Achievable Schemes for Cost/Performance Trade-offs in Networks

A Thesis

Submitted to the Faculty

of

Drexel University

by

Bradford D. Boyle

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

June 2015

Dedications

To Diane whose love and support made this possible

Acknowledgments

During my time at Drexel University, I have had the incredibly good fortune to work with some truly amazing and inspiring people. As my time here comes to a close, I wanted to acknowledge the people who have been instrumental in my studies. I owe a particular debt of gratitude to my advisor Steven Weber. Without his patience and guidance, none of this would have been possible. Aside from my advisor, no other faculty member has had as profound an impact on me and my research as John MacLaren Walsh. With regards to the work presented in this dissertation, the material in chapters three and four were developed and brought to maturity in partnership with him. In that same vein, I would like to thank the remaining members of my committee, Spiros Mancoridis, Harish Sethu, and Matthew Stamm for their invaluable feedback and support in completing this dissertation. If it were not for Moshe Kam, I would have never started my graduate studies and begun this journey. Much of the material contained in this dissertation was greatly facilitated by a collaboration with Leonard Cimini and Javier Garcia-Frias at the University of Delaware.

I would like to thank my fellow students in the Modeling & Analysis of Networks Lab, Data Fusion Lab, and Adaptive Signal Processing & Information Theory Research Group. Among these, I am particularly indebted to Gwanmo Ku and Jie Ren. Gwanmo's work on computing the rate-distortion function for the multiterminal CEO problem with independent sources allowed me to compare my achievable schemes to the fundamental limits and he graciously supplied the data for plotting rate distortion functions. Jie's work in interactive communication motivated my looking at interactive scalar quantization and he assisted me immensely in preparing the related work on interactive communication. More broadly, I would like to thank Pramod Abichandani, Gus Anderson, Raymond Canzanese, David Dorsey, Gabe Ford, Lex Fridman, Tingshan Huang, Ryan Measel, Rich Primerano, Cem Sahin, and Jeffrey Wildman. I have enjoyed working with all of you and hope that I am lucky enough to work with people like you in the future.

I want to say "Thank You" to the ECE staff, especially Kathy Bryant, Tanita Chapelle, Sean

Clark, Chad Morris, Phyllis D. Watson, and Wade Kirkpatrick. Their assistance and patience in helping me to address the non-research aspects of academic life is very much appreciated.

Finally, I want to thank my family and friends for supporting me in this endeavor.

This research has been supported by the Air Force Research Laboratory under agreement number FA9550-12-1-0086. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory or the U.S. Government.

Table of Contents

List of Figures

Abstract

Achievable Schemes for Cost/Performance Trade-offs in Networks
Bradford D. Boyle
Steven Weber, Ph.D.

A common pattern in communication networks (both wired and wireless) is the collection of distributed state information from various network elements. This network state is needed for both analytics and operator policy and its collection consumes network resources, both to measure the relevant state and to transmit the measurements back to the data sink. The design of simple achievable schemes are considered with the goal of minimizing the overhead from data collection and/or trading off performance for overhead. Where possible, these schemes are compared with the optimal trade-off curve.

The optimal transmission of distributed correlated discrete memoryless sources across a network with capacity constraints is considered first. Previously unreported properties of jointly optimal compression rates and transmission schemes are established. Additionally, an explicit relationship between the conditional independence relationships of the distributed sources and the number of vertices for the Slepian-Wolf rate region is given.

Motivated by recent work applying rate-distortion theory to computing the optimal performance-overhead trade-off, the use of distributed scalar quantization is investigated for lossy encoding of state, where a central estimation officer (CEO) wishes to compute an extremization function of a collection of sources. The superiority of a simple heterogeneous (across users) quantizer design over the optimal homogeneous quantizer design is proven.

Interactive communication enables an alternative framework where communicating parties can send messages back-and-forth over multiple rounds. This back-and-forth messaging can reduce the rate required to compute an extremum/extrema of the sources at the cost of increased delay. Again scalar quantization followed by entropy encoding is considered as an achievable scheme for a collection of distributed users talking to a CEO in the context of interactive communication. The

design of optimal quantizers is formulated as the solution of a minimum cost dynamic program. It is established that, asymptotically, the costs for the CEO to compute the different extremization functions are equal. The existence of a simpler search space, which is asymptotically sufficient for minimizing the cost of computing the selected extremization functions, is proven.

Chapter 1: Introduction

Successful and efficient utilization of a communication network requires collecting data about the state of the network, both for analytics and resource allocation/control. Owing to the highly distributed nature of modern networks (especially wireless), collection of relevant state information represents a non trivial cost to the network operator. As an example, approximately 25–30% of LTE downlink transmission bandwidth is used for control signaling [1, 2]. With the increasing demand for and reliance on wireless networks, techniques are needed to minimize the impact of gathering the relevant information. This thesis leverages analytical and algorithmic tools from information theory, linear programming, network flow theory, and combinatorial optimization in addressing this engineering challenge. In particular, the design of simple achievable schemes that minimize the overhead from data collection and/or trade-off performance for overhead are presented. This chapter provides an overview of the problems considered in this work and summarizes the key research contributions.

1.1 Research contributions and thesis outline

As a first step toward addressing the problem of minimizing and/or effecting a trade-off of performance for overhead in communication systems, the present work considers three related problems:

1. lossless transmission of distributed correlated sources across a network with capacity constraints;

2. scalar quantization for lossy distributed extremization, and;

3. interactive scalar quantization for lossless distributed extremization.

The rest of this section describes each of these problems in more detail and summarizes the research contributions made in each.

1.1.1 Lossless transmission of distributed correlated sources across a network with capacity constraints

Efficient minimization of the overhead for lossless transmission of sources across capacity constrained networks requires understanding of the structural properties of the set of feasible source rates and network flow. A key result of this investigation is a comprehensive characterization of the intersection of polymatroids and contrapolymatroids, answering questions about the feasibility of the Pareto-optimal subset of the Slepian-Wolf rate region. Another key result is an explicit relationship between conditional independence relationships and the redundancy of vertices of the Slepian-Wolf rate region, reducing the number of inequalities needed to define the rate region for sources with known structures such as Markov random fields. Combining these new insights about the set of feasible rates demonstrates that a layered solution to the optimization problem may not be optimal when the cost of data compression is in tension with transmission costs.

1.1.2 Scalar quantization for lossy distributed extremization

The previous problem formulation required lossless collection of the relevant sources. In many scenarios, the entirety of the data is not needed; it is sufficient to compute a function of the data. Depending on the exact function to be computed, this result can be obtained losslessly with a lower required rate. In resource allocation contexts, a common class of functions are extremization functions (e.g., max/min). For this class of functions, lossless rate savings are negligible and it is necessary to tolerate bounded distortion to realize an appreciable rate savings [3]. This problem is an indirect distributed lossy source coding problem with the distortion measuring degredation in system performance. Motivated by these insights, the design of a simple achievable scheme based on scalar quantization to trade-off distortion with rate is proposed. For computing an extremization function, scalar quantization has overhead-performance trade-off close to the fundamental limit given by the rate distortion. Additionally, it is shown that a heterogeneous (across users) quantizer design has better performance than an optimal homogeneous quantizer design, even when the source variables are independent and identically distributed (i.i.d.).

1.1.3 Interactive scalar quantization for lossless distributed extremization

An alternative approach to reduce the rate required for function computation is the usage of interactive communication. In the lossy estimator framework, the rate savings were achieved by tolerating distortion. With the interactive approach, the rate savings come at the expense of delay. Mirroring the solution approach of the previous problem, the design of interactive scalar quantizers is considered as a means to effect rate-delay trade-offs. The stages of communication between the CEO and distributed users in this framework suggests dynamic programming as a solution technique to optimizing the proposed scheme. For homogeneous (across users) scalar quantization, it is established that the rate-delay trade-offs of the optimal dynamic program for the different extremization functions are asymptotically equal. Analysis of simple quantization strategies demonstrates that overhead (per user) of scalar quantization decreases as the number of users is increased. Extending the previously mentioned strategies results in a family of quantization strategies that are (asymptotically) sufficient for obtaining the optimal rate-delay trade-offs of scalar quantization.

Chapter 2: Structural and Optimization Properties for Joint Selection of Source Rates and Network Flow

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. We extend previous results concerning the intersection of polymatroids and contrapolymatroids to characterize when all of the vertices of the Slepian-Wolf rate region are feasible for the capacity constrained network. An explicit relationship between the conditional independence relationships of the distributed sources and the number of vertices for the Slepian-Wolf rate region are given. These properties are then applied to characterize the optimal transmission rate and scheme and its connection to the corner points of the Slepian-Wolf rate region. In particular, we demonstrate that when the per-source compression costs are in tension with the per-link flow costs the optimal flow/rate point need *not* coincide with a vertex of the Slepian-Wolf rate region. Finally, we connect results for the single-sink problem to the multi-sink problem by extending structural insights and developing upper and lower bounds on the optimal cost of the multi-sink problem.

2.1 Introduction

2.1.1 Motivation

A class of problems that arises 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 have received much attention in recent years as a result of the development of network coding [4] and can be traced back to the seminal work of Slepian and Wolf [5]. In this chapter, 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 each sink wishes to receive a subset of the sources has an implicit characterization in terms of the

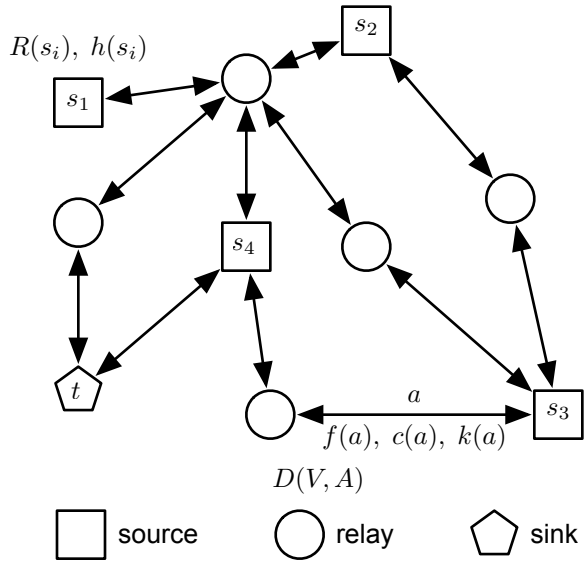


Figure 2.1: Transmission of distributed correlated sources across a network with capacity constraints problem overview: A single sink node losslessly recovers distributed correlated sources over a capacity constrained network. There is a per unit cost associated with activating each link as well as a per unit cost at the sources.

region of entropic vectors and only inner and outer bounds are explicitly known [6, 7].

2.1.2 Related Work

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 [8]. For a single sink, it is known that the min-cut/max-flow bounds can be achieved [7] and in particular, Slepian-Wolf (SW) style source coding [5] followed by routing is sufficient [8]. Han proposes a minimum-cost problem where link activations are charged a per unit cost and cites work by Fujishige [9] as an algorithmic solution to the proposed problem. The proposed algorithm can be applied to problems with both link and source costs; however, it cannot 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 [9]. Barros et al. [10] contains a similar characterization of the set of achievable rates and an identical LP formulation as [8] but no discussion of an efficient algorithm. In the achievability proof of Barros et al. (and Han [8]), a separation between the source encoder rates and the network flows is observed, leading to a natural mapping of this problem into the traditional protocol stack.

When the problem is extended from a single sink to multiple sinks, each sink required to receive all the sources, it is known that i) in general routing is not sufficient for achieving the min-cut/max-flow bounds; ii) network coding is necessary [4], and; iii) in fact linear network coding is sufficient [7]. 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 [6, 11, 12]. These characterizations are a natural extension of the result for a single sink [8]. Earlier work by Cristescu et al. also considers the problem of SW coding across a network with links that were *not* capacity-constrained [13]. 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 [11] and Han [12] are that the former makes the assumption of rational capacities to make use of results from [14] 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 a 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 sources and using the subgradient method to approximate the optimal cost [11]. 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. [15] and Lun et al. [16]. 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 source rates.

2.1.3 Summary of Contributions

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.

The rest of this chapter is organized as follows: In §2.2, we present and discuss relevant supporting material from literature as well as formally pose our optimization problem. In §2.3, we extend existing results concerning the intersection of polymatroid with a contrapolymatroid and characterize their types of intersections. We also relate the conditional independence relationships of the sources to degeneracy of the extreme points of the Slepian-Wolf rate region, reducing the number of inequality constraints needed to describe the polyhedron. In §2.4, we consider the dual of the linear program to develop sufficient conditions for optimal solutions. We are particularly interested in knowing when the optimal solution will coincide with a vertex of the Slepian-Wolf rate region. We demonstrate that when there is an imbalance between the source costs and flow costs (i.e., cheap compression and expensive routing vs. expensive compression and cheap routing), the optimal rate allocation may not coincide with a vertex of the rate region. In §2.5, we partially extend our results to the multi-sink problem and bound the optimal value of the multi-sink problem with the optimal values of related single sink problems. We conclude in §2.6.

2.2 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 network connections between nodes in V. We model the arcs A as capacitated with capacity $c = (c(a), a \in A)$. If $a = (u, v) \in A$, then we define $\text{tail}(a) \triangleq u$ and $\text{head}(a) \triangleq v$ and

$$\delta^{out}(v) \triangleq \{a \in A : \text{tail}(a) = v\} \tag{2.1a}$$

$$\delta^{in}(v) \triangleq \{a \in A : \text{head}(a) = v\}. \tag{2.1b}$$

For an arbitrary set function $f : U \mapsto \mathbb{R}$, we denote $\sum_{u \in B} f(u)$ by $f(B)$ for any subset $B \subseteq U$.

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 $U^{\mathsf{c}} = S \setminus U$.

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 ($0 \leq f(a) \leq c(a)$ 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 of our optimization problem model 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 considers 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 $f_t(a) \leq f(a)$ for all t [16].

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

Theorem 1 (Slepian-Wolf [5])**.** *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*

$$R(U) \geq H(X_U|X_{U^c}) \quad \forall\, U \subseteq S. \tag{2.2}$$

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

$$\sigma_{SW}(U) \triangleq H(X_U|X_{U^c}) \tag{2.3}$$

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

$$\mathcal{R}_{SW} = Q_{\sigma_{SW}} \triangleq \left\{ R \in \mathbb{R}^{|S|} : R(U) \geq \sigma_{SW}(U),\ \forall\, U \subseteq S \right\}. \tag{2.4}$$

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

Theorem 2 (Megiddo [17])**.** *There exists a flow f that supports the rates R iff*

$$R(U) \leq \min\{c(\delta^{out}(X)) : U \subseteq X, t \in V \setminus X\} \quad \forall\, U \subseteq S. \tag{2.5}$$

Paralleling (2.3), define $\rho_c : 2^{|S|} \to \mathbb{R}$ as

$$\rho_c(U) = \min\{c(\delta^{out}(X)) : U \subseteq X, t \in V \setminus X\} \tag{2.6}$$

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 the polymatroid P_{ρ_c} associated with ρ_c:

$$P_{\rho_c} \triangleq \left\{ R \in \mathbb{R}^{|S|} : R \geq 0, R(U) \leq \rho_c(U),\ \forall\, U \subseteq S \right\}. \tag{2.7}$$

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 [8])**.** *Let σ and ρ be supermodular and submodular set*

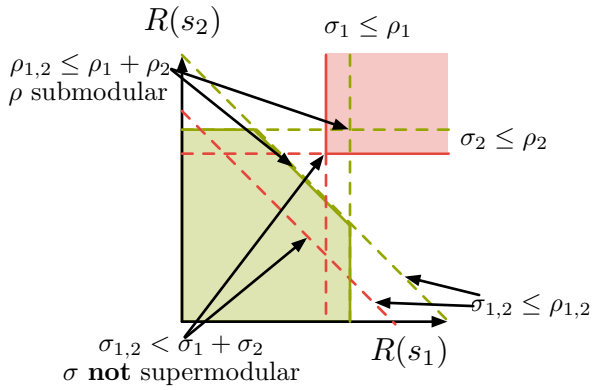


Figure 2.2: Sufficiency of Theorem 3 depends on submodularity and supermodularity. An example of set functions ρ and σ that satisfy $\sigma(U) \leq \rho(U)$ for all $U \subseteq S$. We see that ρ is submodular since $\rho(\{s_1, s_2\}) \leq \rho(\{s_1\}) + \rho(\{s_2\})$, while σ is *not* supermodular since $\sigma(\{s_1, s_2\}) < \sigma(\{s_1\}) + \sigma(\{s_2\})$. In this case, Theorem 3 cannot be used to conclude that P_ρ and Q_σ have a non-empty intersection.

functions, respectively. Then

$$I_{\sigma,\rho} \triangleq Q_\sigma \cap P_\rho \neq \varnothing \tag{2.8}$$

if and only if

$$\sigma(U) \leq \rho(U) \quad U \subseteq S \tag{2.9}$$

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

As mentioned in [8], the proof of the necessity of Theorem 3 is obvious. The proof of the sufficiency of Theorem 3 depends critically on the submodularity of ρ and supermodularity of σ. Figure 2.2 gives an example of generic set functions σ and ρ that satisfy (2.9) for which the set $I_{\sigma,\rho} = \varnothing$ because σ is *not* supermodular. Specializing (2.9) to conditional entropy and min-cut capacity gives

$$H(X_U \mid X_{U^c}) \leq \rho_c(U) \tag{2.10}$$

which has the following interpretation: $H(X_U \mid X_{U^c})$ is the information only available at the set of sources U and the network must be able to at least support a flow of that value from those sources.

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 $k^\intercal f + h^\intercal R$. 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:

$$\begin{aligned}
\underset{f \geq 0, R}{\text{minimize}} \quad & \sum_{a \in A} k(a) f(a) + \sum_{s \in S} h(s) R(s) & \\
\text{subject to} \quad & f(a) \leq c(a) & a \in A \\
& f(\delta^{in}(v)) - f(\delta^{out}(v)) = 0 & v \in N \\
& R(s) + f(\delta^{in}(s)) - f(\delta^{out}(s)) = 0 & s \in S \\
& R(U) \geq H(X_U|X_{U^c}) & U \subseteq S
\end{aligned} \tag{2.11}$$

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 (2.11) 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 (2.11) will satisfy $R^*(S) = H(X_S)$ [8].

2.3 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. The resulting 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 (2.11).

2.3.1 General properties from sub-/supermodularity

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

$$R_\pi(s_{\pi(i)}) = \sigma(U_{\pi(i)}) - \sigma(U_{\pi(i-1)}) \quad i = 1, \ldots, |S| \tag{2.12}$$

where π ranges over all permutations of $[|S|]$[1] and $U_{\pi(i)} = \{s_{\pi(1)}, \ldots, s_{\pi(i)}\}$ [18]. 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

$$R_\pi(s_{\pi(i)}) = \begin{cases} \rho(U_{\pi(i)}) - \rho(U_{\pi(i-1)}) & i \leq k \\ 0 & i > k \end{cases} \tag{2.13}$$

where π ranges over all permutations of $[|S|]$ and where k ranges over $0, \ldots, |S|$ [18]. With these definitions, we can now show that the half-space inequalities for $U_{\pi(i)}$ hold with equality.

Lemma 1. *If R_π is the vertex of Q_σ corresponding to permutation π then*

$$R_\pi(U_{\pi(i)}) = \sigma(U_{\pi(i)}). \tag{2.14}$$

If R_π is the vertex of P_ρ corresponding to permutation π then

$$R_\pi(U_{\pi(i)}) = \rho(U_{\pi(i)}). \tag{2.15}$$

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

Proof. For any supermodular set function we have

$$\begin{aligned} R_\pi(U_{\pi(i)}) &= \sum_{j=1}^{i} R_\pi(s_{\pi(j)}) \\ &= \sum_{j=1}^{i} \sigma(U_{\pi(j)}) - \sigma(U_{\pi(j-1)}) \\ &= \sigma(U_{\pi(i)}) - \sigma(\varnothing). \end{aligned} \tag{2.16}$$

For any submodular set function we have

$$\begin{aligned} R_\pi(U_{\pi(i)}) &= \sum_{j=1}^{i} R_\pi(s_{\pi(j)}) \\ &= \sum_{j=1}^{i} \rho(U_{\pi(j)}) - \rho(U_{\pi(j-1)}) \\ &= \rho(U_{\pi(i)}) - \rho(\varnothing). \end{aligned} \tag{2.17}$$

□

The *base polyhedron* of Q_σ and P_ρ is defined as [19]

$$B_\sigma \triangleq Q_\sigma \cap \{R : R(S) = \sigma(S)\} \tag{2.18a}$$

$$B_\rho \triangleq P_\rho \cap \{R : R(S) = \rho(S)\}. \tag{2.18b}$$

As noted previously, an optimal solution (f^*, R^*) to the LP (2.11) will satisfy $R^*(S) = H(X_S)$ and thus $R^* \in B_{\sigma_{SW}}$.

In general, Han's matching condition (Theorem 3) does not allow us to conclude if the base polyhedron of a contrapolymatroid B_σ is wholly contained in the intersection $Q_\sigma \cap P_\rho$.

Example 1. Consider $S = \{s_1, s_2\}$ and let ρ be submodular and σ supermodular such that $\sigma(U) \leq \rho(U)$ 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 (2.9) that $R(s_1) = \sigma(s_1) \leq \rho(s_1)$ and $R(s_1) + R(s_2) = \sigma(s_1, s_2) \leq \rho(s_1, s_2)$. From the supermodularity of σ, we have that $\sigma(s_2) \leq \sigma(s_1, s_2) - \sigma(s_1)$ and by assumption $\sigma(s_2) \leq \rho(s_2)$; this *does not* allow us to conclude one way or the other if $\sigma(s_1, s_2) - \sigma(s_1) \gtrless \rho(s_2)$ and so we cannot, in

general, determine if $R \in P_\rho$ and therefore $R \in I_{\sigma,\rho}$. □

Our first set of results characterize when B_σ and B_ρ are contained in $I_{\sigma,\rho}$. For generic submodular ρ and supermodular σ set functions we assume, w.l.o.g., that $\sigma(\varnothing) = \rho(\varnothing) = 0$. We begin by combining results from Frank et al. [20] and Fujishige [19, 21] and provide an explicit characterization of the vertices of $I_{\sigma,\rho}$ for certain instances of σ and ρ.

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

$$\sigma(U) - \sigma(U \setminus T) \leq \rho(T) - \rho(T \setminus U) \quad \forall\, T, U \subseteq S \tag{2.19}$$

then the vertices of $I_{\sigma,\rho}$ are given by

$$R^j_\pi(s_{\pi(i)}) = \begin{cases} \rho(U_{\pi(i)}) - \rho(U_{\pi(i-1)}) & i \leq j \\ \sigma(S \setminus U_{\pi(i-1)}) - \sigma(S \setminus U_{\pi(i)}) & i > j \end{cases} \tag{2.20}$$

where π is a permutation and $j \in \{0, \ldots, n\}$.

To proof this theorem, two supporting lemmas are needed.

Lemma 2. *Let $\mathrm{Ext}(P)$ be the set of extreme points of a polyhedron P and $p(P)$ be the projection of P; then $\mathrm{Ext}(p(P)) \subseteq p(\mathrm{Ext}(P))$.*

Proof. Suppose $x \in \mathrm{Ext}(p(P))$; then there exists x' such that $(x, x') \in P$. As x is extreme, there does not exist $y, z \in p(P)$ not equal to x and $\lambda \in (0, 1)$ such that $x = \lambda y + (1 - \lambda)z$. Therefore there is no choice of $(y, y'), (z, z') \in P$ not equal to (x, x') and $\lambda \in (0, 1)$ such that

$$\begin{pmatrix} x \\ x' \end{pmatrix} = \lambda \begin{pmatrix} y \\ y' \end{pmatrix} + (1 - \lambda) \begin{pmatrix} z \\ z' \end{pmatrix} \tag{2.21}$$

and $(x, x') \in \mathrm{Ext}(P)$. □

Lemma 3. *Let $\mathrm{Ext}(P)$ be the set of extreme points of a polyhedron P and $p(P)$ be the projection of P. If p one-to-one, then $\mathrm{Ext}(p(P)) = p(\mathrm{Ext}(P))$.*

Proof. Consider $x \in \text{Ext}(P)$; suppose its projection $p(x) \notin \text{Ext}(p(P))$. W.l.o.g. there exists $y, z \in P$ not equal to x and $\lambda \in (0,1)$ such that

$$p(x) = \lambda p(y) + (1-\lambda)p(z) = p(\lambda y + (1-\lambda)z) \tag{2.22}$$

which follows from projections being affine mappings. Additionally, since the projection is one-to-one we must have

$$x = \lambda y + (1-\lambda)z \tag{2.23}$$

contradicting the assumption of $x \in \text{Ext}(P)$. □

Proof of Theorem 4 . 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

$$f(U) = \begin{cases} \rho(U) & U \in 2^S \\ \gamma - \sigma(S' \setminus U) & U \subset S', s^* \in U \end{cases} \tag{2.24}$$

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 [19]. In fact

$$Q_\sigma \cap P_\rho = \left\{ R \in \mathbb{R}^{|S|} : \exists \alpha \in \mathbb{R} : (R, \alpha) \in B(EP_f) \right\}. \tag{2.25}$$

The vertices of EP_f are given by

$$R_\pi(s_{\pi(i)}) = f(U_{\pi(i)}) - f(U_{\pi(i-1)}) \quad i = 1, \ldots, |S|+1 \tag{2.26}$$

[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.

where π ranges over all permutations of $[|S|+1]$. Let j be the integer such that $s_{\pi(j)} = s'$. Then

$$R_\pi(s_{\pi(i)}) = \begin{cases} \rho(U_{\pi(i)}) - \rho(U_{\pi(i-1)}) & i < j \\ \gamma - \sigma(S' \setminus U_{\pi(i)}) - \rho(U_{\pi(i-1)}) & i = j \\ \sigma(S' \setminus U_{\pi(i-1)}) - \sigma(S' \setminus U_{\pi(i)}) & i > j. \end{cases} \tag{2.27}$$

□

Ignoring the flow costs in the LP of (2.11), we see that an optimal solution corresponds to an extreme point of $Q_{\sigma_{SW}}$.

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

$$\begin{aligned} &\underset{R}{\text{minimize}} \quad \sum_{s \in S} h(s) R(s) \\ &\text{subject to} \quad \sigma(U) \leq R(U) \leq \rho(U) \quad U \subseteq S. \end{aligned} \tag{2.28}$$

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

For σ and ρ that satisfies the conditions of Theorem 4, the set $I_{\sigma,\rho}$ is a *generalized polymatroid* [20], a mathematical object that unifies polymatroids and contrapolymatroids [18]. 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 $p(B_{\rho'})$ is equal to that generalized polymatroid [19] (see Figure 2.3). This insight is half of the proof of Theorem 4; the other half is recognizing that polyhedral properties are preserved by one-to-one affine mappings [21].

We see from (2.13) that the intersection $I_{\sigma,\rho}$ has at most $(n+1)!$ vertices; we can construct trivial examples for which the intersection is a generalized polymatroid and has strictly less than $(n+1)!$ vertices. For a given submodular set function ρ, let $\sigma(X) = \rho(S) - \rho(S \setminus X)$. It can be readily verified that such a σ is supermodular and that (2.19) is always true by the submodularity of ρ. From (2.20), the vertices of $I_{\sigma,\rho}$ are just those of B_ρ (or those of B_σ as $B_\rho = B_\sigma$).

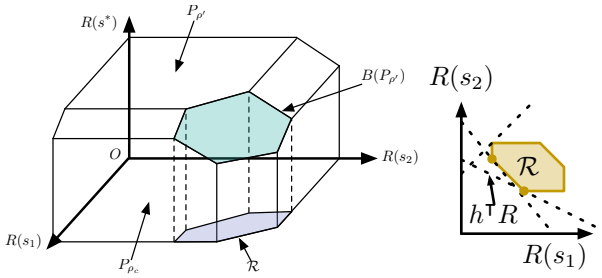


Figure 2.3: Generalized polymatroid as the projection of the base polytope B_f of a polymatroid P_f in a higher dimension. Minimization of a linear objective with sign-definite weight vector h will have a solution at a vertex of B_σ ($h \geq 0$) or at a vertex of B_ρ ($h \leq 0$).

We observe that when (2.19) holds, we have that $B_\sigma \subset I_{\sigma,\rho}$ and $B_\rho \subset I_{\sigma,\rho}$. Motivated by the observation that if (f^*, R^*) is an optimal solution to (2.11), then $R^* \in B_{\sigma_{SW}}$, we loosen the requirement (2.19) of Theorem 4 to characterize when $B_\sigma \subset I_{\sigma,\rho}$.

Theorem 5. $\text{Ext}(B_\sigma) \subseteq P_\rho$ *if and only if*

$$\sigma(T) - \sigma(T \setminus U) \leq \rho(U) \; \forall U \subseteq T \subseteq S. \tag{2.29}$$

Proof. Assume $\sigma(T) - \sigma(T \setminus U) \leq \rho(U)$ for all $U \subseteq T \subseteq S$. Consider an arbitrary permutation π and its associated vertex R_π of Q_σ. For any $U \subseteq S$, define $k \triangleq \min\{k' : U \subseteq U_{\pi(k')}\}$ or equivalently $k \triangleq \max\{k' : s_{\pi(k')} \in U\}$. We have

$$\begin{aligned} R_\pi(U) &= R_\pi(U_{\pi(k)}) - R_\pi(U_{\pi(k)} \setminus U) \\ &= \sigma(U_{\pi(k)}) - R_\pi(U_{\pi(k)} \setminus U) \\ &\leq \sigma(U_{\pi(k)}) - \sigma(U_{\pi(k)} \setminus U) \\ &\leq \rho(U) \end{aligned} \tag{2.30}$$

and therefore $R_\pi \in P_\rho$. This is true for all permutations and we conclude that $B(Q_\sigma) \subseteq P_\rho$.

