Primal-Dual Characterizations of Jointly Optimal Transmission Rate and Scheme for Distributed Sources

Bradford D. Boyle and Steven Weber

Department of Electrical and Computer Engineering
Drexel University, Philadelphia, PA 19104
bradford@drexel.edu, sweber@coe.drexel.edu

Abstract

We consider the optimal transmission of distributed correlated discrete memoryless sources across a network with capacity constraints. We present several previously undiscussed structural properties of the set of feasible rates and transmission schemes. These properties are then applied to develop a characterization of an optimal solution and its connection to the corner points of the Slepian-Wolf rate region.

1 Introduction

A class of problems that arise in many contexts is the transmission of distributed discrete memoryless sources across a capacity-constrained network to a collection of sinks. Information theoretic characterizations of this class of problems has received much attention in recent years as a result of the development of network coding [1], and can be traced back to the seminal work of Slepian and Wolf [2]. In this paper, we consider the design problem of selecting a set of rates and a transmission scheme for a given network that are optimal with respect to known information-theoretic characterizations. A necessary assumption is that all sinks want all sources. The general case where different sinks wish to receive different subsets of the sources has an implicit characterization in terms of the region of entropic vectors and only inner and outer bounds are known explicitly [3, 4].

Han considers the problem of communicating a distributed set of correlated sources to a single sink across a capacity-constrained network and characterizes the set of achievable rates [5]. For a single sink, it is known that the min-cut/max-flow bounds can be achieved [4] and in particular, Slepian-Wolf (SW) style source coding [2] followed by routing is sufficient [5]. Han proposes a minimum-cost problem where link activations are charged a per unit cost and cites work by Fujishige [6] as an algorithmic solution to the proposed problem. The proposed algorithm can be applied to problems with both link and source costs; it cannot, however, be extended to the case of multiple sinks. Additionally, the algorithm is only guaranteed to terminate in finite time if the data are assumed integral [6]. Barros et al. contains a similar characterization of the set of achievable rates and an identical LP formulation as [5] but no discussion of an efficient solution [7].

When the problem is extended from a single sink to multiple sinks, each required to receive all the sources, it is known that in general routing is not sufficient for

This material is based upon work supported by the AFOSR through award FA9550-12-1-0086.

achieving the min-cut/max-flow bounds and that network coding is necessary [1] and in fact linear network coding is sufficient [4]. Identical characterizations of when a distributed correlated source can be multicast across a capacity-constrained network have been given by Song et al., Ramamoorthy, and Han [3, 8, 9]. These characterizations are a natural extension of the result for a single sink [5]. Earlier work by Cristescu et al. also considers the problem of SW coding across a network with links that were *not* capacity-constrained [10]. This allows for an optimal solution to be obtained as the superposition of minimum weight spanning trees. Two key differences between the work of Ramamoorthy [8] and Han [9] are that the former makes the assumption of rational capacities to make use of results from [11] and specifically considers the problem of minimizing the cost to multicast the sources. Focusing on lossless communication and assuming a linear objective, the cost to multicast the sources can be formulated as an linear objective with per unit cost for activating links. By not having a per source cost, the proposed LP can be solved by applying dual decomposition to exploit the combinatorial structure of the SW rate-region associated with the correlated source and using the subgradient method to approximate the optimal cost [8]. In the present work, we consider a more general model by including a per unit rate cost for each source node. The technique of dual decomposition and application of the subgradient method has been used in work by Yu et al. [12] and Lun et al. [13]. Yu et al. considers the problem of *lossy* communication of a set of sources and minimizes a cost function that trades off between the estimation distortion and the transmit power of the nodes in the network. The rate-distortion region is, in general, not polyhedral and the resulting optimization problem is convex. Lun et al. makes the assumption of a single source and therefore does not deal with the interdependencies among the different sources rates.

Previous works have only considered the dual with respect to a subset of the constraints in order to exploit the contrapolymatroidal structure of the SW rate-region. In the present work, we restrict our attention to a single sink and more fully investigate the underlying combinatorial structure of the resulting set of achievable rates. By considering the full dual LP, we demonstrate the application of the additional structural properties towards the development of alternative algorithmic solutions. Extension to the more general case with multiple sinks is left to future work.

2 Preliminaries

We model the network as a simple directed graph $D = (V, A)$ with nodes V representing alternately sources, routers, and destinations, and arcs A representing connections on the network between nodes. We model the arcs A as capacitated with capacity $c = (c(a), a \in A)$. If $a = (u, v) \in A$, then we define $\mathrm{tail}(a) \triangleq u$ and $\mathrm{head}(a) \triangleq v$.

$$\delta^{out}(v) \triangleq \{a \in A : \mathrm{tail}(a) = v\} \tag{1}$$

$$\delta^{in}(v) \triangleq \{a \in A : \mathrm{head}(a) = v\} \tag{2}$$

The distributed sources are located at a subset $S \subset V$ of the network elements and need to be collected at a sink $t \in V \setminus S$. We model the sources as a collection of correlated discrete memoryless random variables $(X_s : s \in S)$. There is a joint distribution $p_{(X_s:s\in S)}$ (shortened to just p_S) on the set of sources which in turn gives rise to a vector of conditional entropies $(H(X_U|X_{U^c}), U \subseteq S)$, where $H(X_U|X_{U^c})$ is the conditional entropy associated with the subset of sources $U \subseteq S$ given the values of the other sources $U^c = S \setminus U$.

The decision variables in our model are both i) the rates for each source, $R = (R(s), s \in S)$, and ii) the flow on each arc, $f = (f(a), a \in A)$. The rate $R(s)$ is the rate at which source s transmits, which must be routed (possibly split over multiple paths) towards the destination t, and the flow $f(a)$ is the superposition over all rates $R(s)$ whose routes traverse arc a. Flows must satisfy: i) capacity constraints ($0 \le f(a) \le c(a)$ for all $a \in A$), and ii) conservation of flow at all non-source, non-sink nodes ($f(\delta^{out}(v)) = f(\delta^{in}(v))$ for all $v \in V \setminus (S \cup \{t\})$). A flow f supports rates R if for all $s \in S$, $R(s) = f(\delta^{out}(s)) - f(\delta^{in}(s))$ The novelty lies in jointly optimizing over both (f, R) simultaneously, since most of the network flow literature assumes the source rates to be an input to the flow problem. While the multi-source network coding problem includes variables for both source rates and edge rates (analogous to our flow variables), much of the network coding literature has focused on characterizing the region obtained by projecting onto either the source rate or edge rate variables. Our work focuses on the cases where rate regions are known and expressly consider the problem of joint optimization *without* the projection onto one set of variables. For the case of multiple sinks, routing will no longer be sufficient and we will need to consider network coding. In this case, there will be a "virtual" flow f_t for each sink t satisfying the normal flow constraints. Under network coding, the physical flow $f(a)$ on an arc a will then satisfy $f_t(a) \le f(a)$ for all t [13].

We begin with the Slepian-Wolf theorem, which characterizes of the set of source rates for which lossless distributed source codes exist.

