

2007

Minimum cost distributed source coding over a network

Aditya Ramamoorthy

Iowa State University, adityar@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/ece_conf



Part of the [Signal Processing Commons](#)

Recommended Citation

Ramamoorthy, Aditya, "Minimum cost distributed source coding over a network" (2007). *Electrical and Computer Engineering Conference Papers, Posters and Presentations*. 36.

http://lib.dr.iastate.edu/ece_conf/36

This Conference Proceeding is brought to you for free and open access by the Electrical and Computer Engineering at Iowa State University Digital Repository. It has been accepted for inclusion in Electrical and Computer Engineering Conference Papers, Posters and Presentations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Minimum cost distributed source coding over a network

Aditya Ramamoorthy

Abstract—This work considers the problem of transmitting multiple compressible sources over a network at minimum cost. The aim is to find the optimal rates at which the sources should be compressed and the network flows using which they should be transmitted so that the cost of the transmission is minimal. We consider networks with capacity constraints and linear cost functions. The problem is complicated by the fact that the description of the feasible rate region of distributed source coding problems typically has a number of constraints that is exponential in the number of sources. This renders general purpose solvers inefficient. We present a framework in which these problems can be solved efficiently by exploiting the structure of the feasible rate regions coupled with dual decomposition and optimization techniques such as the subgradient method and the proximal bundle method.

Index Terms—minimum cost network flow, distributed source coding, network coding, convex optimization, dual decomposition.

I. INTRODUCTION

In recent years the emergence of sensor networks [1] as a new paradigm has introduced a number of issues that did not exist earlier. Sensor networks have been considered among other things by the military for battlefields, by ecologists for habitat monitoring and even for extreme event warning systems. These networks consist of tiny, low-power nodes that are typically energy constrained. In general, they also have low-computing power. Thus, designing efficient sensor networks requires us to address engineering challenges that are significantly different from the ones encountered in networks such as the Internet. One unique characteristic of sensor networks is that the data that is sensed by different sensor nodes and relayed to a terminal is typically highly correlated. As an example consider a sensor network deployed to monitor the temperature or humidity levels in a forest. The temperature is not expected to vary significantly over a small area. Therefore we do expect that the readings corresponding to nearby sensors are quite correlated. It is well-known that the energy consumed in transmission by a sensor is quite substantial and therefore efficient low power methods to transfer the data across the network are of interest. This leads us to investigate efficient techniques for exploiting the correlation of the data while transmitting it across the network. There are multiple ways in which the correlation can be exploited.

- a) The sensor nodes can communicate amongst themselves to inform each other of the similarity of their data

and then transmit only as much data as is required. This comes at the cost of the overhead of inter-sensor communication.

- b) The sensors can choose to act independently and still attempt to transmit the compressed data. This strategy is likely to be more complicated from the receiver's point of view.

Usually the terminal to which the data is transmitted has significantly more resources (energy, computing power etc.). Thus, the latter solution is more attractive from a network resource efficiency point of view. The question of whether the distributed compression of correlated sources can be as efficient as their compression when the sources communicate with each other was first considered and answered in the affirmative by Slepian and Wolf in their famous paper [2]. A number of authors [3][4] have investigated the construction of coding techniques that achieve the Slepian-Wolf bounds and also proposed their usage in sensor networks [5].

New paradigms have also emerged recently in the area of network information transfer. Traditionally information transfer over networks has been considered via routing. Data packets from a source node are allowed to be replicated and forwarded by the intermediate nodes so that terminal nodes can satisfy their demands. However, network coding offers an interesting alternative where intermediate nodes in a network have the ability to forward functions of incoming packets rather than copies of the packets. The seminal work of Ahlswede et al. [6] showed that network coding achieves the capacity of single-source multiple-terminal multicast where all the terminals are interested in receiving the same set of messages from the source. This was followed by a number of works that presented constructions and bounds for multicast network codes [7][8]. More recently, there has been work [9][10] on characterizing rate regions for arbitrary network connections where the demands of the terminals can be arbitrary.

Given these developments in two different fields, a natural question to ask is how can one transmit compressible sources over a network using network coding and whether this can be done efficiently. This problem was considered by Song and Yeung [11] and Ho et al. [12]. They showed that as long as the minimum cuts between all nonempty subsets of sources and a particular terminal were sufficiently large, random linear network coding over the network followed by appropriate decoding at the terminals achieves the Slepian-Wolf bounds. The work of Ramamoorthy et al. [13] investigated the performance of separate source and network codes and showed that separation does not hold in general. Both these papers only

Aditya Ramamoorthy is with the Department of Electrical and Computer Engineering, Iowa State University, Ames IA 50011, USA (email: adityar@iastate.edu). The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Nice, France June 2007

considered capacity constraints on the edges of the network and did not impose any cost associated with edge usage.

In the networking context the problem of minimum cost network flow has been widely investigated. Here, every edge in the network has a cost per unit flow associated with it. The cost of a given routing solution is the sum of the costs incurred over all the links. The problem is one of finding network flows such that the demand of the terminals is satisfied at minimum cost. This problem has been very well investigated in the routing context [14]. The problem of minimum cost multicast using network coding was considered by Lun et al. [15] and they presented centralized and distributed solutions to it.

In this paper we consider the problem of minimum cost joint rate and flow allocation over a network that is utilized for communicating compressible sources. We consider the scenario when the compression is to be performed in a distributed manner. The sources are not allowed to communicate with each other. The main issue with joint rate and flow allocation is that typically the feasible rate region for the recovery of the sources (e.g. the Slepian-Wolf region) is described by a number of inequalities that is exponential in the number of sources. Thus, using a regular LP solver for solving the corresponding linear programming problem will be inefficient. In our work, we only consider networks where the links are independent and where transmission up to the link's capacity is assumed to be error free. In general, the capacity region characterization of more complex networks such as wireless networks will need to take account issues such as interference. Moreover, it would introduce related issues such as scheduling. We do not consider these problems in this work.

A. Main Contributions

The main contributions of this paper are as follows. We present a framework in which minimum cost problems that involve transmitting compressible sources over a network in a distributed manner can be solved efficiently. We consider general linear cost functions, allow capacity constraints on the edges of the network and consider the usage of network coding. The following problems are considered.

- a) *Slepian-Wolf over a network*. The sources are assumed to be discrete and memoryless and they need to be recovered losslessly [2] at the terminals of the network. We address the problem of jointly finding the operating rate vectors for the sources and the corresponding network flows that allow lossless recovery at the terminals at minimum cost.
- b) *Quadratic Gaussian CEO over a network*. A Gaussian source is observed by many noisy sensors and needs to be recovered at the terminal subject to a quadratic distortion constraint [16]. We present a solution to the problem of joint rate and network flow allocation that allows recovery at the terminal at minimum cost.
- c) *Lifetime maximization of sensor networks with distortion constraints*. A Gaussian source observed by many noisy sensors needs to be recovered at the terminal with a certain fidelity. We are interested in finding routing flows that would maximize the lifetime of the network.

We demonstrate that these problems can be solved efficiently by exploiting the structure of the feasible rate regions coupled

with dual decomposition techniques and subgradient methods [17].

B. Related Work

Problems of a similar flavor have been examined in several papers. Cristescu et al. considered the Networked Slepian-Wolf problem [18] and the case of lossy correlated data gathering over a network [19], but did not impose capacity constraints on the edges. Their solutions only considered very specific types of cost functions. The work of Li & Ramamoorthy [20], [21] and Roumy & Gesbert [22] considered a rate allocation under pairwise constraints on the distributed source code used for compression. The work of Liu et al. [23] and [24] considers a related problem, where they seek to minimize the total communication cost of a wireless sensor network with a single sink. They show that when the link communication costs are convex, then the usage of Slepian-Wolf coding and commodity flow routing is optimal. Moreover, they introduce the notion of distance entropy, and show that under certain situations the distance entropy is the minimum cost achieved by Slepian-Wolf coding and shortest path routing. They also propose hierarchical transmission schemes that exploit correlation among neighboring sensor nodes, and do not require global knowledge of the correlation structure. These schemes are shown to be order-optimal in specific situations. The main difference between our work and theirs, is the fact that we consider network coding and networks with multiple terminals. Moreover, in the case of general convex link cost functions, their focus is on showing that Slepian-Wolf coding and commodity flow routing is optimal. They do not consider the problem of actually finding the optimal flows and rates.

A problem formulation similar to ours was introduced by Barros et al. [25] but they did not present an efficient solution to it. The problem of exponentially many constraints has been noted by other authors as well [26][27].

The approach in our work is inspired by the work of Yu et al. [28]. However since our cost functions only penalize the usage of links in the network, we are effectively able to exploit the structure of the feasible rate region to make our overall solution efficient. In addition we explicitly derive the dual function and the corresponding update equations for maximizing it based on the specific structure of the rate region. Furthermore, we consider applications in network coding and lifetime maximization in sensor networks that have not been considered previously. In concurrent and independent work [29] presented some approaches similar to ours (see also [30], where the case of two sources is discussed). However our approach has been applied to the minimum cost quadratic Gaussian CEO problem over a network and lifetime maximization with distortion constraints that were not considered in [29].

A reviewer has pointed out that the problem of generalizing the Slepian-Wolf theorem to the network case was first considered by Han [31] in 1980. However, in [31] only networks with a single terminal were considered. In the single terminal case the corresponding flows can be supported by pure routing. Interestingly, in the same paper, Han references

the work of Fujishige [32] that studies the optimal independent flow problem (this was also pointed by the same reviewer). Fujishige's work considers a network flow problem that has polymatroidal [33] constraints for the source values and the terminal values. In particular, if there is only one terminal, then this algorithm provides an efficient solution to the minimum cost Slepian-Wolf problem over a network. However, it is unclear whether it can be extended to the case of multiple terminals and network coding. We discuss Fujishige's work in more detail in Section III, after the precise problem has been formulated.

This paper is organized as follows. Section II overviews the notation and the broad setup under consideration in this paper. Section III formulates and solves the minimum cost Slepian-Wolf problem over a network, Section IV discusses the quadratic Gaussian CEO problem over a network and Section V presents and solves the problem of lifetime maximization of sensor networks when the source needs to be recovered under distortion constraints. Each of these sections also present simulation results that demonstrate the effectiveness of our method. Section VI discusses the conclusions and future work.

II. PRELIMINARIES

In this section we introduce the basic problem setup that shall be used in the rest of this paper. In subsequent sections we shall be presenting efficient algorithms for solving three different problems that fall under the umbrella of distributed source coding problems over a network. We shall present the exact formulation of the specific problem within those sections. We are given the following.

- a) A directed acyclic graph $G = (V, E, C)$ that represents the network. Here V represents the set of vertices, E the set of edges and $C_{ij}, (i, j) \in E$ is the capacity of the edge (i, j) in bits/transmission. The edges are assumed to be error-free and the capacity of the edges is assumed to be rational. We are also given a set of source nodes $S \subseteq V$ where $|S| = N_S$ and a set of terminal nodes $T \subseteq V$ where $|T| = N_R$. Without loss of generality we assume that the vertices are numbered so that the vertices $1, 2, \dots, N_S$ correspond to the source nodes.
- b) A set of sources X_1, X_2, \dots, X_{N_S} , such that the i^{th} source is observed at source node $i \in S$. The values of the sources are drawn from some joint distribution and can be either continuous or discrete.

Based on these we can define the capacity region of the terminal $T_j \in T$ with respect to S as

$$C_{T_j} = \{(R_1, \dots, R_{N_S}) : \forall B \subseteq S, \sum_{i \in B} R_i \leq \text{min-cut}(B, T_j)\}.$$

Thus, C_{T_j} consists of inequalities that define the maximum flow (or minimum cut) from each subset of S to the terminal T_j . A rate vector $(R_1, \dots, R_{N_S}) \in C_{T_j}$ can be transmitted from the source nodes $1, \dots, N_S$ to terminal T_j via routing [14]. In the subsequent sections we shall consider different minimum cost problems involving the transmission of the sources over the network to the terminals.

III. MINIMUM COST SLEPIAN-WOLF OVER A NETWORK

Under this model, the sources are discrete and memoryless and their values are drawn i.i.d. from a joint distribution $p(x_1, \dots, x_{N_S})$. The i^{th} source node only observes X_i for $i \in S$. The different source nodes operate independently and are not allowed to communicate. The source nodes want to transmit enough information using the network to the terminals so that they can recover the original sources, losslessly.

This problem was first investigated in the seminal paper of Slepian and Wolf [2] where they considered the sources to be connected to the terminal by a direct link and the links did not have capacity constraints. The celebrated result of [2] states that the independent source coding of the sources $X_i, i = 1, \dots, N_S$ can be as efficient as joint coding when the sources need to be recovered error-free at the terminal.