Suppose $\exists U \subseteq T \subseteq S$ such that $\sigma(T) - \sigma(T \setminus U) > \rho(U)$. Let the elements of S be ordered by a permutation π so that $T = \{s_{\pi(1)}, \ldots, s_{\pi(|T|)}\}$ and $U = \{s_{\pi(|T|-|U|+1)}, \ldots, s_{\pi(|T|)}\}$. Then

$T = U_{\pi(|T|)}$ and $T \setminus U = U_{\pi(|T|-|U|)}$. It follows that

$$\begin{aligned} R_\pi(U) &= \sum_{i=|T|-|U|+1}^{|T|} \sigma(U_{\pi(i)}) - \sigma(U_{\pi(i-1)}) \\ &= \sigma(U_{\pi(|T|)}) - \sigma(U_{\pi(|T|-|U|)}) \\ &= \sigma(T) - \sigma(T \setminus U) \\ &> \rho(U) \end{aligned} \tag{2.31}$$

and therefore $R_\pi \notin P_\rho$. We conclude that $B_\sigma \not\subseteq P_\rho$. □

Unsurprisingly, we can loosen (2.19) in a similar manner to characterize when $B_\rho \subset I_{\sigma,\rho}$.

Theorem 6. $\mathrm{Ext}(B_\rho) \subseteq Q_\sigma$ *if and only if*

$$\sigma(U) \leq \rho(T) - \rho(T \setminus U) \; \forall \, U \subseteq T \subseteq S \tag{2.32}$$

Proof. Assume $\sigma(U) \leq \rho(T) - \rho(T \setminus U$ for all $U \subseteq T \subseteq S$. Consider an arbitrary permutation π and its associated vertex R_π of P_ρ. For any $U \subseteq S$, define $k \triangleq \min\{k' : U \subseteq U_{\pi(k')}\}$ or equivalently $k \triangleq \max\{k' : s_{\pi(k')} \in U\}$. We have

$$\begin{aligned} R_\pi(U) &= R_\pi(U_{\pi(k)}) - R_\pi(U_{\pi(k)} \setminus U) \\ &= \rho(U_{\pi(k)}) - R_\pi(U_{\pi(k)} \setminus U) \\ &\geq \rho(U_{\pi(k)}) - \rho(U_{\pi(k)} \setminus U) \\ &\geq \sigma(U) \end{aligned} \tag{2.33}$$

and therefore $R_\pi \in Q_\sigma$. This is true for all permutations and we conclude that $B(P_\rho) \subseteq Q_\sigma$.

Suppose $\exists U \subseteq T \subseteq S$ such that $\sigma(U) > \rho(T) - \rho(T \setminus U)$. Let the elements of S be ordered by a permutation π so that $T = \{s_{\pi(1)}, \ldots, s_{\pi(|T|)}\}$ and $U = \{s_{\pi(|T|-|U|+1)}, \ldots, s_{\pi(|T|)}\}$. Then

$T = U_{\pi(|T|)}$ and $T \setminus U = U_{\pi(|T|-|U|)}$. It follows that

$$\begin{aligned} R_\pi(U) &= \sum_{i=|T|-|U|+1}^{|T|} \rho(U_{\pi(i)}) - \rho(U_{\pi(i-1)}) \\ &= \rho(U_{\pi(|T|)}) - \rho(U_{\pi(|T|-|U|)}) \\ &= \rho(T) - \rho(T \setminus U) \\ &< \sigma(U) \end{aligned} \tag{2.34}$$

and therefore $R_\pi \notin Q_\sigma$. We conclude that $B_\rho \not\subseteq Q_\sigma$. □

Remark. In the proofs of Theorems 5 & 6, we use the existence of T, U that do not satisfy (2.29) (resp. (2.32)) to construct a vertex of Q_σ (resp. P_ρ) that is not retained in the intersection $I_{\sigma,\rho}$. For a given T, U that do not satisfy (2.29), there exists $(|T| - |U|)!\,|U|!\,(|S| - |T|)!$ permutations for which the corresponding vertex of Q_σ is not in P_ρ. Similarly, for a given T, U that do not satisfy (2.32), there exists $(|T| - |U|)!\,|U|!\,(|S| - |T|)!$ permutations for which the corresponding vertex of P_ρ is not in Q_σ.

Observe that Theorems 4, 5, and 6 each imply Theorem 3. To see this, let $T = U$. Figure 2.4 provides an example that illustrates the differences between Theorems 3, 5, & 6. Theorem 3 provides an initial characterization of the structure of $I_{\sigma,\rho}$ by determining when the intersection is empty or not and requires checking 2^n inequalities. It does not provide insight into what the vertices of the intersection are. Theorem 5 and Theorem 6 provide a partial characterization of the vertices of the intersection by characterizing a subset of the vertices of the intersection, but each requires checking 3^n inequalities. If $\text{Ext}(B_\sigma) \subseteq P_\rho$, then $\text{Ext}(B_\sigma) \subseteq \text{Ext}(I_{\sigma,\rho})$ and if $\text{Ext}(B_\rho) \subseteq Q_\sigma$, then $\text{Ext}(B_\rho) \subseteq \text{Ext}(I_{\sigma,\rho})$. However, we know that there are vertices of $I_{\sigma,\rho}$ that do not lie in either B_σ or B_ρ (e.g., Figure 2.4e). Finally, Theorem 4 provides a complete characterization of $\text{Ext}(I_{\sigma,\rho})$, but requires checking 4^n inequalities. Observe that $U \cap T = \varnothing$ the cross inequality (2.19) of Theorem 4 is the tautology $0 \leq 0$ and there are 3^n such pairs T, U of subsets of S. For $T = U$, (2.19) becomes Han's matching condition (2.9). For $T \subseteq U$, (2.19) reduces to (2.29). For $U \subseteq T$, (2.19) reduces to (2.32). The relationship among Theorems 4, 5, & 6 (in terms of pairs of subsets of S) is depicted in

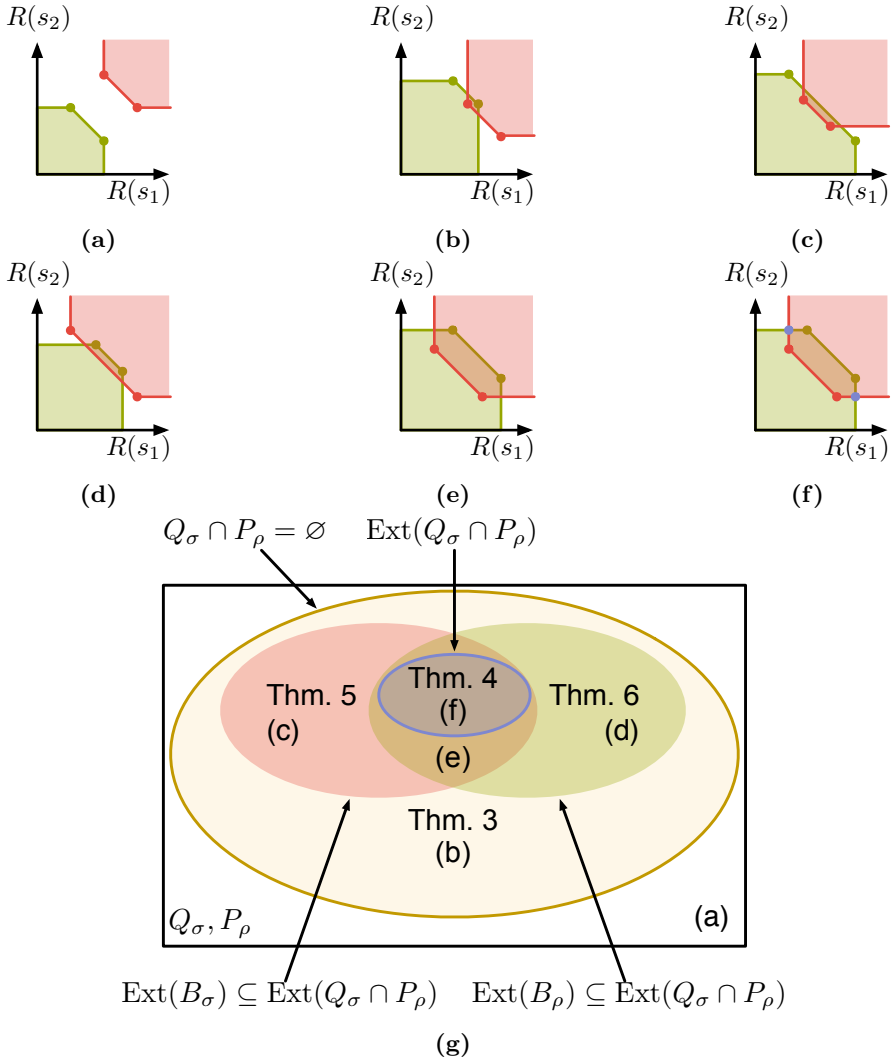


Figure 2.4: An example of Theorems 3, 4, 5, & 6: Theorem 3 differentiates a vs. b–e. If Theorem 3 holds, Theorem 5 differentiates b vs. c; Theorem 6 differentiates b vs. d, and; Theorems 5 & 6 *together* differentiate b vs. e. If Theorem 4 holds, we have a complete characterization of all the vertices of $I_{\sigma,\rho}$; Theorems 5 & 6 are not enough to characterize the vertices (purple, f) that are not in B_σ or B_ρ.

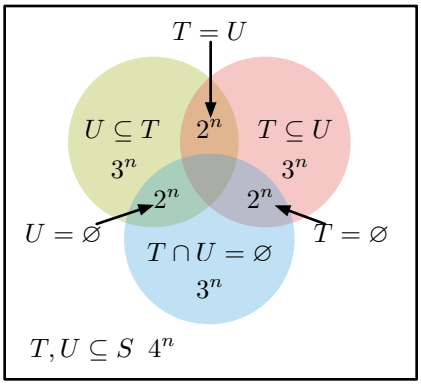


Figure 2.5: Relationship amongst Theorems 3, 4, 5, and 6: Specializing the cross-inequality (2.19) of Theorem 4 for $T = U$ gives Theorem 3, for $T \subseteq U$ gives Theorem 5, and for $U \subseteq T$ gives Theorem 6.

Figure 2.5.

We now show that characterizing $B_\sigma \subset I_{\sigma,\rho}$ only requires checking 2^n inequalities, as opposed to the 3^n inequalities of Theorem 5.

Theorem 7.

$$\sigma(T) - \sigma(T \setminus U) \leq \rho(U) \; \forall U \subseteq T \subseteq S \tag{2.35}$$

if and only if

$$\sigma(S) - \sigma(S \setminus U) \leq \rho(U) \; \forall U \subseteq S. \tag{2.36}$$

Proof. The set of inequalities in (2.35) include (2.36), so the one direction is immediate.

By the supermodularity of σ, we have

$$\sigma(T) + \sigma(S \setminus U) \leq \sigma(S) + \sigma(T \setminus U) \; \forall U \subseteq T \subset S \tag{2.37}$$

which we rearrange to get

$$\sigma(T) - \sigma(T \setminus U) \leq \sigma(S) - \sigma(S \setminus U) \leq \rho(U) \; \forall U \subseteq T \subset S. \tag{2.38}$$

□

While (2.29) (and (2.35)) are readily seen as weaker versions of (2.19), the relationship between (2.36) and (2.19) is not immediate. As before, a similar result holds for characterizing $B_\sigma \subset I_{\sigma,\rho}$.

Theorem 8.

$$\sigma(U) \leq \rho(T) - \rho(T \setminus U) \; \forall \, U \subseteq T \subseteq S \tag{2.39}$$

if and only if

$$\sigma(U) \leq \rho(S) - \rho(S \setminus U) \; \forall \, U \subseteq S. \tag{2.40}$$

Proof. The set of inequalities in (2.39) include (2.40), so the one direction is immediate.

By the submodularity of ρ, we have

$$\rho(S) + \rho(T \setminus U) \leq \rho(T) + \rho(S \setminus U) \; \forall U \subseteq T \subset S \tag{2.41}$$

which we rearrange to get

$$\sigma(U) \leq \rho(S) - \sigma(S \setminus U) \leq \rho(T) - \rho(T \setminus U) \; \forall U \subseteq T \subset S. \tag{2.42}$$

□

One advantage of Theorems 7 and 8 (besides the exponential reduction in inequalities), is the intuitive geometric interpretation of (2.36) and (2.40). For a given supermodular set function σ, let $\bar{\sigma}(U) = \sigma(S) - \sigma(S \setminus U)$, which is a submodular set function. We then have (combining Theorems 5 and 7) $B_\sigma \subset I_{\sigma,\rho}$ if and only if $\bar{\sigma}(U) \leq \rho(U)$ for all $U \subset S$. Equivalently the polymatroid $P_{\bar{\sigma}}$ is a subset of the polymatroid P_ρ, as depicted in Figure 2.6. With a similar argument combining Theorems 6 and 8 we have $B_\sigma \subset I_{\sigma,\rho}$ if and only if the contrapolymatroid $Q_{\bar{\rho}}$ is a subset of the contrapolymatroid Q_σ.

We now specialize (2.36) for the case of conditional entropy σ_{SW} and min-cut capacity ρ_c

$$H(X_U) \leq \rho_c(U) \quad \forall U \subseteq S, \tag{2.43}$$

Figure 2.6: Geometric interpretation of Theorems 7 and 8. Since $B_\sigma = B_{\bar{\sigma}}$, $B_\sigma \subseteq P_\rho$ if and only if $B_{\bar{\sigma}} \subseteq P_\rho$. A similar argument holds for B_ρ.

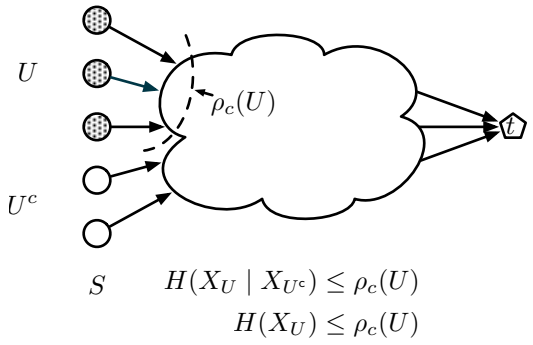


Figure 2.7: Specializing (2.9) to conditional entropy and min-cut capacity requires the network to have enough capacity $\rho_c(U)$ to support the lowest sum-rate $H(X_U \mid X_{U^c})$ from every subset of sources. Specializing (2.36) to conditional entropy and min-cut capacity requires the network to have enough capacity $\rho_c(U)$ to support the "highest" sum-rate $H(X_U)$ from every subset of sources. The sum-rate from a subset of sources could exceed the entropy, but is not needed to support lossless recovery of the sources and would be a suboptimal rate allocation.

which follows from the application of the chain rule for entropy to $H(X_S) - H(X_{S \setminus U} \mid X_U)$. Figure 2.7 illustrates the differences between Theorem 3 and Theorems 7 and 8. Consider a set of sources U. Han's matching condition (2.9) requires the network have enough capacity to support the best-case sum-rate (i.e., minimum) from a set of sources for lossless recovery; in particular $H(X_U \mid X_{U^c}) \leq \rho_c(U)$. The matching condition (2.36) of Theorem 7 requires that the network have enough capacity to support the worst-case sum-rate (i.e., maximum) for all subsets of sources.

Unlike Theorem 3, when we specialize (2.19) to the case of conditional entropy and min-cut capacity there is no immediately obvious intuition for what (2.44) represents.

$$H(X_{U \cap T} \mid X_{U^c}) \leq \rho_c(T) - \rho_c(T \setminus U) \tag{2.44}$$

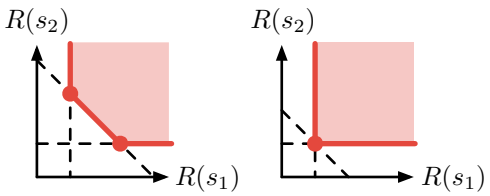


Figure 2.8: Comparison of non-degenerate and degenerate vertices: at a non-degenerate vertex (left), $|S|$ constraints will be active; a degenerate vertex (right) will have more than $|S|$ active constraints.

2.3.2 Properties from conditional entropy

In the previous section, we focused on the properties of general submodular and supermodular set functions in order to more fully characterize the intersection of a polymatroid with a contrapolymatroid. Our next set of results leverage additional properties of the conditional entropy supermodular set function, most notably the chain rule for entropy and the relationship between entropy and conditional independence [22]. To motivate the results of this section, consider the two source SW rate regions in Figure 2.8. In general, the rate region is defined by three inequalities as in Figure 2.8 (left); however, if it is known that the sources are independent (i.e., $X_1 \perp X_2$), the rate region can be defined using only two inequalities. The number of vertices has also been reduced from two non-degenerate vertices to one *degenerate* vertex. We further develop this insight in the remainder of this section.

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 of the LP (2.11) at R.

Lemma 4. *Fix an ordering $s_1, s_2, \ldots, s_n$ 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 (cf. (2.12)) and $U = \{s_{k_1}, \cdots, s_{k_m}\}$ such that $k_1 < k_2 < \ldots < k_m$ then*

$$\begin{aligned} R(U) &= H(X_U | X_{U_{k_1}^c \setminus U}) \\ &\quad + \sum_{j=2}^{m} I(X_{U \setminus U_{k_{j-1}}}; X_{U_{k_j} \setminus U_{k_{j-1}}} | X_{U_{k_j}^c \setminus U}) \end{aligned} \tag{2.45}$$

Recall from Lemma 4 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 5.

$$U_{k_j}^{\mathsf{c}} \setminus U' = U_{k_j}^{\mathsf{c}} \setminus U \tag{2.46}$$

Proof.

$$\begin{aligned} U_{k_j}^{\mathsf{c}} \setminus U &= U_{k_j}^{\mathsf{c}} \cap (\{s_{k_1}\} \cup U')^{\mathsf{c}} \\ &= U_{k_j}^{\mathsf{c}} \cap (\{s_{k_1}\}^{\mathsf{c}} \cap U'^{\mathsf{c}}) \\ &= U_{k_j}^{\mathsf{c}} \cap U'^{\mathsf{c}} \end{aligned} \tag{2.47}$$

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

Lemma 6.

$$U' = U \setminus U_{k_1} \tag{2.48}$$

Proof.

$$U \setminus U_{k_1} = U \cap \{s_{k_1+1}, s_{k_1+2}, \ldots, s_n\} = \{s_{k_2}, \cdots, s_{k_m}\} = U' \tag{2.49}$$

□

Lemma 7.

$$U^{\mathsf{c}} = U_{k_1-1} \cup U_{k_1}^{\mathsf{c}} \setminus U \tag{2.50}$$

Proof.

$$\begin{aligned} U_{k_1-1} \cup U_{k_1}^{\mathsf{c}} \setminus U &= U_{k_1-1} \cup (U_{k_1}^{\mathsf{c}} \cap U^{\mathsf{c}}) \\ &= (U_{k_1-1} \cup U_{k_1}^{\mathsf{c}}) \cap (U_{k_1-1} \cup U^{\mathsf{c}}) \\ &= \{s_{k_1-1}\}^{\mathsf{c}} \cap U^{\mathsf{c}} \\ &= U^{\mathsf{c}} \end{aligned} \tag{2.51}$$

□

Proof of Lemma 4 . Proof by induction on $|U|$. *Base case:* If $|U| = 1$, then $U = \{s_{k_1}\}$ and we have that

$$\begin{aligned} R(s_{k_1}) &= H(X_{U_{k_1}} \mid X_{U_{k_1}^{\mathsf{c}}}) - H(X_{U_{k_1-1}} \mid X_{U_{k_1-1}^{\mathsf{c}}}) \\ &= H(X_{U_{k_1-1}}, X_{s_{k_1}} \mid X_{U_{k_1}^{\mathsf{c}}}) - H(X_{U_{k_1-1}} \mid X_{U_{k_1}^{\mathsf{c}}}, X_{s_{k_1}}) \\ &= H(X_{s_{k_1}} \mid X_{U_{k_1}^{\mathsf{c}}}) \\ &= H(X_{s_{k_1}} \mid X_{U_{k_1}^{\mathsf{c}} \setminus \{s_{k_1}\}}) \end{aligned} \tag{2.52}$$

where the last step follows from the fact that $U_i^{\mathsf{c}} = U_i^{\mathsf{c}} \setminus \{s_i\}$.

Inductive step: Let us define

$$U' \triangleq U \setminus \{s_{k_1}\} = \{s_{k'_1}, \cdots, s_{k'_{m'}}\} \tag{2.53}$$

where $k'_i = k_{i+1}$ and $m' = m - 1$. We have that

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

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

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) $U_{k_1'}^{\mathsf{c}} \setminus U' = U_{k_2}^{\mathsf{c}} \setminus U'$; (e) $U' \subseteq U_{k_1}^{\mathsf{c}}$ so partition $U_{k_1}^{\mathsf{c}}$ into $U_{k_1}^{\mathsf{c}} \setminus U'$ 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. □

Proposition 1. *Fix an ordering $s_1, s_2, \ldots, s_n$ 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 $k_1 < k_2 < \ldots < k_m$. Define $k_0 \triangleq 0$. If $U = U_i$ for some $i \in [n]$, then $R(U) = H(X_U|X_{U^{\mathsf{c}}})$. If $U \neq U_i$ for some i, then $R(U) = H(X_U|X_U^{\mathsf{c}})$ if and only if*

$$(X_{U\setminus U_{k_{j-1}}} \perp X_{U_{k_j-1}\setminus U_{k_{j-1}}})|X_{U_{k_j}^{\mathsf{c}}\setminus U} \quad j = 1, \ldots, m. \tag{2.54}$$

Proof. From the Lemma 4, we have that

$$\begin{aligned}
0 &\leq R(U) - H(X_U|X_{U^c}) \\
&= H(X_U|X_{U^c_{k_1}\setminus U}) - H(X_U|X_{U^c}) \\
&+ \sum_{j=2}^{m} I(X_{U\setminus U_{k_{j-1}}}; X_{U_{k_j-1}\setminus U_{k_{j-1}}}|X_{U^c_{k_j}\setminus U}) \\
&= H(X_U|X_{U^c_{k_1}\setminus U}) - H(X_U|X_{U_{k_1-1}}, X_{U^c_{k_1}\setminus U}) \\
&+ \sum_{j=2}^{m} I(X_{U\setminus U_{k_{j-1}}}; X_{U_{k_j-1}\setminus U_{k_{j-1}}}|X_{U^c_{k_j}\setminus U}) \\
&= I(X_U; X_{U_{k_1-1}}|X_{U^c_{k_1}\setminus U}) \\
&+ \sum_{j=2}^{m} I(X_{U\setminus U_{k_{j-1}}}; X_{U_{k_j-1}\setminus U_{k_{j-1}}}|X_{U^c_{k_j}\setminus U}) \\
&= \sum_{j=1}^{m} I(X_{U\setminus U_{k_{j-1}}}; X_{U_{k_j-1}\setminus U_{k_{j-1}}}|X_{U^c_{k_j}\setminus U}).
\end{aligned} \tag{2.55}$$

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 (2.54). □

Proposition 2. *Let T, U, V, W, Y be a partition of S and let π_T be a permutation of T etc. Define $\pi = (\pi_T, \pi_U, \pi_V, \pi_W, \pi_Y)$ to be the permutation formed from the permutations of the associated partition and $\pi' = (\pi_T, \pi_V, \pi_U, \pi_W, \pi_Y)$. If $X_U \perp X_V \mid X_W$, then $R_\pi = R_{\pi'}$.*

Proof. Denote the elements of S ordered by π as

$$t_1, \ldots, t_{|T|}, u_1, \ldots, u_{|U|}, v_1, \ldots, v_{|V|}, w_1, \ldots, w_{|W|}, y_1, \ldots, y_{|Y|}.$$

Then the elements of S ordered by π' is

$$t_1, \ldots, t_{|T|}, v_1, \ldots, v_{|V|}, u_1, \ldots, u_{|U|}, w_1, \ldots, w_{|W|}, y_1, \ldots, y_{|Y|}. \tag{2.56}$$

We can show the following

$$\begin{aligned} R_\pi(t_i) &= H(X_{t_i} \mid X_{t_{i+1}}, \ldots, X_{t_{|T|}}, X_U, X_V, X_W, X_Y) \\ &= H(X_{t_i} \mid X_{t_{i+1}}, \ldots, X_{t_{|T|}}, X_V, X_U, X_W, X_Y) \\ &= R_{\pi'}(t_i) \end{aligned} \tag{2.57}$$

$$\begin{aligned} R_\pi(u_i) &= H(X_{t_i} \mid X_{u_{i+1}}, \ldots, X_{u_{|U|}}, X_V, X_W, X_Y) \\ &= H(X_{t_i} \mid X_{u_{i+1}}, \ldots, X_{u_{|U|}}, X_W, X_Y) \\ &= R_{\pi'}(u_i) \end{aligned} \tag{2.58}$$

$$\begin{aligned} R_\pi(v_i) &= H(X_{v_i} \mid X_{v_{i+1}}, \ldots, X_{v_{|V|}}, X_W, X_Y) \\ &= H(X_{v_i} \mid X_{v_{i+1}}, \ldots, X_{v_{|V|}}, X_U, X_W, X_Y) \\ &= R_{\pi'}(v_i) \end{aligned} \tag{2.59}$$

$$\begin{aligned} R_\pi(w_i) &= H(X_{w_i} \mid X_{w_{i+1}}, \ldots, X_{w_{|W|}}, X_Y) \\ &= R_{\pi'}(w_i) \end{aligned} \tag{2.60}$$

$$\begin{aligned} R_\pi(y_i) &= H(X_{y_i} \mid X_{y_{i+1}}, \ldots, X_{w_{|Y|}}) \\ &= R_{\pi'}(Y_i) \end{aligned} \tag{2.61}$$

□

Corollary 2. *Let T, U, V, W be a partition of S and let π_T be a permutation of T etc. Define $\pi = (\pi_T, \pi_U, \pi_V, \pi_W)$ to be the permutation formed from the permutations of the associated partition and $\pi' = (\pi_T, \pi_V, \pi_U, \pi_W)$. If $X_U \perp X_V$, then $R_\pi = R_{\pi'}$.*

A polyhedron can be represented as the intersection of half-spaces (*H-rep*) or as the convex combination of its extreme points plus the conic combination of its extreme rays (*V-rep*)[23]. In general, it is more compact to represent polymatroids and contrapolymatroids using half-spaces ($2^{|S|}$) than in terms of the extreme points and extreme rays ($\mathcal{O}(|S|!)$). What the previous two

propositions show is that the size of the representation of the Slepian-Wolf rate region is directly tied to the conditional independence structure of the distributed correlated sources. In turn, this means that the number of inequalities in the LP (2.11) depends on the conditional independence structure of the sources and may have a polynomial (in $|S|$, $|V|$, and $|A|$) number of constraints. For example, if all the sources are independent, then only $|S|$ inequalities are needed to describe the Slepian-Wolf rate region.

Example 2. Consider the following three source discrete memoryless source (DMS):

$$\mathbb{P}(X_1 = x_1, X_2 = x_2) = \begin{cases} \dfrac{1-p}{2} & x_1 = x_2 \\ \dfrac{p}{2} & x_1 \neq x_2 \end{cases} \tag{2.62}$$

and

$$\mathbb{P}(X_3 = 0 \mid X_2 = 0) = \mathbb{P}(X_3 = 1 \mid X_2 = 1) = 1 - q \tag{2.63}$$

with $p, q \neq \frac{1}{2}$. Such a DMS forms the Markov chain $X_1 \leftrightarrow X_2 \leftrightarrow X_3$. For the permutation $\pi = (3, 1, 2)$, we have the three necessarily active constraints

$$R_\pi(s_3) = H(X_3 \mid X_1, X_2) \tag{2.64a}$$

$$R_\pi(\{s_1, s_3\}) = H(X_1, X_3 \mid X_2) \tag{2.64b}$$

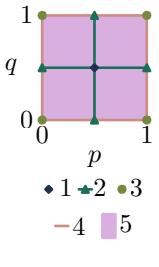
$$R_\pi(\{s_1, s_2, s_3\}) = H(X_1, X_2, X_3). \tag{2.64c}$$

Additionally, because of the Markov structure for this source we have

$$R_\pi(s_1) = H(X_1 \mid X_2) = H(X_1 \mid X_2, X_3) \tag{2.65}$$

active at R_π. Enumerating all of the vertices, we see that the Slepian-Wolf rate region for this *class* of DMSs has only five vertices instead of $3! = 6$. In particular, the permutations $\pi_1 = (1, 3, 2)$ and $\pi_2 = (3, 1, 2)$ map to the same point, i.e., $R_{\pi_1} = R_{\pi_2}$.

If $p = \frac{1}{2}$ or (exclusively) $q = \frac{1}{2}$, then the SW rate region will only have four vertices. If $p = q = \frac{1}{2}$,



p	q	Independence Relationships
$\neq \frac{1}{2}$	$\neq \frac{1}{2}$	$X_1 \perp X_3 \mid X_2$
$\neq \frac{1}{2}$	$= \frac{1}{2}$	$X_1 \perp X_3,\ X_2 \perp X_3$
$= \frac{1}{2}$	$\neq \frac{1}{2}$	$X_1 \perp X_2,\ X_1 \perp X_3$
$= \frac{1}{2}$	$= \frac{1}{2}$	$X_1 \perp X_2,\ X_1 \perp X_3,\ X_2 \perp X_3$

Figure 2.9: The number of vertices the Slepian-Wolf rate region has for the source of Example 2 as a function of the parameters p and q. Depending on the parameter p and q, certain conditional independence relationships hold leading to a reduction in the number of vertices.

then the SW rate region will have one vertex. Figure 2.9 shows the number of vertices that SW rate region has and summarizes the different conditional independence structures as a function of the distribution parameters p and q. □

Proposition 1 will be used in the next section when giving conditions for a feasible solution of the optimization problem (2.11) to be optimal.