Theorem 1 (Slepian-Wolf [2])**.** *The rate region $\mathcal{R}_{SW}$ for distributed lossless source coding the discrete memoryless sources X_S is the set of rate tuples R such that*

$$R(U) \ge H(X_U|X_{S\setminus U}) \quad \forall\, U \subseteq S. \tag{3}$$

For brevity, let us define $\sigma_{SW} : 2^{|S|} \to \mathbb{R}$ as

$$\sigma_{SW}(U) \triangleq H(X_U|X_{U^c}) \tag{4}$$

which is a nonnegative, nondecreasing supermodular set function on the set of sources. Note that the rate region of Theorem 1 is the contrapolymatroid associated with σ_{SW}.

$$\mathcal{R}_{SW} = Q_{\sigma_{SW}} \triangleq \left\{R \in \mathbb{R}^{|S|} : R(U) \ge \sigma_{SW}(U),\ \forall\, U \subseteq S\right\} \tag{5}$$

The following theorem characterizes the set of source rates for which there exists a supporting flow.

Theorem 2 (Megiddo [14])**.** *There exists a flow f that supports the rates R iff*

$$R(U) \le \min\{c(\delta^{out}(X)) : U \subseteq X, t \in V \setminus X\} \quad \forall\, U \subseteq S. \tag{6}$$

Paralleling (4), define $\rho_c : 2^{|S|} \to \mathbb{R}$ as

$$\rho_c(U) = \min\{c(\delta^{out}(X)) : U \subseteq X, t \in V \setminus X\} \tag{7}$$

This is the min-cut capacity/max-flow value from the set U to the sink t, which is a nonnegative, nondecreasing submodular set function on the set of sources. The set of source rates for which there exists a supporting flow is

$$P_{\rho_c} = \left\{R \in \mathbb{R}^{|S|} : R \geq 0, R(U) \leq \rho_c(U),\ \forall\, U \subseteq S\right\} \tag{8}$$

The final theorem in this section characterizes when the intersection of the sets of source rates from the previous two theorems is non-empty.

Theorem 3 (Han's Matching Condition [5])**.** *Let σ and ρ be supermodular and submodular set functions, respectively. Then*

$$Q_\sigma \cap P_\rho \neq \varnothing \tag{9}$$

if and only if

$$\sigma(U) \leq \rho(U) \quad U \subseteq S \tag{10}$$

In particular, there exists distributed lossless source codes for communicating the sources X_S across the network to the sink t iff $\sigma_{SW}(U) \leq \rho_c(U)$ for all $U \subseteq S$.

As mentioned in [5], the proof of Theorem 3 depends critically on the submodularity of ρ and supermodularity of σ.

Our objective is to route the information from the sources S to the sink t as efficiently as possible, which we measure via costs on both the rate of the sources, and the costs of activating the arcs. Specifically, let $h = (h(s), s \in S)$ be the cost per bit per second associated with each source, and $k = (k(a), a \in A)$ be the cost per unit flow associated with each arc.

With this notation, the cost of a solution (f, R) is $k^T f + h^T R$. The constraints are the natural ones given the model description above: i) flows must observe the arc capacity constraints $f \leq c$, ii) flows f and rates R must satisfy conservation of flow at all router nodes $v \in V \setminus (S \cup t)$, iii) the flows and rates must match at the sources, so that the inflow plus the source rate equals the outflow, and iv) the rates must be large enough to fully describe the source entropies $R(U) \geq H(X_U|X_{U^c})$ for all $U \subseteq S$. By only considering a single sink, we only need to find one flow vector f. For the general network coding case, the model can be extended in a natural way to account for the "virtual" flow for each sink and the physical flow on each arc.

The linear program described above is as follows:

$$\begin{array}{lll}
\underset{f \geq 0, R}{\text{minimize}} & \displaystyle\sum_{a \in A} k(a) f(a) + \sum_{s \in S} h(s) R(s) & \\
\text{subject to} & f(a) \leq c(a) & a \in A \\
& f(\delta^{in}(v)) - f(\delta^{out}(v)) = 0 & v \in N \\
& R(s) + f(\delta^{in}(s)) - f(\delta^{out}(s)) = 0 & s \in S \\
& R(U) \geq H(X_U|X_{U^c}) & U \subseteq S
\end{array} \tag{11}$$

where $N \triangleq V \setminus (S \cup \{t\})$, $f(\delta(v)) \triangleq \sum_{a\in\delta(v)} f(a)$, and $R(U) \triangleq \sum_{s\in U} R(s)$, $U \subseteq S$. The linear program in (11) has $|A| + |V| - 1 + 2^{|S|}$ inequalities. If $|S| = \mathcal{O}(|V|)$, then the LP is exponential in the size of the graph. Observe that an optimal solution (f^*, R^*) to (11) will satisfy $R^*(S) = H(X_S)$ [5].

3 Feasible Set Structural Properties

We see from Theorem 1 and Theorem 2 that the set of feasible rates $Q_{\sigma_{SW}} \cap P_{\rho_c}$ is the intersection of a polymatroid with a contrapolymatroid. This polytope can be thought of as being obtained by the projection $p : \mathbb{R}^{|A|+|S|} \to \mathbb{R}^{|S|}$ of the set of feasible (f, R) tuples onto the rate variables R. In this section we present several structural properties of the set of feasible (f, R) and the associated lower dimensional set $Q_{\sigma_{SW}} \cap P_{\rho_c}$ that are *independent* of the assumed objective function in (11).

For any polyhedron P, we denote the set of extreme points as $\mathrm{Ext}(P)$. The extreme points (vertices) of a contrapolymatroid Q_σ are given by

$$R_\pi(s_{\pi(i)}) = \sigma(\{s_{\pi(1)}, \ldots, s_{\pi(i)}\}) - \sigma(\{s_{\pi(1)}, \ldots, s_{\pi(i-1)}\}) \qquad i = 1, \ldots, |S| \tag{12}$$

where π ranges over all permutations of $[|S|]$ [15][1]. The extreme rays of Q_σ are the unit vectors of $\mathbb{R}^{|S|}$. Similarly, the extreme points of a polymatroid P_ρ are given by

$$R_\pi(s_{\pi(i)}) = \begin{cases} \rho(\{s_{\pi(1)}, \ldots, s_{\pi(i)}\}) - \rho(\{s_{\pi(1)}, \ldots, s_{\pi(i-1)}\}) & i \le k \\ 0 & i > k \end{cases} \tag{13}$$

where π ranges over all permutations of $[|S|]$ and where k ranges over $0, \ldots, |S|$ [15]. The *base polyhedron* of Q_σ and P_ρ is defined as [16]