Suppose that for the classical Slepian-Wolf problem, the rate of the i^{th} source is R_i . Let X_B denote the vector of sources $(X_{i_1}, X_{i_2}, \dots, X_{i_{|B|}})$, for $i_k \in B, k = 1, \dots, |B|$. The feasible rate region for this problem is given by

$$\mathcal{R}_{SW} = \{(R_1, \dots, R_{N_S}) : \forall B \subseteq S, \sum_{i \in B} R_i \geq H(X_B | X_{B^c})\}$$

The work of Csiszár [34] showed that linear codes are sufficient to approach the Slepian-Wolf (henceforth S-W) bounds arbitrarily closely.

Note that the original S-W problem does not consider the sources to be communicating with the terminal (or more generally multiple terminals) over a network. Furthermore, there are no capacity constraints on the edges connecting the sources and the terminal. In situations such as sensor networks, where the sensor nodes are typically energy-constrained, we would expect the source nodes to be in communication with the terminal node over a network that is both capacity and cost-limited. Capacity constraints may be relatively strict since a significant amount of power is consumed in transmissions. The costs of using different links could be used to ensure that a certain part of the network is not overused resulting in non-uniform depletion of resources. Thus the problem of transmitting correlated data over a network with capacity constraints at minimum cost is of interest. We define an instance of the S-W problem over a network by $P = \langle \mathcal{R}_{SW}, G, S, T \rangle$.

The transmission schemes based on linear codes (such as those in [34]) are based on block-wise coding, i.e., each source encodes N source symbols at a time. An edge with capacity C_{ij} bits/transmission can transmit $\lfloor NC_{ij} \rfloor$ bits per block. Conceptually, each edge can be regarded as multiple unit capacity edges, with each unit capacity edge capable of transmitting one bit per block. When communicating a block of length N , we consider the graph $G^N = (V, E, \lfloor NC \rfloor)$, or equivalently the graph $(V, E_N, \mathbb{1})$ (where $\mathbb{1}$ denotes a vector of ones) where E_N splits each edge from E into unit capacity edges.

To facilitate the problem formulation we construct an augmented graph G^* where we append a virtual super source node

s^* to G , so that

$$\begin{aligned} V^* &= V \cup \{s^*\}, \\ E^* &= \{(s^*, v) \mid v \in S\} \cup E, \text{ and} \\ C_{ij}^* &= \begin{cases} C_{ij} & (i, j) \in E, \\ H(X_j) & \text{if } i = s^* \text{ and } j \in S. \end{cases} \end{aligned}$$

We let $G^* = (V^*, E^*, C^*)$.

Definition 1: Feasibility. Consider an instance of the S-W problem over a network, $P = \langle \mathcal{R}_{SW}, G, S, T \rangle$. Let C_{T_i} be the capacity region of each receiver $T_i \in T$ with respect to S . If

$$\mathcal{R}_{SW} \cap C_{T_i} \neq \emptyset, \forall i = 1, \dots, N_R,$$

then the feasibility condition is said to be satisfied and P is said to be feasible.

The next theorem (from [12]) implies that as long as the feasibility condition is satisfied, random linear network coding over G^N followed by appropriate decoding at T_i suffices to reconstruct the sources X_1, X_2, \dots, X_{N_S} error-free at T_i .

Theorem 1: Sufficiency of the feasibility condition [12]. Consider an instance of the S-W problem over a network, $P = \langle \mathcal{R}_{SW}, G, S, T \rangle$. If the feasibility condition (Definition 1) is satisfied, then random linear network coding over G^N followed by minimum-entropy [34] or maximum-likelihood decoding can recover the sources at each terminal in T with the probability of decoding error going to 0 as $N \rightarrow \infty$.

The proof of the necessity of the feasibility condition can be found in [31].

It follows that if $C_{T_i} \cap \mathcal{R}_{SW} \neq \emptyset$ for all $T_i \in T$, it is sufficient to perform random linear network coding over a subgraph of G where the feasibility condition continues to be satisfied. The question then becomes, how do we choose appropriate subgraphs?

For this purpose, we now present the formulation of the minimum cost S-W problem over a network.

Let $x_{ij}^{(T_k)}$ represent the flow variable for edge (i, j) over G^* corresponding to the terminal T_k for $T_k \in T$ and z_{ij} represent the max-of-flows variable, $\max_{T_k \in T} x_{ij}^{(T_k)}$ for edge (i, j) . Note that under network coding the physical flow on edge (i, j) will be z_{ij} . The variable $x_{ij}^{(T_k)}$, represents the virtual flow variable over edge (i, j) for terminal T_k [15].

We introduce variables $R_i^{(T_k)}, i = 1, \dots, N_S$ that represent the operating S-W rate variables for each terminal. Thus $R^{(T_k)} = (R_1^{(T_k)}, R_2^{(T_k)}, \dots, R_{N_S}^{(T_k)})$ represents the rate vector for terminal T_k . Let $f_{ij} > 0, (i, j) \in E, f_{ij} = 0, (i, j) \in E^* \setminus E$ represent the cost for transmitting at a unit flow over edge (i, j) . We are interested in the following optimization problem that we call *MIN-COST-SW-NETWORK*.

$$\begin{aligned} &\text{minimize} \quad \sum_{(i,j) \in E} f_{ij} z_{ij} \\ &\text{s. t.} \quad 0 \leq x_{ij}^{(T_k)} \leq z_{ij} \leq C_{ij}^*, \quad (i, j) \in E^*, T_k \in T \end{aligned} \quad (1)$$

$$\sum_{\{j \mid (i,j) \in E^*\}} x_{ij}^{(T_k)} - \sum_{\{j \mid (j,i) \in E^*\}} x_{ji}^{(T_k)} = \sigma_i^{(T_k)}, \quad (2)$$

for $i \in V^*, T_k \in T$,

$$x_{s^*i}^{(T_k)} \geq R_i^{(T_k)}, \quad \text{for } i \in S, T_k \in T \quad (3)$$

$$R^{(T_k)} \in \mathcal{R}_{SW}, \quad \text{for } T_k \in T \quad (4)$$

where

$$\sigma_i^{(T_k)} = \begin{cases} H(X_1, X_2, \dots, X_{N_S}) & \text{if } i = s^* \\ -H(X_1, X_2, \dots, X_{N_S}) & \text{if } i = T_k \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

The constraints in (1), (2) and (5) are precisely the formulation of the minimum cost single-source multiple terminal multicast with network coding for a total rate of $H(X_1, X_2, \dots, X_{N_S})$. The constraint (3) enforces the flow $x_{s^*i}^{(T_k)}$ (corresponding to terminal T_k from s^* through source i) to be at least $R_i^{(T_k)}$. Constraint (4) ensures that the rate vectors $R^{(T_k)}$ belong to the Slepian-Wolf region \mathcal{R}_{SW} . A proof that the total rate can be fixed to be exactly $H(X_1, X_2, \dots, X_{N_S})$ for each terminal can be found in Appendix-I.

Suppose there exists a feasible solution $z, (x^{(T_k)}, R^{(T_k)})$ for $T_k \in T$ to *MIN-COST-SW-NETWORK*. Let

$$\begin{aligned} V^{*z} &= V, \\ E^{*z} &= \{(i, j) \in E^* \mid z_{ij} > 0\}, \text{ and} \\ C_{ij}^{*z} &= \begin{cases} z_{ij} & \text{if } (i, j) \in E^{*z} \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

We define the subgraph of G^* induced by z to be the graph $G_z^* = (V^{*z}, E^{*z}, C^{*z})$ and the corresponding graph over block length N as $G_z^{*N} = (V^{*z}, E_N^{*z}, \mathbb{1})$. The subgraphs induced by $x^{(T_k)}$ can be defined analogously. We now show that if *MIN-COST-SW-NETWORK* is feasible then the subgraph induced by the feasible z continues to satisfy the condition in definition 1 and therefore it suffices to perform random linear network coding over this subgraph followed by appropriate decoding at the terminals to recover the sources.

Lemma 1: Suppose that there exists a feasible solution $z, (x^{(T_k)}, R^{(T_k)})$ for $T_k \in T$ to *MIN-COST-SW-NETWORK*. Then, random linear network coding over the subgraph G_z^{*N} induced by z followed by maximum likelihood decoding at the terminals can recover the sources $X_i, i \in S$ at each terminal in T as $N \rightarrow \infty$.

Proof: To simplify the presentation we assume that all $C_{ij}, (i, j) \in E^*$ and $H(X_B/X_{B^c}), B \subseteq S$ are rational and the block length N is large enough so that NC_{ij} and $NH(X_B/X_{B^c})$ are integral. For each terminal T_k we shall show that $\text{min-cut}(B, T_k) \geq NH(X_B/X_{B^c})$ over G_z^{*N} and then use Theorem 1.

Consider a terminal $T_1 \in T$. We are given the existence of a feasible solution (z, x, R) from which we can find the corresponding flow for T_1 denoted by $x^{(T_1)}$. Now consider the subgraph of G^{*N} induced by $x^{(T_1)}$. Since $x^{(T_1)}$ is feasible, it supports a rate of $NH(X_1, X_2, \dots, X_{N_S})$ from s^* to T_1 which implies (using Menger's theorem [35]) that there exist $NH(X_1, X_2, \dots, X_{N_S})$ edge-disjoint paths from s^* to T_1 . Furthermore at least $NR_i^{(T_1)}$ of those edge-disjoint paths

connect source node i (where $i \in S$) to T_1 . It follows that if $B \subseteq S$ then the number of edge disjoint paths from B to T_1 is greater than or equal to $N \sum_{i \in B} R_i^{(T_1)}$.

Now, note that $R^{(T_1)} \in \mathcal{R}_{SW}$ which implies that for all $B \subseteq S$

$$N \sum_{i \in B} R_i^{(T_1)} \geq NH(X_B|X_{B^c}).$$

This means that there exist at least $NH(X_B|X_{B^c})$ edge-disjoint paths from B to T_1 in the subgraph induced by T_1 which in turn implies that $\min\text{-cut}(B, T_1) \geq NH(X_B|X_{B^c})$ over the subgraph induced by $x^{(T_1)}$. This holds for all $T_k \in T$, as we have a feasible $x^{(T_k)}$ for all the terminals. Finally z induces a subgraph where this property continues to hold true for each terminal since $x_{ij}^{(T_k)} \leq z_{ij}$, for all $(i, j) \in E^*, T_k \in T$. Therefore for each terminal T_k we have shown that $\min\text{-cut}(B, T_k) \geq NH(X_B|X_{B^c})$ over G_z^{*N} for all sufficiently large N . Using Theorem 1 we have the required proof. \blacksquare

The formulation of *MIN-COST-SW-NETWORK* as presented above is a linear program and can potentially be solved by a regular LP solver. However the number of constraints due to the requirement that $\mathbf{R} \in \mathcal{R}_{SW}$ is $|T|(2^{N_S} - 1)$ that grows exponentially with the number of sources. For regular LP solvers the time complexity scales with the number of constraints and variables. Thus, using a regular LP solver is certainly not time-efficient. Moreover even storing the constraints consumes exponential space and thus using a regular LP solver would also be space-inefficient. In the sequel we present efficient techniques for solving this problem.

A. Solving *MIN-COST-SW-NETWORK* via dual decomposition

Suppose that we are given an instance of the S-W problem over a network specified by $P = \langle R_{SW}, G, S, T \rangle$. We assume that P is feasible. The *MIN-COST-SW-NETWORK* optimization problem is a linear program and therefore feasibility implies that strong duality holds [36].

We shall refer to the variables $z, x^{(T_k)}, R^{(T_k)}$ for $T_k \in T$ as the primal variables. To simplify notation we let $x_{s^*}^{(T_k)} = [x_{s^*1}^{(T_k)} \ x_{s^*2}^{(T_k)} \ \dots \ x_{s^*N_S}^{(T_k)}]^T$ denote the vector of flow variables corresponding to terminal T_k on the edges from the virtual super node s^* to the source nodes in S . We form the Lagrangian of the optimization problem with respect to the constraints $R_i^{(T_k)} \leq x_{s^*i}^{(T_k)}$, for $i \in S, T_k \in T$. This is given by

$$\begin{aligned} L(\lambda, z, x^{(T_1)}, \dots, x^{(T_{N_R})}, R^{(T_1)}, \dots, R^{(T_{N_R})}) \\ = f^T z + \sum_{k=1}^{N_R} \lambda_k^T (R^{(T_k)} - x_{s^*}^{(T_k)}), \end{aligned}$$

where $\lambda = [\lambda_1^T \ \lambda_2^T \ \dots \ \lambda_{N_R}^T]^T$ is the dual variable such that $\lambda \succeq 0$ (where \succeq denotes component-wise inequality).