2.4 Sufficient Conditions for Characterizing Optimality

We proceed by finding the dual LP of the primal given in (2.11). In (2.11), 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

$$\begin{aligned} \underset{x\leq 0, y\geq 0, z}{\text{maximize}} \quad & \sum_{a\in A} c(a)x(a) + \sum_{U\subseteq S} H(X_U|X_{U^c})y_U \\ \text{subject to} \quad & x(a) + z(\text{head}(a)) - z(\text{tail}(a)) \leq k(a) \quad a \in A \\ & \sum_{U\ni s} y_U + z(s) - z(t) = h(s) \qquad\qquad s \in S \end{aligned} \tag{2.66}$$

We set $z(t) = 0$ because it is associated with the conservation of flow constraint at the sink, which is omitted from (2.11) 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

$$\bar{k}(a) \triangleq k(a) - (z(\text{head}(a)) - z(\text{tail}(a))) \tag{2.67}$$

and observe that the first set of constraints of (2.66) can be written as $x(a) \leq \bar{k}(a)$ for all $a \in A$ [24]. Combined with the non-positivity constraint on $x(a)$ we have $x(a) \leq \min(0, \bar{k}(a))$. Since we are maximizing in (2.66) and $c(a) > 0$ for all a, we take

$$x(a) = \min(0, \bar{k}(a)) \tag{2.68}$$

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 9. *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*

$$\bar{k}(a) < 0 \implies f^*_{R_i}(a) = c(a) \tag{2.69a}$$

$$\bar{k}(a) > 0 \implies f^*_{R_i}(a) = 0. \tag{2.69b}$$

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

Lemma 8. *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.*

Proof. Omitted for brevity. □

This is a restatement of and proof of Theorem 9.

Theorem 10. *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*

$$x^*(a)(f^*_{R_i}(a) - c(a)) = 0 \tag{2.70a}$$

$$(k(a) - x^*(a) - (z^*(\text{head}(a)) - z^*(\text{tail}(a))))f^*_{R_i}(a) = 0 \tag{2.70b}$$

for all $a \in A$.

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

$$\begin{aligned} \underset{f \geq 0}{\text{minimize}} \quad & \sum_{a \in A} k(a) f(a) \\ \text{subject to} \quad & f(a) \leq c(a) && a \in A \\ & f(\delta^{in}(v)) - f(\delta^{out}(v)) = 0 && v \in N \\ & f(\delta^{in}(v)) - f(\delta^{out}(s)) = -R_i(s) && s \in S \end{aligned} \tag{2.71}$$

and its corresponding dual

$$\begin{aligned} \underset{x \leq 0, z}{\text{maximize}} \quad & \sum_{a \in A} c(a)x(a) - \sum_{s \in S} R_i(s) z(s) \\ \text{subject to} \quad & x(a) + z(\text{head}(a)) - z(\text{tail}(a)) \leq k(a) \quad a \in A. \end{aligned} \tag{2.72}$$

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 8, that f is primal feasible. Checking the

complimentary slackness conditions for f, x^*, and z^*, we have

$$\begin{aligned} x^*(a)(f(a)-c(a)) &= x^*(a)\left(\sum_i \lambda_i f^*_{R_i}(a) - c(a)\right) \\ &= \sum_i \lambda_i \left(x^*(a)(f^*_{R_i}(a)-c(a))\right) \\ &= 0 \end{aligned} \tag{2.73}$$

and similarly

$$(k(a)-x^*(a)-(z^*(\text{head}(a))-z^*(\text{tail}(a))))f^*_{R_i}(a)=0. \tag{2.74}$$

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

Since we are considering fixed rates 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 extreme points of $Q_{\sigma_{SW}}$ are feasible for (2.11). As was mentioned earlier, if (f^*, R^*) is an optimal solution to (2.11) then $R^*(S) = H(X_S)$ [8] 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. In general though, this is not always the case as the next example demonstrates.

Example 3. We consider the relay network with arc capacities and costs as shown in Figure 2.10a. Let the sources X_1, X_2 be binary valued with the following joint distribution

$$\mathbb{P}(X_1=x_1, X_2=x_2) = \begin{cases} \dfrac{1-p}{2} & x_1 = x_2 \\ \dfrac{p}{2} & x_1 \neq x_2 \end{cases}. \tag{2.75}$$

For such a source, the entropies are $H(X_1) = H(X_2) = 1$ and $H(X_1, X_2) = 1 + H(p)$ and the vertices of the Slepian-Wolf rate region are $R_1 = (H(p), 1)$ and $R_2 = (1, H(p))$ [3]. The network of Figure 2.10a has sufficient capacity to support either R_1 or R_2. Fixing $R = R_1$ and solving for the

[3] $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$

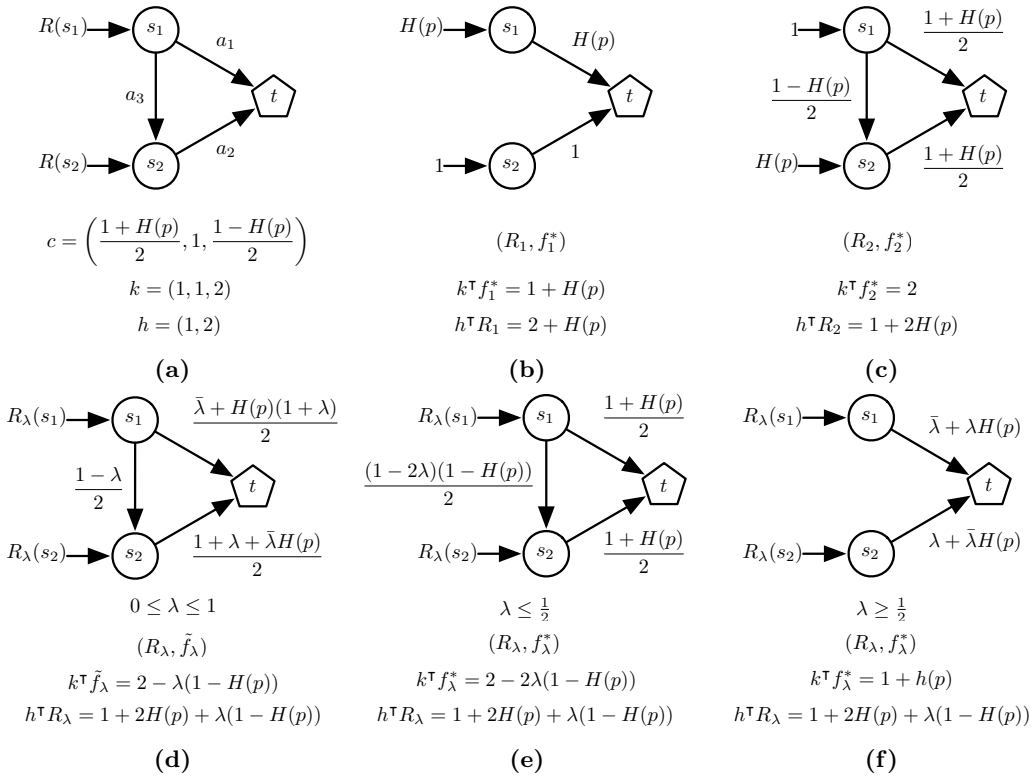


Figure 2.10: Relay network example: a network topology with arc capacities as a function of the source parameter p, arc costs, and source costs; b optimal min-cost flow when the source rates are fixed as $R(s_1) = H(p)$ and $R(s_2) = 1$; c optimal min-cost flow when the source rates are fixed as $R(s_1) = 1$ and $R(s_2) = H(p)$; d convex combination of b and c; e optimal min-cost flow for convex combination of source rates in b and c for $\lambda \le \frac{1}{2}$, and; f optimal min-cost flow for convex combination of source rates in b and c for $\lambda \ge \frac{1}{2}$.

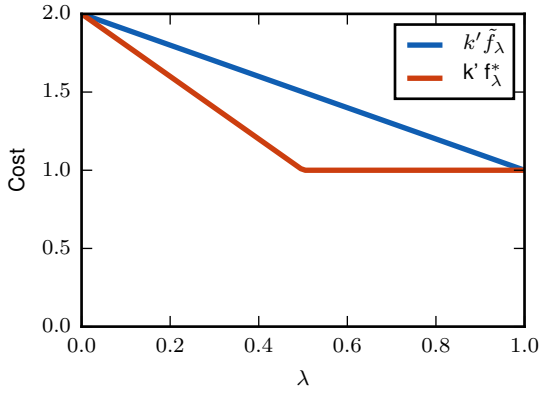


Figure 2.11: Plot of cost of a convex combination of min-cost flows and the cost of a min-cost flow for a convex combination of rates. The source distribution parameter $p = \frac{1}{2}$.

min-cost flow f_1^*, we obtain the solution shown in Figure 2.10b; correspondingly, if we fix $R = R_2$ and solve for the min-cost flow f_2^*, we obtain the solution shown in Figure 2.10c. The feasible solution $\tilde{f}_\lambda = \lambda f_1^* + (1-\lambda) f_2^*$ for $R_\lambda = \lambda R_1 + (1-\lambda) R_2$ is shown in Figure 2.10d. Comparing with optimal min-cost flow f_λ^* for R_λ shown in Figure 2.10e & Figure 2.10f, we see that for $\lambda \in (0,1)$, the convex combination of min-cost flows $\tilde{f}_\lambda$ is *not* a min-cost flow for R_λ. Shown in Figure 2.11 is the cost $k^\intercal \tilde{f}_\lambda$ of the convex combination of min-cost flows as a function of λ compared to the cost $k^\intercal f_\lambda^*$ for the min-cost flow for a convex combination of rates R_λ. Comparing Figure 2.10d with Figure 2.10e & Figure 2.10f, we see immediately why $\tilde{f}_\lambda$ is not optimal: $\tilde{f}_\lambda$ always utilizes the arc a_3 even when arc a_1 (which has a lower cost) has spare capacity. If the same relay network is consider with $k = (2,1,1)$ and all other parameters kept the same, then it can be shown that $k^\intercal \tilde{f}_\lambda = k^\intercal f_\lambda^*$ for $\lambda \in [0,1]$. Observe that with this cost vector k, the cost of the two directed paths $s_1 \to t$ are $k(a_1) = 2$ and $k(a_3) + k(a_2) = 2$ and the cost of any flow supporting $R(s_1)$ is the same. □

A sufficient condition for the existence of z that satisfies the condition of Theorem 9 can be given in terms of the topology of the network and the arc costs k.

Theorem 11. *If for every $v \in V \setminus \{t\}$, the cost of all $v-t$ paths are equal, then there exists a vector*

$(z(v) : v \in V)$ *such that for*

$$\bar{k}(a) < 0 \implies f^*_{R_i}(a) = c(a) \quad \forall i \tag{2.76a}$$

$$\bar{k}(a) > 0 \implies f^*_{R_i}(a) = 0 \quad \forall i. \tag{2.76b}$$

Proof. Define $\mu(v) \triangleq \min_P k(P)$ be the value of a min-cost $v - t$ path in the network and let $v \rightsquigarrow t$ indicate a $v - t$ path. Let $\hat{x}(a) = 0$ for all $a \in A$ and $\hat{z} = -\mu(v)$ for all $v \in V \setminus \{t\}$. At $\hat{x}, \hat{z}$, the constraints of (2.66) are equivalent to

$$\mu(\text{head}(a)) + k(a) \geq \mu(\text{tail}(a)). \tag{2.77}$$

Observe that $\text{tail}(a) \to \text{head}(a) \rightsquigarrow t$ is a directed $\text{tail}(a) - t$ path of cost $\mu(\text{head}(a)) + k(a)$ and therefore $\hat{x}, \hat{z}$ is dual feasible. Furthermore, for every $u \in V \setminus \{t\}$, there exists $v \in V$ such that $(u, v) \in A$ and $\mu(u) = \mu(v) + k((u, v))$. This means that there are at least $|A| + |V| - 1$ *active* constraints at $\hat{x}, \hat{z}$ and it is a vertex. In fact, for a given $v \in V$ all $v - t$ paths have the same cost, *all* constraints are active and $\hat{x}, \hat{z}$ is the *only* vertex of the dual feasible set. The dual feasible set is identical for all choices of source rates R and therefore the optimal solution is give by $x^* = \hat{x}$ and $z^* = \hat{z}$. □

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

$$\bar{h}(s) \triangleq h(s) - (z(s) - z(t)) = h(s) - z(s) \tag{2.78}$$

and rewrite the second set of constraints of (2.66) as

$$\sum_{U \ni s} y_U = \bar{h}(s). \tag{2.79}$$

We seek to express the dual variables y_U as a 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 12. *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*

$$(X_{U \setminus U_{k_{j-1}}} \not\perp X_{U_{k_{j-1}} \setminus U_{k_{j-1}}}) | X_{U_{k_j}^c \setminus U} \tag{2.80}$$

then $y_U^* = 0$.

Proof. Follows immediately from complimentary slackness and Proposition 1. □

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 the LP in (2.11).

Theorem 13. *A feasible solution (f_π^*, R_π) of (2.11) is optimal if there exists vectors $(z(v) : v \in V)$ satisfying*

$$\bar{k}(a) < 0 \implies f_\pi^*(a) = c(a) \quad \forall a \in A \tag{2.81a}$$

$$\bar{k}(a) > 0 \implies f_\pi^*(a) = 0 \quad \forall a \in A \tag{2.81b}$$

and

$$\bar{h}(s_1) \geq \bar{h}(s_2) \geq \cdots \geq \bar{h}(s_n) \geq 0 \tag{2.82}$$

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

$$y_{U_i} = \begin{cases} \bar{h}(s_i) - \bar{h}(s_{i+1}) & i \in [n-1] \\ \bar{h}(s_i) & i = n. \end{cases} \tag{2.83}$$

We construct a dual feasible x from (2.68). Having primal feasible (f_π^*, R_π) and dual feasible (x, y, z), optimality follows from complimentary slackness. □

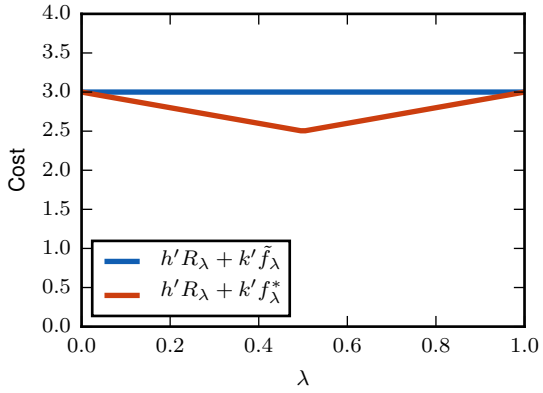


Figure 2.12: Plot of i) source cost plus cost of a convex combination of min-cost flows and ii) source cost plus the cost of a min-cost flow for a convex combination of rates. The source distribution parameter $p = \frac{1}{2}$.

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 (2.68) and $(y^*_U : U \subseteq S)$ according to (2.83). 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. By adding in a per source cost to the previous example, we demonstrate that such a z need not always exist.

Example 4. Looking at Figure 2.11, one might conjecture that an optimal solution (f^*, R^*) to problem in (2.11) would have the property $R^* \in \text{Ext}(B_{\sigma_{SW}})$; i.e., that an optimal rate will always coincide with a vertex of the Slepian-Wolf rate region and that z that satisfies the condition of the previous theorem will always exist. It is certainly true that f^*, R^* will be a vertex of the polyhedron in flow-rate space, the network of Figure 2.10 example demonstrates that for certain choices of source costs h and arc costs k the optimal rate R^* need *not* be a vertex of the Slepian-Wolf rate region. Figure 2.12 shows the cost $h^\intercal R_\lambda + k^\intercal \tilde{f}_\lambda$ of the convex combination of min-cost flows as a function of λ compared to the cost $h^\intercal R_\lambda + k^\intercal f^*_\lambda$ for the min-cost flow for a convex combination of rates R_λ. We see immediately the minimum cost is achieved with $\lambda = \frac{1}{2}$ and $R^*_\lambda \notin \text{Ext}(B_{\sigma_{SW}})$. □

Given the intuitive decomposition of the source coding and routing into different protocol layers noted by Barros et al., it may appear at first glance that a simple decomposed approach to designing

a minimum cost solution might hold [10]. For the case where all extreme points of the Slepian-Wolf rate region are feasible, one might consider a naïve approach of finding a minimum cost (w.r.t. h) source rate $\hat{R}$ and a supporting minimum cost (w.r.t. k) $\hat{f}$. Alternatively, one might try enumerating all extreme points of the Slepian-Wolf rate region (combinatorial complexity aside), solving for a min-cost flow, and keeping track of the best solution. The problem with both of these approaches is that the resulting feasible solution will select a rate $\hat{R}$ that coincides with an extreme point of the Slepian-Wolf rate region. The previous example demonstrates that when there is a imbalance between source costs and flow cost (i.e., cheap compression and expensive routing vs. expensive compression and cheap routing) the optimal rate $R^* \notin \text{Ext}(Q_{\sigma_{SW}})$.

2.5 Extensions to Multiple Sinks

In previous sections, we have focused our attention on the single-sink problem. In many contexts, it may be necessary to recover the source X_S at multiple sinks T. As mentioned earlier, this problem was considered by Ramamoorthy [11]. When there are multiple sinks, routing is no longer sufficient for conveyance of the sources to the sinks; instead network coding is necessary. In the general network coding case, the single flow variable on each edges is replaced by virtual flows, one for each each edge and the traffic carried is represented with a physical flow [16]. Finally, Ramamoorthy augments the original graph by adding in a super source s^* and connecting this vertex to each of the sources $s \in S$ with an edge of zero cost and capacity given by the entropy of the source $H(X_s)$. With this augmentation, the multi-sink linear program can be written as:

$$
\begin{aligned}
&\underset{f,p,R}{\text{minimize}} \qquad \sum_{a \in A} k(a)p(a) \\
&\text{subject to} \\
&0 \leq f^{(t)}(a) \leq p(a) \leq c^*(a) && a \in A, t \in T \\
&f^{(t)}(\delta^{in}(v)) - f^{(t)}(\delta^{out}(v)) = \Delta^{(t)}(v) && v \in V \\
&x^{(t)}((s^*, s)) \geq R^{(t)}(s) && s \in S, t \in T \\
&R^{(t)}(U) \geq H(X_U | X_{U^c}) && U \subseteq S, t \in T.
\end{aligned}
\tag{2.84}
$$

where

$$c^*(a) = \begin{cases} c(a) & a \in A \\ H(X_s) & a = (s^*, s) \end{cases} \tag{2.85}$$

and

$$\Delta(v) = \begin{cases} -H(X_S) & v = s^* \\ H(X_S) & v = t \\ 0 & \text{otherwise} \end{cases} \tag{2.86}$$

We can view the multi-sink problem as being the intersection of multiple single-sink problems. Looking at (2.84), we can see that $f^{(t)}$ is a valid flow that supports $R^{(t)}$ for the sink t. We can define a min-cost capacity set function for each of the sinks

$$\rho_c^{(t)}(U) = \min\{c(\delta^{out}(X)) : U \subseteq X, t \in V \setminus X\} \tag{2.87}$$

and we then see that $R^{(t)} \in Q_{\sigma_{SW}} \cap P_{\rho_c^{(t)}}$.

Although the multi-sink scenario requires network coding (as compared to distributed source coding and routing for the single-sink scenario), many of the insights developed for the single-sink case carry over. The first is characterizing the feasibility of (2.84) [6, 11, 12].

Theorem 14 (Han's Matching Condition for Multiple Sinks [12])**.** *The sources X_S are transmissible across the network to the sinks T if and only if*

$$\sigma_{SW}(U) \leq \min_{t \in T} \rho_c^{(t)}(U) \forall U \subseteq S. \tag{2.88}$$

Comparing (2.9) and (2.88), we see that Theorem 14 implies Theorem 3 for every sink $t \in T$. We can extend Theorem 5 in a similar manner.

Theorem 15. *All of the vertices of the Slepian-Wolf rate region $Q_{\sigma_{SW}}$ are feasible for the multi-sink problem* (2.84) *if and only if*

$$H(X_U) \leq \min_{t \in T} \rho_c^{(t)}(U) \ \forall U \subseteq S. \tag{2.89}$$

Proof. The condition (2.89) implies $B_{\sigma_{SW}} \subseteq P_{\rho_c^{(t)}}$ for all $t \in T$, which means that for any vertex of $Q_{\sigma_{SW}}$ and any sink $t \in T$ there exists a supporting virtual flow $f^{(t)}$. If on the other hand (2.89) does not hold, this implies there exists a sink $t \in T$ and a vertex of $Q_{\sigma_{SW}}$ for which no supporting flow $f^{(t)}$ exists. □

We can bound the optimal value of the multi-sink min-cost flow problem in (2.84) in terms of the optimal values for a collection of single-sink min-cost flow problems.

Theorem 16. *Let* (f^*, p^*, R^*) *be an optimal solution to* (2.84) *and* $(\hat{f}^{(t)}, \hat{R}^{(t)})$ *be an optimal solution to the single sink problem for* $t \in T$*. We have*

$$\max_{t \in T} \sum_{a \in A} k(a) \hat{f}^{(t)}(a) \leq \sum_{a \in A} k(a) p^*(a) \tag{2.90}$$

and

$$\sum_{a \in A} k(a) p^*(a) \leq \sum_{a \in A} k(a) \max_{t \in T} \hat{f}^{(t)}(a). \tag{2.91}$$

Proof. Denote an optimal solution to (2.84) as (f^*, p^*, R^*). The optimal value of a min-cost flow for the single sink t is given by

$$\sum_{a \in A} k(a) \hat{f}^{(t)}(a) \tag{2.92}$$

and this must be a lower bound for the multi-sink problem with $t \in T$. Suppose that it was not; then,

$$\begin{aligned} \sum_{a \in A} k(a) \hat{f}^{(t)}(a) &\geq \sum_{a \in A} k(a) p(a) \\ &\geq \sum_{a \in A} k(a) f^{*(t)}(a) \end{aligned} \tag{2.93}$$

which follows from the constraints of (2.84). Note that the virtual flow $f^{*(t)}(a)$ is a feasible solution to the single sink problem for sink t, contradicting the assumption of $\hat{f}^{(t)}$ as an optimal solution to the single sink problem for t. Since (2.92) is a lower bound for every $t \in T$, it must be true for

$$\max_{t \in T} \sum_{a \in A} k(a) \hat{f}^{(t)}(a). \tag{2.94}$$

To produce an upper bound for the optimal value of (2.84), we form a feasible solution from $\left\{(\hat{f}^{(t)}, \hat{R}^{(t)}) : t \in T\right\}$ the set of optimal solution to single-sink problems:

$$f^{(t)}(a) = f^{*(t)}(a) \; \forall a \in A, t \in T \tag{2.95a}$$

$$R^{(t)}(s) = R^{*(t)}(s) \; \forall s \in S, t \in T \tag{2.95b}$$

$$p(a) = \max_t f^{(t)}(a). \tag{2.95c}$$

Finally, the cost of this feasible solution is

$$\sum_{a \in A} k(a) \max_{t \in T} \hat{f}^{(t)}(a). \tag{2.96}$$

□

Example 5. As an example of the above bounds, we consider a multi-sink problem instance formulated by Ramamoorthy [11]. In this scenario, the network consists of 50 nodes and 286 edges; 10 of the nodes are sources and 3 are sinks, with the rest of the nodes acting as relays. The edge capacities are either 20 or 40 depending upon the distance between the connected nodes and the edge costs are all 1. A more detailed description of the network topology and source model can be found in [11], §3-B. Shown in Figure 2.13 is a comparison of the cost of the solution found using a partial dual decomposition and application of the subgradient method, the optimal value as computed from building and solving the LP in (2.84), and the lower and upper bounds of Theorem 16.[4] We observe, for this problem, that neither bound is tight; the lower bound has a relative difference of 47.5% while the upper bound has a relative difference of 34.7%. We see that the subgradient method does converge to the optimal value rather quickly, realizing a relative error of 0.19%, 0.13%, and 0.06% after 10, 100, and 1000 iterations, respectively. □

[4] In Fig. 1 (b) of [11], the optimal value is reported as 646.69 which is different from the optimal value of 573.45 shown in Figure 2.13. Through personal correspondence with the author of [11] it was determined that in computing the solution costs Fig. 1 (b) of [11] non-zero costs were being assigned to the edges between the super source s^* and the sources $s \in S$. Correcting for this gives the subgradient results in Figure 2.13

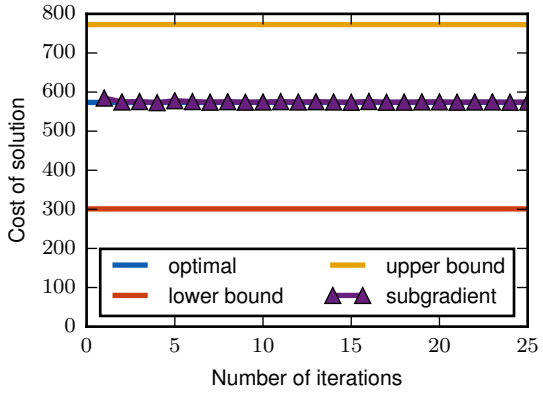


Figure 2.13: Comparison of subgradient method [11] with optimal value and upper and lower bound.

2.6 Conclusion

In this chapter, 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 characterize when the SW vertices are all feasible and give an explicit characterization of all the vertices of the intersection of polymatroid with a contrapolymatroid for certain sub-/supermodular set functions. The size of the representation of the SW rate region is related to the conditional independence relationships among the sources and in some cases may require a sub-exponential number of inequalities to describe the rate region. We have shown that these properties lead to a characterization relating optimal solutions and the corner points of the SW rate region. Through a simple, but natural counter-example we demonstrate that an optimal rate allocation may not be a vertex of the SW rate region. Our result concerning the feasibility of all the SW rate region vertices naturally extends from the single sink problem to the multi-sink setting. The optimal value of the multi-sink is bounded from above and below in terms of the optimal solutions to a collection of related single-sink problems.

Acknowledgment

The authors would like to thank A. Ramamoorthy for graciously providing data utilized in producing Figure 2.13.

Chapter 3: Distributed Scalar Quantization for Resource Allocation

Efficient downlink resource allocation (e.g., subbands in OFDMA/LTE) requires channel state information (e.g., subband gains) local to each user be transmitted to the base station (BS). Lossy encoding of the relevant state may result in suboptimal resource allocations by the BS, the performance cost of which may be captured by a suitable distortion measure. This problem is an indirect distributed lossy source coding problem with the function to be computed representing the optimal resource allocation, and the distortion measuring the cost of suboptimal allocations. In this chapter we investigate the use of distributed scalar quantizers for lossy encoding of state, where the BS wishes to compute the index of the user with the largest gain on each subband. We prove the superiority of a heterogeneous (across users) quantizer design over the optimal homogeneous quantizer design, even though the source variables are i.i.d.

3.1 Introduction

In a multiuser OFDMA system, each of the available subbands may be assigned to a user. Once this assignment is done, the BS can select an appropriate modulation and coding scheme to approach the capacity of the channel on the subband between the BS and the selected user. The BS needs to know, for each subband, the user with the best channel gain and the value of the best gain; the other users' channel gains are irrelevant. Further, if the BS utilizes a *rateless* code on the subbands then it only needs to compute the index of the user with the best channel gain and not the actual value of the channel gain. A distinguishing feature of our approach is to model resource allocation problems as a chief estimating officer (CEO) (i.e., an indirect distributed lossy source coding) problem [25], by capturing the resource allocation decision as a function of the state variables, and representing the cost of suboptimal allocations via a suitable distortion measure. The associated rate-distortion (R-D) function gives the fundamental overhead-performance trade-offs in the resource allocation problem, where the overhead is the sum rate of the messages sent to the BS, and the performance cost is the

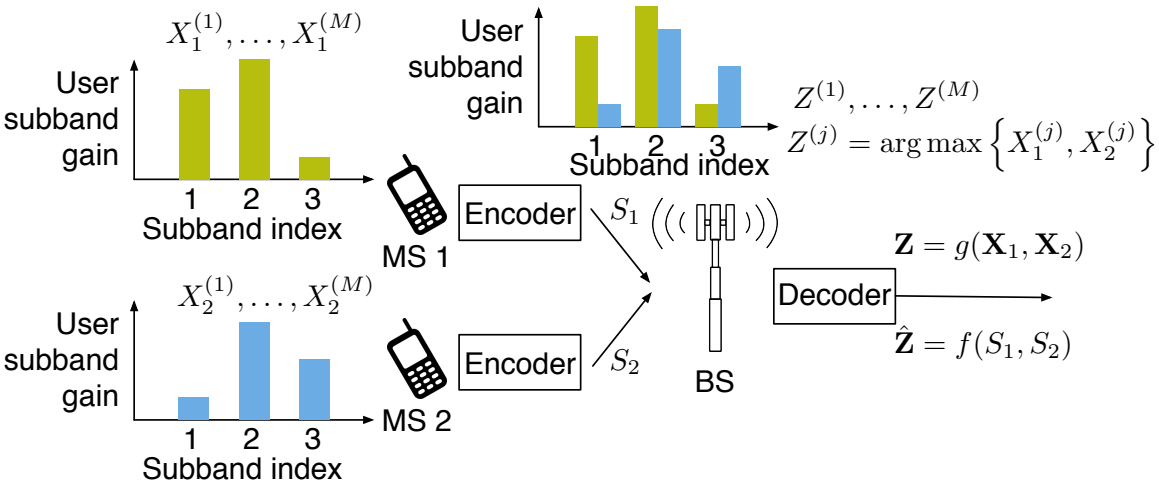


Figure 3.1: Distributed scalar quantization for resource allocation problem overview. The BS wishes to compute the index (arg max) of the user with the largest gain on each subband. The users encode their local gains across subbands using SQs.

subband gain lost due to a suboptimal allocation.