$$B(Q_\sigma) \triangleq Q_\sigma \cap \{R : R(S) = \sigma(S)\} \tag{14}$$

$$B(P_\rho) \triangleq P_\rho \cap \{R : R(S) = \rho(S)\}. \tag{15}$$

In general, (10) does not allow us to conclude if the base polyhedron of a contrapolymatroid $B(Q_\sigma)$ is wholly contained in the intersection $Q_\sigma \cap P_\rho$. To see this, let us consider $S = \{s_1, s_2\}$ and let ρ be submodular and σ supermodular such that $\sigma(U) \le \rho(U)$ for all $U \subseteq S$. Consider the vertex $R = (\sigma(s_1), \sigma(s_1, s_2) - \sigma(s_1))$ of Q_σ. We have, by the assumption of (10) that $R(s_1) = \sigma(s_1) \le \rho(s_1)$ and $R(s_1) + R(s_2) = \sigma(s_1, s_2) \le \rho(s_1, s_2)$. From the supermodularity of σ, we have that $\sigma(s_2) \le \sigma(s_1, s_2) - \sigma(s_1)$ and by assumption $\sigma(s_2) \le \rho(s_2)$; this *does not* allow us to conclude one way or the other if $\sigma(s_1, s_2) - \sigma(s_1) \gtreqless \rho(s_2)$ and so we cannot, in general, determine if $R \in P_\rho$ and therefore $R \in Q_\sigma \cap P_\rho$. Our first theorem provides a sufficient condition for $B(Q_\sigma)$ and $B(P_\rho)$ to be contained in $Q_\sigma \cap P_\sigma$.

Theorem 4. *Let σ be supermodular set function and ρ be a submodular set function. If*

$$\sigma(Y) - \sigma(Y \setminus X) \le \rho(X) - \rho(X \setminus Y) \quad \forall\, X, Y \subseteq S \tag{16}$$

then $\mathrm{Ext}(B(Q_\sigma)) \subseteq Q_\sigma \cap P_\rho$ *and* $\mathrm{Ext}(B(P_\rho)) \subseteq Q_\sigma \cap P_\rho$.

[1] For an integer i, the set $\{1, \ldots, i\}$ is denoted by $[i]$.

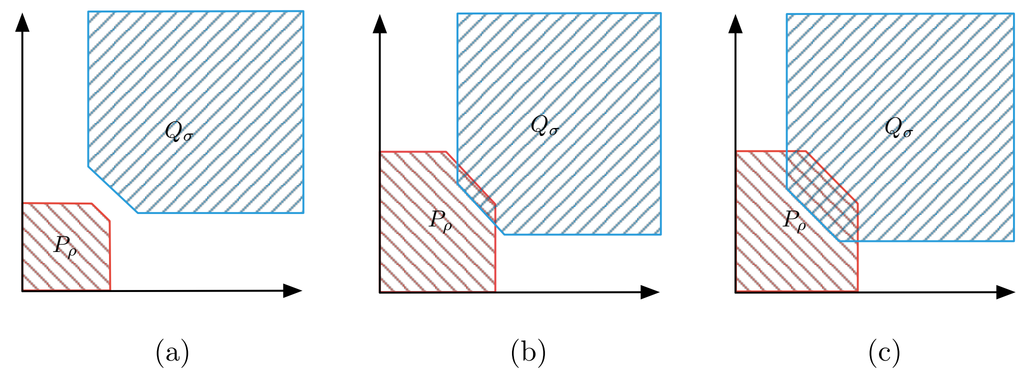

Figure 1: An example of Theorems 3 & 4: (a) A (σ, ρ) pair that does not satisfy (10); (b) A (σ, ρ) pair that satisfies (10); and; (c) A (σ, ρ) pairs that satisfies (16).

Proof. See Appendix A.1 $\square$

Remark. Observe that $\mathrm{Ext}(Q_\sigma) = \mathrm{Ext}(B(Q_\sigma))$. The requirement of this lemma (16) implies Han's matching condition (10). To see this, let $X = Y = U$ in (16). Figure 1 provides an example that illustrates the differences between Theorems 3 & 4. Theorem 4, beyond identifying a subset of the extreme points of the set of feasible rates, allows us to prove the optimality of the greedy algorithm for certain instances of (11).

For σ and ρ that satisfies the conditions of Theorem 4, the set $Q_\sigma \cap P_\rho$ is a *generalized polymatroid* [17], a mathematical object that unifies polymatroids and contrapolymatroids [15]. For every generalized polymatroid in $\mathbb{R}^{|S|}$, there exists a submodular set function $\rho' : 2^{|S|+1} \to \mathbb{R}$ and a projection $p : \mathbb{R}^{|S|+1} \to \mathbb{R}^{|S|}$ such that $p(B(P_{\rho'}))$ is equal to that generalized polymatroid [16]. This implies that optimizing a linear objective over a generalized polymatroid is no more difficult than maximizing a linear objective over the associated polymatroid $P_{\rho'}$, and the greedy algorithm for polymatroids can be used to solve LPs over generalized polymatroids in a straightforward manner. This property is at the heart of the proof of the following proposition.

Proposition 1. *Let σ and ρ satisfy the conditions of Theorem 4 and consider the LP given by*

$$\begin{aligned} \underset{R}{\textit{minimize}} \quad & \sum_{s \in S} w(s) R(s) \\ \textit{subject to} \quad & \sigma(U) \leq R(U) \leq \rho(U) \quad U \subseteq S. \end{aligned} \tag{17}$$

If $w(s) \geq 0$ for all $s \in S$, then there exists $R^* \in \mathrm{Ext}(Q_\sigma)$ that is an optimal solution to the given LP. If $w(s) \leq 0$ for all $s \in S$, then there exists $R^* \in \mathrm{Ext}(B(P_\rho))$ that is an optimal solution to the given LP.

Proof. See Appendix A.2 $\square$

The next lemma establishes that a convex combination of rates can be supported by a convex combination of supporting flows.

Lemma 1. *Suppose $R_i \in Q_{\sigma SW} \cap P_{\rho_c}$ and let f_i be a flow that supports R_i. If $R_\lambda = \sum_i \lambda_i R_i$ for $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$ then $f_\lambda = \sum_i \lambda_i f_i$ is a flow that supports R_λ.*

For an extreme point R of $Q_{\sigma_{SW}}$, we provide an expression for the sum rate for an arbitrary set of sources and then use this to characterize the active inequalities at R.