For a given λ , let $g(\lambda)$ denote the dual function obtained by minimizing $L(\lambda, z, x^{(T_1)}, \dots, x^{(T_{N_R})}, R^{(T_1)}, \dots, R^{(T_{N_R})})$ over $z, x^{(T_1)}, \dots, x^{(T_{N_R})}, R^{(T_1)}, \dots, R^{(T_{N_R})}$. Since strong duality holds in our problem we are guaranteed that the optimal value of *MIN-COST-SW-NETWORK* can be equivalently found

by maximizing $g(\lambda)$ subject to $\lambda \succeq 0$ [36]. Thus, if $g(\lambda)$ can be determined in an efficient manner for a given λ then we can hope to solve *MIN-COST-SW-NETWORK* efficiently.

Consider the optimization problem for a given $\lambda \succeq 0$.

$$\begin{aligned} \text{minimize} \quad & f^T z + \sum_{k=1}^{N_R} \lambda_k^T (R^{(T_k)} - x_{s^*}^{(T_k)}) \\ \text{s. t.} \quad & 0 \leq x_{ij}^{(T_k)} \leq z_{ij} \leq C_{ij}, \quad (i, j) \in E^*, T_k \in T \\ & \sum_{\{j|(i,j) \in E^*\}} x_{ij}^{(T_k)} - \sum_{\{j|(j,i) \in E^*\}} x_{ji}^{(T_k)} = \sigma_i^{(T_k)}, \quad i \in V^*, T_k \in T \\ & R^{(T_k)} \in \mathcal{R}_{SW}, T_k \in T. \end{aligned}$$

We realize on inspection that this minimization decomposes into a set of independent subproblems shown below.

$$\begin{aligned} \text{minimize} \quad & f^T z - \sum_{k=1}^{N_R} \lambda_k^T x_{s^*}^{(T_k)} \\ \text{s. t.} \quad & 0 \leq x_{ij}^{(T_k)} \leq z_{ij} \leq C_{ij}, \quad (i, j) \in E^*, T_k \in T \\ & \sum_{\{j|(i,j) \in E^*\}} x_{ij}^{(T_k)} - \sum_{\{j|(j,i) \in E^*\}} x_{ji}^{(T_k)} = \sigma_i^{(T_k)}, \quad i \in V^*, T_k \in T \end{aligned} \quad (6)$$

and for each $T_k \in T$,

$$\begin{aligned} \text{minimize} \quad & \lambda_k^T R^{(T_k)} \\ \text{subject to} \quad & R^{(T_k)} \in \mathcal{R}_{SW}. \end{aligned} \quad (7)$$

The optimization problem in (6) is a linear program with variables z and $x^{(T_k)}$ for $k = 1, \dots, N_R$ and a total of $(2|T| + 1)|E^*| + |T||V^*|$ constraints that can be solved efficiently by using a regular LP solver. It can also be solved by treating it as a minimum cost network flow problem with fixed rates for which many efficient techniques have been developed [14].

However each of the subproblems in (7) still has $2^{N_S} - 1$ constraints and therefore the complexity of using an LP solver is still exponential in N_S . However using the supermodularity property of the conditional entropy function $H(X_B|X_B^c)$, it can be shown that \mathcal{R}_{SW} is a contra-polymatroid with rank function $H(X_B|X_{B^c})$ [37]. In addition, the form of the objective function is also linear. It follows that the solution to this problem can be found by a greedy allocation of the rates as shown in [33]. We proceed as follows.

- 1) Find a permutation π such that $\lambda_{k,\pi(1)} \geq \lambda_{k,\pi(2)} \geq \dots \geq \lambda_{k,\pi(N_S)}$.
- 2) Set

$$\begin{aligned} R_{\pi(1)}^{(T_k)} &= H(X_{\{\pi(1)\}}|X_{\{\pi(1)\}^c}) \quad \text{and} \\ R_{\pi(i)}^{(T_k)} &= H(X_{\{\pi(1), \dots, \pi(i)\}}|X_{\{\pi(1), \dots, \pi(i)\}^c}) \\ &\quad - H(X_{\{\pi(1), \dots, \pi(i-1)\}}|X_{\{\pi(1), \dots, \pi(i-1)\}^c}) \\ &\quad \text{for } 2 \leq i \leq N_S. \end{aligned} \quad (8)$$

The previous algorithm presents us a technique for finding the value of $g(\lambda)$ efficiently. It remains to solve the maximization

$$\max_{\lambda \succeq 0} g(\lambda).$$

For this purpose we use the fact that the dual function is concave (possibly non-differentiable) and can therefore be

maximized by using the projected subgradient algorithm [17]. The subgradient for λ_k can be found as $R^{(T_k)} - x_{s^*}^{(T_k)}$ [17].

Let λ^i represent the value of the dual variable λ at the i^{th} iteration and θ_i be the step size at the i^{th} iteration. A step by step algorithm to solve *MIN-COST-SW-NETWORK* is presented below.

- 1) Initialize $\lambda^0 \geq 0$.
- 2) For given λ^i solve

$$\begin{aligned} & \text{minimize } f^T z - \sum_{k=1}^{N_R} (\lambda_k^i)^T x_{s^*}^{(T_k)} \\ & \text{s. t. } 0 \leq x_{ij}^{(T_k)} \leq z_{ij} \leq C_{ij}, \quad (i, j) \in E^*, T_k \in T \\ & \quad \sum_{\{j|(i,j) \in E^*\}} x_{ij}^{(T_k)} - \sum_{\{j|(j,i) \in E^*\}} x_{ji}^{(T_k)} = \sigma_i^{(T_k)}, \\ & \quad \text{for } i \in V^*, T_k \in T \end{aligned}$$

using an LP solver and for each $T_k \in T$,

$$\begin{aligned} & \text{minimize } (\lambda_k^i)^T R^{(T_k)} \\ & \text{subject to } R^{(T_k)} \in \mathcal{R}_{SW} \end{aligned} \quad (9)$$

using the greedy algorithm presented in (8).

- 3) Set $\lambda_k^{i+1} = [\lambda_k^i + \theta_i(R^{(T_k)} - x_{s^*}^{(T_k)})]^+$ for all $T_k \in T$. Goto step 2 and repeat until convergence.

While subgradient optimization provides a good approximation on the optimal value of the primal problem, a primal optimal solution or even a feasible, near-optimal solution is usually not available. In our problem, we seek to jointly find the flows and the rate allocations that support the recovery of the sources at the terminals at minimum cost. Thus, finding the appropriate flows and rates specified by the primal-optimal or near primal-optimal $z, x^{(T_1)}, \dots, x^{(T_{N_R})}, R^{(T_1)}, \dots, R^{(T_{N_R})}$ is important. Towards this end we use the method of Sherali and Choi [38].

We now briefly outline the primal recovery procedure of [38]. Let μ_j^k for $j = 1, \dots, k$ be a set of convex combination weights for each $k \geq 1$. This means that

$$\sum_{j=1}^k \mu_j^k = 1, \text{ and } \mu_j^k \geq 0.$$

We define

$$\gamma_{jk} = \mu_j^k / \theta_k, \text{ for } 1 \leq j \leq k, \text{ and } k \geq 1,$$

and let

$$\Delta \gamma_k^{\max} \triangleq \max\{\gamma_{jk} - \gamma_{(j-1)k} : j = 2, \dots, k\}.$$

Let the primal solution returned by subgradient optimization at iteration k be denoted by the vector $(z, x, R)_k$ and let the k^{th} primal iterate be defined as

$$(\tilde{z}, \tilde{x}, \tilde{R})_k = \sum_{j=1}^k \mu_j^k (z, x, R)_j \text{ for } k \geq 1. \quad (10)$$

Suppose that the sequence of weights μ_j^k for $k \geq 1$ and the sequence of step sizes $\theta_k, k \geq 1$ are chosen such that

- 1) $\gamma_{jk} \geq \gamma_{(j-1)k}$ for all $j = 2, \dots, k$ for each k .
- 2) $\Delta \gamma_k^{\max} \rightarrow 0$, as $k \rightarrow \infty$, and

- 3) $\gamma_{1k} \rightarrow 0$ as $k \rightarrow \infty$ and $\gamma_{kk} \leq \delta$ for all k , for some $\delta > 0$.

Then an optimal solution to the primal problem can be obtained from any accumulation point of the sequence of primal iterates $\{(\tilde{z}, \tilde{x}, \tilde{R})\}$.

Some useful choices for the step sizes θ_k and the convex combination weights μ_j^k that satisfy these conditions are given below (see [38]).

- 1) $\theta_k = a/(b + ck)$, for $k \geq 1$ where $a > 0, b \geq 0$, and $c \geq 0$ and $\mu_j^k = 1/k$ for all $j = 1, \dots, k$.
- 2) $\theta_k = k^{-\alpha}$, for $k \geq 1$ where $0 < \alpha < 1$ and $\mu_j^k = 1/k$ for all $j = 1, \dots, k$.

The strategy for obtaining a near-optimal primal solution for the *MIN-COST-SW-NETWORK* problem is now quite clear. We run the subgradient algorithm in the manner outlined above and keep computing the sequence of primal iterates $\{(\tilde{z}, \tilde{x}, \tilde{R})\}_{k \geq 1}$ and stop when the primal iterates have converged.

B. Results

In this section we present results on the performance of our proposed algorithm. We generated graphs at random by choosing the position of the nodes uniformly at randomly from the unit square. Two nodes were connected with an edge of capacity 40.0 if they were within a distance of $\frac{0.3}{\sqrt{2}}$ of each other and were connected with an edge of capacity 20.0 if they were within a distance of 0.3 of each other. The orientation of the edges is from left to right. A certain number of nodes were declared to be sources, a certain number to be terminals and the remaining nodes were used for relaying.

Let the random vector at the sources be denoted by $\mathbf{Z} = (Z_1, Z_2, \dots, Z_{N_S})$. As in [39], a jointly Gaussian model was assumed for the data sensed at the sources. Thus the pdf of the observations is assumed to be

$$\begin{aligned} f(z_1, z_2, \dots, z_{N_S}) &= \frac{1}{\sqrt{2\pi}^{N_S} \sqrt{\det(C_{ZZ})}} \\ &\times \exp\left(-\frac{1}{2}(\mathbf{z} - \mu)^T C_{ZZ}^{-1}(\mathbf{z} - \mu)\right), \end{aligned}$$

where C_{ZZ} is the covariance of the observations. We assumed a correlation model where $C_{ZZ}(i, i) = \sigma_i^2$ and $C_{ZZ}(i, j) = \sigma^2 \exp(-cd_{ij}^\beta)$ when $i \neq j$ (where c and β are positive constants and d_{ij} is the distance between nodes i and j). It is further assumed that the samples are quantized independently at all source nodes with the same quantization step Δ that is sufficiently small. Under these conditions, the quantized random vector $\mathbf{X} = (X_1, X_2, \dots, X_{N_S})$ is such that

$$H(\mathbf{X}) \approx h(\mathbf{Z}) - N_S \log \Delta$$

as shown in [40] where $H(X)$ represents the entropy of \mathbf{X} and $h(\mathbf{Z}) = \frac{1}{2} \log(2\pi e)^{N_S} \det(C_{ZZ})$ represents the differential entropy of \mathbf{Z} . We can also express the conditional entropy

$$\begin{aligned} H(X_B | X_{B^c}) &\approx \frac{1}{2} \log \left((2\pi e)^{N_S - |B^c|} \frac{\det(C_{ZZ})}{\det(C_{Z_{B^c} Z_{B^c}})} \right) \\ &\quad - (N_S - |B^c|) \log \Delta \end{aligned}$$

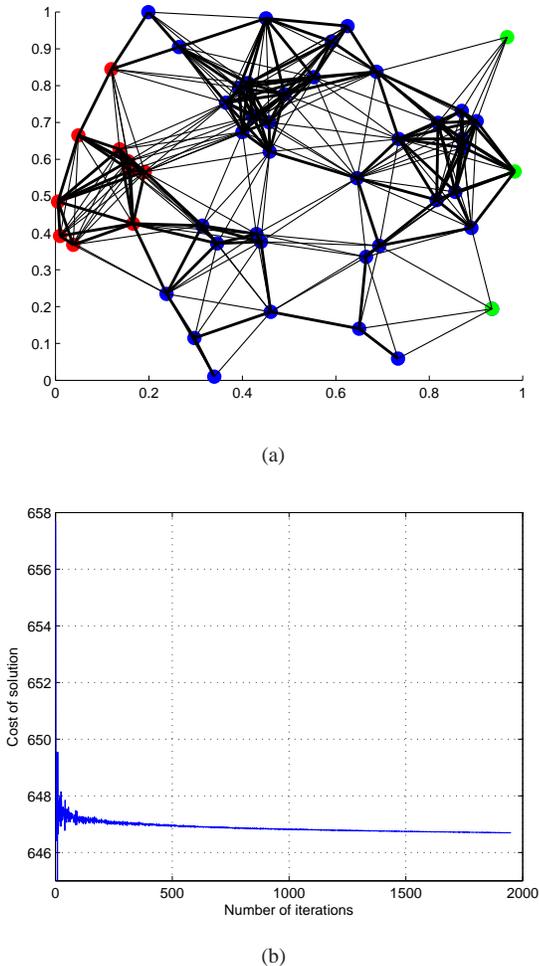


Fig. 1. (a) Network with ten source nodes (in red) and three terminal nodes (in green). The light edges have capacity 20 and the dark edges have capacity 40. All edges are directed from left to right and have unit cost. (b) Convergence of the subgradient algorithm to the optimal cost.