In this work, we consider the design of distributed scalar quantizers (SQs) as an achievable scheme (i.e., realizable code) for encoding state for the allocation of subbands across users. Traditionally, quantizers are designed to minimize the reproduction distortion when using a finite-rate code for a continuous random source. For the resource allocation problem, what needs to be recovered at the decoder is a function of the sources instead of the sources directly. Additionally, the distributed nature of the sources greatly complicates the code design process. For example, if N mobile stations (MSs) were *centrally* located, then $\log_2 N$ bits would suffice to communicate which user has the best channel to the BS.

Recent work by Misra et al. considered the problem of distributed functional scalar quantization (DFSQ) [26]. By focusing on the high-rate regime and assuming a mean squared error (MSE) distortion, the authors are able to make several approximations to obtain distortion expressions that are optimal asymptotically (i.e., as the rate goes to infinity). We assume a different distortion measure, derive an exact expression for the distortion as a function of the quantizer parameters, and derive necessary conditions for optimal parameters. Moreover, our results hold for all rates.

Our focus on the use of SQs as an achievable scheme is motivated by several results concerning

the optimality of a layered architecture of quantization followed by entropy coding. Zamir et al. considered the distributed encoding and centralized decoding of continuous valued sources and established that *lattice* quantization followed by Slepian-Wolf (SW) encoding is optimal asymptotically in rate [27]. When the sources are Gaussian and the distortion is MSE, local vector quantizers followed by SW coding is optimal, not just asymptotically [28]. For discrete valued random variables, *scalar* quantization with block entropy encoding is optimal [29]. Each of the problem models considered in [27–29] can be understood as an instance of indirect distributed lossy source coding for the identity function.

Our main result is a proof of the superiority of a simple and natural heterogeneous scalar quantizer (HetSQ) design over the optimal homogeneous scalar quantizer (HomSQ) design for computing the optimal user index under our given distortion measure. Superiority means the HetSQ achieves the same distortion with lower rate as the best HomSQ. Our result is surprising in that HetSQ are seen to achieve significant rate gains over optimal HomSQ even when the state variables are i.i.d. For a uniform discrete distribution we show numerically that our HetSQ is both significantly better than the optimal HomSQ, and moreover very close to the fundamental limit.

3.2 Problem Model

We focus our attention on the subband allocation problem for a single BS with two MSs as depicted in Figure 3.1. At the i-th MS, the local state $\mathbf{X}_i$ is a vector of downlink channel capacities, which we model as random variables that are i.i.d. across users. Having observed its local state, the i-th MS sends a message $S_i \in \{1, \ldots, 2^{nR_i}\}$ at rate R_i to the BS. If the BS had direct access to the local state at each MS, it could compute the optimal subband allocation $\mathbf{Z} = g(\mathbf{X}_1, \ldots, \mathbf{X}_N)$. Instead, the BS has to compute the subband allocation $\hat{\mathbf{Z}} = f(S_1, \ldots, S_N)$ based on the messages it has received from the MSs. The performance of the system is measured with a distortion function $d(\boldsymbol{Z}, \hat{\boldsymbol{Z}})$.

If we assume the BS utilizes a rateless code then the only information the BS needs is which MS has the better subband; the optimal resource allocation function is naturally

$$Z^{(j)} = \arg\max_i \{X_i^{(j)} : i = 1, \ldots, N\}. \tag{3.1}$$

The BS estimates which user has largest gain on each subband and uses a rateless code (e.g., Hybrid ARQ) to find the coding rate on each subband. This is contrasted with the conventional adaptive modulation and coding (AMC), where the BS needs to estimate both the best user and the max gain on each subband. The distortion is the difference between the max rate for the MS with the actual arg max index and the rate for the MS with estimated arg max

$$d(Z, \hat{Z}) = \frac{1}{M} \sum_{j=1}^{M} \left(X^{(j)}_{Z^{(j)}} - X^{(j)}_{\hat{Z}^{(j)}} \right) \tag{3.2}$$

where M is the number of available subbands. Observe that by definition, the quantity $X^{(j)}_{Z^{(j)}} - X^{(j)}_{\hat{Z}^{(j)}}$ is always non-negative.

3.3 Optimal Scalar Quantizer Design

In this section, we consider the design of SQs as an achievable scheme and compare their performance to computed rate-distortion functions. We first consider the case where all users are using the same quantizer and derive an expression for the resulting distortion. Using this expression, we pose two non-linear optimization problems: first, minimize distortion for a given number of bins, and; second, minimize distortion for a given number of bins subject to a constraint on the entropy of the quantizer output. We provide first order necessary conditions for the optimal quantizer for both non-linear optimizations. We then argue that the same distortion performance can be achieved with a smaller sum rate by utilizing different quantizers at each user. We show that the design of the HetSQ can be accomplished via the same design procedure as for the HomSQ.

Let $X_i : i = 1, \ldots, N$ be the sources for the N users and let Z_A be the index of the user with maximum value. Unlike previous sections, we assume *continous* (instead of discrete) random variables $X_i : i = 1, \ldots, N$. As before, we still assume they are i.i.d. with common PDF $f(x)$, CDF $F(x)$, and support set $\mathcal{X} \subseteq \mathbb{R}_+$.

3.3.1 Homogeneous Scalar Quantizers

Normally, a SQ is specified as a set of *decision boundaries* and *reconstruction levels* [30]. For the estimating the arg max, we do not need the CEO to produce estimates for $X_i : i = 1, \ldots, N$ or even X_{Z_A} (i.e., the value of the maximum source). We can therefore specify the quantizer with just a set of decision boundaries $\{\ell_k : k = 0, \ldots, K\}$ which divide the support set $\mathcal{X}$ into K intervals

$$\mathcal{L}_k = [\ell_{k-1}, \ell_k] \quad k = 1, \ldots, K \tag{3.3}$$

where $\ell_0 \triangleq \inf \mathcal{X}$ and $\ell_K \triangleq \sup \mathcal{X}$. Let $U_i \in \{1, \ldots, K\}$ indicate the interval in which user i's observed value lies. The CEO will pick user i if $U_i > U_{i'}$ for all $i' \neq i$ and will randomly pick a user from $\arg\max_i U_i$ otherwise; we denote the estimate so obtained as $X_{\hat{Z}_A}$.

For notational brevity, we define the following: $E_j \triangleq \mathbb{E}[X \mid \ell_{j-1} \leq X \leq \ell_j]$, $f_j = f(\ell_j)$, $F_j \triangleq F(\ell_j)$, and $p_j \triangleq \mathbb{P}(\ell_{j-1} \leq X \leq \ell_j)$.

Lemma 9. *Let $(X_i : i \in [N])$ be a collection of i.i.d. random variables with cdf $F(x)$ and pdf $f(x)$ and*

$$Z_A \triangleq \{i | X_i = \max\{X_1, \ldots, X_N\}, i \in [N]\}. \tag{3.4}$$

The expected value of the max is

$$\mathbb{E}[X_i | i \in Z_A] = \int_{\inf \mathcal{X}}^{\sup \mathcal{X}} x N F^{N-1}(x) f(x) \, \mathrm{d}x. \tag{3.5}$$

Theorem 17. *Let $(X_i : i \in [N])$ be a collection of i.i.d. random variables with cdf $F(x)$ and pdf $f(x)$ and*

$$Z_A \triangleq \{i | X_i = \max\{X_1, \ldots, X_N\}, i \in [N]\}. \tag{3.6}$$

The expected value of the estimated arg max *when using HomSQs with K intervals is*

$$\mathbb{E}\left[X_{\hat{Z}_A}\right] = \sum_{j=1}^{K} \left[E_j \left(F_j^N - F_{j-1}^N\right)\right]. \tag{3.7}$$

Proof. The optimal Bayes estimator will select one of the users that reports being in the highest interval.

$$n_j \triangleq \sum_{i=1}^{N} \mathbb{1}_{\ell_{j-1} \leq X_i \leq \ell_j}. \tag{3.8}$$

We then have

$$\begin{aligned}
\mathbb{E}\left[X_{\hat{Z}_A}\right] &= \sum_{j=1}^{K} \left[E_j \sum_{\substack{\sum_k^j n_k = N \\ n_j > 0}} \binom{N}{n_1, \ldots, n_j} p_1^{n_1} \cdots p_j^{n_j} \right] \\
&= \sum_{j=1}^{K} \left[E_j \left(\left(\sum_{k=1}^{j} p_k \right)^N - \left(\sum_{k=1}^{j-1} p_k \right)^N \right) \right] \\
&= \sum_{j=1}^{K} \left[E_j \left(F_j^N - F_{j-1}^N \right) \right].
\end{aligned} \tag{3.9}$$

The last step follows from observing

$$\begin{aligned}
\sum_{\sum_k^j n_k = N} \binom{N}{n_1, \ldots, n_j} p_1^{n_1} \cdots p_j^{n_j} = &\sum_{\substack{\sum_k^j n_k = N \\ n_j > 0}} \binom{N}{n_1, \ldots, n_j} p_1^{n_1} \cdots p_j^{n_j} \\
&+ \sum_{\substack{\sum_k^j n_k = N \\ n_j = 0}} \binom{N}{n_1, \ldots, n_j} p_1^{n_1} \cdots p_j^{n_j};
\end{aligned} \tag{3.10}$$

rearranging and applying the multinomial theorem yeilds

$$\sum_{\substack{\sum_k^j n_k = N \\ n_j > 0}} \binom{N}{n_1, \ldots, n_j} p_1^{n_1} \cdots p_j^{n_j} = \left(\sum_{k=1}^{j} p_k \right)^N - \left(\sum_{k=1}^{j-1} p_k \right)^N. \tag{3.11}$$

□

Recall that for a collection of i.i.d. random variables $X_i : i \in [N]$), the CDF of maximum $Z = \max_i X_i$ is given as

$$F_Z(z) = F_X^N(z). \tag{3.12}$$

We see then that an alternative and more intuitive way to view (3.7) is given as

$$\mathbb{E}\left[X_{\hat{Z}_A}\right] = \sum_{j=1}^{K} E_j \mathbb{P}\left(\ell_{j-1} \leq X_{Z_A} \leq \ell_j\right). \tag{3.13}$$

Lemma 10.

$$\frac{\partial E_k}{\partial \ell_{k-1}} = f_{k-1}\frac{E_k - \ell_{k-1}}{p_k} \tag{3.14a}$$

$$\frac{\partial E_k}{\partial \ell_k} = f_k\frac{\ell_k - E_k}{p_k} \tag{3.14b}$$

Proof. Follows from application of the quotient rule and Leibniz's rule. □

Lemma 11.

$$\frac{\partial \mathbb{E}\left[X_{\hat{Z}_A}\right]}{\partial \ell_k} = f_k\left[\frac{(F_{k+1}^N - F_k^N)(E_{k+1} - \ell_k)}{p_{k+1}} + \frac{(F_k^N - F_{k-1}^N)(\ell_k - E_k)}{p_k} - NF_k^{N-1}(E_{k+1} - E_k)\right] \tag{3.15}$$

Proof. We re-write (3.7) as

$$\begin{aligned}\mathbb{E}\left[X_{\hat{Z}_A}\right] &= \sum_{j=1}^{K}\left[E_j\left(F_j^N - F_{j-1}^N\right)\right] \\ &= F_K^N E_K - \sum_{j=1}^{K-1} F_j^N(E_{j+1} - E_j) - F_0^N E_1 \\ &= E_K - \sum_{j=1}^{K-1} F_j^N(E_{j+1} - E_j)\end{aligned} \tag{3.16}$$

and take derivatives

$$\frac{\partial \mathbb{E}\left[X_{\hat{Z}_A}\right]}{\partial \ell_k} = \frac{\partial}{\partial \ell_k}E_K - \sum_{j=1}^{K-1}\frac{\partial}{\partial \ell_k}F_j^N(E_{j+1} - E_j). \tag{3.17}$$

If $k \neq K-1$, the above becomes

$$\begin{aligned}
\frac{\partial \mathbb{E}\left[X_{\hat{Z}_A}\right]}{\partial \ell_k} &= -\sum_{j=1}^{K-1} \frac{\partial}{\partial \ell_k} F_j^N (E_{j+1} - E_j) \\
&= -\frac{\partial}{\partial \ell_k} F_{k-1}^N (E_k - E_{k-1}) - \frac{\partial}{\partial \ell_k} F_k^N (E_{k+1} - E_k) - \frac{\partial}{\partial \ell_k} F_{k+1}^N (E_{k+2} - E_{k+1}) \\
&= -F_{k-1}^N \frac{\partial E_k}{\partial \ell_k} - N F_k^{N-1} f_k (E_{k+1} - E_k) - F_k^N (\frac{\partial E_{k+1}}{\partial \ell_k} - \frac{\partial E_k}{\partial \ell_k}) + F_{k+1}^N \frac{\partial E_{k+1}}{\partial \ell_k} \\
&= f_k \left[-F_{k-1}^N \frac{\ell_k - E_k}{p_k} - N F_k^{N-1} (E_{k+1} - E_k) - F_k^N \left(\frac{E_{k+1} - \ell_k}{p_{k+1}} - \frac{\ell_k - E_k}{p_k} \right) \right] \\
&+ f_k \left[F_{k+1}^N \frac{E_{k+1} - \ell_k}{p_{k+1}} \right]
\end{aligned} \tag{3.18}$$

If $k = K-1$, the above becomes

$$\begin{aligned}
\frac{\partial \mathbb{E}\left[X_{\hat{Z}_A}\right]}{\partial \ell_{K-1}} &= \frac{\partial}{\partial \ell_{K-1}} E_K - \frac{\partial}{\partial \ell_{K-1}} F_{K-2}^N (E_{K-1} - E_{K-2}) - \frac{\partial}{\partial \ell_{K-1}} F_{K-1}^N (E_K - E_{K-1}) \\
&= -F_{K-2}^N \frac{\partial E_{K-1}}{\partial \ell_{K-1}} - N F_{K-1}^{N-1} f_{K-1} (E_K - E_{K-1}) - F_{K-1}^N (\frac{\partial E_K}{\partial \ell_{K-1}} - \frac{\partial E_{K-1}}{\partial \ell_{K-1}}) \\
&+ F_K^N \frac{\partial E_K}{\partial \ell_{K-1}}
\end{aligned} \tag{3.19}$$

The above follows from recognizing that $F_K = 1$ and we see that the expression for $k \neq K-1$ holds for $k = K-1$. □

Corollary 3. *For $N = 2$, the above simplifies to*

$$\frac{\partial \mathbb{E}\left[X_{\hat{Z}_A}\right]}{\partial \ell_k} = f_k \left[\int_{\ell_{k-1}}^{\ell_{k+1}} (x - \ell_k) f(x) \, \mathrm{d}x \right]. \tag{3.20}$$

Proof.

$$\begin{aligned}
\frac{\partial \mathbb{E}\left[X_{\hat{Z}_A}\right]}{\partial \ell_k} &= f_k\left[\frac{(F_{k+1}^2 - F_k^2)(E_{k+1} - \ell_k)}{p_{k+1}} + \frac{(F_k^2 - F_{k-1}^2)(\ell_k - E_k)}{p_k} - 2F_k(E_{k+1} - E_k)\right] \\
&= f_k\left[\frac{(F_{k+1}^2 - F_k^2)(E_{k+1} - \ell_k)}{F_{k+1} - F_k} + \frac{(F_k^2 - F_{k-1}^2)(\ell_k - E_k)}{F_k - F_{k-1}} - 2F_k(E_{k+1} - E_k)\right] \\
&= f_k\left[(F_{k+1} + F_k)(E_{k+1} - \ell_k) + (F_k + F_{k-1})(\ell_k - E_k) - 2F_k(E_{k+1} - E_k)\right] \\
&= f_k\left[(F_{k+1} - F_k)E_{k+1} + (F_k - F_{k-1})E_k - (F_{k+1} - F_{k-1})\ell_k\right] \\
&= f_k\left[\int_{\ell_k}^{\ell_{k+1}} x f(x)\,\mathrm{d}x + \int_{\ell_{k-1}}^{\ell_k} x f(x)\,\mathrm{d}x - \ell_k \int_{\ell_{k-1}}^{\ell_{k+1}} f(x)\,\mathrm{d}x\right]
\end{aligned} \tag{3.21}$$

□

Minimum Distortion

For a given number of intervals K, the decision boundaries $\{\ell_k : k = 0, \ldots, K\}$ that minimize the expected distortion are given by the solution to the following non-linear optimization:

$$\begin{aligned}
&\underset{\boldsymbol{\ell}}{\text{minimize}} \quad && D(\boldsymbol{\ell}) \\
&\text{subject to} \quad && \ell_{k-1} \leq \ell_k \quad k = 1, \ldots, K.
\end{aligned} \tag{3.22}$$

Theorem 18. *If $\{\ell_k^* : k = 0, \ldots, K\}$ is an optimal solution to (3.22) then there exists $\mu_K^* \geq 0$ for $k = 1, \ldots, K$ such that*

$$f_k\left[\frac{(F_{k+1}^N - F_k^N)(\ell_k^* - E_{k+1})}{p_{k+1}} + \frac{(F_k^N - F_{k-1}^N)(E_k - \ell_k^*)}{p_k} - NF_k^{N-1}(E_k - E_{k+1})\right] - \mu_k^* + \mu_{k+1}^* = 0 \tag{3.23a}$$

$$\mu_k^*(\ell_{k-1}^* - \ell_k^*) = 0. \tag{3.23b}$$

Proof. The Lagrangian associated with this problem is

$$L(\boldsymbol{\ell}, \boldsymbol{\mu}) = D(\boldsymbol{\ell}) + \sum_{k=1}^{K} \mu_k(\ell_{k-1} - \ell_k) \tag{3.24}$$

Taking the derivative w.r.t. ℓ_i gives

$$\frac{\partial L(\boldsymbol{\ell}, \boldsymbol{\mu})}{\partial \ell_k} = \frac{\partial D(\boldsymbol{\ell})}{\partial \ell_k} - \mu_k + \mu_{k+1} \tag{3.25}$$

where

$$\frac{\partial D(\boldsymbol{\ell})}{\partial \ell_k} = -\frac{\partial \mathbb{E}\left[X_{\hat{Z}_A}\right]}{\partial \ell_k}. \tag{3.26}$$

The result follows from setting the above equal to zero and complementary slackness. □

Corollary 4. *For $N = 2$, the above simplifies to*

$$f_k \left[\int_{\ell_{k-1}^*}^{\ell_{k+1}^*} (\ell_k^* - x) f(x) \, \mathrm{d}x \right] - \mu_k^* + \mu_{k+1}^* = 0 \tag{3.27a}$$

$$\mu_k^* (\ell_{k-1}^* - \ell_k^*) = 0. \tag{3.27b}$$

Remark. In §3.4, we solved for the optimal decision boundaries by setting all the Lagrange multipliers to zero and solving (3.23a). Depending upon the distribution, (3.23a) can be solved exactly or with a non-linear solver.

Entropy-constrained minimum distortion

The interval U_i that the i-th user's observed value lies in is a discrete random variable with probability mass function given by $\boldsymbol{p} = (p_k : k = 1, \ldots, K)$ and the entropy of U_i is $H(U_i) = -\sum_{k=1}^{K} p_k \log_2 p_k$. The total rate needed for the N users to report their intervals is then

$$R_{HomSQ}(\boldsymbol{\ell}) \triangleq \sum_{i=1}^{N} H(U_i) = NH(U) \tag{3.28}$$

by the i.i.d. assumption of the sources and the homogeneity of the quantizers.

Lemma 12.

$$\frac{\partial R_{HomSQ}(\boldsymbol{\ell})}{\partial \ell_k} = N f_k \log_2 \left(\frac{p_{k+1}}{p_k} \right) \tag{3.29}$$

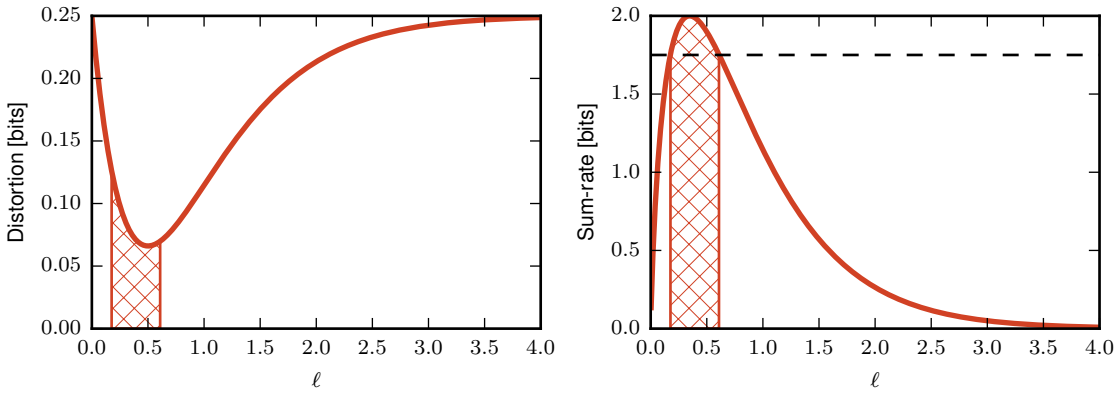


Figure 3.2: Plots of $D(\ell)$ and $R(\ell)$ as functions of ℓ for exponentially distributed source. For $R(\ell) \leq 1.75$, the set of feasible ℓ is seen to be non-convex.

Proof.

$$\begin{aligned}\frac{\partial R_{HomSQ}(\boldsymbol{\ell})}{\partial \ell_k} &= N \sum_{j=1}^{K} \frac{\partial}{\partial \ell_k} p_j \log\left(\frac{1}{p_j}\right) \\ &= N\left(\frac{\partial}{\partial \ell_k} p_k \log\left(\frac{1}{p_k}\right) + \frac{\partial}{\partial \ell_k} p_{k+1} \log\left(\frac{1}{p_{k+1}}\right)\right) \\ &= N\left(-f_k + f_k \log\left(\frac{1}{p_k}\right) + f_k - f_k \log\left(\frac{1}{p_{k+1}}\right)\right)\end{aligned} \tag{3.30}$$

□

We now consider the problem of minimizing the distortion subject to an upper limit on the sum rate.

$$\begin{aligned}\underset{\boldsymbol{\ell}}{\text{minimize}} \quad & D(\boldsymbol{\ell}) \\ \text{subject to} \quad & R_{HomSQ}(\boldsymbol{\ell}) \leq R_0 \\ & \ell_{k-1} \leq \ell_k \quad k = 1, \ldots, K\end{aligned} \tag{3.31}$$

In general, this problem is *not* convex. To see this, consider $X_i \sim$ Exponential(λ) and a single threshold ℓ (two intervals: $[0, \ell), [\ell, \infty)$). Figure 3.2 shows a plot of $D(\ell)$ (top) and $R(\ell)$ (bottom) as ℓ is swept from $\inf \mathcal{X}$ to $\sup \mathcal{X}$. For $R_0 = 1.75$ bits, the range of *infeasible* ℓ is shown as a filled area under the rate and distortion curves and we see that the set of feasible ℓ is non-convex.

Theorem 19. *If* $\{\ell_k^* : k = 0, \ldots, K\}$ *is an optimal solution to* (3.31)*, then there exists* $\mu_K^* \geq 0$ *for*

$k = 1, \ldots, K$ *and* $\mu_R \geq 0$ *such that*

$$f_k \left[\frac{(F_{k+1}^N - F_k^N)(\ell_k^* - E_{k+1})}{p_{k+1}} + \frac{(F_k^N - F_{k-1}^N)(E_k - \ell_k^*)}{p_k} - N F_k^{N-1}(E_k - E_{k+1}) \right] + f_k \mu_R^* N \log_2 \left(\frac{p_{k+1}}{p_k} \right) - \mu_k^* + \mu_{k+1}^* = 0 \tag{3.32a}$$

$$\mu_i^*(\ell_{i-1}^* - \ell_i^*) = 0 \text{ and } \mu_R^*(R_{HomSQ}(\boldsymbol{\ell}^*) - R_0) = 0. \tag{3.32b}$$

Proof. The Lagrangian associated with this problem is

$$L(\boldsymbol{\ell}, \boldsymbol{\mu}) = D(\boldsymbol{\ell}) + \mu_R(R_{HomSQ}(\boldsymbol{\ell}) - r) + \sum_{k=1}^{K} \mu_k(\ell_{k-1} - \ell_k) \tag{3.33}$$

Taking the derivative w.r.t. ℓ_i gives

$$\frac{\partial L(\boldsymbol{\ell}, \boldsymbol{\mu})}{\partial \ell_i} = \frac{\partial D(\boldsymbol{\ell})}{\partial \ell_i} + \mu_R \frac{\partial R_{HomSQ}(\boldsymbol{\ell})}{\partial \ell_i} - \mu_i + \mu_{i+1}. \tag{3.34}$$

The result follows from setting the above equal to zero and complementary slackness. □

Remark. Solving for the optimal entropy constrained quantizer is more difficult than solving for the minimum distortion quantizer. Depending upon the given values of R_0 and K, the decision boundaries may collapse and the associated Lagrange multipliers need no longer be identically zero. A general solution technique for (3.32) is beyond the scope of the present work; generalizations to both Lloyd's and Max's algorithms for entropy constrained quantizer design are presented in [31].

We conclude with some observations about the rate-distortion curve for entropy-constrained quantizers. For a given K, suppose $\boldsymbol{\ell}^*$ is a solution to (3.22). If $R_0 \geq R_{HomSQ}(\boldsymbol{\ell}^*)$, then the rate constraint in (3.31) is not active and $\boldsymbol{\ell}^*$ is also a solution to (3.31) for the same K. On the other hand, if $R_0 < R(\boldsymbol{\ell}^*)$ then the rate constraint in (3.31) is active and $\boldsymbol{\ell}^*$ is infeasible for (3.31) [31]. Next, consider the rate-distortion curve for a N-level entropy-constrained quantizer and the sequence of rate-distortion points given by (3.22) for $K = 1, \ldots, N$. These rate-distortion points all lie in the rate-distortion curve for the N-level entropy-constrained quantizer.

3.3.2 Heterogeneous Scalar Quantizers

It is somewhat intuitive to suppose that because the sources are i.i.d., the quantizers at each user should be identical. For symmetric functions (e.g., max), Misra et al. consider only the design of the quantizer for a single user [26]. When the function is *not* symmetric (e.g., arg max as in our case), the assumption of HomSQ is in fact not true.

Theorem 20. *For an optimal HomSQ $\boldsymbol{\ell}^*$ that achieves a distortion $D(\boldsymbol{\ell}^*)$, there exists a HetSQ that achieves the same distortion with rate*

$$R_{HetSQ}(\boldsymbol{\ell}) = (N-2)H(U) + \delta \tag{3.35}$$

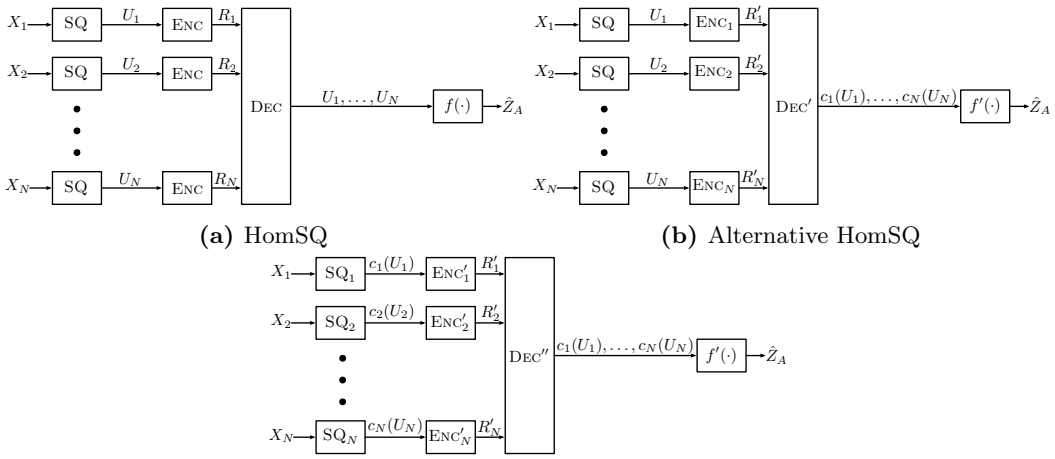
where

$$\delta = \sum_{k=1}^{K} p_k \log \frac{1}{(p_{k-1}+p_k)(p_k+p_{k+1})} \leq 2H(U) \tag{3.36}$$

and $p_0 = 0$ and $p_{K+1} = 0$.