Lemma 2. *Fix an ordering $s_1, s_2, \dots, s_n$ of the elements of S and define $U_i \triangleq \{s_j : j \in [i]\}$. If R is the vertex in $Q_{\sigma_{SW}}$ corresponding to this ordering and $U = \{s_{k_1}, \cdots, s_{k_m}\}$ such that $k_1 < k_2 < \dots < k_m$ then*

$$R(U) = H(X_U | X_{U^c_{k_1} \setminus U}) + \sum_{j=2}^{m} I(X_{U \setminus U_{k_{j-1}}}; X_{U_{k_j - 1} \setminus U_{k_{j-1}}} | X_{U^c_{k_j} \setminus U}) \tag{18}$$

Proof. See Appendix A.3 □

Corollary 1. *If R is the vertex corresponding to permutation π then*

$$R(U_i) = H(X_{U_i} | X_{U_i^c}). \tag{19}$$

Proof. See Appendix A.4 □

Proposition 2. *Fix an ordering $s_1, s_2, \dots, s_n$ of the elements of S and define $U_i \triangleq \{s_j : j \in [i]\}$ and $U_0 = \varnothing$. Let R be the vertex in $Q_{\sigma_{SW}}$ that corresponds to this ordering and $U = \{s_{k_1}, \cdots, s_{k_m}\}$ such that $k_1 < k_2 < \dots < k_m$. Define $k_0 \triangleq 0$. If $U = U_i$ for some $i \in [n]$, then $R(U) = H(X_U | X_{U^c})$. If $U \neq U_i$ for some i, then $R(U) = H(X_U | X_U^c)$ if and only if*

$$(X_{U \setminus U_{k_{j-1}}} \perp X_{U_{k_j - 1} \setminus U_{k_{j-1}}}) | X_{U^c_{k_j} \setminus U} \quad j = 1, \dots, m. \tag{20}$$

Proof. From the Lemma 2, we have that

$$R(U) - H(X_U | X_{U^c}) = \sum_{j=1}^{m} I(X_{U \setminus U_{k_{j-1}}}; X_{U_{k_j - 1} \setminus U_{k_{j-1}}} | X_{U^c_{k_j} \setminus U}) \geq 0.$$

The above is a sum of conditional mutual informations which is zero iff each of the terms is equal to zero. This happens when the random variables X_U satisfies (20). □

The previous proposition will be used in the next section when giving conditions for a feasible solution to (11) to be optimal.

4 Sufficient Conditions for Characterizing Optimality

We proceed by finding the dual LP of the primal given in (11). In (11), we have three types of constraints: i) a capacity constraint for each edge, ii) flow conservation for each node, and iii) rate requirements for each subset of sources. The dual, then, will have three types of dual variables: i) $(x(a) : a \in A)$, ii) $(z(v) : v \in V)$, and iii) $(y_U : U \subset S)$. The dual LP is given as

$$\begin{aligned}
\underset{x\leq 0, y\geq 0, z}{\text{maximize}} \quad & \sum_{a\in A} c(a)x(a) + \sum_{U\subseteq S} H(X_U|X_{U^c})y_U \\
\text{subject to} \quad & x(a) + z(\text{head}(a)) - z(\text{tail}(a)) \leq k(a) \quad\quad a \in A \\
& \sum_{U\ni s} y_U + z(s) - z(t) = h(s) \quad\quad\quad\quad s \in S
\end{aligned} \tag{21}$$

We set $z(t) = 0$ because it is associated with the conservation of flow constraint at the sink, which is omitted from (11) as it is a consequence of the equality constraints at every other node. Observe that the number of dual variables is exponential in $|S|$. We now show that, in a certain sense, the dual variables $x(a)$ for $a \in A$ and y_U for $U \subseteq S$ are unnecessary.

Let us define the *reduced cost* of $a \in A$ as

$$\bar{k}(a) \triangleq k(a) - (z(\text{head}(a)) - z(\text{tail}(a))) \tag{22}$$

and observe that the first set of constraints of (21) can be written as $x(a) \leq \bar{k}(a)$ for all $a \in A$ [18]. Combined with the non-positivity constraint on $x(a)$ we have $x(a) \leq \min(0, \bar{k}(a))$. Since we are maximizing in (21) and $c(a) > 0$ for all a, we take

$$x(a) = \min(0, \bar{k}(a)) \tag{23}$$

and see that the dual variable $x(a)$ can be expressed in terms of $(z(v) : v \in V)$. As we show in the next theorem, characterizations of optimal solutions do not need to explicitly consider the dual variables $(x(a) : a \in A)$.

Theorem 5. *Let $f^*_{R_i}$ be a min-cost flow that supports rate R_i. Let $R = \sum_i \lambda_i R_i$. The flow $f = \sum_i \lambda_i f^*_{R_i}$ is a flow that supports R of minimum cost if there exists a vector $(z(v) : v \in V)$ such that for all i*

$$\bar{k}(a) < 0 \implies f^*_{R_i}(a) = c(a) \tag{24a}$$

$$\bar{k}(a) > 0 \implies f^*_{R_i}(a) = 0. \tag{24b}$$

Proof. See Appendix A.6 □

Since we are considering the rates to be fixed in the previous theorem, there are no dual variables $(y_U : U \subseteq S)$. If the conditional entropies of the sources and the min-cut capacities satisfy the requirements of Theorem 4, then all of the extreme points of $Q_{\sigma_{SW}}$ are feasible for (11). As was mentioned earlier, if (f^*, R^*) is an optimal solution to (11) then $R^*(S) = H(X_S)$ [5] and therefore R^* can be written as a convex combination of the extreme points of $Q_{\sigma_{SW}}$. The previous theorem shows that in certain cases, f^* can be found as a convex combination of the min-cost flows for the extreme points of the SW rate region.

We now define the reduced cost of $s \in S$ as

$$\bar{h}(s) \triangleq h(s) - (z(s) - z(t)) = h(s) - z(s) \tag{25}$$

and rewrite the second set of constraints of (21) as

$$\sum_{U \ni s} y_U = \bar{h}(s). \tag{26}$$

We seek to express the dual variables y_U as function of the dual variables $z(s)$ as we did for the dual variables $x(a)$. The following theorem provides a characterization of which of the dual variables y_U must be zero as a function of the correlation structure of the source random variables.

Theorem 6. *Suppose R^* is primal optimal and y^* is dual optimal and let $U = \{s_{k_1}, \cdots, s_{k_m}\}$ such that $k_1 < k_2 < \ldots < k_m$. If R^* is a vertex of $Q_{\sigma_{SW}}$ and there exists $j \in [m]$ such that*

$$(X_{U \setminus U_{k_{j-1}}} \not\perp X_{U_{k_j-1} \setminus U_{k_{j-1}}}) | X_{U_{k_j}^c \setminus U} \tag{27}$$

then $y_U^* = 0$.

Proof. Follows immediately from complimentary slackness and Proposition 2. □

This characterization suggests the following sufficient condition for an extreme point R_π of the SW rate region $Q_{\sigma_{SW}}$ and its associated min-cost flow f_π^* to be a solution to (11).

Theorem 7. *A feasible solution (f_π^*, R_π) of (11) is optimal if there exists vectors $(z(v) : v \in V)$ satisfying for each $a \in A$*

$$\bar{k}(a) < 0 \implies f_\pi^*(a) = c(a) \tag{28a}$$
$$\bar{k}(a) > 0 \implies f_\pi^*(a) = 0 \tag{28b}$$

and

$$\bar{h}(s_1) \geq \bar{h}(s_2) \geq \cdots \geq \bar{h}(s_n) \geq 0 \tag{29}$$

where the elements of S are ordered according the permutation π.