We used these conditional entropies for the Slepian-Wolf region of the sources.

Figure 1(a) shows a network consisting of 50 nodes, with 10 source nodes, 37 relay nodes and 3 terminals. We chose $\sigma^2 = 1$, $c = 1$ and $\beta = 1$ for this example. The quantization step size was chosen to be $\Delta = 0.01$. The cost of all the edges in the graph was set to 1.0.

For the subgradient algorithm, we chose $\mu_j^k = 1/k$ for all j, k and the step size $\theta_k = 8/k^{0.8}$. The averaging process ignored the first 50 primal solution due to their poor quality. We observe a gradual convergence of the cost of our solution to the optimal in Fig. 1(b).

1) *Remark 1:* If one uses a regular LP solver to solve the *MIN-COST-SW-NETWORK* problem, as noted above, the complexity would scale with the number of variables and constraints, that grow exponentially with the number of sources. However, one is guaranteed that the LP solver will terminate in a finite number of steps eventually. Our proposed algorithm uses the subgradient method with step sizes such that the recovered solution will converge to the optimal as the number of iterations go to infinity [17].

In general, it does not seem possible to claim convergence in a finite number of steps for this method. A discussion around convergence issues of the subgradient method can be found in Chap. 6 of [17]. We point out that in practice, we found the algorithms to converge well. Note also that even the description of the LP requires space that increases very quickly, therefore using the LP formulation becomes impractical even with a moderate number of sources.

C. Discussion about Fujishige's algorithm

We now discuss the work of Fujishige [32]. Towards this end we need to define the following quantities. A polymatroid P is defined as a pair (A, ρ) where A is a finite set and ρ is a function from 2^A to the positive reals, \mathbb{R}_+ , that satisfies the axioms of a rank function. A vector $\alpha \in \mathbb{R}_+^{|A|}$, with entries indexed by the elements of A is called an independent vector of $P(A, \rho)$ if $\sum_{e \in B} \alpha(e) \leq \rho(B)$, for all $B \subseteq A$.

Suppose that we have a directed graph $G = (V, E, C)$, with linear costs $f_{ij}, \forall (i, j) \in E$, (as defined above) with two vertex subsets V_1 (source nodes) and V_2 (terminal nodes). Suppose that for each $V_i, i = 1, 2$, a polymatroid $P_i(V_i, \rho_i)$ is defined. An independent flow is a triple (α, δ, β) , such that: (i) α is an independent vector of P_1 and is the vector of flows entering the network at the source nodes, (ii) β is an independent vector of P_2 and is the vector of flows absorbed at the terminal nodes, and (iii) δ is a flow vector such that flow balance is satisfied at each node in G . The algorithm in [32], returns an independent flow of maximum value, whose cost $f^T \delta$ is minimum.

In the case of a single terminal, this algorithm can be used to solve the *MIN-COST-SW-NETWORK* problem (as noted in [31]) as follows. The set of source nodes and the conditional entropy function, specify a contra-polymatroid. An equivalent polymatroidal representation can be found without much difficulty (see [31]). Thus, these specify V_1 and P_1 . If there is only one terminal, then one can simply define a (trivial) polymatroid on it. This specifies V_2 and P_2 .

The situation is different when one considers network coding and multiple terminals. The algorithm in [32], is only guaranteed to return one set of flows that satisfies flow balance at all source nodes, internal nodes and the terminals. It is possible to show the existence of instances where one set of flows will not simultaneously satisfy all the terminals, when one considers the *MIN-COST-SW-NETWORK* problem. An example can be found in Figure 6 in [13]. Moreover, the objective function in *MIN-COST-SW-NETWORK*, penalizes the maximum of the flows across the terminals at each node, which is different from the one in [32]. Thus, it is unclear whether this algorithm can be adapted to our problem in a straightforward manner.

IV. QUADRATIC GAUSSIAN CEO OVER A NETWORK

In general, the problem of transmitting compressible sources over a network need not have the requirement of lossless reconstruction of the sources. This maybe due to multiple reasons. The terminal may be satisfied with a low resolution reconstruction of the sources to save on network resources or lossless reconstruction may be impossible because of the

nature of sources. If a source is continuous then perfect reconstruction would theoretically require an infinite number of bits. Thus the problem of lossy reconstruction has also been an active area of research. In this section we shall consider the quadratic Gaussian CEO problem [16] over a network. We start by outlining the original problem considered by [16]. We then present the minimum cost formulation in the network context and present efficient solutions to it.

Consider a data sequence $\{X(t)\}_{t=1}^{\infty}$ that cannot be observed directly. Instead, independent corrupted versions of the data sequence are available at a set of L agents who are in communication with the Chief Estimation Officer (CEO) over different communication channels. The agents are not allowed to cooperate in any fashion. Suppose that the CEO requires the reconstruction of $\{X(t)\}_{t=1}^{\infty}$ at an average distortion level of at most D . Here, the distortion level is a metric of the fidelity of the reconstruction. Suppose agent i communicates with the CEO at rate R_i . The CEO problem [41] is one of studying the region of feasible rate vectors (R_1, R_2, \dots, R_L) that allow the reconstruction of the data sequence under the prescribed distortion constraint. As in the Slepian-Wolf case there is a direct link between the agents and the terminal (or the CEO). The quadratic Gaussian CEO problem is the particular instance of the CEO problem when the data source $\{X(t)\}_{t=1}^{\infty}$ is Gaussian and the distortion metric is mean squared error. A formal description of the problem follows.

Let $\{X(t)\}_{t=1}^{\infty}$ represent a sequence of i.i.d. Gaussian random variables and $\{Y_i(t)\}_{t=1}^{\infty} = \{X(t) + N_i(t)\}_{t=1}^{\infty}$, $i = 1, \dots, N_S$ where $\{N_i(t)\}_{t=1}^{\infty}$ are i.i.d. Gaussian independent of $\{X(t)\}_{t=1}^{\infty}$ with $E(N_i(t)) = 0$, $Var(N_i(t)) = \sigma_i^2$. Furthermore $\{N_i(t)\}_{t=1}^{\infty}$ and $\{N_j(t)\}_{t=1}^{\infty}$ are independent when $i \neq j$.

Let $\epsilon > 0$ be a small real number. The i^{th} agent encodes a block of length n from his observations $\{y_i(t)\}_{t=1}^{\infty}$ (here, $y_i(t)$ denotes a particular realization of the random variable $Y_i(t)$) using an encoding function $f_n^i : \mathbb{R}^n \rightarrow \{1, 2, \dots, \lfloor 2^{n(R_i + \epsilon)} \rfloor\}$ of rate $R_i + \epsilon$. The codewords from the N_S sources are sent to the CEO who seeks to recover an estimate of the source message over n time instants $(x(1), x(2), \dots, x(n))$ using a decoding function $g_n : \{1, 2, \dots, \lfloor 2^{n(R_1 + \epsilon)} \rfloor\} \times \dots \times \{1, 2, \dots, \lfloor 2^{n(R_{N_S} + \epsilon)} \rfloor\} \rightarrow \mathbb{R}^n$.

Definition 2: A rate vector (R_1, \dots, R_{N_S}) is said to be achievable for a distortion level D if for $\epsilon > 0$, there exists n_0 such that for all $n > n_0$, there exist encoding functions $f_n^i : \mathbb{R}^n \rightarrow \{1, 2, \dots, \lfloor 2^{n(R_i + \epsilon)} \rfloor\}$ and a decoding function $g_n : \{1, 2, \dots, \lfloor 2^{n(R_1 + \epsilon)} \rfloor\} \times \dots \times \{1, 2, \dots, \lfloor 2^{n(R_{N_S} + \epsilon)} \rfloor\} \rightarrow \mathbb{R}^n$ such that $\frac{1}{n} E \sum_{t=1}^n (X(t) - \hat{X}(t))^2 \leq D + \epsilon$ where $\hat{X}^n = g_n(f_n^1(Y_1^n), \dots, f_n^{N_S}(Y_{N_S}^n))$.

A complete characterization of the feasible rate region for a given distortion level D , denoted by $\mathcal{R}(D)$ has been obtained in [42][43] and is given below.

$$\mathcal{R}(D) = \bigcup_{(r_1, \dots, r_{N_S}) \in \mathcal{F}(D)} \mathcal{R}_{\mathcal{D}}(r_1, \dots, r_{N_S}) \quad (11)$$

where

$$\begin{aligned} \mathcal{R}_{\mathcal{D}}(r_1, \dots, r_{N_S}) &\triangleq \left\{ (R_1, \dots, R_{N_S}) : A \subseteq \{1, \dots, N_S\}, \right. \\ A \neq \phi, \sum_{k \in A} R_k &\geq \sum_{k \in A} r_k + \frac{1}{2} \log \left(\frac{\frac{1}{\sigma_X^2} + \sum_{r_i=1}^{N_S} \frac{1 - e^{-2r_i}}{\sigma_i^2}}{\frac{1}{\sigma_X^2} + \sum_{i \in A^c} \frac{1 - e^{-2r_i}}{\sigma_i^2}} \right) \left. \right\}, \\ \mathcal{F}(D) &= \left\{ (r_1, \dots, r_{N_S}) : r_i \geq 0, \frac{1}{\sigma_X^2} + \sum_{i=1}^{N_S} \frac{1 - e^{-2r_i}}{\sigma_i^2} \geq \frac{1}{D} \right\}. \end{aligned}$$

It is important to note that $\mathcal{R}(D)$ is convex [43]. Thus, in principle the minimization of a convex function of $(R_1, R_2, \dots, R_{N_S})$ can be performed efficiently.

We are interested in the quadratic Gaussian CEO problem over a network. In line with our general setup presented in Section II, the i^{th} source node in S observes the process $\{Y_i(t)\}_{t=1}^{\infty} = \{X(t) + N_i(t)\}_{t=1}^{\infty}$ and encodes the observations at a rate R_i . Once again we are interested in the minimum cost network flow problem with rates such that they permit the recovery of the source X at the terminals with the desired level of fidelity, which in this case shall be measured by mean squared error.

We start by highlighting the differences between this problem and the minimum cost Slepian-Wolf over a network. In the previous subsection we observed that the work of Ho et al. shows that random linear network coding over a subgraph such that $C_{T_i} \cap \mathcal{R}_{SW} \neq \phi, \forall T_i \in T$ allows the lossless recovery of the sources at the terminals in T and essentially the Slepian-Wolf theorem holds even in the network case with multiple terminals i.e. any rate vector that can be obtained by joint coding can be obtained by distributed coding even when there are multiple terminals. However, an analogous result in the case of the quadratic Gaussian CEO problem does not exist. Furthermore, the rate region for the quadratic Gaussian CEO problem over a general network is unknown. As a simple example, we may have a network where two source nodes are connected to a common intermediate node. The intermediate node can then combine the quantized observations from these source nodes to generate a new quantized observation such that a lower rate is possible. Thus the rate region given by the classical Gaussian CEO problem may not hold as the two codewords may be fused to produce a new codeword that enables lower rate transmission.

