\mathcal{U}|$ variables I_i^a , $|\mathcal{P}_{|N(j)|}(|\mathcal{U}|)$ variables $\alpha_{j,T}$ and $|\mathcal{U}||N(j)||\mathcal{P}_{|N(j)|}(|\mathcal{U}|)$ variables $z_{i,j,T,a}$, there are $O(|N(j)| + |\mathcal{P}_{|N(j)|}(|\mathcal{U}|) + |N(j)||\mathcal{P}_{|N(j)|}(|\mathcal{U}|))$ variables. Since in general $|N(j)| \leq n$, we have that there $O(n^{|\mathcal{U}|})$ variables. By similar arguments, the constraints involving $z_{i,j,T,a}$ dominate the characterization of $\mathcal{H}(\tilde{\mathcal{W}}_j^r(H_j^r, s_j^r))$ and thus $\mathcal{H}(\tilde{\mathcal{W}}_j^r(H_j^r, s_j^r)) = O(n^{|\mathcal{U}|})$. Let $\tilde{\mathcal{Q}}_j^r(H_j^r, s_j^r)$ represent the projection of $\tilde{\mathcal{W}}_j^r(H_j^r, s_j^r)$ onto the I variables, i.e.,

$$\tilde{\mathcal{Q}}_j^r(H_j^r, s_j^r) = \{I : (I, \alpha, z) \in \tilde{\mathcal{W}}_j^r(H_j^r, s_j^r)\}.$$

Before proving the equivalence of $\tilde{\mathcal{Q}}_j^r(H_j^r, s_j^r)$ and $\tilde{\mathcal{B}}_j^r(H_j^r, s_j^r)$, we briefly note that the polytope $\tilde{\mathcal{B}}_j^r(H_j^r, s_j^r)$ can be expressed as the set of all I such that $(I, \omega) \in \tilde{\mathcal{Z}}_j^r(H_j^r, s_j^r)$ for some ω where $\tilde{\mathcal{Z}}_j^r(H_j^r, s_j^r)$ has the following constraints:

$$\forall I \in \mathcal{I}(H_j^r, s_j^r), \quad 0 \leq w_{j,I} \leq 1 \quad (19)$$

$$\sum_{I \in \mathcal{I}(H_j^r, s_j^r)} w_{j,I} = 1 \quad (20)$$

$$\forall i \in N(j) \quad I_i^a = \sum_{\substack{I \in \mathcal{I}(H_j^r, s_j^r), \\ I: \mu(I)=a}} w_{j,I} \quad (21)$$

This follows directly from (11).

We now show that $\tilde{\mathcal{Q}}_j^r(H_j^r, s_j^r) = \tilde{\mathcal{B}}_j^r(H_j^r, s_j^r)$. Suppose $I \in \tilde{\mathcal{B}}_j^r(H_j^r, s_j^r)$. Set variables $\{w_{j,I}\}_{I \in \mathcal{I}(H_j^r, s_j^r)}$ such that $(I, w) \in \tilde{\mathcal{Z}}_j^r(H_j^r, s_j^r)$. Set

$$\alpha_{j,T} = \sum_{\substack{I \in \mathcal{I}(H_j^r, s_j^r), \\ |N(j)|P_{\mu(I)}=T}} w_{j,I} \quad (22)$$

$$z_{i,j,T,a} = \sum_{\substack{I \in \mathcal{I}(H_j^r, s_j^r), \\ |N(j)|P_{\mu(I)}=T, \\ \mu(I)_i=a}} w_{j,I} \quad (23)$$

Note that (16),(17),(18) are satisfied. Constraint (13) is implied by (21) and (23). Constraint (14) is implied by (20) and (22).

Finally, we have

$$\begin{aligned}
\sum_{i \in N(j)} z_{i,j,T,a} &= \sum_{i \in N(j)} \sum_{\substack{I \in \mathcal{I}(H_j^r, s_j^r), \\ |N(j)|_{P_{\mu(I)}=T}, \\ \mu(I)_i=a}} w_{j,I} \quad (\text{owing to (23)}) \\
&= \sum_{\substack{I \in \mathcal{I}(H_j^r, s_j^r), \\ |N(j)|_{P_{\mu(I)}=T}}} T(a) w_{j,I} \\
&= T(a) \sum_{\substack{I \in \mathcal{I}(H_j^r, s_j^r), \\ |N(j)|_{P_{\mu(I)}=T}}} w_{j,I} \\
&= T(a) \alpha_{j,T} \quad (\text{owing to (22)})
\end{aligned}$$

which gives constraint (15). It thus follows that $\tilde{\mathcal{B}}_j^r(H_j^r, s_j^r) \subseteq \tilde{\mathcal{Q}}_j^r(H_j^r, s_j^r)$.