Proof. Ordering the elements of S according to the permutation π induces a nested family of subsets $U_i \triangleq \{s_j : j \in [i]\}$. We construct a dual feasible y by setting $y_U = 0$ for U not in the nested family and

$$y_{U_i} = \begin{cases} \bar{h}(s_i) - \bar{h}(s_{i+1}) & i \in [n-1] \\ \bar{h}(s_i) & i = n. \end{cases} \tag{30}$$

We construct a dual feasible x from (23). Having primal feasible (f_π^*, R_π) and dual feasible (x, y, z), optimality follows from complimentary slackness. □

The impact of the previous two theorems is that even though the dual has an exponential number of variables, we need only consider a linear (in $|V|$) number of them. Given $(z^*(v) : v \in V)$, we can compute $(x^*(a) : a \in A)$ according to (23) and $(y_U^* : U \subseteq S)$ according to (30). The extreme points of the SW rate region are significant because codes that satisfy $R(S) = H(X_S)$ can be constructed from codes for these points via time sharing.

5 Conclusion

In this paper, we have considered the transmission of distributed sources across a network with capacity constraints. Previous works have only made use of the fact that SW rate region is a contrapolymatroid as part of an iterative subgradient method. The set of achievable rates is the intersection of the SW rate region with the polymatroid defined by the min-cut capacities. We have established several structural properties of this set that have not been previously reported. We have shown that these properties lead to a characterization relating optimal solutions and the corner points of the SW rate region. Unsurprisingly, this characterization is connected with the correlation structure of the sources. Future work will more fully explore this connection and the implications of different correlation structures (e.g., Markov random fields) on the optimal solution and generalize to multiple sinks.

6 References

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, 2000.

[2] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 4, 1973.

[3] L. Song and R. W. Yeung, "Network information flow—multiple sources," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2001.

[4] R. W. Yeung, *Information Theory and Network Coding.* Springer, 2008.

[5] T. S. Han, "Slepian-Wolf-Cover theorem for networks of channels," *Information and Control*, vol. 47, no. 1, 1980.

[6] S. Fujishige, "Algorithms for solving the independent-flow problems," *J. Oper. Res. Soc. Japan*, vol. 21, 1978.

[7] J. Barros and S. D. Servetto, "Network information flow with correlated sources," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, 2006.

[8] A. Ramamoorthy, "Minimum cost distributed source coding over a network," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, Jan. 2011.

[9] T. S. Han, "Multicasting multiple correlated sources to multiple sinks over a noisy channel network," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, Jan. 2011.

[10] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "Networked Slepian-Wolf: theory, algorithms, and scaling laws," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, 2005.

[11] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, 2006.

[12] W. Yu and J. Yuan, "Joint source coding, routing and resource allocation for wireless sensor networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 2, 2005.

[13] D. S. Lun, N. Ratnakar, M. Médard, R. Koetter, D. R. Karger, T. Ho, E. Ahmed, and F. Zhao, "Minimum-cost multicast over coded packet networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, 2006.

[14] N. Megiddo, "Optimal flows in networks with multiple sources and sinks," *Mathematical Programming*, vol. 7, no. 1, 1974.

[15] A. Schrijver, *Combinatorial Optimization: Polyhedra and Efficiency.* Springer, 2003.

[16] S. Fujishige, *Submodular Functions and Optimization*, 2nd ed. Elsevier, 2005.

[17] A. Frank and É. Tardos, "Generalized polymatroids and submodular flows," *Mathematical Programming*, vol. 42, no. 1–3, 1988.
[18] W. J. Cook, W. H. Cunningham, W. R. Pulleyblank, and A. Schrijver, *Combinatorial Optimization.* Wiley Interscience, 2011.

A Omitted Proofs

A.1 Proof of Theorem 4

For the following set of lemmas, order the elements of S as $s_1, s_2, \ldots, s_n$ arbitrarily and define

$$U_i \triangleq \{s_1, \ldots, s_i\}. \tag{31}$$

For an arbitrary $U \subseteq S$, let k be the largest index such that $s_k \in U$ and k' be the largest index such that $s_{k'} \in U \setminus \{s_k\}$.

Lemma 3.

$$U_{k'} \cap (U_k \setminus U) = U_{k'} \setminus (U \setminus \{s_k\}) \tag{32}$$

Proof.

$$\begin{aligned} U_{k'} \cap (U_k \setminus U) &= U_{k'} \cap U_k \cap U^c \\ &= U_{k'} \setminus U \\ &= U_{k'} \setminus (U \setminus \{s_k\}) \end{aligned}$$

Where the last step follows from $s_k \notin U_{k'}$. □

Lemma 4.

$$U_{k'} \cup (U_k \setminus U) = U_{k-1} \tag{33}$$

Proof. By definition, $s_i \notin U$ for $i = k'+1, \ldots, k-1$ and therefore $\{s_{k'+1}, \ldots, s_{k-1}\} \subseteq U_k \setminus U$. Therefore $U_{k'} \cup (U_k \setminus U) = U_{k-1}$. □

Lemma 5. *If σ is supermodular, then*

$$\sigma(U_{k-1}) - \sigma(U_{k'}) + \sigma(U_{k'} \setminus (U \setminus \{s_k\})) \geq \sigma(U_k \setminus U) \tag{34}$$

Proof. Follows immediately from the supermodularity of σ and the previous two lemmas. □

Corollary 2. *If ρ is submodular, then*

$$\rho(U_{k-1}) - \rho(U_{k'}) + \rho(U_{k'} \setminus (U \setminus \{s_k\})) \leq \rho(U_k \setminus U) \tag{35}$$

Proof. If ρ is submodular, then $-\rho$ is supermodular. □

Lemma 6. *If $R(\{s_i\}) = \sigma(U_i) - \sigma(U_{i-1})$ for all $i = 1, \ldots, n$, then for all non-empty $U \subseteq S$ we have*

$$R(U) \leq \sigma(U_k) - \sigma(U_k \setminus U) \tag{36}$$

where k is the largest index such that $s_k \in U$.

Proof. Proof by induction on $|U|$. *Base case*: If $|U| = 1$,then we have $U = \{s_k\}$ and $U_k \setminus U = U_{k-1}$. By the assumption of R being a vertex, we have that $R(\{s_k\}) \triangleq \sigma(U_k) - \sigma(U_{k-1})$ and the assertion is trivially true.

Inductive step: We have that

$$\begin{aligned} R(U) &= R(\{s_k\}) + R(U \setminus \{s_k\}) \\ &= \sigma(U_k) - \sigma(U_{k-1}) + R(U \setminus \{s_k\}) \\ &\le \sigma(U_k) - \sigma(U_{k-1}) + \sigma(U_{k'}) - \sigma(U_{k'} \setminus (U \setminus \{s_k\})) \\ &\le \sigma(U_k) - \sigma(U_k \setminus U) \end{aligned}$$

where the second to last step follows from the inductive hypothesis and the last step from the previous lemma. □

The point R in the previous lemma is the vertex of Q_σ corresponding to the order of the elements of S $s_1, \ldots, s_n$. Our choice of this ordering was arbitrary, so this lemma means that the given inequality is true for any vertex.