The first issue can be handled by assuming that there is only one terminal, i.e., $N_R = 1$ and $C_{T_1} \cap \mathcal{R}(D) \neq \phi$ so that routing will suffice to transmit a rate vector belonging to $C_{T_1} \cap \mathcal{R}(D)$ to the terminal T_1 . Thus in this problem we shall not consider network coding. For the second issue we assume that the network operates in a separate compression and information transfer mode. The set of source nodes quantize their observations as they would in the original quadratic Gaussian CEO problem. After this source coding step, the network ignores any data correlations and routes the data as though it were incompressible. In general this separation of the compression and the information transfer is suboptimal, however it is likely to be a simple way of operating the network.

It is more convenient to cast this optimization in terms of the original graph rather than the augmented graph. The *MIN-COST-QUAD-CEO-NETWORK* problem becomes

$$\begin{aligned} & \text{minimize} \quad \sum_{(i,j) \in E} f_{ij} x_{ij} \\ & \text{subject to} \quad 0 \leq x_{ij} \leq C_{ij}, \quad (i,j) \in E \end{aligned}$$

$$\sum_{\{j|(i,j) \in E\}} x_{ij} - \sum_{\{j|(j,i) \in E\}} x_{ji} = 0, i \in (S \cup \{T_1\})^c \quad (12)$$

$$\sum_{\{j|(i,j) \in E\}} x_{ij} - \sum_{\{j|(j,i) \in E\}} x_{ji} \geq R_i, \quad i \in S \quad (13)$$

$$\sum_{\{j|(i,j) \in E\}} x_{ij} - \sum_{\{j|(j,i) \in E\}} x_{ji} \leq -\sum_{i \in S} R_i, \quad i = T_1 \quad (14)$$

$$R \in \mathcal{R}(\mathcal{D})$$

Here (12) enforces the flow balance at all nodes in V except those in $S \cup \{T_1\}$, (13) enforces the constraint that at least R_i units of flow is injected at each source node $i \in S$ and (14) ensures that at least $\sum_{i \in S} R_i$ is received at the terminal T_1 . For the *MIN-COST-SW-NETWORK* problem the total rate to be transmitted to T_1 could be fixed to $H(X_1, \dots, X_N)$ as shown in the Appendix. However for the problem presented above fixing the total rate is not possible because of the nature of the inequalities specifying $\mathcal{R}(\mathcal{D})$. A feasible solution to the optimization presented above would yield a routing solution such that the delivery of a rate vector belonging to $\mathcal{R}(\mathcal{D})$ is possible at terminal T_1 . The proof is similar to the one presented in the proof of Lemma 1. However even though the optimization under consideration above is convex, the number of constraints specifying $\mathcal{R}(\mathcal{D})$ is exponential in N_S that would make a regular convex program solver inefficient.

A. Solving *MIN-COST-QUAD-CEO-NETWORK* via dual decomposition

We assume that the *MIN-COST-QUAD-CEO-NETWORK* problem is strictly feasible so that strong duality holds [36]. The Lagrangian with respect to the set of flow balance constraints that contain terms dependent on $R_i, i \in S$ for a given λ is given by

$$\begin{aligned} L(\lambda, \mathbf{x}, R) = & \mathbf{f}^T \mathbf{x} + \sum_{i \in S} \lambda_i \left(R_i - \sum_{\{j|(i,j) \in E\}} x_{ij} \right. \\ & \left. + \sum_{\{j|(j,i) \in E\}} x_{ji} \right) + \lambda_{T_1} \left(\sum_{i \in S} R_i + \sum_{\{j|(T_1,j) \in E\}} x_{T_1 j} \right. \\ & \left. - \sum_{\{j|(j,T_1) \in E\}} x_{j T_1} \right). \end{aligned}$$

It is easy to see that finding the dual function $g(\lambda) = \min_{\mathbf{x}, R} L(\lambda, \mathbf{x}, R)$ subject to the remaining constraints decom-

poses as

$$\begin{aligned} & \text{minimize} \quad f^T \mathbf{x} - \sum_{i \in S} \lambda_i \left(\sum_{\{j|(i,j) \in E\}} x_{ij} - \sum_{\{j|(j,i) \in E\}} x_{ji} \right) \\ & \quad + \lambda_{T_1} \left(\sum_{\{j|(T_1,j) \in E\}} x_{T_1 j} - \sum_{\{j|(j,T_1) \in E\}} x_{j T_1} \right) \\ & \text{subject to} \quad 0 \leq x_{ij} \leq C_{ij}, \quad \forall (i,j) \in E, \\ & \quad \sum_{\{j|(i,j) \in E\}} x_{ij} - \sum_{\{j|(j,i) \in E\}} x_{ji} = 0, \quad i \in V \setminus S \cup \{T_1\}, \end{aligned} \quad (15)$$

and

$$\begin{aligned} & \text{minimize} \quad \sum_{i \in S} \lambda_i R_i + \lambda_{T_1} \sum_{i \in S} R_i \\ & \text{subject to} \quad R \in \mathcal{R}(\mathcal{D}). \end{aligned} \quad (16)$$

The optimization in (15) is a linear program that can be solved efficiently. To solve (16) we note that for a given (r_1, \dots, r_{N_S}) it can be shown that the region $\mathcal{R}_{\mathcal{D}}(r_1, \dots, r_{N_S})$ (defined in (11)) is a contra-polymatroid [37]. Therefore an optimization problem such as

$$\begin{aligned} & \text{minimize} \quad w^T R \\ & \text{subject to} \quad R \in \mathcal{R}_{\mathcal{D}}(r_1, \dots, r_{N_S}) \end{aligned} \quad (17)$$

can be solved in closed form by using the greedy allocation algorithm presented earlier. Using this fact we show (see Appendix-II) that the optimization in (16) reduces to a convex optimization problem with $N_S + 1$ constraints that can be solved efficiently via Lagrange multiplier methods. It is important to note that the optimal value of the above optimization problem is $-\infty$ if $w_i < 0$ for any i . This is because the inequalities defining $\mathcal{R}(\mathcal{D})$ do not impose any upper bounds on the individual rates R_i for $i \in S$. Consequently the optimization in (16) has a finite optimal value only if $\lambda_i + \lambda_{T_1} \geq 0, \forall i \in S$.

It is clear based on the previous arguments that we can evaluate $g(\lambda)$ efficiently. We now need to solve the optimization

$$\begin{aligned} & \max g(\lambda) \\ & \text{subject to} \quad \lambda_i \geq 0, i \in S, \lambda_{T_1} \geq 0. \end{aligned}$$

For solving this maximization we use the projected subgradient method [17]. As noted in Section III the subgradient algorithm may not return a primal optimal or primal near-optimal solution. For primal recovery for the *MIN-COST-QUAD-CEO-NETWORK* problem we use the technique proposed by Larsson et al. [44] that generalizes the method of [38] to the case of general convex programs. We point out some differences between the two methods below.

The method of Larsson et al. considers an optimization of the following form.

$$\begin{aligned} & \min f(x) \\ & \text{subject to} \quad h_i(x) \leq 0, i \in \mathcal{I} \\ & \quad x \in X \end{aligned}$$

where the functions f and $h_i, i \in \mathcal{I}$ are convex and the set X is convex and compact. It assumes the Slater constraint

qualification condition, i.e., the existence of a strictly feasible point x_1 such that $\{x_1 \in X \mid h_i(x_1) < 0, i \in \mathcal{I}\}$ and considers the dual function with respect to the constraints $h_i, i \in \mathcal{I}$,

$$\theta(u) = \min_{x \in X} f(x) + \sum_{i \in \mathcal{I}} u_i h_i(x)$$

and then solves the maximization of the dual function

$$\begin{aligned} & \max \theta(u) \\ & \text{subject to } u \succeq 0 \end{aligned}$$

by using the projected subgradient algorithm. Let $x^{(k)}$ denote the primal solution obtained at the k^{th} iteration. Section 3.2 of [44] shows that for step sizes $\alpha_k \in [\frac{\mu}{b+k}, \frac{M}{b+k}]$, $k > 0, 0 < \mu \leq M < \infty, k \geq 0$ the sequence of averages defined by

$$\hat{x}^n = \frac{1}{n} \sum_{k=0}^n x^{(k)} \quad (18)$$

converges to the primal optimal solution as $n \rightarrow \infty$. The choice of step sizes is more limited in this method as compared to [38]. In our problem the h_i functions are the linear inequality constraints in (13) and (14) and the existence of a strictly feasible solution is assumed. However, the set $\mathcal{R}(\mathcal{D})$ is convex but not compact. The condition of compactness is however rather technical and can be enforced by imposing a loose upper bound on the rates. In practice, while running the subgradient algorithm the dual variables $\lambda_i, i \in S$ and λ_{T_1} were constrained to be larger than or equal to 10^{-10} at any iteration to ensure that the optimized rates were bounded. Averaging the solutions as in (18) we observed a steady convergence of the algorithm to the optimal cost in our simulations.

B. Results

As in the previous section we generated the graphs randomly. However, there is only one terminal in the *MIN-COST-QUAD-CEO-NETWORK* problem since we want a solution based on routing. Figure 2(a) shows an example of a network with 10 source nodes, 39 relay nodes and 1 terminal node. In this particular example we chose the variance of the source to be $\sigma_X^2 = 0.01$, the sensing noise variance of the source nodes to be $\sigma_i^2 = 0.005$ for all $i \in S$ and the required distortion level $D = 0.003$. The capacity of the light edges is 11 and the capacity of the dark edges is 22. The cost of each edge was set to 1.0. For the subgradient algorithm the step size was chosen to be $\alpha_k = \frac{10}{1+k}$. The averaging process ignored the first 100 primal solutions. As demonstrated in Fig. 2(b) there is a steady convergence of the subgradient algorithm to the optimal solution.

V. LIFETIME MAXIMIZATION FOR SENSOR NETWORKS WITH DATA DISTORTION CONSTRAINTS

We now consider the problem of maximizing the lifetime of a sensor network when the terminal node needs to be able to reconstruct the data at a particular distortion level (related problems were studied in [26]). This problem is important in the context of sensor networks where nodes are typically battery-limited and are sensing correlated phenomena that

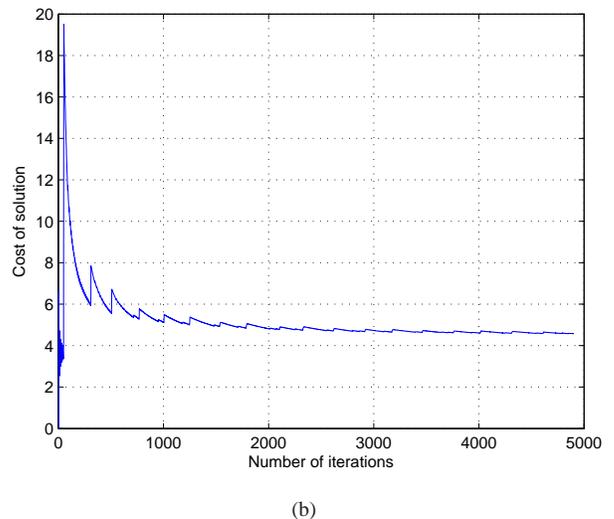
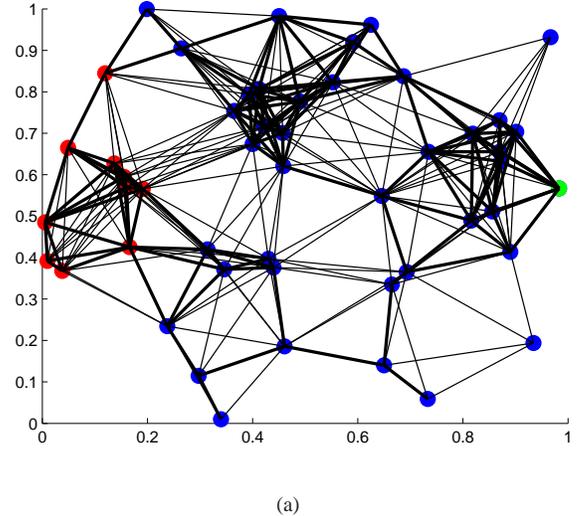


Fig. 2. (a) Network with ten source nodes (in red) and one terminal node (in green). The light edges have capacity 11 and the dark edges have capacity 22. All edges are directed from left to right and have unit cost. (b) Convergence of the subgradient algorithm to the optimal cost.

need to be reconstructed at the terminal node. It has been studied in [45] when the rates of the sensors are fixed. As explained in the previous section, the sensor nodes observe independent corrupted versions of an i.i.d. Gaussian data sequence $\{X(t)\}_{t=1}^{\infty}$ and communicate at a particular rate to the terminal node. The operating rate vector, whose components consists of the operating rates of each sensor should be such that the terminal should be able to reconstruct $\{X(t)\}_{t=1}^{\infty}$ subject to a mean squared error distortion constraint. We are interested in finding routes over which the data should be routed so that the reconstruction (with an acceptable level of fidelity) can be ensured at the terminal for the longest period of time before a node runs out of energy.