Proof. We think of HomSQ as approximating the continous distribution with a discrete one and then losslessly computing the arg max of the quantization bin indices. This is exactly the problem considered in Section II of [3]. From [3], we know that fewer than $R_{HomSQ}(\boldsymbol{\ell})$ bits are needed to enable the CEO to losslessly determine arg max of the bin indices. In [3], a code is constructed by coloring the vertices of the associated characteristic graphs for each user and entropy coding the vertex colors. The rate savings comes by allowing a pair of consecutive bin indices for a user to be assigned the same color, provided the pair of indices are assigned different colors for every other user. We can compute the colors directly, by observing that if a pair of consecutive bin indices are being assigned the same color we are merging the underlying bins into one larger bin for that user *only*. □

Remark. As was shown in [3], the total rate savings for losslessly determinging the arg max of a discrete distribution is at most 2 bits. Therefore, the rate savings of HetSQs versus HomSQs is also at most 2 bits and the savings per user goes to zero as the number of users is increased.



(a) HomSQ **(b)** Alternative HomSQ

(c) HetSQ

Figure 3.3: Block diagram of possible scalar quantizers for arg max. For HomSQ a, $U_1, \ldots, U_N$ are i.i.d. and the encoder is the same for each user. In the alternative HomSQ scheme b, the reduction in rate comes from coloring the vertices in a characteristic graph associated with each source. In general, this graph is different for each user and therefore the encoder will be different for each user. Finally, by having the quantization operation at each source determine the "vertex color", $U'_1, \ldots, U'_N$ are independent but *not* identically distributed.

For HetSQ, when $N = 2$ and $K = 2$ only one of the sources is sending back a bit. We can use results from rate-distortion for the Bernoulli(p) source with Hamming distortion to trace out the low-rate/high-distortion segment of the trade-off curve.

Lemma 13. *The expected value of the estimator when a lossy source code is used to communicate the output HetSQ for $N = 2$ and $K = 2$ to the CEO is given by*

$$\mathbb{E}\left[X_{\hat{Z}_A}\right] = (1-\hat{p})\mathbb{E}\left[X\right] + \hat{p}(D_H\mathbb{E}\left[X|X \leq \ell\right] + (1-D_H)\mathbb{E}\left[X|\ell \leq X\right]) \tag{3.37}$$

where

$$\hat{p} = \frac{p_2 - D_H}{1 - 2D_H} \tag{3.38}$$

and the rate is given by

$$R(\ell, D_H) = \begin{cases} h_2(p_2) - h_2(D_H) & D_H \leq \min\{p_2, 1-p_2\} \\ 0 & D_H > \min\{p_2, 1-p_2\} \end{cases} \tag{3.39}$$

Proof. We assume that user 1 is sending the single indicator bit to the CEO w.l.o.g. and model this as a Bernoulli(p_2) source with $p_2 = \mathbb{P}(X_1 \geq \ell)$ and Hamming distortion D_H. The rate-distortion function for this subproblem is given by (3.39). The test channel that achieves this is a binary symmetric channel (BSC)(D_H) with input $\hat{X} \sim$ Bernoulli($\hat{p}$). From this we obtain an expression for the joint probability mass function (pmf) $\mathbb{P}\left(X = x, \hat{X} = \hat{x}\right)$ from which we can derive (3.37). □

Remark. Observer that for $D_H = 0$, we obtain the same expression as (3.7) for $N = 2$ and $K = 2$ and for $D_H = \min\{p_2, 1 - p_2\}$, we get $\mathbb{E}\left[X_{\hat{Z}_A}\right] = \mathbb{E}[X]$.

3.4 Examples & Results

Example 6 (arg max SQ for $N \geq 2$ Uniform$(0,1)$)**.** We now consider the design of a HomSQ for estimating the arg max from $N > 2$ sources i.i.d. Uniform$(0,1)$. From (3.7) we obtain the following expression for the expected distortion

$$D(\boldsymbol{\ell}) = \frac{N}{N+1} - \sum_{j=1}^{K} \frac{(\ell_{j-1} + \ell_j)(\ell_j^N - \ell_{j-1}^N)}{2} \tag{3.40}$$

from which we solve for optimal quantizer parameter $\boldsymbol{\ell}^*$

$$\ell_{j-1}^{*N-1} = \frac{\ell_{j+1}^{*N} - \ell_{j-1}^{*N}}{N(\ell_{j+1}^* - \ell_{j-1}^*)}. \tag{3.41}$$

Figure 3.4 shows the per user rate and normalized distortion for HomSQ and the staggered HetSQ derived from the optimal HomSQ for estimating the arg max of a collection of distributed users with sources i.i.d. Uniform$(0,1)$. The left subplot is for $N = 2$ users, the middle subplot for $N = 4$ users, and $N = 8$ users. We observe immediately that the performance gains of the staggered HetSQ over HomSQ diminish as the number of users increases. Additionally, while the zero rate distortion is increasing in the number of users, we observe that the required rate per user to achieve a specified normalized distortion is non-montonic in the number of users. For example, fixing $D = 0.01$ we observe per user rate for HomSQ is 2.32 bits for $N = 2$, 2.32 bits for $N = 4$, and 1.86 bits for $N = 8$. The per user rate for HetSQ is 1.52 bits for $N = 2$, 1.95 bits for $N = 4$, and 1.71 bits for $N = 8$.

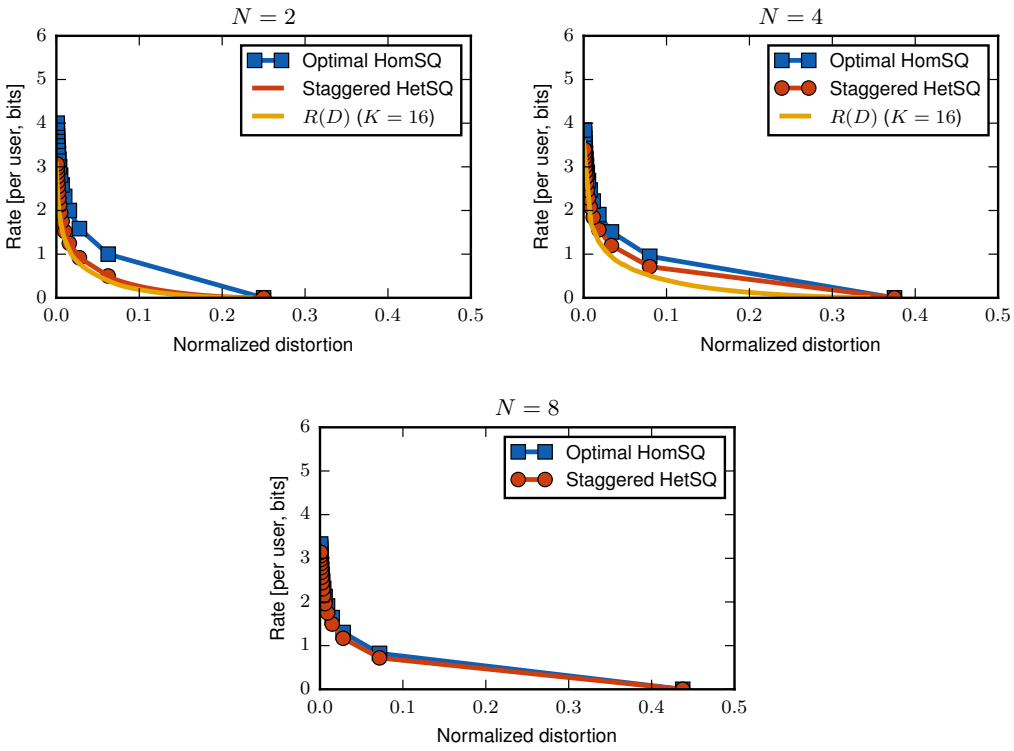


Figure 3.4: Rate-distortion trade-offs for HomSQ and HetSQ and rate-distortion function for varying numbers of users: $N = 2$ users (top, left); $N = 4$ users (top right), and; $N = 8$ users (bottom). The source distribution was assumed Uniform(0, 1).

The performance gain of HetSQ over HomSQ does come at a cost in terms of fairness. To demonstrate this, consider the optimal HomSQ when using three intervals at each user, for which the optimal thresholds are given as $\ell^{(1)*} = \frac{2a+b}{3}$ and $\ell^{(2)*} = \frac{a+2b}{3}$. The HetSQ bins for X_1 will be $[a, \frac{2a+b}{3})$ and $[\frac{2a+b}{3}, b]$; the HetSQ bins for X_2 will be $[a, \frac{a+2b}{3})$ and $[\frac{a+2b}{3}, b]$. Observe that $S_1 \sim$ Bernoulli(2/3) and $S_2 \sim$ Bernoulli(1/3) independent of a and b and therefore

$$R_{HomSQ}(\boldsymbol{\ell}) = 2\log_2 3 \approx 3.170$$
$$R_{HetSQ}(\boldsymbol{\ell}) = \frac{2}{3}\log_2\left(\frac{27}{4}\right) \approx 1.837.$$

User one is selected as the arg max with probability $\mathbb{P}(S_1 = 0, S_2 = 0) + \mathbb{P}(S_1 = 1, S_2 = 0) + \mathbb{P}(S_1 = 1, S_2 = 1) = \frac{5}{9}$ while user two is selected with probability $\frac{4}{9}$. If we use three total thresholds (i.e., $K = 4$) with one thresholds at user one and two threshold at user two, then user one and user two are selected as the arg max with equal probability. However, user one's rate is $R_1 = 1.0$ bits while user two's rate is $R_2 = 1.5$ bits.

Example 7 (arg max SQ for Exponential(λ))**.** We now consider the scenario where each user's source is distributed Exponential(λ). If we define $w_j \triangleq \lambda \ell_j$, then expected distortion can be written as

$$D(\boldsymbol{\ell}) = \frac{1}{\lambda}\left(H_N - \sum_{j=1}^{K} \frac{\left(e^{w_j}(w_{j-1}+1) - e^{w_{j-1}(w_j+1)}\right)\left((1-e^{-w_j})^n - (1-e^{-w_{j-1}})^n\right)}{e^{w_j} - e^{w_{j-1}}}\right) \tag{3.42}$$

where H_N is the N^{th} harmonic number. When $N = 2$, we have from (3.23a) that an optimal solution will satisfy

$$w_k^* = \frac{-e^{-w_{k-1}^*}(1 + w_{k-1}^*) + e^{-w_{k+1}^*}(1 + w_{k+1}^*)}{(e^{-w_{k+1}^*} - e^{-w_{k-1}^*})} \tag{3.43}$$

with boundary conditions

$$w_1^* = \frac{1 - e^{-w_2^*}(1 + w_2^*)}{(1 - e^{-w_2^*})},\ w_{K-1}^* = 1 + w_{K-2}^*. \tag{3.44}$$

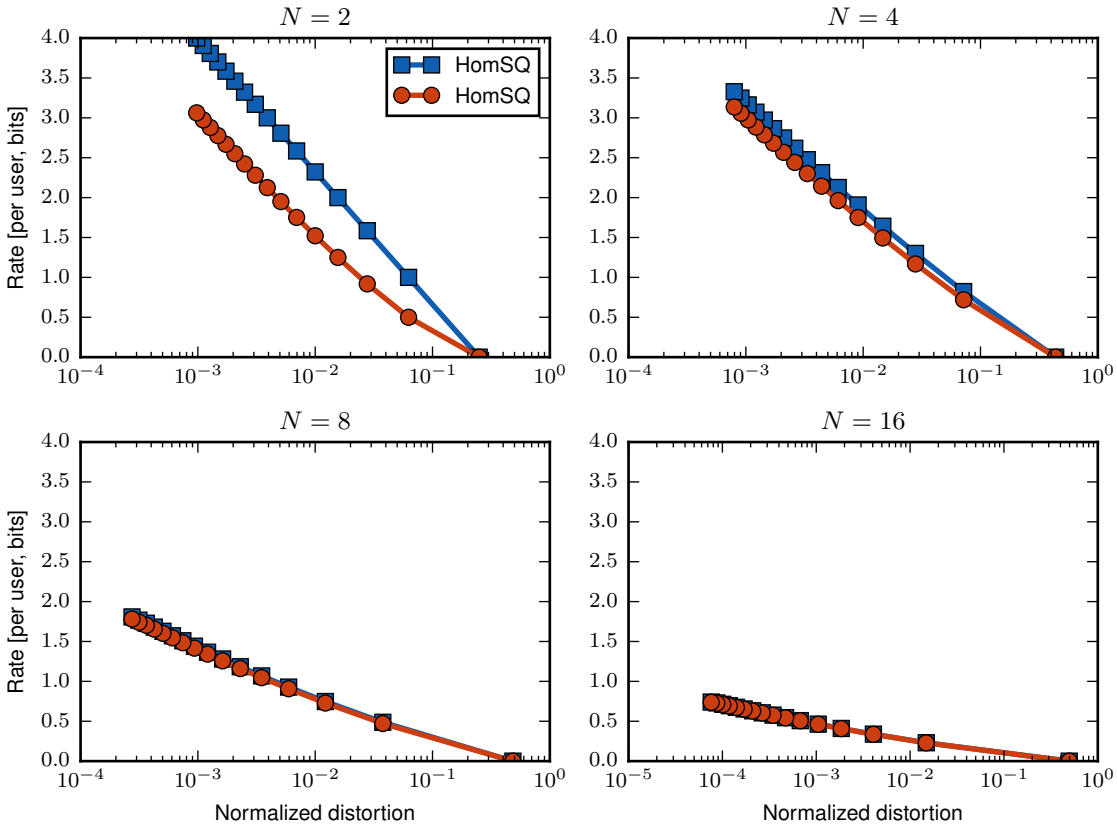


Figure 3.5: Rate-distortion trade-offs for HomSQ and HetSQ for Uniform(0, 1) as the number of users is increased.

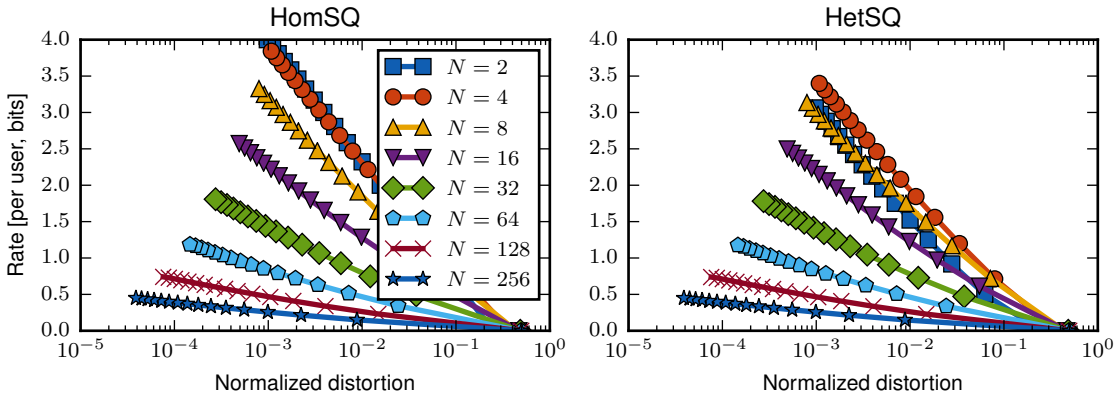


Figure 3.6: Rate-distortion trade-offs for HomSQ (left) and HetSQ (right) for Uniform(0, 1) as a function of the number of users.

Such an expression can be found for $N > 2$, but the details are omitted here. We observe that the optimal quantizer for $X_1, \ldots, X_N \sim \text{Exponential}(\lambda)$ is given by an appropriate scaling of the optimal quantizer for $X_1, \ldots, X_N \sim \text{Exponential}(1)$.

Figure 3.7 shows the per user rate and normalized distortion for HomSQ and the staggered HetSQ derived from the optimal HomSQ for estimating the arg max of a collection of distributed users with sources i.i.d. Exponential(2). The left subplot is for $N = 2$ users, the middle subplot for $N = 4$ users, and $N = 8$ users. As with the previous example, we observe that the performance gains of the staggered HetSQ over HomSQ diminish as the number of users increases. The zero rate distortion is increasing in the number of users, but we observe that the required rate per user to acheive a specified normalized distortion is non-montonic in the number of users. For example, fixing $D = 0.01$ we observe per user rate for HomSQ is 2.51 bits for $N = 2$, 2.51 bits for $N = 4$, and 2.07 bits for $N = 8$. The per user rate for HetSQ is 1.68 bits for $N = 2$, 2.12 bits for $N = 4$, and 1.91 bits for $N = 8$. We conclude by noting that for a fixed normalize distortion, the SQs for exopentially distributed sources require a higher per user rate than SQs for uniformly distribute sources. This makes sense intuitively as the support set of $\text{Uniform}(0, 1)$ is bounded, while it is unbounded for $\text{Exponential}(\lambda)$.

As above, consider the optimal HetSQ when using two intervals at each user, for which the optimal thresholds are given as $w_1^* = 1 + W(-2e^{-2})$ and $w_2^* = 2 + W(-2e^{-2})$ where $W(x)$ is the Lambert W function. Observe that $S_1 \sim \text{Bernoulli}(1 - e^{-w_1^*}) \approx 0.448$ and $S_2 \sim \text{Bernoulli}(1 - e^{-w_2^*}) \approx 0.797$ and therefore

$$R_1 + R_2 = H(S_1) + H(S_2) \approx 1.720$$

We get the same distortion as for identical quantizers with $K = 3$ which requires 3.033 bits. User two is selected as the arg max with probability $\mathbb{P}(S_1 = 0, S_2 = 0) + \mathbb{P}(S_1 = 1, S_2 = 0) + \mathbb{P}(S_1 = 1, S_2 = 1) \approx 0.560$ while user one is selected with probability 0.440.

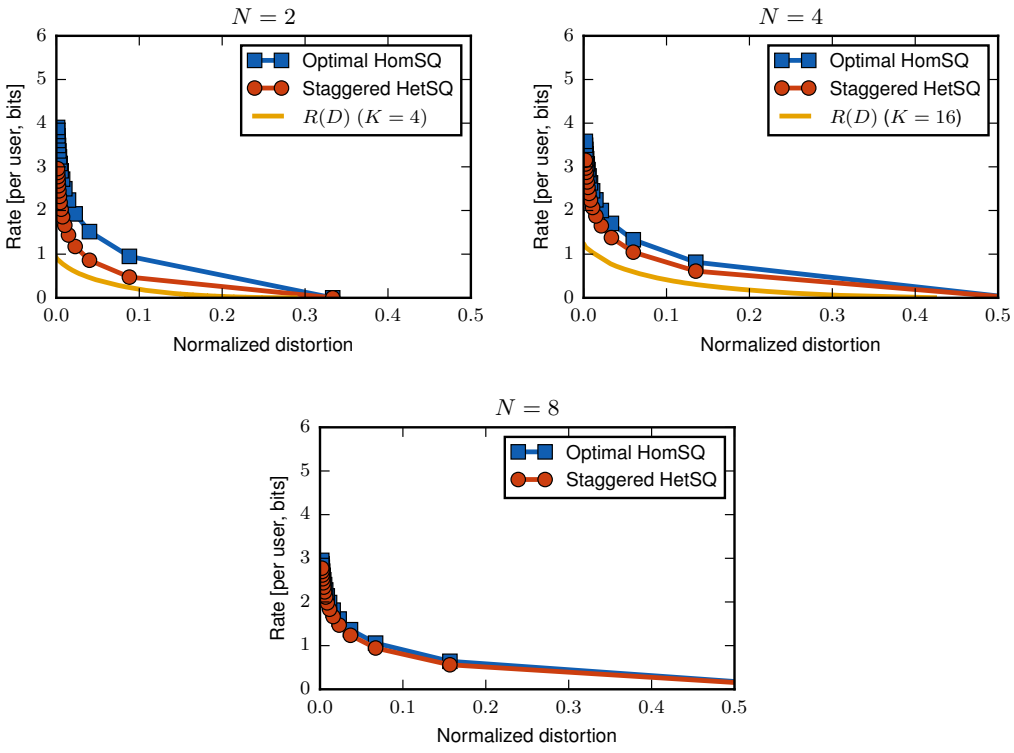


Figure 3.7: Rate-distortion trade-offs for HomSQ and HetSQ and rate-distortion function for varying numbers of user: $N = 2$ users (top left); $N = 4$ users (top right), and; $N = 8$ users (bottom). The source distribution was assumed Exponential(2).

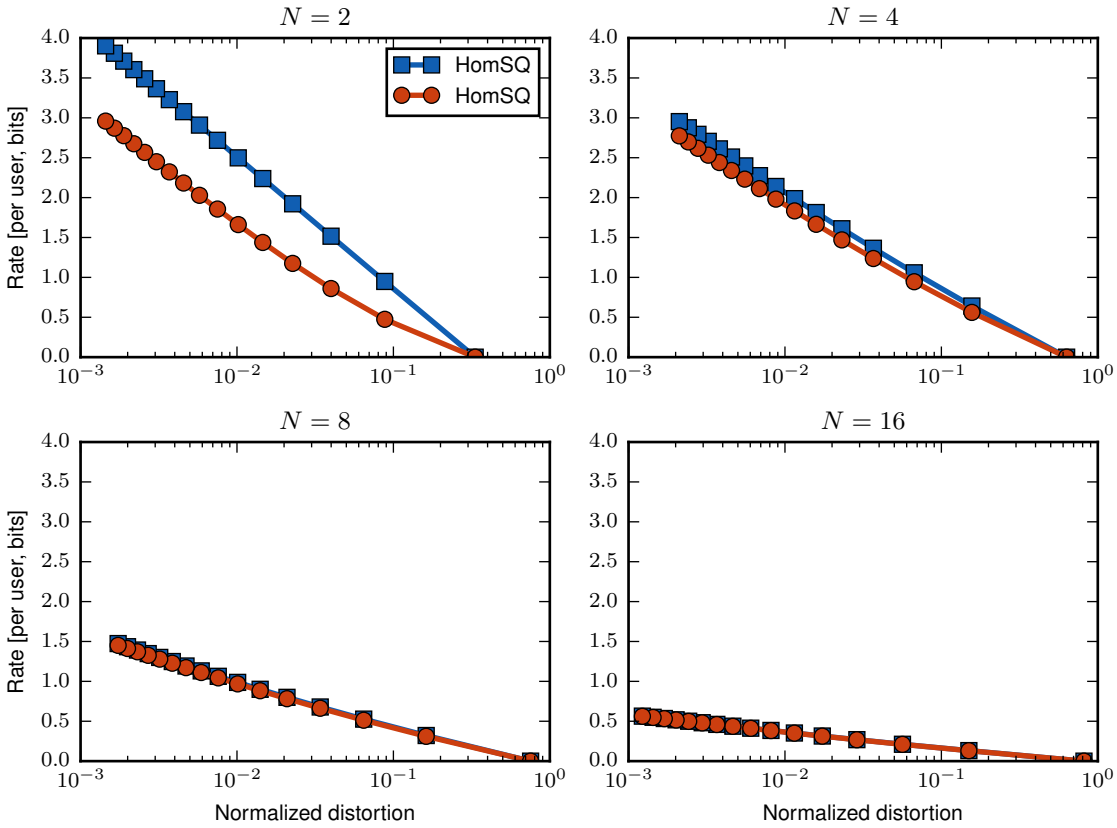


Figure 3.8: Rate-distortion trade-offs for HomSQ and HetSQ for Exponential(2) as the number of users is increased.

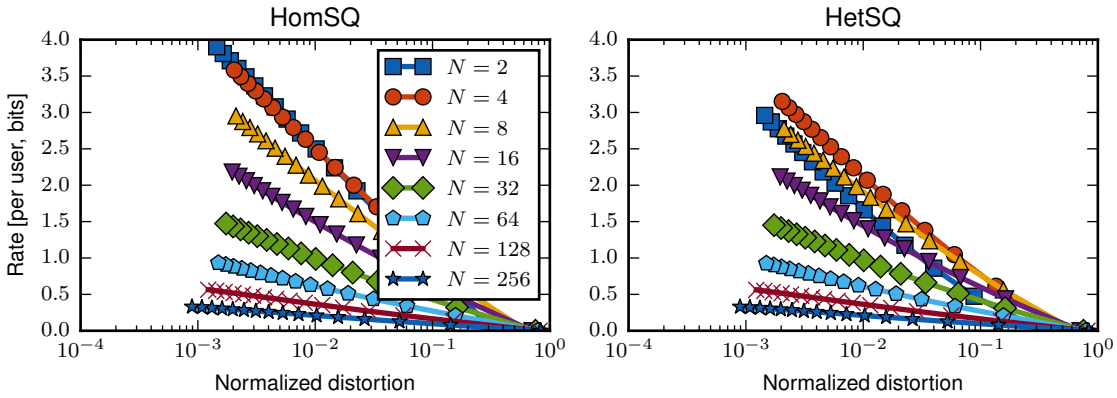


Figure 3.9: Rate-distortion trade-offs for HomSQ (left) and HetSQ (right) for Exponential(2) as a function of the number of users.

3.5 Conclusions

In this chapter, we have considered the design of SQs for distributed function computation in the context of a resource allocation problem. We proposed a natural, non-quadratic distortion measure and provided an exact expression as a function of the quantizer parameters. We have argued that HetSQ are more efficient than HomSQ by achieving the same distortion at a lower rate. We considered several example capacity distributions, demonstrating that the performance of HetSQ *may* be close to fundamental limits.

We restricted our focus to SQs for a single subband. It is known that vector quantizers are more efficient than SQs, even when the source outputs being blocked into vectors are independent [22, 30]. As mentioned previously, recent results have shown that local vector quantizers followed by SW encoding is optimal for certain two-terminal problems with continuous distributions [27, 28]. This motivates the consideration of vector quantization in future work and to handle the case with multiple subbands.

Chapter 4: Interactive Scalar Quantization for Distributed Extremization

In many resource allocation problems, a centralized controller needs to award some resource to a user selected from a collection of distributed users with the goal of maximizing the utility the user would receive from the resource. This can be modeled as the controller computing an extremization of the distributed users' utilities. The overhead rate necessary to enable the controller to reproduce the users' local state can be prohibitively high. Two approaches to reduce this overhead are lossy estimation and interactive communication. In the lossy estimator framework, rate savings are achieved by tolerating a bounded expected reduction in utility. In interactive communication, rate savings come at the expense of delay. In this chapter, we consider the design of a simple achievable scheme based on successive refinements of scalar quantization at each user. The optimal quantization policy is computed via a dynamic program and we demonstrate that tolerating a small increase in delay can yield significant rate savings. We then consider two simpler quantization policies to investigate the scaling properties of the rate-delay trade-offs. Using a combination of these simpler policies, the performance of the optimal policy can be closely approximated with lower computational costs.

4.1 Introduction

4.1.1 Motivation

A common pattern in resource limited systems is the allocation of the resource by a controller among a set of competing consumers/users. In an effort to make the most efficient usage of the resource, the controller awards the resource to the user that derives the most utility from it. Often, these users are not colocated and must communicate some local state to the centralized controller. This communication presents an overhead. Hence techniques for minimizing the required rate are needed. In some cases, the rate can be reduced by incurring only a small penalty. Rate-distortion theory is an example where the rate required to communicate/reconstruct a signal can be reduced if small errors in the reconstruction are tolerated. Interactive communication allows an alternative

framework where communicating parties can send messages back-and-forth over multiple rounds [32]. This back-and-forth messaging can reduce the rate required to compute an extremum/extrema of the sources at the cost of increased delay. One achievable scheme for a collection of distributed users talking to central controller (multiterminal CEO) is quantization followed by entropy (i.e., Huffman) encoders [33, 34]. In this chapter, we consider this same technique in the context of interactive communication. We formulate the design of the multi-round quantization as the solution to a dynamic program and demonstrate a substantial reduction in overhead communication rate can be obtained for a small increase in delay. Because of the high computational complexity of the dynamic programming problem, we identify two simpler schemes and investigate their performance over a range of parameters. By combining these two schemes, we demonstrate a close approximation of the dynamic programming solution at a lower computational cost.

4.1.2 Contributions

This chapter is organized as follows. We review related literature from information theory, signal processing, and communication complexity in §4.2. We present the basic problem model and establish the mathematical notation used throughout the chapter in §4.3. In §4.4, we formulate the optimal rate-delay trade-off of scalar quantization as the solution of a minimum cost dynamic program. We characterize the set of terminating states for the dynamic program when computing the different extremization functions under consideration (cf. Proposition 3 & Proposition 4). We prove that, asymptotically in the number of users, the cost for the CEO to compute the different extremization functions are equal (cf. Proposition 6). In §4.5, we restrict our attention to the case of users' utilities being distributed uniformly. This assumption allows us to provide analytical expressions for the rate and delay of simple quantization policies (cf. Proposition 7 & Proposition 8). We then provide an extension to these simple schemes, which is asymptotically (in the number of users) sufficient for minimizing the cost of computing the selected extremization functions (cf. Proposition 10). The proposed family of quantizers is significantly smaller compared to the space of all possible quantizers. Hence, with these results the dynamic program can be solved more quickly with a relatively small incurred penalty relative to the optimal dynamic program. In §4.6, we show the rate-delay

trade-offs for different distributions and varying numbers of users. We also present a comparison of the minimum cost from searching over all quantizers to the minimal cost from searching over our proposed family of quantizers. We present directions for future work in §4.7.

4.2 Related Work

In the present work, we consider a distributed model where all users communicate to a central node. In particular, the users can not overhear each other but can design their codes (with knowledge of the source distributions of other users) to communicate with the central node—called "cooperative design, separate encoding" [35]. These models are referred as the chief estimation officer (CEO) problem [36]. We begin by reviewing fundamental limits and achievable schemes for non-interactive variants of CEO-type problems. We then review results for interactive variants that demonstrate significant rate savings may be possible. The non-interactive CEO problem has received considerably more attention than the interactive variant, and, for the cases where fundamental limits are known, quantization followed by entropy coding closely approximates these limits. This motivates our study of interactive quantization as a means to realize further rate savings at the expense of an incurred delay.