Corollary 3. *If $R(\{s_i\}) = \rho(U_i) - \rho(U_{i-1})$ for all $i = 1, \ldots, n$, then for all non-empty $U \subseteq S$ we have*

$$R(U) \ge \rho(U_k) - \rho(U_k \setminus U) \tag{37}$$

where k is the largest index such that $s_k \in U$.

Proof. Proof by induction on $|U|$. *Base case*: If $|U| = 1$, then we have $U = \{s_k\}$ and $U_k \setminus U = U_{k-1}$. By the assumption of R being a vertex, we have that $R(\{s_k\}) \triangleq \rho(U_k) - \rho(U_{k-1})$ and the assertion is trivially true.

Inductive step: We have that

$$\begin{aligned} R(U) &= R(\{s_k\}) + R(U \setminus \{s_k\}) \\ &= \rho(U_k) - \rho(U_{k-1}) + R(U \setminus \{s_k\}) \\ &\ge \rho(U_k) - \rho(U_{k-1}) + \rho(U_{k'}) - \rho(U_{k'} \setminus (U \setminus \{s_k\})) \\ &\ge \rho(U_k) - \rho(U_k \setminus U) \end{aligned}$$

where the second to last step follows from the inductive hypothesis and the last step from the previous corollary. □

Proof of Theorem 4 . We will show that all the vertices of Q_σ (which have the property that $R(S) = \sigma(S)$) are elements of P_ρ. All the vertices of Q_σ are the vertices of $B(Q_\sigma)$. Then by the convexity of P_ρ, we will have that $B(Q_\sigma) \subseteq P_\rho$ and therefore $B(Q_\sigma) \subseteq P_\rho \cap Q_\sigma$. Note that we assume w.l.o.g. that $\rho(\varnothing) = \sigma(\varnothing) = 0$.

Let R be a vertex of Q_σ and apply strong compliance inequality result to the result of the previous lemma; we have

$$\begin{aligned} R(T) &\le \sigma(U_k) - \sigma(U_k \setminus T) \\ &\le \rho(T) - \rho(T \setminus U_k) \\ &= \rho(T) \end{aligned}$$

We conclude then that every vertex of Q_σ is an element of P_ρ. As P_ρ is convex, the convex hull of the vertices of Q_σ, which is $B(Q_\sigma)$, is contained in P_ρ.

Let R be a vertex of $B(P_\rho)$ and apply strong compliance inequality result to the result of the previous corollary; we have

$$\begin{aligned} R(T) &\geq \rho(U_k) - \rho(U_k \setminus T) \\ &\geq \sigma(T) - \sigma(T \setminus U_k) \\ &= \sigma(T) \end{aligned}$$

We conclude then that every vertex of $B(P_\rho)$ is an element of Q_σ. As Q_σ is convex, the convex hull of the vertices of $B(P_\rho)$, which is $B(P_\rho)$, is contained in Q_σ. □

A.2 Proof of Proposition 1

Proof. Assuming σ and ρ satisfy the condition of Theorem 4, we have that $Q_\sigma \cap P_\rho$ is non-empty. Let us define $S' = S \cup \{s^*\}$ and

$$f(U) = \begin{cases} \rho(U) & U \in 2^S \\ \gamma - \sigma(S' \setminus U) & U \subset S', s^* \in U \end{cases} \tag{38}$$

where $\gamma \in \mathbb{R}$ is arbitrary but fixed. Such a f is a submodular function on $2^{S'}$ and $B(EP_f) = EP_f \cap \{R : R(S) = f(S)\}$ [2] is non-empty [16]. In fact

$$Q_\sigma \cap P_\rho = \left\{ R \in \mathbb{R}^{|S|} : \exists \alpha \in \mathbb{R} : (R, \alpha) \in B(EP_f) \right\}. \tag{39}$$

We now rewrite (17) as

$$\begin{aligned} \underset{R}{\text{minimize}} \quad & \sum_{s \in S} w(s) R(s) + 0 \cdot \alpha \\ \text{subject to} \quad & (R, \alpha) \in B(EP_f) \end{aligned}$$

and proceed by applying the greedy algorithm [16]. Assuming $w(s) \geq 0$ for all $s \in S$, we order the elements of S as $s_1, \ldots, s_n$ such that $0 \triangleq w(s^*) \leq w(s_1) \cdots \leq w(s_n)$. Let

$$\begin{aligned} \alpha &= \gamma - \sigma(S) \\ R^*(s_i) &= \sigma(\{s_i, \ldots, s_n\}) - \sigma(\{s_{i+1}, \ldots, s_n\}) \quad i = 1, \ldots, n; \end{aligned}$$

such an R^* is an optimal solution to the given LP. This is the vertex of Q_σ corresponding to the permutation of the elements of S given by $\pi(i) = n + 1 - i$. If instead $w(s) \leq 0$ for all $s \in S$, we order the elements of S as $s_1, \ldots, s_n$ such that $\leq w(s_1) \cdots \leq w(s_n) \leq w(s^*) = 0$. Let

$$\begin{aligned} R^*(s_i) &= \rho(\{s_1, \ldots, s_i\}) - \rho(\{s_1, \ldots, s_{i-1}\}) \quad i = 1, \ldots, n \\ \alpha &= \gamma - \rho(S); \end{aligned}$$

such an R^* is an optimal solution to the given LP. This is the vertex of $B(P_\rho)$ corresponding to the permutation of the elements of S given by $\pi(i) = i$. □

[2] The set $EP_f = \{R \in \mathbb{R}^{|S'|} : R(U) \leq f(U)\}$ is the extended polymatroid associated with f while $P_f = \{x \in \mathbb{R}^{|S'|} : R \geq 0, R(U) \leq f(U)\}$ is the polymatroid associated with f.

Remark. The function f constructed in the proof satisfies the requirement $f(\varnothing) \geq 0$ and therefore EP_f is nonempty. If we take $\gamma \geq \sigma(S)$, then $f(U') \geq 0$ for all $U' \subseteq S'$ and P_f is nonempty. There is a one-to-one correspondence between nonempty polymatroids and nondecreasing submodular functions [15]. That is, we can find a nondecreasing submodular $\bar{f}$ with $\bar{f}(\varnothing) = 0$ such that $P_f = P_{\bar{f}}$. These extra steps are unnecessary for the proof, as simply appyling the greedy algorithm to $B(EP_f)$ yields the required result.

A.3 Proof of Lemma 2

Recall from Lemma 2 that $U_i \triangleq \{s_j : j \in [i]\}$ and $U = \{s_{k_1}, \cdots, s_{k_m}\}$ such that $k_1 < k_2 < \ldots < k_m$. Let us define $U' \triangleq U \setminus \{s_{k_1}\} = \{s_{k'_1}, \cdots, s_{k'_{m'}}\}$ where $k'_i = k_{i+1}$ and $m' = m - 1$. We begin with three supporting lemmas.