There is an inherent trade-off between the choice of the operating rate for a given sensor and the energy consumption that occurs when the data from the sensor is transmitted to the terminal since the amount of energy consumption roughly

depends on the amount of distance or hops that the data has to travel.. This is best illustrated in Fig. 3. The sensor node $S2$ that is closer to the phenomenon of interest has a better sensing SNR, but is far from the terminal. Therefore $S2$ requires less bits for quantization, however the data needs to travel a longer distance. On the other hand, sensor node $S1$ that is further away from the source but closer to the terminal has a lower sensing SNR and requires more bits for quantization, but its data needs to travel a smaller distance. Thus there is a clear tradeoff in how we would want to perform the rate and the flow allocation if we wanted to maximize the network lifetime.

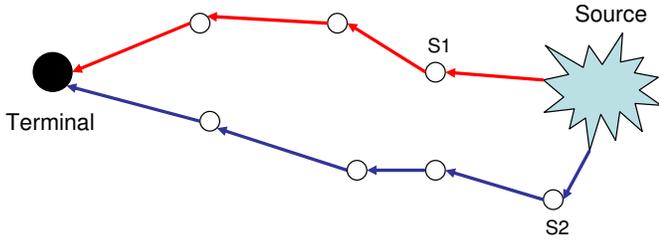


Fig. 3. The figure shows sensor nodes $S1$ and $S2$ at different distances from the phenomenon of interest. The red path corresponds to the data from $S1$ and the blue path corresponds to the data from $S2$.

In this section we formulate the problem of maximizing the lifetime of a network in the quadratic Gaussian CEO setting. As in the previous section we assume that the network operates by separate compression and network information transfer. We are given the following.

- a) An i.i.d. Gaussian data source sequence $\{X(t)\}_{t=1}^{\infty}$ and set of source nodes S that observe independent corrupted versions of $\{X(t)\}$ given by $\{Y_i(t)\}_{t=1}^{\infty} = \{X(t) + N_i(t)\}$ for $i = 1, \dots, N_S$ where $\{N_i(t)\}_{t=1}^{\infty}$ are i.i.d. Gaussian independent of $\{X(t)\}_{t=1}^{\infty}$ with $E(N_i(t)) = 0, \text{Var}(N_i(t)) = \sigma_i^2$. There is a single terminal node T_1 that seeks to reconstruct the source such that the distortion under the mean squared error metric is at most D . The feasible rate region for this problem denoted $\mathcal{R}(\mathcal{D})$ is given by (11).
- b) The initial battery level, E_i of each node $i \in V$ and the following power consumption figures of interest.
 - i) $P_{tx}(i, j)$ - the power consumed when i transmits data to j at unit rate for all $i, j \in V$,
 - ii) $P_{rx}(i, j)$ - the power consumed when j receives data from i at unit rate for all $i, j \in V$ and,
 - iii) $P_{sense}(i)$ - the power consumed when i uses an additional bit for quantizing its observations.

We define the network lifetime to be the time until the first node runs out of energy. The resultant optimization problem that we call *MAX-LIFETIME-DISTORTION-CONSTRAINT*

can be expressed as

$$\begin{aligned}
& \text{minimize } \Gamma \\
& \text{s.t. } 0 \leq x_{ij} \leq C_{ij}, \quad \forall (i, j) \in E \\
& \sum_{\{j|(i,j) \in E\}} x_{ij} - \sum_{\{j|(j,i) \in E\}} x_{ji} = 0, \quad i \in V \setminus S \cup T_1 \\
& \sum_{\{j|(i,j) \in E\}} x_{ij} - \sum_{\{j|(j,i) \in E\}} x_{ji} \geq R_i, \quad i \in S \\
& \sum_{\{j|(i,j) \in E\}} x_{ij} - \sum_{\{j|(j,i) \in E\}} x_{ji} \leq -\sum_{j \in S} R_j, \quad i = T_1 \\
& \sum_{\{j|(i,j) \in E\}} P_{tx}(i, j)x_{ij} + \sum_{\{j|(j,i) \in E\}} P_{rx}(j, i)x_{ji} \\
& + P_{sense}R_i \leq E_i\Gamma, \quad i \in S \\
& \sum_{\{j|(i,j) \in E\}} P_{tx}(i, j)x_{ij} \\
& + \sum_{\{j|(j,i) \in E\}} P_{rx}(j, i)x_{ji} \leq E_i\Gamma, \quad i \in V \setminus S \\
& R \in \mathcal{R}(\mathcal{D})
\end{aligned}$$

where Γ denotes the reciprocal of the network lifetime. Note that [45] considers the lifetime maximization problem when the operating rate for each node $i \in V$ is fixed. Therefore the problem becomes a linear program that can be solved efficiently. However we are interested in jointly optimizing the operating rates and the lifetime of the network. In the formulation above we note that the specification of the region $\mathcal{R}(\mathcal{D})$ is non-linear (although convex) with exponentially (in the number of sources) many inequalities. In [26], the authors considered a problem similar to *MAX-LIFETIME-DISTORTION-CONSTRAINT* and proposed suboptimal solutions for by approximating the constraints of $\mathcal{R}(\mathcal{D})$ by linear inequalities. The authors presented approximate linear programs that were obtained by strengthening and weakening these constraints and concluded that the true network lifetime was between the results obtained by solving these linear programs. However the number of constraints was still exponential in the number of sources that precluded solving large instances of the problem. In this section we present a solution to this problem based on dual decomposition.

We point out that we are considering a strategy where the rate allocation is static. In practice, it may be beneficial to adapt the rate allocation over time to extend the lifetime of the network.

A. Solving *MAX-LIFETIME-DISTORTION-CONSTRAINT* by dual decomposition

We note that as in section IV we can dualize the appropriate flow balance and energy consumption constraints and compute the dual function for this problem efficiently by exploiting the contra-polymatroidal structure of $\mathcal{R}(\mathcal{D})$. However, in practice we observed that the simple projected subgradient algorithm for maximizing the dual function for this problem is far too slow to be practical. Therefore we pursue an alternate line of attack here. We actually minimize Γ^2 instead of Γ and use the proximal bundle method [46] and perform primal recovery

as explained in [46]. The Lagrangian for a given λ_1 and λ_2 becomes

$$\begin{aligned} L(\lambda_1, \lambda_2, x, R) = & \Gamma^2 + \sum_{i \in S} \lambda_{1i} \left(R_i - \sum_{\{j|(i,j) \in E\}} x_{ij} + \sum_{\{j|(j,i) \in E\}} x_{ji} \right) \\ & + \lambda_{1T_1} \left(\sum_{k \in S} R_k + \sum_{\{j|(T_1,j) \in E\}} x_{T_1j} - \sum_{\{j|(j,T_1) \in E\}} x_{jT_1} \right) \\ & + \sum_{i \in S} \lambda_{2i} \left(\sum_{\{j|(i,j) \in E\}} P_{tx}(i,j)x_{ij} + \sum_{\{j|(j,i) \in E\}} P_{rx}(j,i)x_{ji} \right. \\ & \left. + P_{sense}R_i - E_i\Gamma \right) \end{aligned}$$

and finding $g(\lambda_1, \lambda_2) = \min_{\Gamma, x, R} L(\lambda_1, \lambda_2, x, R)$ subject to the remaining constraints decomposes as

$$\begin{aligned} \text{minimize } & \Gamma^2 - \Gamma \sum_{i \in S} \lambda_{2i} E_i \\ & + \sum_{i \in S} \lambda_{1i} \left(- \sum_{\{j|(i,j) \in E\}} x_{ij} + \sum_{\{j|(j,i) \in E\}} x_{ji} \right) \\ & + \lambda_{1T_1} \left(\sum_{\{j|(T_1,j) \in E\}} x_{T_1j} - \sum_{\{j|(j,T_1) \in E\}} x_{jT_1} \right) \\ & + \sum_{i \in S} \lambda_{2i} \left(\sum_{\{j|(i,j) \in E\}} P_{tx}(i,j)x_{ij} + \sum_{\{j|(j,i) \in E\}} P_{rx}(j,i)x_{ji} \right) \\ \text{subject to } & 0 \leq x_{ij} \leq C_{ij}, \quad \forall (i,j) \in E \\ & \sum_{\{j|(i,j) \in E\}} x_{ij} - \sum_{\{j|(j,i) \in E\}} x_{ji} = 0, \quad i \in V \setminus S \cup T_1 \\ & \sum_{\{j|(i,j) \in E\}} P_{tx}(i,j)x_{ij} + \sum_{\{j|(j,i) \in E\}} P_{rx}(j,i)x_{ji} \leq E_i\Gamma, \quad i \in V \setminus S \end{aligned} \quad (19)$$

and

$$\begin{aligned} \text{minimize } & \sum_{i \in S} (\lambda_{1i} + \lambda_{1T_1} + P_{sense}\lambda_{2i}) R_i \\ \text{subject to } & R \in \mathcal{R}(\mathcal{D}). \end{aligned} \quad (20)$$

The optimization problem in (19) can be solved by using a quadratic programming solver and the optimization problem in (20) can be solved as shown in Section IV. We used the quadratic programming package offered by MOSEK¹ for this part of the work.

From the above decomposition we have an efficient method to evaluate $g(\lambda_1, \lambda_2)$. It remains to evaluate

$$\begin{aligned} & \text{maximize } g(\lambda_1, \lambda_2) \\ & \text{subject to } \lambda_1, \lambda_2 \geq 0 \end{aligned}$$

for which we used the method of [46]. We now briefly overview the proximal bundle technique.

Consider the convex optimization problem

$$\begin{aligned} & \min \psi_0(z) \\ & \text{subject to } \psi_j(z) \leq 0, j = 1 \dots n \\ & z \in Z \end{aligned}$$

¹<http://www.mosek.com>

where Z is a compact and convex set, ψ_j is a convex function for $j = 0 \dots n$. Let f be the dual function of this optimization problem with respect to the constraints $\psi_j(z) \leq 0, j = 1 \dots n$. Note that the dual function is always concave and that we may not know f in its functional form. However we assume f and a subgradient of f at any given point is available via an oracle. We are interested in solving the original convex optimization by finding $\max_{x \in A} f(x)$ where A is a non-empty, closed convex set and performing primal recovery. Toward this end we use the following algorithm.

Proximal Bundle Algorithm

- **Step 1:** Let $\bar{\delta} > 0, m \in (0, 1)$. Choose an initial point \hat{x}^0 , set $y^0 = \hat{x}^0$, and let $k = 0$. Compute $f(\hat{x}^0)$ and a subgradient s_0 at \hat{x}^0 . Define the 0th polyhedral approximation to f as $\hat{f}_0(y) = f(\hat{x}^0) + s_0^T(y - \hat{x}^0)$.
- **Step 2:** At the k^{th} iteration, compute

$$y^{k+1} \in \arg \max_{y \in A} \left[\hat{f}_k(y) - \frac{\mu_k}{2} \|y - \hat{x}^k\|^2 \right],$$

where μ_k is a proximity weight. Store the Lagrange multipliers corresponding to this optimization denoted by $\nu_j^k \geq 0, 1 \leq j \leq k$ such that $\sum_j \nu_j^k = 1$.

- **Step 3:** Define $\delta_k = \hat{f}_k(y) - f(\hat{x}^k)$. If $\delta_k < \bar{\delta}$ STOP.
- **Step 4:** Compute $f(y^{k+1})$ and a subgradient s_{k+1} at y^{k+1} . Also store the value of the primal variables corresponding to y^{k+1} , denoted by z^{k+1} .
- **Step 5:** If $f(y^{k+1}) - f(\hat{x}^k) \geq m\delta_k$, perform a *SERIOUS STEP (SS)* $\hat{x}^{k+1} = y^{k+1}$, else perform a *NULL STEP (NS)* $\hat{x}^{k+1} = \hat{x}^k$.
- **Step 6:** Update the model

$$\hat{f}_{k+1}(y) = \min\{\hat{f}_k(y), f(y^{k+1}) + s_{k+1}^T(y - y^{k+1})\}.$$

- **Step 7:** Set $k = k + 1$ and goto Step 2

The work of [46] shows that the aggregate primal solution obtained by computing $\sum_{j=1}^k \nu_j^k z^j$ produces an asymptotically optimal primal solution as $k \rightarrow \infty$. For more details (such as the choice of the μ_k sequence) and for techniques for reducing the storage requirements associated with this method, we refer the reader to [46].