4.2.1 Non-interactive communication: fundamental limits

In information theory, the interest is usually on characterizing inner/outer bounds for the rate region. Berger et al. introduced the generic CEO problem, wherein the CEO wants to reproduce the source from the received signals [36]. The rate region for the problem of source reproduction with constrained distortion remains unknown, except for the cases of Gaussian distribution with quadratic distortion [37, 38] or i.i.d. discrete source distributions [39].

In our work, we are interested in having the CEO compute a function of all sources; this is referred to as the *distributed function computation (DFC) problem*, and was considered in [40–42]. This general formulation contains the specialized problem of function computation with side information; in this problem, the CEO knows all but one of the sources [43, 44]. When the problem requires error-free computation, it was shown that the minimum worst case rate is related to the

chromatic number of the characteristic graph of the source [43]. In the case of lossless (in the Shannon sense) computation, it was shown that the minimum average rate is the conditional graph entropy of the characteristic graph of the source [44, 45]. Building upon this line of research, the rate region for the lossless DFC problem was characterized for certain problem instances [41, 42]. Sefidgaran et al. derived inner and outer bounds to the rate region for a class of tree structured networks (which includes the CEO problem) and showed that the inner and outer bound coincide with each other if the sources obey certain Markov properties [41]. Doshi et al. gave the rate region for the DFC problem under a different constraint that they referred to as the "zig-zag" condition [42]. They showed that any achievable rate point can be realized by graph coloring at each user and Slepian-Wolf (SW) [5] encoding the colors. Han and Kobayashi partitioned all DFC problems based on whether their achievable rate region coincides with the SW region [40].

The aforementioned literature provides insightful outer bounds for comparing the performance of distributed quantizer designs [3, 46], but the achievable schemes used in the proofs usually require block coding with infinite block length, which is not practical. For use in a real system, simpler achievable schemes with low computational complexity and performance close to the limits are needed.

4.2.2 Non-interactive communication: achievability

A concern of signal processing is to provide optimized practical quantization algorithms for the DFC system with performance close to the rate-distortion limits [26, 33–35, 47–49]. There are asymptotic results for sufficiently high-rate and low-distortion that are derived by applying high rate quantization theory [50], while there are also non-asymptotic results derived from generalizations of Lloyd's algorithm.

For the high-rate and low-distortion scenario, Misra et al. considered a quantization scheme for the analysis of distributed scalar quantization [26]. It was shown that, with certain constraints on the objective function and source distributions, the high-resolution approach can asymptotically achieve the rate-distortion limits, and the optimized quantization is regular[1]. Sun et al. used a

[1] A quantizer is called regular if each partition cell is an interval and each output level is within the corresponding interval.

similar high-resolution approach, but with a simpler decoder design and relaxed source distribution requirements [48].

For the general rate-distortion problem, an algorithm for building optimized distributed quantizers was given wherein the CEO uses the quantized observations to perform hypothesis testing [35]. A two-stage distributed scheme was proposed for the case when the users each have a noisy observation on the same source, and the CEO needs to reproduce the source with a bounded expected distortion [33, 34]. A first stage of local quantization is followed by a second stage of encoding the quantized signals based on Slepian-Wolf coding using syndrome codes [34] or index reuse techniques [33].

Most of the provided distributed quantization schemes are non-interactive, which means the users each communicate with the CEO once, and no feedback is allowed from the CEO. For the problems in which the non-interactive fundamental limits are known, distributed quantization shows satisfactory performance as compared with the limit. However, very little work has been done in the interactive distributed cases.

4.2.3 Interactive communication

Interactive communication is a scheme that allows message passing over multiple rounds. At each round, the communicating parties are allowed to send messages based on what they have received in previous rounds as well as their local source observation [32]. The interactive communication literature is roughly divided into two categories: communication complexity and interactive information theory.

The communication complexity literature is concerned with finding communication protocols that minimize the sum-rate subject to different sets of constraints. Overviews of communication complexity can be found in [51, 52]. Communication complexity is defined as the sum-rate cost minimized over all protocols and maximized over all possible input pairs (worst case cost). Average cost has also been studied for randomized coding protocols. Much of the communication complexity literature is focused on 2 users. Models with an N-terminal setup were considered in [53, 54], where the authors focused on providing communication complexity bounds with the restrictions that the function must be Boolean and the message sent at each round must be binary. However, in our

work, we are interested in providing achievable schemes for problems without the limitation to 1 bit of communication in each round or the restriction to Boolean functions.

Kaspi determined the two party information theoretic limit for lossy compression via interactive communication [32]. This line of research was continued by Ma and Ishwar, who showed (by an example) that the minimum rate for a given distortion constraint can be arbitrarily smaller than the non-interactive minimum rate obtaining the same distortion [55]. In follow up work, Ma and Ishwar showed (by an example) that for the DFC problem, the minimum sum-rate for losslessly computing a function can be smaller than the non-interactive rate; even infinitely-many rounds of interaction may still improve the rate-region [56].

These results motivate us to consider interaction for the DFC problem. In earlier work, we considered the non-interactive DFC problem of computing an extremization of independent users. We developed distributed scalar quantizers with rate-distortion performance close to the rate-distortion limits [3, 46]. We provided an achievable interactive communication scheme where the CEO communicates a threshold to the users at each round and the users reply with a single bit indicating if its value is above or below the threshold [3, 57]. This scheme can be thought of as a simple two-bin quantizer selected by the CEO at the beginning of each round; in the present work we extend this by allowing the CEO to select a multi-bin quantizer in each round.

4.3 Problem Model

We assume that we have N users, each with local state X_i for $i = 1, \ldots, N$, vying for a resource to be allocated by the controller (i.e., CEO). We model the users' local states as i.i.d. (across users) discrete random variables with support set $\mathcal{X}$ and PMF $p_X(x)$, and interpret state as a proxy for the users' utility. Maximum utility of the resource is obtained by allocating it to the user with the largest local state value. Without loss of generality, we will take $\mathcal{X} = \{1, \ldots, L\}$ where L is the size of support set. To select a user to award the resource to, the CEO wishes to compute one of the following functions[2]: 1. $\arg\max_i X_i$; 2. $\max_i X_i$, or; 3. $(\arg\max_i X_i, \max_i X_i)$. By computing

[2]Though we focus exclusively on the case of maximization in this chapter, similar results hold for the case of minimization. In later sections, we enforce a decreasing order on certain parameters; to obtain results for minimization, the order should be increasing.

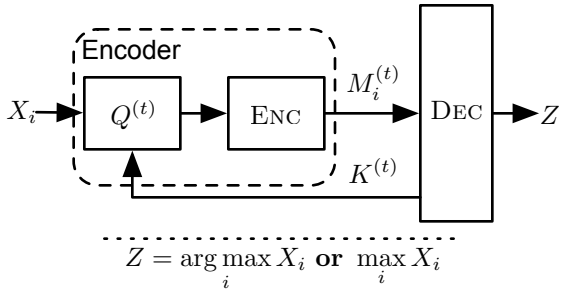


Figure 4.1: Interactive quantization system diagram. The users' utilities X_i are quantized using the quantization function $Q^{(t)}(\cdot)$ giving the quantized utilities $Q_i^{(t)}$ which are then entropy encoded (ENC) before being sent to the CEO (DEC). Based on the received quantized utilities, the CEO updates the quantization function at the users until the desired function has been computed.

the arg max, the CEO can allocate the resource to an appropriate user without knowing exactly its value. Computing the max of the users' local state allows the CEO to make a global decision, e.g., turning on building-wide air conditioning based on local temperature readings. Computing both enables the CEO to know the local value of the user being awarded the resource.

We view quantization as a function that maps the support set $\mathcal{X} \subset \mathbb{R}$ onto some finite set $\hat{\mathcal{X}}$, i.e., $Q : \mathcal{X} \to \hat{\mathcal{X}}$. Traditionally, quantization is used as a lossy compression scheme for representing sources with values drawn from a continuous support set. In this case, the quantizer is specified by partitioning the support set into intervals that are mapped to representative values. Implicit is the assumption the quantizer is monotonically increasing $x, y \in \mathcal{X}$ s.t. $x \leq y \implies Q(x) \leq Q(y)$. When $\mathcal{X}$ is finite (as we assume in this work), the quantizer can still be specified in terms of intervals but this representation may not be unique. With the assumptions of $\mathcal{X}$ finite and Q order preserving, we can alternately specify a K-level quantizer as a K-tuple of integers $\boldsymbol{n} = n_1, \ldots, n_K$ that sum to L with the following interpretation: the first n_1 elements of $\mathcal{X}$ ($\{1, \ldots, n_1\}$) are mapped to 1, the next n_2 elements of $\mathcal{X}$ ($\{n_1 + 1, \ldots, n_2\}$) are mapped to 2, and so on. This representation is unique (in that different integer tuples correspond to different quantization functions) and the set of possible scalar quantizers is isomorphic to the compositions of the integer L. For brevity, we define

$n_{1:k} = n_1 + \cdots + n_k$. The induced PMF on $\mathcal{Q}$ is then

$$p_Q(k) = \sum_{i=n_{1:k-1}+1}^{n_{1:k}} p_X(i). \tag{4.1}$$

We assume that time is slotted into rounds of sufficient length that the CEO can communicate to the users and receive their responses in a single slot. We indicate the time slot/round of interaction by t. In our analysis, we assume that feedback from the CEO provides all users with the same knowledge as the CEO. We further assume the cost of communication from users to the CEO is more expensive than communication from the CEO to the users. Therefore we omit the cost of dissemination on the downlink from the CEO to the users in our analysis.

Suppose that at the beginning of the t-th round of interaction, the CEO observes that there are $N^{(t)}$ active users and the support set has size $L^{(t)}$. The CEO will select a quantization function $Q^{(t)}$ (homogeneous across users) by selecting the number of quantization levels $K^{(t)}$ and the quantization bin sizes $\boldsymbol{n}^{(t)} = (n_1^{(t)}, \ldots, n_{K^{(t)}}^{(t)})$ and communicate this to the users. Let $Q_i^{(t)} \triangleq Q^{(t)}(X_i)$ denote the quantization bin in which X_i lies and let $\boldsymbol{Q}^{(t)} = (Q_i^{(t)} : i = 1, \ldots, N^{(t)})$ denote the length $N^{(t)}$ tuple of response received by the CEO from the users. Define

$$N_k^{(t)} = \sum_{i=1}^{N^{(t)}} \mathbb{1}_{Q_i^{(t)}=k}, \quad k = 1, \ldots, K^{(t)} \tag{4.2}$$

as the number of sources with state in bin k and let $k_t^* = \max\{k : N_k^{(t)} > 0\} = \max_i Q_i^{(t)}$ be the index of the largest non-empty quantization bin. Based on the responses $\boldsymbol{Q}^{(t)}$ from the active users, the CEO performs the following updates in each round of interaction:

$$N^{(t+1)} = N_{k_t^*}^{(t)}, \qquad L^{(t+1)} = n_{k_t^*}^{(t)} \tag{4.3a}$$

$$p_X^{(t+1)}(i) = \begin{cases} \frac{p_X^{(t)}(i)}{p_Q^{(t)}(k_t^*)} & Q^{(t)}(i) = k_t^* \\ 0 & \text{o.w.} \end{cases} \tag{4.3b}$$

The first equation captures the fact that if a user's quantized value is not in the highest reported

bin, the user's utility cannot be a maximizing value. Only the users with values in the highest reported bin need to continue interacting with the CEO; the remaining users become inactive and do not participate in subsequent rounds. The second equation updates the cardinality of the support set; the maximizing value X_i corresponds to one of the values of $\mathcal{X}^{(t)}$ that maps to the maximum reported quantization bin. The final equation updates the PMF for the remaining range of user values. At this point, the CEO is ready to begin the $t+1$-th round of interaction. In the following, we will omit the time superscript when the time instance t is not relevant to the discussion and/or is clear from context.

4.4 Optimal Solution via Dynamic Programming

At time t, the CEO observes the state of the system $s^{(t)} = \left(N^{(t)}, L^{(t)}, p_X^{(t)}\right)$ and wishes to compute a quantization policy $\boldsymbol{a}^{(t)} \triangleq \left(K^{(t)}, \boldsymbol{n}^{(t)}\right) \in \mathcal{A}^{(t)}$ that minimizes the cost of computing the desired function f:

$$C_f\left(\boldsymbol{s}^{(t)}\right) = \min_{\boldsymbol{a}^{(t)} \in \mathcal{A}^{(t)}} \left[(1-\lambda) R\left(\boldsymbol{a}^{(t)}, \boldsymbol{s}^{(t)}\right) + \lambda \tau\left(\boldsymbol{s}^{(t)}, \boldsymbol{a}^{(t)}\right) + \mathbb{E}\left[C_f\left(\boldsymbol{S}^{(t+1)}\right)\right]\right]. \tag{4.4}$$

The first part of the term inside the minimization consists of a weighting (λ) of the rate and delay incurred by choosing the quantizer given by $\boldsymbol{a}^{(t)}$ when in state $\boldsymbol{s}^{(t)}$. The parameter λ is fixed throughout and sets the relative importance of minimizing the rate ($R(\cdot)$) versus the delay ($\tau(\cdot)$). Given $\boldsymbol{s}^{(t)}$ and $\boldsymbol{a}^{(t)}$, the rate and delay are:

$$R\left(\boldsymbol{a}^{(t)}, \boldsymbol{s}^{(t)}\right) = \begin{cases} N^{(t)} H\left(p_Q^{(t)}\right) & \boldsymbol{s}^{(t)} \notin \mathcal{S}_f^* \\ 0 & \text{o.w.} \end{cases} \tag{4.5a}$$

$$\tau\left(\boldsymbol{a}^{(t)}, \boldsymbol{s}^{(t)}\right) = \begin{cases} 1 & \boldsymbol{s}^{(t)} \notin \mathcal{S}_f^* \\ 0 & \text{o.w.} \end{cases} \tag{4.5b}$$

where $p_Q^{(t)}$ is given by (4.1) and the set $\mathcal{S}^*$ represents terminating states (from which f may be computed). In general, the set $\mathcal{S}^*$ will depend on the particular extremization function being considered.

If $\boldsymbol{s}^{(t)} \notin \mathcal{S}^*$, then the CEO will select a quantizer and the rate incurred is the entropy of the induced PMF on the quantized values times the number of active users $N^{(t)}$; the delay is an additional round of interaction. If $\boldsymbol{s}^{(t)} \in \mathcal{S}^*$, the CEO can compute f and the interaction is over.

The second part of the expression inside of the minimization consists of the "cost to go". Depending on the particular state $\boldsymbol{s}^{(t)}$ and particular action $\boldsymbol{a}^{(t)}$ chosen, the system state transitions to state $\boldsymbol{s}^{(t+1)}$ with some probability. The "cost to go" is an expectation of the optimal cost function taken over all possible next time step states $\boldsymbol{s}^{(t+1)}$ and is given by

$$\begin{aligned} \mathbb{E}\left[C_f\left(\boldsymbol{S}^{(t+1)}\right)\right] &= \sum_{k=1}^{K^{(t)}}\left[\sum_{i=1}^{N^{(t)}} \rho(k,i) C_f\left(\left(i, n_k^{(t)}, p_X^{(t+1)}\right)\right)\right] \\ \rho(k,i) &= \left(p_Q^{(t)}(k)\right)^i \left(\sum_{j=1}^{k-1} p_Q^{(t)}(j)\right)^{N^{(t)}-i} \binom{N^{(t)}}{i} \end{aligned} \tag{4.6}$$

The outer summation conditions on the largest reported quantization bin ($k_t^* = k$) while the inner summation conditions on there being i users in the k_t^*-th bin. Given these two outcomes, the state in the next time step is given as $\boldsymbol{s}^{(t+1)} = \left(i, n_k^{(t)}, p_X^{(t)}\right)$.

The next result characterizes the set $\mathcal{S}^*$ of terminating states for the functions arg max, max, or the pair (arg max, max).

Proposition 3. *When the CEO wishes to determine the* arg max *of the set of users' values,* $\boldsymbol{s}^{(t)} \in \mathcal{S}_A^*$ *iff* $N^{(t)} = 1$ *or* $L^{(t)} = 1$.

Proof. If at some time t there is only one user still contending for the resource it must be the unique maximizer. If the set of possible values consists of a single value, then all remaining users' values equal this value and they are all maximizers. In either case, the CEO has losslessly determined the set of arg max users. □

Proposition 4. *When the CEO wishes to determine either the* max *or the pair* (arg max, max) *of the set of users' values,* $\boldsymbol{s}^{(t)} \in \mathcal{S}_M^*$ *iff* $L^{(t)} = 1$.

Proof. If the set of possible values consists of a single value, then all remaining users' values equal this value and they are all maximizers. □

The set of terminating states is larger when the CEO wishes to determine the arg max because communication can stop when a single user is left, regardless of the set of remaining possible values. When determining the max, the CEO still needs subsequent rounds of communication with the single remaining user to determine its value. The following gives the optimal cost $C_M^{(t)}$ for $\boldsymbol{s}^{(t)} \in \mathcal{S}_A^* \setminus \mathcal{S}_M^*$.

Proposition 5. *For $\boldsymbol{s}^{(t)} \in \mathcal{S}_A^* \setminus \mathcal{S}_M^*$, the optimal quantization strategy is for the single remaining user to entropy code the state , and thus complete the communication in one additional round, i.e., $C_M^{(t)}\left(\boldsymbol{s}^{(t)}\right) = (1-\lambda)\log_2 L^{(t)} + \lambda$.*

Proof. First, we show that for any quantizer $\boldsymbol{n}$ the expected rate is $\log_2 L^{(t)}$. We proceed by induction. The base case of $L^{(t)} = 2$ is immediate. For the inductive step, suppose we have a quantizer $\boldsymbol{n} = (n_1, \ldots, n_K)$; then

$$R(L^{(t)}) = H(p_Q) + \mathbb{E}\left[R(L^{(t+1)})\right] \tag{4.7}$$

$$H(p_Q) = \log_2 L^{(t)} - \frac{1}{L^{(t)}} \sum_{k=1}^{K} n_k \log_2 n_k \tag{4.8}$$

and

$$\mathbb{E}\left[R(L^{(t+1)})\right] = \frac{1}{L^{(t)}} \sum_{k=1}^{K} n_k \log_2 n_k. \tag{4.9}$$

We conclude that $R(L^{(t)}) = \log_2 L^{(t)}$, and the cost depends only on the delay. A minimum delay of 1 is achieved by $\boldsymbol{n} = (1, \ldots, 1)$. □

The next proposition shows how the cost of computing the arg max is related to the cost of computing the max.

Proposition 6. *The cost for the CEO to compute the* max *exceeds the cost of computing the* arg max*, but the difference between the two costs goes to zero as the number of users N increases. We have*

$$C_A^{(t)}\left(\boldsymbol{s}^{(t)}\right) \leq C_M^{(t)}\left(\boldsymbol{s}^{(t)}\right) \leq C_A^{(t)}\left(\boldsymbol{s}^{(t)}\right) + \bar{\Delta} \tag{4.10}$$

where $\bar{\Delta} \to 0$ *as* $N \to \infty$.

Proof. Recall that $\mathcal{S}_M^* \subset \mathcal{S}_A^*$ (Proposition 3 & Proposition 4). Therefore, an optimal quantization policy for computing the max is a feasible quantization policy for computing the arg max and the lower bound is immediate.

To establish the upper bound, an optimal quantization policy for the arg max can be extended into a feasible quantization policy for the max. Starting at time $t = 0$, the CEO follows the optimal quantization policy for arg max until some time $t = \delta$ such that $\boldsymbol{s}^{(\delta)} \in \mathcal{S}_A^*$. If $\boldsymbol{s}^{(\delta)} \in \mathcal{S}_M^*$, then the CEO has determined the max. If $\boldsymbol{s}^{(\delta)} \in \mathcal{S}_A^* \setminus \mathcal{S}_M^*$, then the CEO has determined who the unique maximizer is but not what their value is. Since there is a single user left, the minimum cost $C^{(\delta)}\left(\boldsymbol{s}^{(\delta)}\right) = (1-\lambda)H\left(p_X^{(\delta)}\right) + \lambda$ (cf. Proposition 5). We have

$$H\left(p_X^{(\delta)}\right) \leq \log_2 L^{(\delta)} \leq \log_2\left(L^{(0)} - 1\right) \tag{4.11}$$

where the last inequality follows from observing that the size of the support set decreases by at least one at each round and $\delta \geq 1$. As this is a *feasible* quantization policy for the max policy we have

$$\begin{aligned} C_M^{(t)} &\leq C_A^{(t)} + \sum_{\boldsymbol{s}^{(\delta)} \in \mathcal{S}_A^* \setminus \mathcal{S}_M^*} \mathbb{P}\left(\boldsymbol{s}^{(\delta)}\right)\left[(1-\lambda)H\left(p_X^{(\delta)}\right) + \lambda\right] \\ &\leq C_A^{(t)} + \mathbb{P}\left(\mathcal{E}\right)\left[(1-\lambda)\log_2\left(L^{(0)} - 1\right) + \lambda\right] \end{aligned} \tag{4.12}$$

where $\mathcal{E}$ is the event $\left\{\boldsymbol{s}^{(\delta)} \in \mathcal{S}_A^* \setminus \mathcal{S}_M^*\right\}$ and $\mathbb{P}\left(\boldsymbol{s}^{(\delta)}\right)$ is the probability of the quantization policy ending in the state $\boldsymbol{s}^{(\delta)}$. We have that $\mathcal{E} \subseteq \{|\arg\max_i X_i| = 1\}$. Suppose that there are two or more maximizers. Then, for any sequence of quantizers, the set of maximizers will respond with the same quantized utility and the only possible terminating state is $\boldsymbol{s}^{(\delta)} \in \mathcal{S}_M^*$.

To show that $\bar{\Delta} \to 0$ as $N \to \infty$ it suffices to show that

$$\lim_{N\to\infty} \mathbb{P}\left(\left|\arg\max_i X_i\right| = 1\right) = 0. \tag{4.13}$$

Let A denote the event that $|\arg\max_i X_i| = 1$ and B_k denote the event $\max_i X_i = k$. By the law of

total probability we have

$$\mathbb{P}(A) = \sum_{k=1}^{L} \mathbb{P}(A \cap B_k) = \sum_{k=1}^{L} N F_X(k-1)^{N-1} p_X(k). \tag{4.14}$$

Since

$$\lim_{N\to\infty} N F_X^{N-1}(k-1) p_X(k) = 0 \quad \forall\, k = 1, \ldots, L, \tag{4.15}$$

it follows that $\mathbb{P}(A) \to 0$ as $N \to \infty$ □

To trace out the rate-delay trade-offs of interactive scalar quantization, the dynamic program of (4.4) is solved for multiple values of λ and the corresponding rate-delay values are plotted parametrically. This is done in §4.6 for representative source distributions. We conclude this section with an assumption concerning the search space $\mathcal{A}^{(t)}$ in (4.4).

Assumption 1. *We take the set of quantizers in solving* (4.4) *to be the set of all partitions of the integer L instead of the set of all compositions.*

Remark. In §4.3, we showed that set of quantizers is isomorphic to the set of compositions of the integer L. Thus there are 2^{L-1} possible quantizers to search over in solving (4.4). A related combinatorial object is a partition of the integer L. Unlike compositions, order does not matter for partitions. For example, consider the integer 3: there are 4 compositions (namely $\{(3), (2,1), (1,2), (1,1,1)\}$) while there are only 3 partitions ($\{(3), (2,1), (1,1,1)\}$). As the integer L gets larger, the number of partitions grows slower than the number of compositions (cf. Figure 4.2, left hand side). For computational tractability, we take the set of quantizers to be the set of all partitions instead of compositions. Our justification for doing so is shown in the right hand side of Figure 4.2. For a small initial number of users N and initial support set size $L = 16$, we computed the rate and delay cost components for every composition and every partition assuming subsequent rounds are solved optimally. The Pareto optimal boundary for compositions (markers) matches the boundary for partitions (no markers) for these parameters. For a given source distribution, as the number of users increases the distribution of the maximum becomes more and mored "peaked" about the

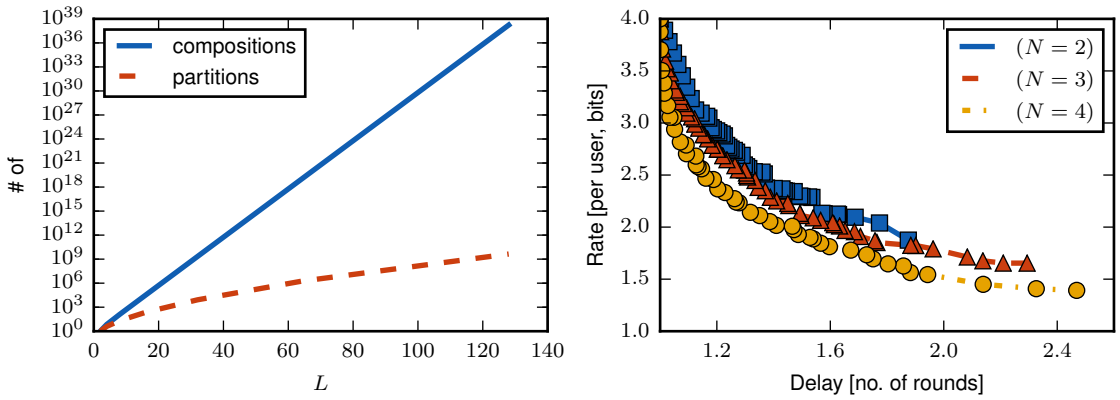


Figure 4.2: Comparison of compositions and partitions. The left hand side shows how the number of compositions and partitions grows as a function of the support set size L and the right hand side shows the rate-delay trade-off obtained by optimizing over the set of compositions (*markers*) and partitions (*no markers*).

largest possible value. Since the CEO is seeking to identify either the arg max or max (or both), we expect that quantizers that more finely quantize the larger values of the support set will outperform those quantizers that do not.

4.5 Analysis of Suboptimal Schemes

The dynamic programming formulation of the previous section is amenable to computing the minimum cost, and therefore the rate-delay trade-off, of interactive scalar quantization as an achievable scheme. A drawback with computing the solution to the dynamic program is that it does not provide insight into how the minimum cost scales in the number of users N and/or support set size L. Additionally, the computation provides little insight about the structure of the optimal quantizers. In this section, we consider two simple quantization strategies and derive expressions for the associated rate and delay. We then generalize to a family of strategies and prove their near-optimality; the significant reduction in the size of the search space results in faster computation of (4.4) with only a small penalty.

In this section, we assume the users' utilities have a uniform distribution. This assumption is motivated by the results of §4.6.1 and the analytical tractability of the resulting rate and delay expressions.

4.5.1 Binary search

We first consider a quantization strategy for computing the arg max inspired by binary search. At each round, the remaining support set is divided in half and the users indicate whether their values lie in the lower or upper half. This process is repeated until either a single user remains or the support set has been reduced to one.

We assume L is a power of two to repeatedly divide in half. The rate for this scheme is given by

$$R_b(N, 2L) = N + \frac{R_b(N, L) + \sum_{i=2}^{N} \binom{N}{i} R_b(i, L)}{2^N} \tag{4.16}$$

with base cases $R_b(1, \cdot) = R_b(\cdot, 1) = 0$. Since the initial support set size $(2L)$ is being halved, each user replies with a single bit for a total of N bits. Depending on the users' values, all N users could be in the lower half (L) which happens with probability 2^{-N}; or i users could be in the upper half which happens with probability $\binom{N}{i}2^{-N}$. Following a similar line of reasoning, delay is given by

$$\tau_b(N, 2L) = 1 + \frac{\tau_b(N, L) + \sum_{i=2}^{N} \binom{N}{i} \tau_b(i, L)}{2^N}, \tag{4.17}$$

with base cases $\tau_b(1, \cdot) = \tau_b(\cdot, 1) = 0$.

Proposition 7. *If L is a power of two, the expected rate of computing the* arg max *with binary search is*

$$R_b(N, L) = 2N \left(1 - \frac{1}{L}\right) \tag{4.18}$$

and the expected delay is bounded by

$$\tau_b(N, L) \leq \min\{\log_2 N + 1, \log_2 L\} \tag{4.19}$$

for $N \geq 2$ and $L \geq 2$. It follows that

$$\lim_{L\to\infty} \frac{R_b(N, L)}{N} = 2, \qquad \lim_{L\to\infty} \tau_b(N, L) \leq \log_2 N + 1. \tag{4.20}$$

Proof. We prove the expression for expected rate by induction. The base case of $N = 2$ and $L = 2$ is immediate. For the inductive step, we have

$$\begin{aligned} & R_b(N, 2L) \\ & \overset{(a)}{=} N + 2^{-N} \left[2N \left(1 - \frac{1}{L} \right) + \sum_{i=2}^{N} \binom{N}{i} \left(2i \left(1 - \frac{1}{L} \right) \right) \right] \\ & \overset{(b)}{=} N + 2^{-(N-1)} \left(1 - \frac{1}{L} \right) 2^{(N-1)} N \end{aligned} \tag{4.21}$$

where (a) follows by the inductive assumption and (4.16) & (4.18), and (b) follows from a standard identity.