Lemma 7.

$$U^c_{k_j} \setminus U' = U^c_{k_j} \setminus U \tag{40}$$

Proof.

$$\begin{aligned} U^c_{k_j} \setminus U &= U^c_{k_j} \cap (\{s_{k_1}\} \cup U')^c \\ &= U^c_{k_j} \cap (\{s_{k_1}\}^c \cap U'^c) \\ &= U^c_{k_j} \cap U'^c \end{aligned}$$

The first step follows from the definition of U' and the last step from recognizing that $U^c_{k_j} \subseteq \{s_{k_1}\}^c$. □

Lemma 8.

$$U' = U \setminus U_{k_1} \tag{41}$$

Proof.

$$U \setminus U_{k_1} = U \cap \{s_{k_1+1}, s_{k_1+2}, \ldots, s_n\} = \{s_{k_2}, \cdots, s_{k_m}\} = U'$$

□

Lemma 9.

$$U^c = U_{k_1-1} \cup U^c_{k_1} \setminus U \tag{42}$$

Proof.

$$\begin{aligned} U_{k_1-1} \cup U^c_{k_1} \setminus U &= U_{k_1-1} \cup (U^c_{k_1} \cap U^c) \\ &= (U_{k_1-1} \cup U^c_{k_1}) \cap (U_{k_1-1} \cup U^c) \\ &= \{s_{k_1-1}\}^c \cap U^c \\ &= U^c \end{aligned}$$

□

Proof of Lemma 2 . Proof by induction on $|U|$. *Base case:* If $|U| = 1$, then $U = \{s_{k_1}\}$ and we have that

$$\begin{aligned}
R(s_{k_1}) &= H(X_{U_{k_1}}|X_{U^c_{k_1}}) - H(X_{U_{k_1-1}}|X_{U^c_{k_1-1}}) \\
&= H(X_{U_{k_1-1}}, X_{s_{k_1}}|X_{U^c_{k_1}}) - H(X_{U_{k_1-1}}|X_{U^c_{k_1}}, X_{s_{k_1}}) \\
&= H(X_{s_{k_1}}|X_{U^c_{k_1}}) \\
&= H(X_{s_{k_1}}|X_{U^c_{k_1}\setminus\{s_{k_1}\}})
\end{aligned}$$

where the last step follows from the fact that $U^c_i = U^c_i \setminus \{s_i\}$.

Inductive Step: Let us define $U' \triangleq U \setminus \{s_{k_1}\} = \{s_{k'_1}, \cdots, s_{k'_{m'}}\}$ where $k'_i = k_{i+1}$ and $m' = m - 1$. We have that

$$\begin{aligned}
R(U) &= R(s_{k_1}) + R(U') \\
&\overset{(a)}{=} H(X_{s_{k_1}}|X_{U^c_{k_1}}) + R(U') \\
&\overset{(b)}{=} H(X_{s_{k_1}}|X_{U^c_{k_1}}) + H(X_{U'}|X_{U^c_{k'_1}\setminus U'}) \\
&+ \sum_{i=1}^{m'-1} I(X_{U'\setminus U_{k'_i}}; X_{U_{k'_{i+1}-1}\setminus U_{k'_i}}|X_{U^c_{k'_{i+1}}\setminus U'}) \\
&\overset{(c)}{=} H(X_{s_{k_1}}|X_{U^c_{k_1}}) + H(X_{U'}|X_{U_{k_2-1}\setminus U_{k_1}}, X_{U^c_{k'_1}\setminus U'}) + I(X_{U'}; X_{U_{k_2-1}\setminus U_{k_1}}|X_{U^c_{k'_1}\setminus U'}) \\
&+ \sum_{i=1}^{m'-1} I(X_{U'\setminus U_{k'_i}}; X_{U_{k'_{i+1}-1}\setminus U_{k'_i}}|X_{U^c_{k'_{i+1}}\setminus U'}) \\
&\overset{(d)}{=} H(X_{s_{k_1}}|X_{U^c_{k_1}}) + H(X_{U'}|X_{U_{k_2-1}\setminus U_{k_1}}, X_{U^c_{k_2}\setminus U'}) + I(X_{U'}; X_{U_{k_2-1}\setminus U_{k_1}}|X_{U^c_{k_2}\setminus U'}) \\
&+ \sum_{i=1}^{m'-1} I(X_{U'\setminus U_{k'_i}}; X_{U_{k'_{i+1}-1}\setminus U_{k'_i}}|X_{U^c_{k'_{i+1}}\setminus U'}) \\
&\overset{(e)}{=} H(X_{s_{k_1}}|X_{U^c_{k_1}\setminus U'}, X_{U'}) + H(X_{U'}|X_{U^c_{k_1}\setminus U'}) + I(X_{U'}; X_{U_{k_2-1}\setminus U_{k_1}}|X_{U^c_{k_2}\setminus U'}) \\
&+ \sum_{i=1}^{m'-1} I(X_{U'\setminus U_{k'_i}}; X_{U_{k'_{i+1}-1}\setminus U_{k'_i}}|X_{U^c_{k'_{i+1}}\setminus U'}) \\
&\overset{(f)}{=} H(X_U|X_{U^c_{k_1}\setminus U'}) + I(X_{U'}; X_{U_{k_2-1}\setminus U_{k_1}}|X_{U^c_{k_2}\setminus U'}) \\
&+ \sum_{i=1}^{m'-1} I(X_{U'\setminus U_{k'_i}}; X_{U_{k'_{i+1}-1}\setminus U_{k'_i}}|X_{U^c_{k'_{i+1}}\setminus U'}) \\
&\overset{(g)}{=} H(X_U|X_{U^c_{k_1}\setminus U'}) + I(X_{U'}; X_{U_{k_2-1}\setminus U_{k_1}}|X_{U^c_{k_2}\setminus U'}) \\
&+ \sum_{i=2}^{m-1} I(X_{U\setminus U_{k_i}}; X_{U_{k_{i+1}-1}\setminus U_{k_i}}|X_{U^c_{k_{i+1}}\setminus U})
\end{aligned}$$

$$\stackrel{(h)}{=} H(X_U|X_{U^c_{k_1}\setminus U'}) + I(X_{U\setminus U_{k_1}}; X_{U_{k_2-1}\setminus U_{k_1}}|X_{U^c_{k_2}\setminus U})$$
$$+ \sum_{i=2}^{m-1} I(X_{U\setminus U_{k_i}}; X_{U_{k_{i+1}-1}\setminus U_{k_i}}|X_{U^c_{k_{i+1}}\setminus U})$$
$$\stackrel{(i)}{=} H(X_U|X_{U^c_{k_1}\setminus U}) + \sum_{i=1}^{m-1} I(X_{U\setminus U_{k_i}}; X_{U_{k_{i+1}-1}\setminus U_{k_i}}|X_{U^c_{k_{i+1}}\setminus U})$$

where: (a) follows from the definition of a vertex; (b) follows from the application of the inductive hypothesis; (c) follows from the definition of conditional mutual information; (d) $U^c_{k'_1} \setminus U' = U^c_{k_2} \setminus U'$; (e) $U' \subseteq U^c_{k_1}$ so partition $U^c_{k_1}$ into $U^c_{k_1} \setminus U'$ and U'; (f) follows from the chain rule for conditional entropy; (g) follows from a change of variable for the sum index; (h) follows from expressing the conditional mutual information in terms of the original set, and; (i) follows from moving the first conditional mutual information into the sum. □