We applied this method to our problem since we can efficiently evaluate the dual function and can compute a subgradient at each point as well. Note that in this method as well, there is the technical compactness condition on Z . In our problem since there are no upper bounds on the rates, our region is not compact. However as in the previous section we impose loose upper bounds on the rates by enforcing the dual variables to be larger than or equal to 10^{-10} .

B. Results

We ran the previous algorithm on the same topology shown in Fig. 2(a). We chose the variance of the source to be $\sigma_X^2 = 0.01$, the sensing noise variance of the source nodes to be $\sigma_i^2 = 0.005$ for all $i \in S$ and the required distortion level $D = 0.003$. The battery levels for all the nodes were chosen to be 200 and we set $P_{tx}(i, j) = 1.0, P_{rx}(i, j) = 0.5$ and $P_{sense} = 0.001$. As demonstrated in Fig. 4 there is a steady

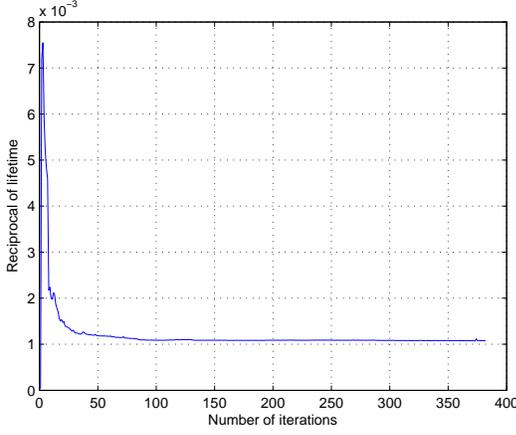


Fig. 4. Convergence of the proximal bundle algorithm to the optimal value of Γ .

convergence of the proximal bundle algorithm to the optimal solution.

We note that the lifetime maximization problem can also be solved in a similar manner if one consider the case of lossless reconstruction with multiple terminals.

VI. CONCLUSIONS AND FUTURE WORK

We considered the problem of jointly allocating rates and flows at minimum cost for distributed source coding problems over a network. In particular, we considered (a) the Slepian-Wolf problem, (b) the minimum cost quadratic Gaussian CEO problem and (c) the problem of maximizing network lifetime when a source needs to be reconstructed within a quadratic distortion constraint. These problems are of interest in domains such as sensor networks where the data that is sensed by different sensors is typically highly correlated. The feasible rate region of distributed source coding problems is typically specified by a number of inequalities that is exponential in the number of sources that makes these problems hard to solve. We presented an approach based on dual decomposition that uses the special structure of the rate regions to efficiently compute the dual function. Finally, we demonstrated approaches for maximizing the dual function using the subgradient algorithm and the proximal bundle algorithm.

It would be interesting to investigate algorithms along the lines of those considered by Fujishige [32] (that do not use dual decomposition), for the problems considered in this paper and study whether they have lower complexity. In all the problems we considered, we were able to decouple the rate allocations from the flow allocations. This essentially happens because the two sets of variables have a limited interaction via simple linear inequalities. As pointed out by a reviewer there may be other problems that may lend themselves to this kind of decomposition, where the interaction between these variables is more complex.

VII. ACKNOWLEDGEMENTS

The author would like to thank the anonymous reviewers whose suggestions greatly improved the quality and presenta-

tion of the paper.

VIII. APPENDIX-I

Theorem 2: Consider a vector (R_1, R_2, \dots, R_n) such that

$$\sum_{i \in S} R_i \geq H(X_S | X_{S^c}), \text{ for all } S \subset \{1, 2, \dots, n\}, \text{ and}$$

$$\sum_{i=1}^n R_i > H(X_1, X_2, \dots, X_n).$$

Then there exists another vector $(R'_1, R'_2, \dots, R'_n)$ such that $R'_i \leq R_i$ for all $i = 1, 2, \dots, n$ and

$$\sum_{i \in S} R'_i \geq H(X_S | X_{S^c}), \text{ for all } S \subset \{1, 2, \dots, n\}, \text{ and}$$

$$\sum_{i=1}^n R'_i = H(X_1, X_2, \dots, X_n).$$

Proof. We claim that there exists a $R_{j^*} \in \{R_1, R_2, \dots, R_n\}$ such that all inequalities in which R_{j^*} participates are loose. The proof of this claim follows.

Suppose that the above claim is not true. Then for all R_i where $i \in \{1, 2, \dots, n\}$, there exists at least one subset $S_i \subset \{1, 2, \dots, n\}$ such that,

$$\sum_{k \in S_i} R_k = H(X_{S_i} | S_i^c).$$

i.e. each R_i participates in at least one inequality that is tight.

Consider the subsets S_1 and S_2 , i.e. the subsets for which the inequalities are tight for R_1 and R_2 respectively. We have by assumption,

$$\begin{aligned} \sum_{k \in S_1} R_k + \sum_{k \in S_2} R_k &= \sum_{k \in S_1 \cap S_2} R_k + \sum_{k \in S_1 \cup S_2} R_k \\ &= H(X_{S_1} | X_{S_1^c}) + H(X_{S_2} | X_{S_2^c}) \\ &\leq H(X_{S_1 \cup S_2} | X_{(S_1 \cup S_2)^c}) + H(X_{S_1 \cap S_2} | X_{(S_1 \cap S_2)^c}) \end{aligned} \quad (21)$$

where in the second step we have used the supermodularity property of conditional entropy. Now we are also given that

$$\sum_{k \in S_1 \cap S_2} R_k \geq H(X_{S_1 \cap S_2} | X_{(S_1 \cap S_2)^c}). \quad (22)$$

Therefore we can conclude that

$$\sum_{k \in S_1 \cup S_2} R_k \leq H(X_{S_1 \cup S_2} | X_{(S_1 \cup S_2)^c}). \quad (23)$$

Now let $S_{12} = S_1 \cup S_2$. We have two cases

- a) $S_{12} = \{1, 2, \dots, n\}$.

In this case we have a contradiction since the conclusion above implies $\sum_{i=1}^n R_i \leq H(X_1, X_2, \dots, X_n)$.

- b) $S_{12} \subset \{1, 2, \dots, n\}$.

In this case consider applying a similar argument as before with S_{12} and S_3 . i.e.

$$\begin{aligned} \sum_{k \in S_{12}} R_k + \sum_{k \in S_3} R_k &= \sum_{k \in S_{12} \cup S_3} R_k + \sum_{k \in S_{12} \cap S_3} R_k \\ &\leq H(X_{S_{12} \cup S_3} | X_{(S_{12} \cup S_3)^c}) + H(X_{S_3} | X_{S_3^c}) \\ &\leq H(X_{S_{12} \cup S_2 \cup S_3} | X_{(S_{12} \cup S_2 \cup S_3)^c}) \\ &\quad + H(X_{S_{12} \cap S_2 \cap S_3} | X_{(S_{12} \cap S_2 \cap S_3)^c}). \end{aligned} \quad (24)$$

Now since

$$\sum_{k \in S_1 \cup S_2 \cup S_3} R_k \geq H(X_{S_1 \cup S_2 \cup S_3} | X_{(S_1 \cup S_2 \cup S_3)^c})$$

we obtain

$$\sum_{S_1 \cup S_2 \cup S_3} R_k \leq H(X_{(S_1 \cup S_2 \cup S_3)} | X_{(S_1 \cup S_2 \cup S_3)^c}).$$

If $S_1 \cup S_2 \cup S_3 = \{1, 2, \dots, n\}$ we have the required contradiction otherwise we can argue recursively to arrive at the contradiction. Note that the process terminates since $S_1 \cup S_2 \dots \cup S_n = \{1, 2, \dots, n\}$.

The above argument shows that there exists some j^* such that all inequalities in which R_{j^*} participates are loose. Therefore we can reduce R_{j^*} to a new value $R_{j^*}^{red}$ until one of the inequalities in which it participates is tight. If the sum-rate constraint is met with equality then we can set $R_{j^*}^{red} = R_{j^*}^{red}$ otherwise we can recursively apply the theorem to arrive at a new vector that is component-wise smaller than the original vector (R_1, R_2, \dots, R_n) . ■

We refer the S-W constraint $\sum_j R_j \geq H(X_1, \dots, X_{N_S})$ as the sum rate constraint. From theorem 2 we realize that if there exists a rate vector $(R_1, R_2, \dots, R_n) \in \mathcal{R}_{SW}$ that does not meet the sum rate constraint with equality then we can always find another vector $(R'_1, R'_2, \dots, R'_n)$ that is component-wise smaller and meets the sum-rate constraint with equality. Now consider the constraint (2) in the *MIN-COST-SW-NETWORK* problem. Instead of setting the flow-balance at s^* to $H(X_1, \dots, X_{N_S})$ and at T_k to $-H(X_1, \dots, X_{N_S})$ we could have introduced the constraints

$$\begin{aligned} \sum_{\{j|(i,j) \in E^*\}} x_{ij}^{(T_k)} - \sum_{\{j|(j,i) \in E^*\}} x_{ji}^{(T_k)} &\geq \sum_j R_j^{T_k}, i = s^* \text{ and,} \\ \sum_{\{j|(i,j) \in E^*\}} x_{ij}^{(T_k)} - \sum_{\{j|(j,i) \in E^*\}} x_{ji}^{(T_k)} &\leq -\sum_j R_j^{T_k}, i = T_k \end{aligned}$$

and attempted to solve the resulting linear program (without forcing the sum rate constraint to be satisfied with equality). Suppose that this optimization has a feasible point (z, x, R) where the rate allocation for some terminal does not satisfy the sum rate constraint. Then based on the previous observation we can conclude that we can replace R by a new set of rate allocations R' such that $R \succeq R'$ and (z, x, R') continues to be feasible with the same cost. In fact one can possibly find a new set of flows that may have a lower cost. To conclude this shows that it is sufficient to consider rate allocations that satisfy the sum rate constraint.

IX. APPENDIX-II

Consider the quadratic Gaussian CEO problem with n sources. For a given (r_1, \dots, r_n) , it can be seen that the rank function of the contra-polymatroid specified by $\mathcal{R}(D)$ is given by

$$f(A) = \sum_{k \in A} r_k + \frac{1}{2} \log \frac{\frac{1}{\sigma_X^2} + \sum_{k=1}^n \frac{1-e^{-2r_k}}{\sigma_k^2}}{\frac{1}{\sigma_X^2} + \sum_{k \in A^c} \frac{1-e^{-2r_k}}{\sigma_k^2}}$$

where $A \subseteq \{1, 2, \dots, n\}$, $A \neq \emptyset$. Therefore the minimizer of (17) can be written as

$$\begin{aligned} R_{\pi(1)} &= f(\{\pi(1)\}) \\ R_{\pi(i)} &= f(\{\pi(1), \dots, \pi(i)\}) - f(\{\pi(1), \dots, \pi(i-1)\}) \\ &\text{for } 2 \leq i \leq n \end{aligned}$$

where π is a permutation such that $w_{\pi(1)} \geq w_{\pi(2)} \geq \dots \geq w_{\pi(n)}$. The optimization problem in (17) becomes

$$\begin{aligned} &\text{minimize } \sum_{i=1}^n w_{\pi(i)} R_{\pi(i)} \\ &\text{subject to } \frac{1}{\sigma_X^2} + \sum_{i=1}^{N_S} \frac{1-e^{-2r_i}}{\sigma_i^2} \geq \frac{1}{D}, r_i \geq 0, \forall i \end{aligned} \quad (25)$$

Let A_i^π denote the set $\{\pi(1), \dots, \pi(i)\}^c$. The objective function simplifies to

$$\sum_{i=1}^n w_{\pi(i)} r_{\pi(i)} + \frac{1}{2} \sum_{i=1}^n w_{\pi(i)} \log \frac{\frac{1}{\sigma_X^2} + \sum_{k \in A_{i-1}^\pi} \frac{1-e^{-2r_k}}{\sigma_k^2}}{\frac{1}{\sigma_X^2} + \sum_{k \in A_i^\pi} \frac{1-e^{-2r_k}}{\sigma_k^2}} \quad (26)$$