To prove the upper bound for delay, we first consider the case of L fixed and show that

$$\tau_b(N, L) \leq \log_2 L \quad \forall\, N. \tag{4.22}$$

We proceed by induction in L; the base case of $L = 1$ is immediate. For the inductive step, we have

$$\begin{aligned} & \tau_b(N, 2L) \\ & \overset{(a)}{\leq} 1 + 2^{-N} \left[\log_2 L + \sum_{i=2}^{N} \binom{N}{i} \log_2 L \right] \\ & \overset{(b)}{=} 1 + \left(1 - N 2^{-N} \right) \log_2 L \leq \log_2 (2L) \end{aligned} \tag{4.23}$$

where (a) follows by the inductive assumption and (4.17) & (4.19), and (b) follows from a standard identity.

We now consider the case of N fixed and show that

$$\tau_b(N, L) \leq \log_2 N + 1 \quad \forall\, L. \tag{4.24}$$

We proceed by induction in N; the base case of $N = 1$ is immediate. For the inductive step, let

$$g(i) \triangleq \log_2(2i) = \log_2 i + 1. \tag{4.25}$$

Observe that the recurrence for $\tau_b(\cdot, \cdot)$ can be written as

$$\tau_b(N, 2L) = 1 + \mathbb{E}\left[\tau_b(Z, L)\right] \tag{4.26}$$

where Z is a random variable with PMF given as

$$\mathbb{P}\left(Z = i\right) = \begin{cases} \binom{N}{i} 2^{-N} & i \in \{1, \ldots, N-1\} \\ 2^{-(N-1)} & i = N \end{cases} \tag{4.27}$$

and expected value

$$\mathbb{E}\left[Z\right] = N\left(\frac{1}{2} + 2^{-N}\right). \tag{4.28}$$

Applying the law of total expectation we have

$$\tau_b(N, 2L) = 1 + \mathbb{E}\left[\tau_b(Z, L) | Z > 1\right] \mathbb{P}\left(Z > 1\right) \tag{4.29}$$

which follows from $\tau_b(1, L) = 0$. Substituting the upper bound gives

$$\tau_b(N, 2L) \leq 1 + \mathbb{E}\left[g(Z) | Z > 1\right] \left(1 - N2^{-N}\right). \tag{4.30}$$

We have

$$\begin{aligned} \mathbb{E}\left[g(Z) | Z > 1\right] &\overset{(a)}{\leq} g\left(\mathbb{E}\left[Z | Z > 1\right]\right) \\ &\overset{(b)}{=} g\left(\frac{N}{2\left(1 - N2^{-N}\right)}\right) = \log_2\left(\frac{N}{1 - N2^{-N}}\right) \end{aligned} \tag{4.31}$$

where (a) follows by Jensen's inequality and (b) follows from

$$\begin{aligned}\mathbb{E}\left[Z|Z>1\right] &= \frac{\mathbb{E}\left[Z\right] - \mathbb{E}\left[Z|Z=1\right]\mathbb{P}\left(Z=1\right)}{\mathbb{P}\left(Z>1\right)} \\ &= \frac{N\left(\frac{1}{2}+2^{-N}\right) - N2^{-N}}{1-N2^{-N}} = \frac{N}{2\left(1-N2^{-N}\right)}.\end{aligned} \tag{4.32}$$

It follows that

$$\mathbb{E}\left[g(Z)|Z>1\right]\left(1-N2^{-N}\right) \leq \log_2 N \tag{4.33}$$

and we conclude $\tau_b(N, 2L) \leq 1 + \log_2 N$. □

Binary search is an attractive quantization strategy because it requires at most 2 bits per user on average to compute the arg max and the delay remains bounded (by the number of users) as the support set size grows. With the support set being halved at each round of interaction, the support set size does not equal one until after $\log_2 L$ rounds of interaction. Using binary search as a quantization strategy for computing the max has a delay of exactly $\log_2 L$ rounds (c.f. Proposition 4).

4.5.2 Max search

As described at the end of the previous subsection, binary search has a constant delay of $\log_2 L$ rounds when the CEO wishes to compute the max. We propose the following search strategy, which we refer to as *max search*: at each round, the users indicate to the CEO whether their source observation is the largest possible value in the current support. If at least one user replies in the affirmative, the interaction stops and the max has been found. Only if none of the users have the largest value does the interaction continue. This strategy is motivated by the following two observations: 1. for a fixed support set size, as the number of users increases, the PMF of the max becomes more and more peaked about the larger support set values (cf. Figure 4.3), and; 2. at every round of interaction, there is a quantization bin of size one and therefore a non-zero probability of the interaction ending.

The rate for max search can be written recursively as

$$R_m(N, L) = Nh(p_L) + (1-p_L)^N R_m(N, L-1) \tag{4.34}$$

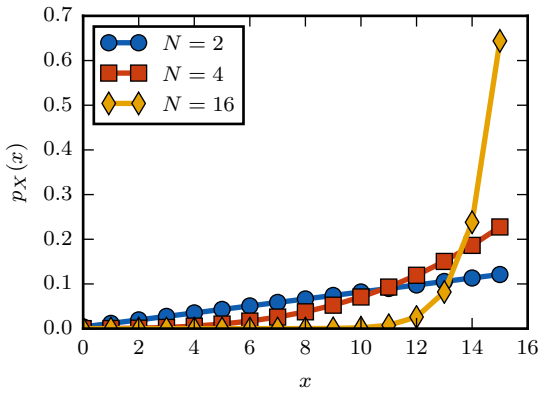


Figure 4.3: Probability mass function for $\max_i X_i$ when X_i's are i.i.d. uniform ($L = 16$) as a function of the number of users. As N increases, the probability becomes more concentrated around the larger values.

where $p_L = 1/L$; the delay can be written recursively as

$$\tau_m(N, L) = 1 + (1 - p_L)^N \tau_m(N, L - 1). \tag{4.35}$$

Proposition 8. *We have*

$$R_m(N, L) = N \sum_{i=2}^{L} \left(\frac{i}{L}\right)^N h(1/i) \tag{4.36a}$$

$$\tau_m(N, L) = \sum_{i=2}^{L} \left(\frac{i}{L}\right)^N. \tag{4.36b}$$

It follows that

$$\lim_{N\to\infty} \frac{R_m(N, L)}{N} = h\left(\frac{1}{L}\right), \quad \lim_{N\to\infty} \tau_m(N, L) = 1. \tag{4.37}$$

Proof. By induction on L. The base case of $L = 2$ is immediate. For the inductive step we have

$$\begin{aligned} &R_m(N, L) \\ &= Nh(1/L) + \left(\frac{L-1}{L}\right)^N N \sum_{i=1}^{L-1} \left(\frac{i}{L-1}\right)^N h(1/i) \end{aligned} \tag{4.38}$$

which follows from the inductive assumption and (4.34) & (4.36a). The proof for τ_m follows the

same arguments. □

This scheme has very low rate and delay as the number of users N gets larger. Using this scheme, however, the CEO does not have the ability to select a desired rate-delay trade-off. In the next subsection, we extend this simple search strategy to a family of search strategies that give the CEO the ability to operate at a desired rate-delay trade-off. We conclude by noting that even though max search is designed to enable the CEO to interactively compute the max, Proposition 6 establish it as a quantization strategy for computing the arg max as well.

4.5.3 Extended max search

We extend the quantization strategy of the previous section into a family of quantization strategies. The previous strategy worked by asking the users to indicate whether or not their state $X_i = \max \mathcal{X}^{(t)}$ at each iteration t, terminating when at least one user replied in the affirmative. We extend this strategy by asking the users to indicate which of the $K-1$ largest values of $\mathcal{X}^{(t)}$ they have or indicating their state is not one of these values. For example: consider $L = 5$ and $K = 4$. The quantizer for this parameter set would be $\boldsymbol{n} = (2, 1, 1, 1)$. Like the quantization strategy of the previous section, this family of strategies has the property that iteration continues if and only if all N users' states are in the first bin. Unlike the previous strategy, we are not able to write a closed form expression for the rate and delay components when selecting quantizers from this family of quantizers. However, we are able to prove several non-trivial and import properties of this family.

We begin by giving a formal description of the family. For notational compactness, let $\mathbf{1}_k$ be the k-tuple of all ones and denote tuple concatenation as $\oplus$. Suppose we are in a state $\boldsymbol{s} = (N, L)$. Define

$$\mathcal{L}(L) = \{(L-k+1) \oplus \mathbf{1}_{k-1} : k = 2, \ldots, L\}. \tag{4.39}$$

For example: $\mathcal{L}(4) = \{(3,1), (2,1,1), (1,1,1,1)\}$. Our first result shows that for any quantizer $\boldsymbol{n} \in \mathcal{L}(L)$, permutation of the bin sizes results in a quantizer with a higher cost. Recall Assumption 1 of §4.4 where we took the search space of quantizers to be the set of partitions instead of compositions. Our justifications for this assumption were concerns of computational complexity (Figure 4.2, left)

and the observation that the optimal solution was still found when searching over partitions for small problem instances (Figure 4.2, right). Since permutation of an integer partition gives (in general) a composition, the next result is a further justification of Assumption 1.

Proposition 9. *For computing the* max *of the users' values, the K-level quantizer*

$$\boldsymbol{n} = (L - K + 1) \oplus \mathbf{1}_{K-1} \tag{4.40}$$

has lower cost than any other quantizer obtained by permutation.

Proof. Let $\boldsymbol{n}_m$ be the quantizer obtained from $\boldsymbol{n}$ by shifting the bin of size $L - K + 1$ m locations to the right. For example $\boldsymbol{n}_3 = \mathbf{1}_3 \oplus (L - K + 1) \oplus \mathbf{1}_{K-4}$. The stage cost of the quantizer

$$(1 - \lambda) R\left(\boldsymbol{a}^{(t)}, \boldsymbol{s}^{(t)}\right) + \lambda \tau\left(\boldsymbol{a}^{(t)}, \boldsymbol{s}^{(t)}\right) \tag{4.41}$$

is invariant to permutation. The expected cost to go of the quantizer $\boldsymbol{n}$ is

$$\left(\frac{L - K + 1}{L}\right)^N C_M^{(t+1)}\left(N, L - K + 1\right) \tag{4.42}$$

and the expected cost to go of the quantizer $\boldsymbol{n}_m$ is

$$\begin{aligned} &\sum_{i=1}^{N} \rho(i) C_M^{(t+1)}\left(i, L - K + 1\right) \\ \rho(i) &= \binom{N}{i} \left(\frac{m}{L}\right)^{N-i} \left(\frac{L - K + 1}{L}\right)^i \end{aligned} \tag{4.43}$$

Taking the difference we see

$$\sum_{i=1}^{N-1} \rho(i) C_M^{(t+1)}\left(i, L - K + 1\right) \geq 0 \tag{4.44}$$

and conclude that $\boldsymbol{n}$ has better cost than $\boldsymbol{n}_m$. As the choice of m was arbitrary, the result holds for any permutation. □

An attractive property of $\mathcal{L}(L)$ is that $|\mathcal{L}(L)| = L - 1$ where as the number of all quantizers is exponential in L. If we could show that $\mathcal{L}(L)$ was *sufficient* for solving (4.4) (instead of the set of all quantizers) optimally, it would represent a significant reduction in computational complexity. As a first step, we show for a given quantizer $\boldsymbol{n} \notin \mathcal{L}(L)$ how to select a quantizer $\boldsymbol{n}_r \in \mathcal{L}(L)$ that asymptotically in N has performance no worse than $\boldsymbol{n}$.

Consider a K-bin quantizer $\boldsymbol{n} = (n_1, \ldots, n_K)$; if $n_1 = L - K + 1$, then $\boldsymbol{n} \in \mathcal{L}(L)$. Otherwise $n_1 < L - K + 1$ and $\boldsymbol{n} \notin \mathcal{L}(L)$. Let $\boldsymbol{n}_r = (L - K + 1) \oplus \mathbf{1}_{K-1} \in \mathcal{L}(L)$ be the quantizer with the same number of bins as $\boldsymbol{n}$. For example: if $\boldsymbol{n} = (2, 2, 1)$ then $\boldsymbol{n}_r = (3, 1, 1)$.

For notational compactness in the rest of the section, denote the current state as $\boldsymbol{s} = (N, L)$ (which is fixed and known) and the state in the next iterations as $\boldsymbol{s}' = (N', L')$ (which are discrete random variables whose PMF depends on the select quantizer). The cost when using $\boldsymbol{n}$ is

$$C_{\boldsymbol{n}}(N, L) = (1 - \lambda) N H(\boldsymbol{n}) + \lambda + \mathbb{E}_{\boldsymbol{n}} \left[C(N', L') \right] \tag{4.45}$$

where

$$\mathbb{E}_{\boldsymbol{n}} \left[C(N', L') \right] = \sum_{j=1}^{k} \sum_{i=1}^{N} \mathbb{P}_{\boldsymbol{n}} \left[N' = i, L' = n_j \right] C\left(i, n_j\right), \tag{4.46a}$$

and

$$\begin{aligned} &\mathbb{P}_{\boldsymbol{n}} \left[N' = i, L' = n_j \right] = \\ &\left[\left(\frac{n_{1:j}}{L} \right)^N - \left(\frac{n_{1:j-1}}{L} \right)^N \right] \frac{\binom{N}{i} \rho_j^i \left(1 - \rho_j\right)^{N-i}}{1 - \left(1 - \rho_j\right)^N}, \end{aligned} \tag{4.46b}$$

and $\rho_j = \frac{n_j}{n_{1:j}}$. The cost when using $\boldsymbol{n}_r$ is

$$C_{\boldsymbol{n}_r}(N, L) = (1 - \lambda) N H(\boldsymbol{n}_r) + \lambda + \mathbb{E}_{\boldsymbol{n}_r} \left[C(N', L') \right] \tag{4.47}$$

where

$$\mathbb{E}_{\boldsymbol{n}_r} \left[C(N', L') \right] = \left(\frac{L - K + 1}{L} \right)^N C(N, L - K + 1). \tag{4.48}$$

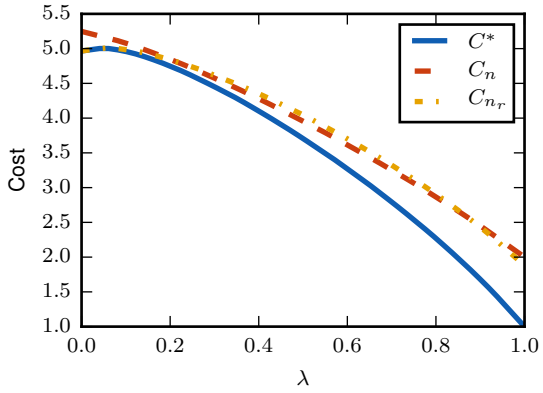


Figure 4.4: Comparison of the cost for $\boldsymbol{n} = (11, 5)$ and $\boldsymbol{n}_r = (15, 1)$ for the state $\boldsymbol{s} = (N = 2, L = 16)$. $\boldsymbol{n}_r$ does not outperform $\boldsymbol{n}$ in terms of cost for all λ.

Taking the difference we have

$$\begin{aligned}\Delta_{\boldsymbol{n},\boldsymbol{n}_r} &= C_{\boldsymbol{n}}(N, L) - C_{\boldsymbol{n}_r}(N, L) \\ &= (1-\lambda)N\left(H(\boldsymbol{n}) - H(\boldsymbol{n}_r)\right) \\ &+ \mathbb{E}_{\boldsymbol{n}}\left[C(N', L')\right] - \mathbb{E}_{\boldsymbol{n}_r}\left[C(N', L')\right].\end{aligned} \tag{4.49}$$

The difference in the stage costs of quantizer $\boldsymbol{n}$ and $\boldsymbol{n}_r$ is expressed in terms of the difference in entropies of the induced probability mass functions $H(\boldsymbol{n}) - H(\boldsymbol{n}_r) \geq 0$ where the inequality follows from the fact that $\boldsymbol{q}$ is majorized by $\boldsymbol{b}$ and $x \log x$ is convex [58].

Unfortunately, the difference in the expected cost to go is not always positive. In fact, the difference in the expected cost to go can be negative enough to offset the positive difference in the quantizer rates. In Figure 4.4 we consider the case of $N = 2$ and $L = 16$ and compare the costs of the non-$\mathcal{L}(16)$ quantizer $\boldsymbol{n} = (11, 5)$ to the $\boldsymbol{n}_r \in \mathcal{L}(16)$ quantizer $(15, 1)$. For certain values of λ, $C_{\boldsymbol{n}}(N, L)$ is less than $C_{\boldsymbol{n}_r}(N, L)$. This does *not* disprove the sufficiency of $\mathcal{L}(16)$ in solving (4.4) optimally; the range of λ for which $C_{\boldsymbol{n}}(N, L) < C_{\boldsymbol{n}_r}(N, L)$ does not coincide with the range of λ for which $C(N, L) = C_{\boldsymbol{n}_r}(N, L)$. The above derivation allows us to prove that $\mathcal{L}(L)$ is asymptotically sufficient for solving (4.4) optimally.

Proposition 10. *$\mathcal{L}(L)$ is asymptotically sufficient for minimizing the cost associated with interac-*

tively computing the max. *The set* $\mathcal{L}(L)$ cannot *be made smaller without losing this property. For a given value of* L *and* $\boldsymbol{n} \notin \mathcal{L}(L)$

$$\lim_{N\to\infty} \Delta_{\boldsymbol{n},\boldsymbol{n}_r} \geq 0 \tag{4.50}$$

where $\boldsymbol{n}_r \in \mathcal{L}(L)$ *has the same number of bins as* $\boldsymbol{n}$*.*

Proof. The proof proceeds by induction with an inductive assumption that only quantizers from $\mathcal{L}(L^{(t)})$ for $t \geq 1$ are used for subsequent rounds.

For the state (N, L), $\Delta_{\boldsymbol{n},\boldsymbol{n}_r}$ is asymptotically (in N) non-negative. By assumption $\boldsymbol{n} \notin \mathcal{L}(L)$, therefore $L-K+1 > n_1 \geq n_2 \geq \cdots \geq n_K \geq 1$. This, in turn, implies that $\hat{\boldsymbol{n}}_1 = (n_1) \oplus \mathbf{1}_{L-K+1-n_1} \in \mathcal{L}(L-K+1)$ is a valid quantizer for the state $\boldsymbol{s}' = (N^{(t)}, L^{(t)} - K + 1)$ and

$$\begin{aligned} C_M(\boldsymbol{s}') &\leq (1-\lambda) N^{(t+1)} H(\hat{\boldsymbol{n}}_1) + \lambda \\ &+ \left(\frac{n_1}{L^{(t)} - K + 1}\right)^n C_M(N^{(t)}, n_1). \end{aligned} \tag{4.51}$$

Next, there exists some k^* such that $n_2 \geq \cdots \geq n_{k^*} > 1$; if not then $\boldsymbol{n} \in \mathcal{L}(L)$. We have that

$$C_M(i, n_j) \geq \lambda \quad j \leq k^* \quad C_M(i, n_j) = 0 \quad j > k^*. \tag{4.52}$$

Finally, for notational compactness let $\nu = \frac{L-K+1}{L} < 1$.

We then have

$$\begin{aligned}
\Delta_{\boldsymbol{n},\boldsymbol{n}_r} &\overset{(a)}{\geq} (1-\lambda)N\left(H(\boldsymbol{n}) - H(\boldsymbol{n}_r) - \nu^N H(\hat{\boldsymbol{n}}_1)\right) \\
&+ \sum_{j=1}^{k^*}\sum_{i=1}^{N} \mathbb{P}_{\boldsymbol{n}}\left[N^{(t+1)} = i, L^{(t+1)} = n_j\right] C\left(i, n_j\right) \\
&- \nu^n \left(\left(\frac{q_1}{L-K+1}\right)^N C(N, n_1) + \lambda\right) \\
&\overset{(b)}{=} (1-\lambda)N\left(H(\boldsymbol{n}) - H(\boldsymbol{n}_r) - \nu^N H(\hat{\boldsymbol{n}}_1)\right) \\
&+ \sum_{j=2}^{k^*}\sum_{i=1}^{N} \mathbb{P}_{\boldsymbol{n}}\left[N^{(t+1)} = i, L^{(t+1)} = n_j\right] C\left(i, n_j\right) - \nu^N\lambda \qquad (4.53) \\
&\overset{(c)}{\geq} (1-\lambda)N\left(H(\boldsymbol{n}) - H(\boldsymbol{n}_r) - \nu^n H(\hat{\boldsymbol{n}}_1)\right) \\
&+ \sum_{j=2}^{k^*}\sum_{i=1}^{N} \mathbb{P}_{\boldsymbol{n}}\left[N^{(t+1)} = i, L^{(t+1)} = n_j\right] \lambda - \nu^N\lambda \\
&= (1-\lambda)N\left(H(\boldsymbol{n}) - H(\boldsymbol{n}_r) - \nu^N H(\hat{\boldsymbol{n}}_1)\right) \\
&+ \lambda\left(\frac{n_{1:k^*}}{L}\right)^N \left[1 - \left(\frac{n_1}{n_{1:k^*}}\right)^N - \left(\frac{L-K+1}{n_{1:k^*}}\right)^N\right]
\end{aligned}$$

where

(a) follows from (4.46), (4.48), (4.49), (4.51), and (4.52)

(b) follows from

$$\mathbb{P}_{\boldsymbol{n}}\left[N^{(t+1)} = i, L^{(t+1)} = q_1\right] = \begin{cases} 0, & i < n \\ \left(\frac{q_1}{L}\right)^n, & i = n \end{cases} \qquad (4.54)$$

and

(c) follows from (4.52)

$n_{1:k^*} = L - k + k^* > K - k + 1$ and therefore the right hand side of (4.53) has a non-negative limit because

$$\lim_{N\to\infty} \lambda\left(\frac{n_{1:k^*}}{L}\right)^N \left[1 - \left(\frac{n_1}{n_{1:k^*}}\right)^N - \left(\frac{L-K+1}{n_{1:k^*}}\right)^N\right] = 0. \qquad (4.55)$$

The set $\mathcal{L}(L)$ contains one and only one quantizer for each possible bin size. If this set were

smaller, then for a given $\boldsymbol{n}$ the quantizer $\boldsymbol{n}_r$ (which has the same number of bins as $\boldsymbol{n}$) may not be in $\mathcal{L}'(L) \subset \mathcal{L}(L)$. □

In general, there exists values of N and L for which $\mathcal{L}(L)$ is not sufficient. Let $\mathcal{Q}(L)$ be the set of all quantizers for support set of size L. For $\mathcal{U} \subset \mathcal{Q}$, define

$$\begin{gathered}\Delta(\mathcal{U}, \lambda) = \min_{q \in \mathcal{U}} C(\lambda, q) - \min_{q \in \mathcal{Q}} C(\lambda, q) \\ \overline{\Delta}(\mathcal{U}) = \max_{\lambda} \Delta(\mathcal{U}, \lambda) \quad \underline{\Delta}(\mathcal{U}) = \min_{\lambda} \Delta(\mathcal{U}, \lambda).\end{gathered} \tag{4.56}$$

Here $\Delta(\mathcal{U}, \lambda)$ is understood as the λ-dependent "gap to optimality" when using quantizers from $\mathcal{U}$ only instead of all quantizers $\mathcal{Q}$; $\overline{\Delta}(\mathcal{U})$ and $\underline{\Delta}(\mathcal{U})$ are (respectively) the worst-case and best-case gap to optimality. From these definitions, we have $0 \leq \underline{\Delta}(\mathcal{U}) \leq \overline{\Delta}(\mathcal{U})$.

The top left side of Figure 4.5 is a plot of $\overline{\Delta}(\mathcal{L}(L))$ and $\underline{\Delta}(\mathcal{L}(\mathcal{L}))$ and the bottom left side is a plot of $\overline{\Delta}(\mathcal{Q}(L) \setminus \mathcal{L}(L))$ and $\underline{\Delta}(\mathcal{Q}(L) \setminus \mathcal{L}(L))$ as a function of L. When $N = 2$ and $L \leq 34$, $\overline{\Delta}(\mathcal{L}) = 0$ and $\underline{\Delta}(\mathcal{Q} \setminus \mathcal{L}) > 0$. This means that for all λ, the optimal quantizer for $N = 2$, L can be taken from $\mathcal{L}$ and there is no value of λ for which the optimal quantizer *can be* taken from $\mathcal{Q} \setminus \mathcal{L}$. These two inequalities imply that $\mathcal{L}$ is necessary and sufficient for minimizing the cost. When $N = 2$ and $35 \leq L \leq 76$, from the figure $0 = \underline{\Delta}(\mathcal{L}) < \overline{\Delta}(\mathcal{L})$ and $0 = \underline{\Delta}(\mathcal{Q} \setminus \mathcal{L}) < \overline{\Delta}(\mathcal{Q} \setminus \mathcal{L})$. This means that there exists a value of λ such that the optimal quantizer $n^* \notin \mathcal{L}$. The right side of Figure 4.5 plots the gap to optimality for $\mathcal{L}(L)$ as a function of λ for representative values of L. The top right of the figure plots the absolute magnitude of this gap, whereas the bottom right shows the magnitude of this gap as a relative percentage of the optimal value. The worst-case gap is growing in L (as Figure 4.5 shows as well), but the gap is still very small, at less than 0.35%, when $L = 76$. There is a range of $\lambda > 0.6$ for which gap to optimality is identically zero—this range is where the delay component of cost is weighted more heavily than the rate component. A minimum delay of 1 can be achieved with the quantizer $(1, \ldots, 1)$ which is in $\mathcal{L}(L)$. Finally, for $L_1 < L_2$ the interval of λ for which $\mathcal{L}(L_1)$ is not sufficient is a subset of the interval for which $\mathcal{L}(L_2)$ is not sufficient.

The previous counterexample was for the case of $N = 2$ and we needed $L \geq 35$ for $\mathcal{L}(L)$ to no longer be necessary and sufficient for optimally solving (4.4). When N gets larger, the value of L at

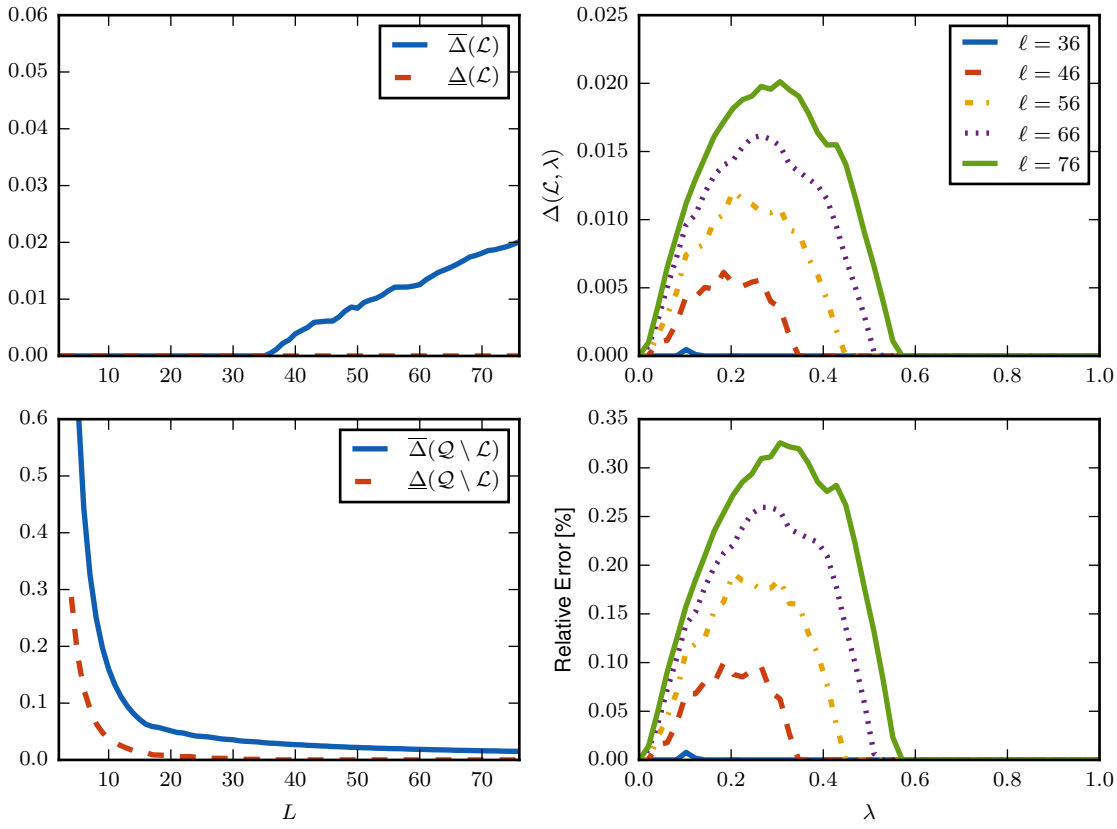


Figure 4.5: For $N = 2$: (left) $\overline{\Delta}(\cdot)$ and $\underline{\Delta}(\cdot)$ for $\mathcal{L}$ & $\mathcal{Q} \setminus \mathcal{L}$. $\overline{\Delta}(\mathcal{L}) > 0$ & $\underline{\Delta}(\mathcal{Q} \setminus \mathcal{L}) = 0$ for $L \geq 35$. (right) $\Delta(\mathcal{L}, \lambda)$ $L \in \{36, 46, 56, 66, 76\}$: (top) absolute error and (bottom) relative error. Max. relative error for $L = 76$ is $\approx 0.33\%$.

which $\mathcal{L}(L)$ is no longer necessary and sufficient gets larger as well.

In summary, when computing the max interactively we know that $\mathcal{L}(L)$:

1. is not an optimal search space for solving (4.4) in general;
2. is asymptotically (L fixed, N increasing) sufficient for solving (4.4);
3. has *linear* growth (vs. exponential for all quantizers), and;
4. incurs a small decrease in performance when it is not optimal.