The following lemma (provided in [9, pp. 80-81]) will aid to prove that $\tilde{\mathcal{Q}}_j^r(H_j^r, s_j^r) \subseteq \tilde{\mathcal{B}}_j^r(H_j^r, s_j^r)$:

Lemma 1.1: Let $X = \{x_1, \dots, x_n\}$, $x_i \leq M$, and $\sum_i x_i = kM$ where k, n, M and all x_i are nonnegative integers. Then X can be expressed as the sum of sets of size k . In other words, there exists a setting of the variables $\{w_S : \mathcal{S} \subseteq \{1, \dots, n\}, |\mathcal{S}| = k\}$ to non-negative integers such that $\sum_{\mathcal{S}} w_{\mathcal{S}} = M$, and for all $i \in \{1, \dots, n\}$, $x_i = \sum_{\mathcal{S} \ni i} w_{\mathcal{S}}$

Proof: By induction on M , The base case ($M = 1$) holds because all x_i are either 0 or 1 and thus k of them are equal to 1. Let $\mathcal{S} = \{i : x_i = 1\}$ and set $w_{\mathcal{S}} = 1$.

For the induction step, we will greedily choose \mathcal{S} consisting of the indices of the k largest x_i . After increasing $w_{\mathcal{S}}$ by 1, our induction step will be complete.

Assume WLOG that $x_1 \geq x_2 \geq \dots \geq x_n$. Set $X' = (x'_1, \dots, x'_n)$ where $x'_i = x_i - 1$ for $i \leq k$ and $x'_i = x_i$ otherwise. Since $\sum_i x_i = kM$ and $x_i \leq M$, it follow sthat the largest k values x_1, \dots, x_k are at least 1. Also, we must have that $x_i \leq M - 1$ for all $i > k$. Thus $0 \leq x'_i \leq M - 1$ for all i . We also have $\sum_i x'_i = \sum_i x_i - k = (M - 1)k$. Therefore, by induction, X' can be expressed as the sum of $w'_{\mathcal{S}}$ where \mathcal{S} has size k . Set $w = w'$ and increase $w_{\{1, \dots, k\}}$ by 1. ■

Now suppose $I \in \mathcal{V}(\tilde{\mathcal{Q}}_j^r(H_j^r, s_j^r))$. Set the variables α and z such that $(I, \alpha, z) \in \mathcal{W}_j^r(H_j^r, s_j^r)$. For all $T \in T_j$, consider the set

$$X_1 = \left\{ \frac{z_{i,j,T,a}}{\alpha_{j,T}} \right\}.$$

Note that by (18) all members of X_1 are between 0 and 1. Let $\frac{1}{\beta}$ be a common divisor of the numbers in X_1 so that β is an integer. Let

$$X_2 = \left\{ \beta \frac{z_{i,j,T,a}}{\alpha_{j,T}} \right\}.$$

The set X_2 contains integers between 0 and β . By (15) we have that the sum of the elements in X_2 equals $T(a)\beta$. Thus by Lemma 1.1, the set X_2 can be expressed as the sum of sets of size $T(a)$. Set the variables $\{w_I, I \in \mathcal{I}, |N(j)|_{P_{\mu(I)}=T} = T(a)\}$ according to Lemma 1.1. Now set $w_{j,I} = \frac{\alpha_{j,T}}{\beta} w_I$, for

all $I \in \mathcal{I}, |N(j)|_{P_{\mu(I)}=T} = T(a)$. We immediately satisfy (19). By Lemma 1.1 we have:

$$z_{i,j,T,a} = \sum_{\substack{I \in \mathcal{I}(H_j^r, s_j^r), \\ |N(j)|_{P_{\mu(I)}=T}, \\ \mu(I)_i=a}} w_{j,I} \quad (24)$$

$$\alpha_{j,T} = \sum_{\substack{I \in \mathcal{I}(H_j^r, s_j^r), \\ |N(j)|_{P_{\mu(I)}=T}}} w_{j,I} \quad (25)$$

By (13) we have

$$\begin{aligned}
I_i^a &= \sum_{T \in T_j} z_{i,j,T,a} \\
&= \sum_{T \in T_j} \sum_{\substack{I \in \mathcal{I}(H_j^r, s_j^r), \\ |N(j)|_{P_{\mu(I)}=T}, \\ \mu(I)_i=a}} w_{j,I} \quad (\text{owing to (24)}) \\
&= \sum_{\substack{I \in \mathcal{I}(H_j^r, s_j^r), \\ \mu(I)_i=a}} w_{j,I}
\end{aligned}$$

which gives (21). By (14) we have

$$\begin{aligned}
1 &= \sum_{T \in T_j} \alpha_{j,T} = \sum_{T \in T_j} \sum_{\substack{I \in \mathcal{I}(H_j^r, s_j^r), \\ |N(j)|_{P_{\mu(I)}=T}}} w_{j,I} \quad (\text{owing to (25)}) \\
&= \sum_{I \in \mathcal{I}} w_{j,I}
\end{aligned}$$

which gives (20). Thus $(I, w) \in \tilde{\mathcal{Z}}_j^r(H_j^r, s_j^r)$ and so $I \in \tilde{\mathcal{B}}_j^r(H_j^r, s_j^r)$. We have shown that all $I \in \mathcal{V}(\tilde{\mathcal{Q}}_j^r(H_j^r, s_j^r))$ are contained in $\tilde{\mathcal{B}}_j^r(H_j^r, s_j^r)$ and thus $\tilde{\mathcal{Q}}_j^r(H_j^r, s_j^r) \subseteq \tilde{\mathcal{B}}_j^r(H_j^r, s_j^r)$. ■