We note that the first constraint in (25) has to be tight. To see this suppose $r_{\pi(1)} > 0$ and that the constraint is not tight. Then we can reduce $r_{\pi(1)}$ so that the objective function (26) is reduced. If $r_{\pi(1)} = 0$, then the argument can be applied to $r_{\pi(2)}$ by realizing that the term corresponding to $i = 1$ in the second summation above is zero and so on. Using this observation the optimization in (25) can be rewritten as

$$\begin{aligned} &\text{minimize } \sum_{i=1}^n w_{\pi(i)} r_{\pi(i)} + w_{\pi(1)} \log(1/D) - w_{\pi(n)} \log(1/\sigma_X^2) \\ &\quad + \sum_{i=0}^{n-1} (w_{\pi(i+1)} - w_{\pi(i)}) \log \left(\frac{1}{\sigma_X^2} + \sum_{k \in A_i^\pi} \frac{1-e^{-2r_k}}{\sigma_k^2} \right) \end{aligned} \quad (27)$$

$$\begin{aligned} &\text{subject to } \frac{1}{D} - \frac{1}{\sigma_X^2} - \sum_{i=1}^n \frac{1-e^{-2r_i}}{\sigma_i^2} = 0, \\ &\quad -r_i \leq 0, \forall i. \end{aligned}$$

Next, we form the Lagrangian of the optimization problem in (27) above with respect to the equality constraint while treating the positivity constraint on the r_i 's to be implicit and obtain the KKT conditions [36].

$$\begin{aligned} L(r, \nu) &= \sum_{i=1}^n w_{\pi(i)} r_{\pi(i)} + w_{\pi(1)} \log(1/D) - w_{\pi(n)} \log(1/\sigma_X^2) \\ &\quad + \sum_{i=1}^{n-1} (w_{\pi(i+1)} - w_{\pi(i)}) \log \left(\frac{1}{\sigma_X^2} + \sum_{k \in A_i^\pi} \frac{1-e^{-2r_k}}{\sigma_k^2} \right) \\ &\quad - \nu \left(\frac{1}{\sigma_X^2} + \sum_{i=1}^n \frac{1-e^{-2r_i}}{\sigma_i^2} - \frac{1}{D} \right) \end{aligned}$$

Differentiating with respect to $r_{\pi(i)}$ for $i = 1, \dots, n$ and

setting to zero, we obtain the following equations

$$\begin{aligned}\frac{\partial L}{\partial r_{\pi(1)}} &= w_{\pi(1)} - \nu \frac{2}{\sigma_{\pi(1)}^2} e^{-2r_{\pi(1)}} = 0 \\ \frac{\partial L}{\partial r_{\pi(k)}} &= w_{\pi(k)} - \frac{2\nu}{\sigma_{\pi(k)}^2} e^{-2r_{\pi(k)}} \\ &\quad + \frac{e^{-2r_{\pi(k)}}}{\sigma_{\pi(k)}^2} \sum_{i=1}^{k-1} \frac{w_{\pi(i+1)} - w_{\pi(i)}}{\frac{1}{D} - \sum_{j \in \{\pi(1), \dots, \pi(i)\}} \frac{1 - e^{-2r_j}}{\sigma_j^2}} = 0\end{aligned}$$

Solving these equations, we obtain

$$\begin{aligned}r_{\pi(1)} &= \left[\frac{1}{2} \log \left(\frac{2\nu}{w_{\pi(1)} \sigma_{\pi(1)}^2} \right) \right]^+ \\ r_{\pi(k)} &= \left[\frac{1}{2} \log \left(\frac{2\nu}{w_{\pi(k)} \sigma_{\pi(k)}^2} - \right. \right. \\ &\quad \left. \left. \frac{\sum_{i=1}^{k-1} (w_{\pi(i+1)} - w_{\pi(i)}) \left[\frac{1}{D} - \sum_{j \in \{\pi(1), \dots, \pi(i)\}} \frac{1 - e^{-2r_j}}{\sigma_j^2} \right]^{-1}}}{w_{\pi(k)} \sigma_{\pi(k)}^2} \right) \right]^+\end{aligned}$$

for $k \geq 2$

where $x^+ = \max(x, 0)$. Note that the form of the equations is such that they can be solved recursively for a given value of ν . Furthermore ν is uniquely determined by the equality constraint. Therefore a simple grid search on ν suffices to solve these equations quickly. We note that a derivation similar to the one above has been performed in a completely different context in [47].

REFERENCES

- [1] G. Pottie and W. Kaiser, "Wireless Integrated Sensor Networks," *Communications of the ACM*, vol. 43, no. 5, pp. 51–58, 2000.
- [2] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Info. Th.*, vol. 19, pp. 471–480, Jul. 1973.
- [3] S. S. Pradhan and K. Ramchandran, "Distributed Source Coding using Syndromes (DISCUS): Design and Construction," *IEEE Trans. on Info. Th.*, vol. 49, pp. 626–643, Mar. 2003.
- [4] V. Stankovic, A. D. Liveris, Z. Xiong, and C. N. Georghiades, "On code design for the Slepian-Wolf problem and lossless multiterminal networks," *IEEE Trans. on Info. Th.*, vol. Volume: 52, Issue: 4, pp. 1495–1507, April 2006.
- [5] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed Source Coding for Sensor Networks," in *IEEE Signal Processing Magazine*, Sept. 2004.
- [6] R. Ahlswede, N. Cai, S.-Y. Li, and R. W. Yeung, "Network Information Flow," *IEEE Trans. on Info. Th.*, vol. 46, no. 4, pp. 1204–1216, 2000.
- [7] S.-Y. Li, R. W. Yeung, and N. Cai, "Linear Network Coding," *IEEE Trans. on Info. Th.*, vol. 49, no. 2, pp. 371–381, 2003.
- [8] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. M. G. M. Tolhuizen, "Polynomial time algorithms for multicast network code construction," *IEEE Trans. on Info. Th.*, vol. 51, no. 6, pp. 1973–1982, June 2005.
- [9] L. Song and R. W. Yeung, "Zero-error network coding for acyclic networks," *IEEE Trans. on Info. Th.*, vol. 49, no. 12, pp. 3129–3139, Dec. 2003.
- [10] N. Harvey, R. Kleinberg, and A. R. Lehman, "On the Capacity of Information Networks," *IEEE Trans. on Info. Th.*, vol. 52, no. 6, pp. 2345–2364, 2006.
- [11] L. Song and R. W. Yeung, "Network information flow - multiple sources," in *IEEE Intl. Symposium on Info. Th.*, 2001, pp. 102–102.
- [12] T. Ho, M. Médard, M. Effros, and R. Koetter, "Network Coding for Correlated Sources," in *Conf. on Information Sciences and Systems*, 2004.
- [13] A. Ramamoorthy, K. Jain, P. A. Chou, and M. Effros, "Separating Distributed Source Coding from Network Coding," *IEEE Trans. on Info. Th.*, vol. 52, no. 6, pp. 2785–2795, 2006.
- [14] R. K. Ahuja, T. L. Maganti, and J. B. Orlin, *Network Flows: Theory, Algorithms and Applications*. Prentice-Hall, 1993.
- [15] D. S. Lun, N. Ratnakar, M. Médard, R. Koetter, D. R. Karger, T. Ho, E. Ahmed, and F. Zhao, "Minimum-Cost Multicast over Coded Packet Networks," *IEEE Trans. on Info. Th.*, vol. 52, pp. 2608–2623, June 2006.
- [16] H. Viswanathan and T. Berger, "The Quadratic Gaussian CEO Problem," *IEEE Trans. on Info. Th.*, vol. 43, no. 5, pp. 1549–1559, 1997.
- [17] D. Bertsekas, *Nonlinear Programming*, 2nd ed. Athena Scientific, 1999.
- [18] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "Networked Slepian-Wolf: theory, algorithms, and scaling laws," *IEEE Trans. on Info. Th.*, vol. 51, no. 12, pp. 4057–4073, 2005.
- [19] R. Cristescu and B. Beferull-Lozano, "Lossy network correlated data gathering with high-resolution coding," *IEEE Trans. on Info. Th.*, vol. 52, no. 6, pp. 2817–2824, 2006.
- [20] S. Li and A. Ramamoorthy, "Rate and power allocation under the pairwise distributed source coding constraint," in *IEEE Intl. Symposium on Info. Th.*, 2008, pp. 2312–2316.
- [21] —, "Rate and power allocation under the pairwise distributed source coding constraint," *IEEE Trans. on Comm.*, vol. to appear.
- [22] A. Roumy and D. Gesbert, "Optimal matching in wireless sensor networks," in *IEEE Intl. Symposium on Info. Th.*, 2007, pp. 2116–2120.
- [23] J. Liu, M. Adler, D. Towsley, and C. Zhang, "On optimal communication cost for gathering correlated data through wireless sensor networks," in *Proceedings of the 12th ACM Annual Intl. Conf. on Mobile Comp. and Netw. (MobiCom)*, 2006.
- [24] —, "Collecting Correlated Data through a network with minimum cost: Distance entropy and a practical order optimal design," *IEEE Trans. on Info. Th.*, accepted.
- [25] J. Barros and S. Servetto, "Network Information Flow with Correlated Sources," *IEEE Trans. on Info. Th.*, vol. 52, pp. 155–170, Jan. 2006.
- [26] A. Kansal, A. Ramamoorthy, M. B. Srivastava, and G. Pottie, "On sensor network lifetime and data distortion," in *IEEE Intl. Symposium on Info. Th.*, 2005, pp. 6–10.
- [27] J. C. F. Li and S. Dey, "Lifetime Optimization for Multi-hop Wireless Sensor Networks with Rate Distortion Constraints," in *Signal Processing Advances in Wireless Communications (SPAWC)*, 2006, pp. 1–5.
- [28] W. Yu and J. Yuan, "Joint source coding, routing and resource allocation for wireless sensor networks," in *IEEE Intl. Conf. Comm.*, 2005, pp. 737–741.
- [29] T. Cui, T. Ho, and L. Chen, "On Distributed Distortion Optimization for Correlated Sources with Network Coding," in *IEEE Intl. Symposium on Info. Th.*, 2007, pp. 2731–2735.
- [30] A. Lee, M. Médard, K. Z. Haigh, S. Gowan, and P. Rubel, "Minimum-Cost Subgraphs for Joint Distributed Source and Network Coding," in *Third Workshop on Network Coding, Theory and Applications (NET-COD)*, 2007.
- [31] T. S. Han, "Slepian-Wolf-Cover Theorem for a Network of Channels," *Inform. Contr.*, vol. 47, no. 1, pp. 67–83, 1980.
- [32] S. Fujishige, "Algorithms for solving the independent-flow problems," *J. Operations Res. Soc. Japan*, vol. 21, pp. 189–204, 1978.
- [33] D. Tse and S. Hanly, "Multi-Access Fading Channels: Part I: Polymatroid Structure, Optimal Resource Allocation and Throughput Capacities," *IEEE Trans. on Info. Th.*, vol. 44, pp. 2796–2815, Nov. 1998.
- [34] I. Csiszár, "Linear Codes for Sources and Source Networks: Error Exponents, Universal Coding," *IEEE Trans. on Info. Th.*, vol. 28, no. 4, pp. 585–592, 1982.
- [35] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*. Cambridge University Press, 1993.
- [36] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [37] J. Chen, X. Zhang, T. Berger, and S. B. Wicker, "An Upper Bound on the Sum-Rate Distortion Function and Its Corresponding Rate Allocation Schemes for the CEO Problem," *IEEE J. Select. Areas Comm.*, vol. 22, no. 6, pp. 977–987, 2004.
- [38] H. D. Sherali and G. Choi, "Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs," *Oper. Res. Letters*, vol. 19, pp. 105–113, 1996.
- [39] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On Network Correlated Data Gathering," in *IEEE Infocom*, 2004.
- [40] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley Series, 1991.
- [41] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO Problem," *IEEE Trans. on Info. Th.*, vol. 42, no. 3, pp. 887–902, May 1996.
- [42] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. on Info. Th.*, vol. 51, pp. 2577–2593, 2005.
- [43] V. Prabhakaran, D. Tse, and K. Ramchandran, "Rate Region of the quadratic Gaussian CEO problem," in *IEEE Intl. Symposium on Info. Th.*, 2004, pp. 119–119.

- [44] T. Larsson, M. Patriksson, and A.-B. Stromberg, "Ergodic, primal convergence in dual subgradient schemes for convex programming," *Math. Program.*, vol. 86, pp. 283–312, 1999.
- [45] J.-H. Chang and L. Tassiulas, "Energy conserving routing in wireless ad-hoc networks," in *IEEE Infocom*, 2000.
- [46] K. C. Kiwiel, "Approximations in proximal bundle methods and decomposition of convex programs," *J. Optim. Theory Appl.*, vol. 84, pp. 529–548, 1995.
- [47] J. Chen and T. Berger, "Successive Wyner-Ziv Coding Scheme and its Application to the Quadratic Gaussian CEO Problem," *IEEE Trans. on Info. Th.*, vol. 54, no. 4, pp. 1586–1603, 2008.