As we show in §4.6, adding binary-search quantizer to $\mathcal{L}(L)$ results in a simplified search space for computing the arg max with little to no incurred penalty. Depending upon the system where this interactive quantization strategy is employed, the large reduction in computation costs may more than make up for the small increase in cost that occurs when using $\mathcal{L}(L)$ for selecting quantizers.

4.6 Results

In this section, we investigate the rate-delay trade-offs, both for the optimized scalar quantizer scheme (§4.4) and the proposed heuristics (§4.5). For brevity, when we refer to optimal rate-delay trade-offs, we are referring to the rate-delay trade-offs of the optimized scalar quantizer scheme. We begin by considering the optimal rate-delay trade-offs for a collection of representative distributions and show that the uniform distribution represents a worst-case distribution. This makes sense as the uniform distribution is entropy-maximizing for a given support set size. We then investigate the rate-delay trade-offs for binary and max search, demonstrating that these schemes can closely approximate the optimal rate-delay trade-off.

4.6.1 Optimized interactive quantization rate-delay trade-offs

For a comparison of the rate-delay trade-offs for various distributions, we consider the following representative distributions parameterized with L and p which effects the concentration of the distribution.

$$g_X(x; L, p) = \frac{(1-p)^{L-x-1}p}{1-(1-p)^L}, \quad x = 0, \ldots, L-1 \tag{4.57}$$

and,

$$b_X(x; L, p) = \binom{L}{x} p^x (1-p)^{L-x}, \quad x = 0, \ldots, L-1. \tag{4.58}$$

The distribution $g_X(x; L, p)$ is shown in Figure 4.6 (top left); the effect of varying p is to vary the "distance" from a uniform distribution. Also plotted in Figure 4.6 is the optimal rate-delay trade-offs for both the arg max (top center) and the max (top right) functions. As p is made smaller, the trade-offs get worse in that a larger delay is incurred for smaller rates. The distribution $b_X(x; L, p)$ is shown in Figure 4.6 (bottom left); unlike $g_X(x; L, p)$ the "spread" of the distribution is not sensitive to the parameter p. Looking at the rate-delay trade-offs (bottom center & left), p has little effect on the performance of the optimized scheme.

Figure 4.7 (top left) shows the rate-delay trade-offs for these distributions for both arg max and max as computed by solving (4.4) and finding the optimal homogeneous quantizer at each round. The trade-offs for uniform are worse than for the other distributions. For a given upper limit on delay, uniformly distributed sources will require more rate than the other two distributions. For a fixed alphabet size, as the number of users is increased (upper right), the trade-offs for the uniform distribution gets better. Figure 4.7 (bottom row) shows how the rate-delay trade-offs for computing arg max and max become identical as the number of users increases (cf. Proposition 6). We see that when N is small, the CEO is able to compute the arg max with either a lower rate (fixed delay) or lower delay (fixed rate) than it would require for computing the max with the same fixed rate or delay. This difference is especially large in the low rate/high delay regime. Doubling the number of users from 2 to 4 significantly reduces this difference.

4.6.2 Extended max search rate-delay trade-offs

As noted in §4.5, as the size of the support set L increases the number of possible quantizers gets large quickly. Based on the simple quantization strategies of binary search and max search, we proposed the extended max search family $\mathcal{L}(L)$ of quantizers. Figure 4.8 shows the rate-delay trade-off when the search space of (4.4) is taken to be 1. binary & max search with the entropy coding quantizer $(1, \ldots, 1)$ (solid line); 2. binary & extended max search (which includes the entropy coding

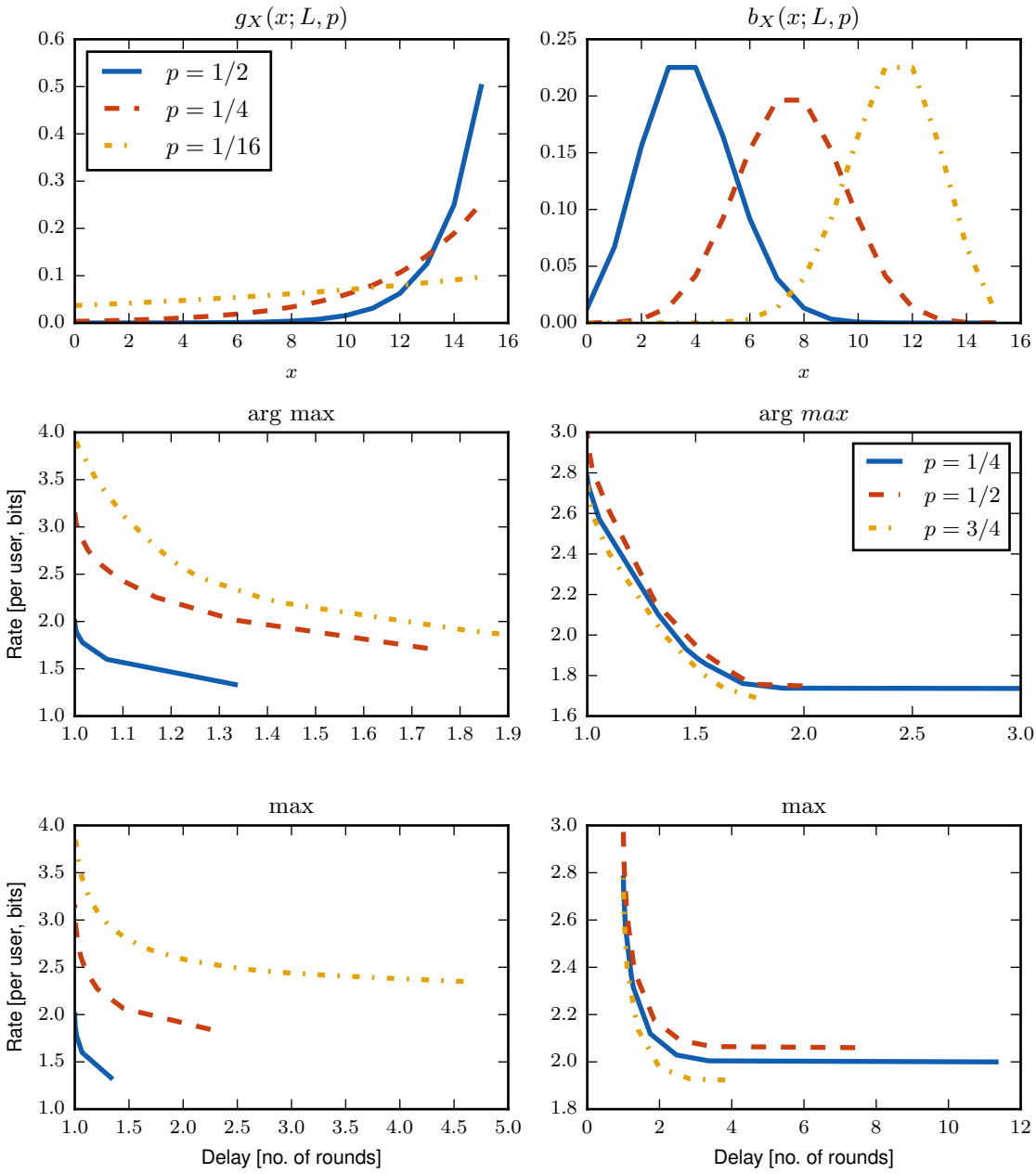


Figure 4.6: Rate-delay trade-offs for $g_X(x; L, p)$ (left column) and $b_X(x; L, p)$ (right column) for various values of p. PMFs are shown in the top row, rate-delay trade-offs for arg max in the center row, and max in the bottom row.

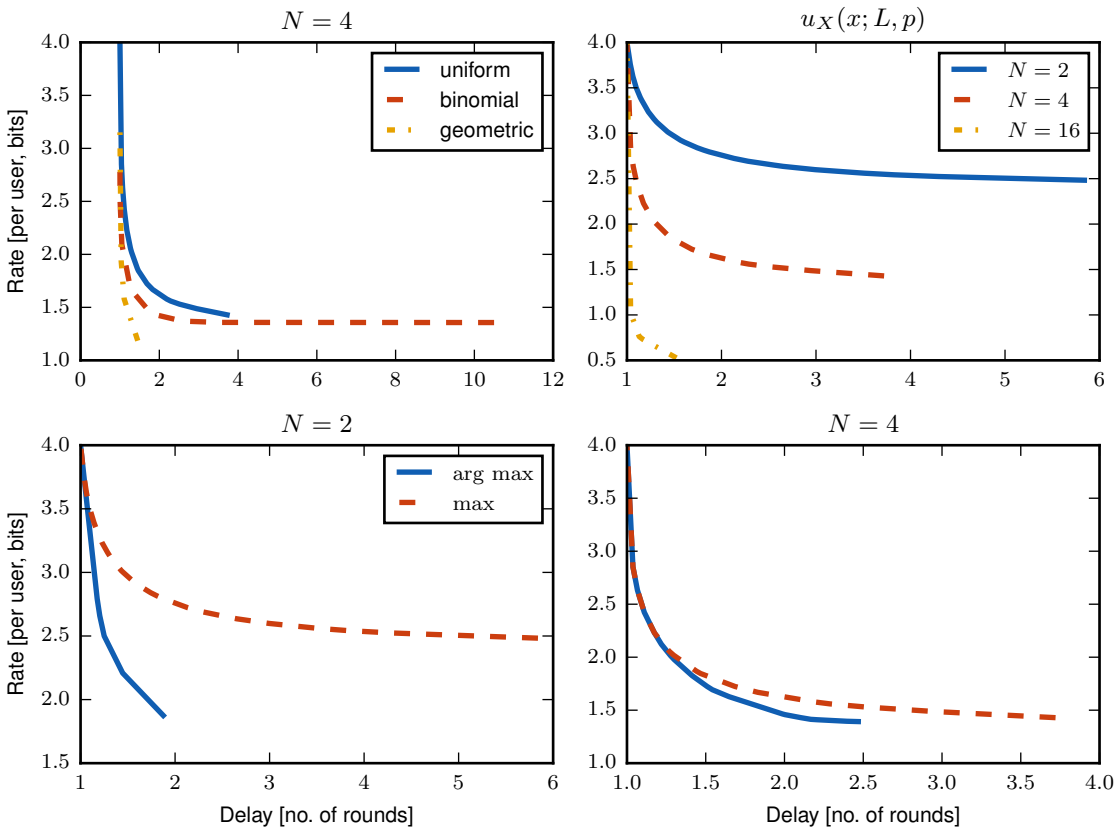


Figure 4.7: Rate-delay trade-offs of optimized interactive scalar quantization: (top, left) for uniform, $g_X(x; L, p)$, and $b_X(x; L, p)$ ($L = 16$), (top, right) when computing the max for varying number of users (N), and (bottom) when computing arg max vs. max for $N = 2$ (right) and $N = 4$. The source distribution was uniform with support set size $L = 16$.

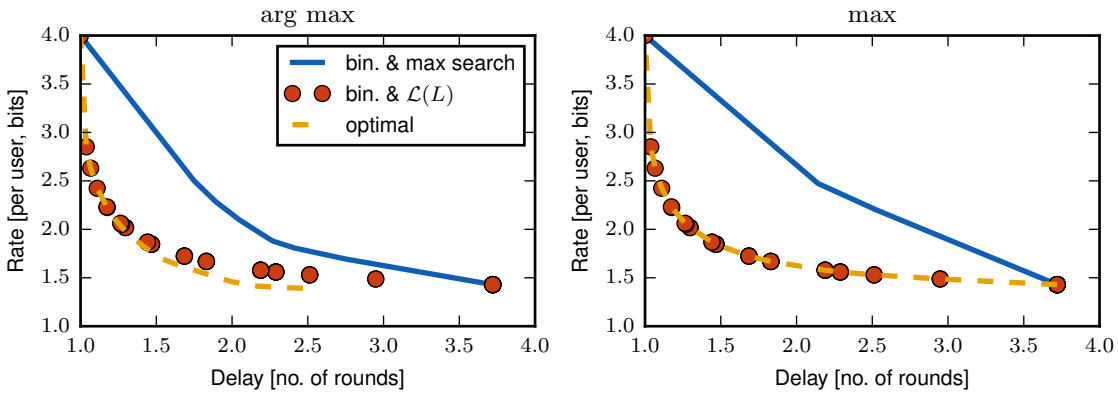


Figure 4.8: Rate-delay trade-off for various quantizer search spaces when computing arg max (left) and max (right). The source distribution was uniform with support set size $L = 16$ and the number of users was $N = 4$.

quantizer) (circle markers), and; 3. all possible quantizers (dashed line). The left side is for the case of computing the arg max and the right side is for the case of computing the max. Binary & max search together can achieve the minimum rate and minimum delay ends of the trade-off curve for both functions, but performs poorly in efficiently trading off delay for rate. For computing the arg max, extended max search is almost equal to the optimal trade-off curve, deviating in the low rate/high delay regime. For computing the max, extended max search equals the optimal trade-off curve. By Proposition 6 and Figure 4.7, we know that as N increases the cost of computing the arg max is equal to the cost of computing the max. Even though extended max search is designed with computation of the max in mind, Figure 4.8 and Proposition 6 show that it is an effective quantizer search space for computing the arg max.

4.7 Conclusion

In this chapter, we considered the problem of a CEO computing a function of distributed users' state as a model for distributed resource allocation. We proposed interactive scalar quantization as an achievable scheme for reducing the required rate to enable the CEO to compute the desired function losslessly at the expense of an increase in delay. We solved for optimal rate-delay trade-off of scalar quantization via a dynamic program. We established that asymptotically (in the number of users

N), the cost to compute arg max is the same as the cost to compute max. By considering simple quantization schemes based on binary search and max search, we designed a family of quantization strategies that is nearly optimal with a significantly reduced computational cost.

In the present work, we assumed that every user was using the same quantizer (i.e. *homogeneous quantization*). We know that in the rate-distortion problem, *heterogeneous quantization* can achieve a lower rate for a given distortion than homogeneous quantization [3, 46]. A future direction for the present work would be to extend the model to incorporate different quantizers at the different users. Small numerical experiments have demonstrated that substantial further reduction in the rate required to calculate the extremum at a bounded expected delay can be obtained by switching from homogeneous to heterogeneous designs. A potential obstacle is the dramatic increase in the size of the search space; the size of the search space is equal to the size of the search space of homogeneous quantization raised to N.

The current work demonstrates that the required rate can be reduced by tolerating a small increase in delay; in a similar manner, the required rate can be reduced by tolerating a small increase in distortion [3, 46]. The model of these two lines of inquiry could be combined into a single framework to quantify the rate savings that could be realized by tolerating both delay and distortion. This would require suitably modifying the cost function of (4.4) to include a term for distortion. At each round, the CEO would decide if the distortion is low enough to stop or if communication should continue.

Chapter 5: Conclusions

This study sought to investigate techniques for minimizing the overhead from data collection and/or trade off performance for reductions in overhead in communication networks. Three distinct problem instances were considered:

1. lossless transmission of distributed correlated sources across a network with capacity constraints;
2. scalar quantization for lossy distributed extremization, and;
3. interactive scalar quantization for lossless distributed extremization.

5.1 Lossless transmission of distributed correlated sources across a network with capacity constraints

In the first problem instance, the optimal lossless transmission of distributed sources across a network with capacity constraints was formulated as the solution to a linear program with an exponential (in the number of sources) number of constraints. Previous works have only made use of the fact that SW rate region is a contrapolymatroid as part of an iterative subgradient method [11]. Building upon earlier work that established the set of achievable rates as the intersection of the SW rate region with the polymatroid defined by the min-cut capacities, the present work identifies when the SW vertices are all feasible and gives an explicit characterization of all the vertices of the intersection of polymatroid with a contrapolymatroid for certain sub-/supermodular set functions. It is further shown that the size of the representation of the SW rate region is related to the conditional independence relationships among the sources and in some cases (e.g., Markov random fields) may require significantly fewer inequalities to describe the rate region. These properties were leveraged to develop a relationship between optimal solutions and the corner points of the SW rate region. A simple, but natural counter-example demonstrated that an optimal rate allocation may not be a

vertex of the SW rate region. The results concerning the feasibility of all the SW rate region vertices naturally extends from the single sink problem to the multi-sink setting.

5.2 Scalar quantization for lossy distributed extremization

In many resource allocation problems, a centralized controller needs to award some resource to a user selected from a collection of distributed users with the goal of maximizing the utility the user would receive from the resource. This can be modeled as the controller computing an extremization of the distributed users' utilities. The overhead rate necessary to enable the controller to reproduce the users' local state can be prohibitively high. Two approaches to reduce this overhead are lossy estimation and interactive communication.

In the lossy estimator framework, rate savings are achieved by tolerating a bounded expected reduction in utility. A layered approach of scalar quantization followed by entropy encoding was designed for distributed function extremization in the context of a resource allocation problem. A natural, non-quadratic distortion that measured the reduction in system performance was proposed and an exact expression as a function of the quantizer parameters was provided. The Karush-Kuhn-Tucker necessary conditions for optimal HomSQ are given. The existence of a more efficient (i.e., achieving the same distortion at a lower rate) HetSQ scheme was proven by a suitable modification of an optimal HomSQ scheme. By considered several example source distributions, it was demonstrated that the performance of HetSQ may be close to fundamental limits.

It is known that vector quantizers are more efficient than SQs, even when the source outputs being blocked into vectors are independent [22, 30]. Recent results have shown that local vector quantizers followed by SW encoding is optimal for certain two-terminal problems with continuous distributions [27, 28]. This motivates the consideration of vector quantization in future work.

5.3 Interactive scalar quantization for lossless distributed extremization

In interactive communication, rate savings come at the expense of delay. Motivated by the performance of scalar quantization in the lossy estimator framework, interactive scalar quantization is utilized as an achievable scheme for reducing the required rate to enable the CEO to compute the

desired function losslessly at the expense of an increase in delay. The optimal rate-delay trade-offs for scalar quantization were formulated as the minimum cost solution of a dynamic program and it is established that asymptotically (in the number of sources), the costs to compute the different extremization functions are equal. A family of quantization strategies is formulated that is nearly optimal with a significantly reduced computational complexity to search over.

For the interactive communication framework, it was assumed that every user was using the same quantizer (i.e. HomSQ). A future direction for the present work would be to extend the model to incorporate different quantizers at the different users. Small numerical experiments have demonstrated that substantial further reduction in the rate required at a bounded expected delay can be obtained by switching from HomSQ to HetSQ.

This study has demonstrated that the overhead can be reduced by tolerating a small increase in delay; in a similar manner, the overhead can be reduced by tolerating a small increase in distortion [3, 46]. The model of these two lines of inquiry could be combined into a single framework to quantify the overhead reduction that could be realized by tolerating both delay and distortion.

Bibliography

[1] Gwanmo Ku and John MacLaren Walsh. Resource allocation and link adaptation in lte and lte advanced: A tutorial. *IEEE Commun. Surveys Tuts.*, PP(99):1–1, 2014.

[2] Evolved universal terrestrial radio access (E-UTRA); physical channels and modulation (release 8), 2009.

[3] Jie Ren, Bradford D. Boyle, Gwanmo Ku, Steven Weber, and John MacLaren Walsh. Overhead performance tradeoffs—a resource allocation perspective. *IEEE Trans. Inf. Theory*, submitted Aug. 2014. URL `http://arxiv.org/abs/1408.3661`.

[4] Rudolf Ahlswede, Ning Cai, Shuo-Yen Robert Li, and Raymond W. Yeung. Network information flow. *IEEE Trans. Inf. Theory*, 46(4):1204–1216, 2000.

[5] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory*, 4:471–480, 1973.

[6] Lihua Song and Raymond W. Yeung. Network information flow—multiple sources. In *Int. Symp. Inf. Theory*, ISIT, page 102, IEEE, 2001.

[7] Raymond W. Yeung. *Information Theory and Network Coding*. Springer, 2008.

[8] Te Sun Han. Slepian-Wolf-Cover theorem for networks of channels. *Information and Control*, 47(1):67–83, 1980.

[9] Satoru Fujishige. Algorithms for solving the independent-flow problems. *J. Oper. Res. Soc. Japan*, 21:189–204, 1978.

[10] João Barros and Sergio D. Servetto. Network information flow with correlated sources. *IEEE Trans. Inf. Theory*, 52(1):155–170, 2006.

[11] Aditya Ramamoorthy. Minimum cost distributed source coding over a network. *IEEE Trans. Inf. Theory*, 57(1):461–475, 2011.

[12] Te Sun Han. Multicasting multiple correlated sources to multiple sinks over a noisy channel network. *IEEE Trans. Inf. Theory*, 57(1):4–13, 2011.

[13] Răzvan Cristescu, Baltasar Beferull-Lozano, and Martin Vetterli. Networked Slepian-Wolf: theory, algorithms, and scaling laws. *IEEE Trans. Inf. Theory*, 51(12):4057–4073, 2005.

[14] Tracey Ho, Muriel Médard, Ralf Koetter, David R. Karger, Michelle Effros, Jun Shi, and Ben Leong. A random linear network coding approach to multicast. *IEEE Trans. Inf. Theory*, 52 (10):4413–4430, 2006.

[15] Wei Yu and Jun Yuan. Joint source coding, routing and resource allocation for wireless sensor networks. In *Int. Conf. Commun. (ICC)*, ICC, pages 737–741, IEEE, 2005.

[16] Desmond S. Lun, Niranjan Ratnakar, Muriel Médard, Ralf Koetter, David R. Karger, Tracey Ho, Ebad Ahmed, and Fang Zhao. Minimum-cost multicast over coded packet networks. *IEEE Trans. Inf. Theory*, 52(6):2608–2623, 2006.

[17] Nimrod Megiddo. Optimal flows in networks with multiple sources and sinks. *Mathematical Programming*, 7(1):97–107, 1974.

[18] Alexander Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency*, volume B. Springer, 2003.

[19] Satoru Fujishige. *Submodular Functions and Optimization*. Elsevier, second edition, 2005.

[20] András Frank and Éva Tardos. Generalized polymatroids and submodular flows. *Mathematical Programming*, 42(1–3):489–563, 1988.

[21] Satoru Fujishige. A note on Frank's generalized polymatroids. *Discrete Applied Mathematics*, 7(1):105–109, 1984.

[22] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Interscience, second edition, 2006.

[23] Dimitris Bertsimas and John N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, 1997.

[24] William J. Cook, William H. Cunningham, William R. Pulleyblank, and Alexander Schrijver. *Combinatorial Optimization*. Wiley Interscience, 2011.

[25] Abbas El Gamal and Young-Han Kim. *Network Information Theory*. Cambridge University Press, 2011.

[26] Vinith Misra, Vivek K. Goyal, and Lav R. Varshney. Distributed scalar quantization for computing: High-resolution analysis and extensions. *IEEE Trans. Inf. Theory*, 57(8):5298–5325, 2011.

[27] Ram Zamir and Toby Berger. Multiterminal source coding with high resolution. *IEEE Trans. Inf. Theory*, 45(1):106–117, 1999.

[28] Aaron B. Wagner, Saurabha Tavildar, and Pramod Viswanath. Rate region of the quadratic gaussian two-encoder source-coding problem. *IEEE Trans. Inf. Theory*, 54(5):1938–1961, 2008.

[29] Sergio D. Servetto. Achievable rates for multiterminal source coding with scalar quantizers. In *Asilomar Conf. Signals, Systems and Computers*, pages 1762–1766, 2005.

[30] Khalid Sayood. *Introduction to Data Compression*. Elsevier, fourth edition, 2012.

[31] Nariman Farvardin and James W. Modestino. Optimum quantizer performance for a class of non-Gaussian memoryless sources. *IEEE Trans. Inf. Theory*, 30(3):485–497, 1984.

[32] Amiram H. Kaspi. Two-way source coding with a fidelity criterion. *IEEE Trans. Inf. Theory*, 31(6):735–740, 1985.

[33] Thomas J. Flynn and Robert M. Gray. Encoding of correlated observations. *IEEE Trans. Inf. Theory*, 33(6):773–787, 1987.

[34] S. Sandeep Pradhan, Julius Kusuma, and Kannan Ramchandran. Distributed compression in a dense microsensor network. *IEEE Signal Process. Mag.*, 19(2):51–60, 2002.

[35] Maurizio Longo, Tom D. Lookabaugh, and Rober M. Gray. Quantization for decentralized hypothesis testing under communication constraints. *IEEE Trans. Inf. Theory*, 36(2):241–255, 1990.

[36] Toby Berger, Zhen Zhang, and Harish Viswanathan. The CEO problem [multiterminal source coding]. *IEEE Trans. Inf. Theory*, 42(3):887–902, 1996.

[37] Aaron B. Wagner and Venkat Anantharam. An improved outer bound for multiterminal source coding. *IEEE Trans. Inf. Theory*, 54(5):1919–1937, 2008.

[38] Vinod Prabhakaran, David Tse, and Kannan Ramachandran. Rate region of the quadratic gaussian CEO problem. In *Int. Symp. Inf. Theory Proceedings*, ISIT, page 117, IEEE, 2004.

[39] Gwanmo Ku, Jie Ren, and John MacLaren Walsh. Computing the rate distortion region for the CEO problem with independent sources. *IEEE Trans. Signal Process.*, 63(3):567–575, 2015.

[40] Te Sun Han and Kingo Kobayashi. A dichotomy of functions $f(x, y)$ of correlated sources (x, y) from the viewpoint of the achievable rate region. *IEEE Trans. Inf. Theory*, 33(1):69–76, 1987.

[41] Milad Sefidgaran and Aslan Tchamkerten. Distributed function computation over a rooted directed tree. *IEEE Trans. Inf. Theory*, submitted Dec. 2013. URL `http://arxiv.org/pdf/1312.3631v1.pdf`.

[42] Vishal Doshi, Devavrat Shah, Muriel Médard, and Sidharth Jaggi. Functional compression through graph coloring. *IEEE Trans. Inf. Theory*, 56(8):3901–3917, 2010.

[43] Hans S. Witsenhausen. The zero-error side information problem and chromatic numbers (corresp.). *IEEE Trans. Inf. Theory*, 22(5):592–593, 1976.

[44] Alon Orlitsky and James R. Roche. Coding for computing. *IEEE Trans. Inf. Theory*, 47(3): 903–917, 2001.

[45] János Körner. Coding of an information source having ambiguous alphabet and the entropy of graphs. In *Prague Conf. Inf. Theory*, pages 411–425, 1973.

[46] Bradford D. Boyle, John MacLaren Walsh, and Steven Weber. Distributed scalar quantizers for subband allocation. In *Conf. Information Sciences and Systems (CISS)*, CISS, 2014.

[47] Robert M. Gray. Quantization in task-driven sensing and distributed processing. In *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pages 1049–1052, 2006.

[48] John Z. Sun, Vinith Misra, and Vivek K. Goyal. Distributed functional scalar quantization simplified. *IEEE Trans. Signal Process.*, 61(14):3495–3508, 2013.

[49] Michael Fleming, Qian Zhao, and Michelle Effros. Network vector quantization. *IEEE Trans. Inf. Theory*, 50(8):1584–1604, 2004.

[50] Tom D. Lookabaugh and Robert M. Gray. High-resolution quantization theory and the vector quantizer advantage. *IEEE Trans. Inf. Theory*, 35(5):1020–1033, 1989.

[51] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[52] Laszlo Lovasz. *Communication Complexity: A Survey*. Springer, 1990.

[53] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *Symp. Theory of Computing*, pages 94–99, ACM, 1983.

[54] László Babai, Noam Nisant, and Márió Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Computer and System Sciences*, 45(2):204–232, 1992.

[55] Nan Ma and Prakash Ishwar. Interaction strictly improves the Wyner-Ziv rate-distortion function. In *Int. Symp. Inf. Theory Proceedings*, ISIT, pages 61–65, IEEE, 2010.

[56] Nan Ma and Prakash Ishwar. Some results on distributed source coding for interactive function computation. *IEEE Trans. Inf. Theory*, 57(9):6180–6195, 2011.

[57] Jie Ren and John MacLaren Walsh. Interactive communication for resource allocation. In *Conf. Inf. Sciences and Systems (CISS)*, pages 1–6, 2014.

[58] Albert W. Marshall, Ingram Olkin, and Barry C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer Science & Business Media, second edition, 2011.

Vita

Bradford D. Boyle was born in Tacoma, Washington on 30 August 1984, the son of Thomas J. and Barbara M. Boyle. After completing his schoolwork at Mannheim American High School in Mannheim, Germany, in 2002, Bradford enrolled at Drexel University in Philadelphia, Pennsylvania. He received a Bachelor of Science and a Master of Science in electrical engineering in 2006 and 2008, respectively. In 2015 he completed his Doctorate of Philosophy. He is a U.S. citizen and current resident of Philadelphia, Pennsylvania.

Publications

1. B. D. Boyle and S. Weber, "Primal-Dual Characterizations of Jointly Optimal Transmission Rate and Scheme for Distributed Sources," *Data Compression Conf. (DCC)*, March 2014.

2. B. D. Boyle, J. M. Walsh, and S. Weber, "Distributed Scalar Quantizers for Subband Allocation," *Conf. Information Sciences and Systems (CISS)*, March 2014.

3. J. Hummel, A. McDonald, V. Shah, R. Singh, B. D. Boyle, T. Huang, N. Kandasamy, H. Sethu, and S. Weber, "A Modular Multi-Location Anonymized Traffic Monitoring Tool for a WiFi Network," *ACM Conf. Data and Application Security and Privacy (CODASPY)* (Outstanding Poster Award), March 2014.

4. B. D. Boyle, J. M. Walsh, and S. Weber, "Channel Dependent Adaptive Modulation and Coding Without Channel State Information at the Transmitter," IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), May 2013.