Remark. We interpret the sets in the above expression

$$\begin{gathered} U^c_{k_1} \setminus U = \{s_{k_1+1}, \ldots, s_{k_2-1}, s_{k_2+1}, \ldots, s_{k_m-1}, s_{k_m+1}, \ldots, s_n\} \\ U \setminus U_{k_{j-1}} = \{s_{k_j}, \ldots, s_{k_m}\} \\ U_{k_j-1} \setminus U_{k_{j-1}} = \{s_i : i = k_j, \ldots, k_j - 1\} \\ U^c_{k_j} \setminus U = \{s_{k_j+1}, \ldots, s_{k_{j+1}-1}, s_{k_{j+1}+1}, \ldots, s_{k_m-1}, s_{k_m+1}, \ldots, s_n\} \end{gathered}$$

For concreteness, suppose $n = 6$ and $U = \{s_2, s_4, s_6\}$. We have

$$\begin{gathered} U^c_2 \setminus U = \{s_3, s_5\} \\ U \setminus U_{k_1} = U \setminus U_2 = \{s_4, s_6\} \\ U \setminus U_{k_2} = U \setminus U_4 = \{s_6\} \\ U_{k_2-1} \setminus U_{k_1} = U_3 \setminus U_2 = \{s_i : i = k_1 + 1, \ldots, k_2 - 1\} = \{s_3\} \\ U_{k_3-1} \setminus U_{k_2} = U_5 \setminus U_4 = \{s_i : i = k_2 + 1, \ldots, k_3 - 1\} = \{s_5\} \\ U^c_{k_2} \setminus U = \{s_5\} \\ U^c_{k_3} \setminus U = \varnothing \end{gathered}$$

and $R(U) = H(X_{s_2}, X_{s_4}, X_{s_6}|X_{s_3}, X_{s_5}) + I(X_{s_4}, X_{s_6}; X_{s_3}|X_{s_5}) + I(X_{s_6}; X_{s_5})$.

A.4 Alternative proof of Corollary 1

Proof. For any supermodular set function we have

$$R(U_i) = \sum_{j=1}^{i} R(s_{\pi(j)}) = \sum_{j=1}^{i} \sigma(U_j) - \sigma(U_{j-1}) = \sigma(U_i) - \sigma(\varnothing). \tag{43}$$

If $\sigma(U) = H(X_U|X^c_U)$, the above is $R(U_i) = H(X_{U_i}|X_{U^c_i})$. □

A.5 Full Proof of Proposition 2

Proof.

$$\begin{aligned}
0 &\le R(U) - H(X_U|X_{U^c}) \\
&= H(X_U|X_{U^c_{k_1}\backslash U}) - H(X_U|X_{U^c}) + \sum_{j=2}^{m} I(X_{U\backslash U_{k_{j-1}}}; X_{U_{k_j-1}\backslash U_{k_{j-1}}}|X_{U^c_{k_j}\backslash U}) \\
&= H(X_U|X_{U^c_{k_1}\backslash U}) - H(X_U|X_{U_{k_1-1}}, X_{U^c_{k_1}\backslash U}) + \sum_{j=2}^{m} I(X_{U\backslash U_{k_{j-1}}}; X_{U_{k_j-1}\backslash U_{k_{j-1}}}|X_{U^c_{k_j}\backslash U}) \\
&= I(X_U; X_{U_{k_1-1}}|X_{U^c_{k_1}\backslash U}) + \sum_{j=2}^{m} I(X_{U\backslash U_{k_{j-1}}}; X_{U_{k_j-1}\backslash U_{k_{j-1}}}|X_{U^c_{k_j}\backslash U}) \\
&= \sum_{j=1}^{m} I(X_{U\backslash U_{k_{j-1}}}; X_{U_{k_j-1}\backslash U_{k_{j-1}}}|X_{U^c_{k_j}\backslash U}).
\end{aligned}$$

□

A.6 Proof of Theorem 5

This is a restatement of and proof of Theorem 5.

Theorem 8. *Alternative to Conjecture Let $f^*_{R_i}$ be a min-cost flow that supports rate R_i. Let $R = \sum_i \lambda_i R_i$. The flow $f = \sum_i \lambda_i f^*_{R_i}$ is a flow that supports R of minimum cost if there exists a vectors $(x^*(a) : a \in A)$ and $(z^*(v) : v \in V)$ such that for all i*

$$x^*(a)(f^*_{R_i}(a) - c(a)) = 0 \qquad \forall\, a \in A \tag{44a}$$

$$(k(a) - x^*(a) - (z^*(\text{head}(a)) - z^*(\text{tail}(a))))f^*_{R_i}(a) = 0 \qquad \forall\, a \in A. \tag{44b}$$

Proof. Having fixed a rate vector R_i, we can solve for the min-cost flow for that rate with following LP

$$\begin{array}{lll}
\underset{f\ge 0}{\text{minimize}} & \sum_{a\in A} k(a)f(a) & \\
\text{subject to} & f(a) \le c(a) & a \in A \\
& f(\delta^{in}(v)) - f(\delta^{out}(v)) = 0 & v \in N \\
& f(\delta^{in}(v)) - f(\delta^{out}(s)) = -R_i(s) & s \in S
\end{array} \tag{45}$$

and its corresponding dual

$$\begin{array}{lll}
\underset{x\le 0, z}{\text{maximize}} & \sum_{a\in A} c(a)x(a) - \sum_{s\in S} R_i(s)z(s) & \\
\text{subject to} & x(a) + z(\text{head}(a)) - z(\text{tail}(a)) \le k(a) & a \in A.
\end{array} \tag{46}$$

If R_i is a feasible rate vector, then there exists a min-cost flow $f^*_{R_i}$ for this R_i and therefore optimal dual variables $(x^*_{R_i}, z^*_{R_i})$. Observe that the set of feasible dual variables does not depend on the rates R_i, only on the edge costs k. By assumption $x^*_{R_i} = x^*$ and $z^*_{R_i} = z^*$ for all i and therefore (x^*, z^*) is dual feasible for R. We have that by Lemma 1, that f is primal feasible. Checking the complimentary slackness conditions for f, x^*, and z^*, we have

$$x^*(a)(f(a) - c(a)) = x^*(a)\left(\sum_i \lambda_i f^*_{R_i}(a) - c(a)\right) = \sum_i \lambda_i \left(x^*(a)(f^*_{R_i}(a) - c(a))\right) = 0 \tag{47}$$

and similarly

$$(k(a) - x^*(a) - (z^*(\text{head}(a)) - z^*(\text{tail}(a))))f^*_{R_i}(a) = 0. \tag{48}$$

We conclude that f is primal optimal and x^*, z^* are dual optimal solutions for a min-cost flow that supports R. $\square$