Convergence Rates of Min-Cost Subgraph Algorithms for Multicast in Coded Networks

Fang Zhao*, Muriel Médard*, Desmond Lun†, and Asuman Ozdaglar*
* Lab for Information and Decision Systems
MIT, Cambridge, MA 02139, USA
E-mail: zhaof/medard/asuman@mit.edu
†The Broad Institute of MIT and Harvard
Cambridge, MA 02142, USA
E-mail: dlun@broad.mit.edu

***Abstract*— The problem of establishing minimum-cost multicast connections in coded networks can be viewed as an optimization problem, and decentralized algorithms were proposed by Lun et. al. to compute the optimal subgraph using the subgradient method on the dual problem. However, the convergence rate problem for these algorithms remains open. In general, there are results available for the convergence rate of the subgradient method on the dual problem, but it is hard to find the convergence rate of the primal solutions. In this paper, we analyze the convergence rates of the min-cost subgraph algorithms for both the dual and the primal sides. We show that using the incremental subgradient method on the dual problem with appropriately chosen step sizes yields linear convergence rate to a neighborhood of the optimal solution. Also, if we use constant step sizes in the subgradient method and simple averaging for primal recovery, the primal solutions recovered from the dual iterations converge to a neighborhood of the optimal solution with rate $O(1/n)$.**

I. Introduction

When network coding is used to perform multicast, the problem of establishing minimum-cost multicast connection is equivalent to two effectively decoupled problems [1]: one of determining the subgraph to code over, and the other of determining the code to use over that subgraph. The latter problem has been studied in [2], [3], [4], [5], and a variety of methods have been proposed, which include employing simple random linear coding at every node. Such random linear coding schemes are completely decentralized, requiring no coordination between nodes, and can operate under dynamic conditions [6].

The former problem has been studied in [7], [8], and [9]. While [7] looked at centralized solutions, [8] proposed approaches to find minimum-cost multicast subgraphs for both linear and strictly convex cost functions in a decentralized manner. Reference [9] took a closer look at the decentralized algorithm in [8], and proposed different variations of it to improve the convergence performance. In addition, simulation results are presented in [8] showing the performance of these algorithms in both static and dynamic wireless networks.

However, the convergence rate problem for these algorithms remain open, and it needs to be studied before any algorithm can be implemented with confidence. Although simulation results in [8] show that the distributed algorithms produce significant reductions in multicast cost as compared to centralized routing algorithms just after a few iterations, and it is robust to changes in the network, it is desirable to have theoretical bounds on the performance, and this is what we aim to solve in this paper. For simplicity, we focus on the problem of single multicast session in wireline networks in this paper, but the proofs can be easily modified to work for the wireless case.

In general, the subgradient method used on the dual problem has been extensively studied, and convergence results are available under various step size rules. However, in practice, the interest is in solving the primal problem, and recovering, from the dual iterations, feasible or near-feasible primal solutions that converge to the optimal solution as the number of iterations goes to infinity. There are special cases where the primal solutions computed as the by-product of the dual iterations are feasible, such as in [14], but this is not the case in general. There are only a few papers studying this aspect of the problem, for example [15] and [16], and even fewer studying the convergence rate of the primal solutions. In this paper, we provide theoretical bounds on the convergence rate of both the dual and the primal problems for the subgraph optimization algorithms.

The remaining part of this paper is organized as follows. Section II describes the network model used and reviews the formulation of the subgraph optimization problem and the decentralized algorithm. In Section III, we analyze convergence rate performance of the dual problem using the incremental subgradient method studied in [12], and these results can help us in the study of convergence rate of the primal problem, which is of main interest here. The results for primal convergence analysis are presented in Section IV, and finally, we conclude in Section V.

II. Distributed Algorithm for Min-Cost Subgraph Optimization

In this section, we give an overview of the distributed algorithm proposed in [8], and introduce the notations that will be used throughout this paper. We focus on the problem of single multicast in wireline networks, and model the network with a directed graph $G = (N, A)$, where N is the set of nodes and A is the set of links in the network. Each link (i,j), $i \in N, j \in N$, $(i,j) \in A$, is associated with a non-negative number a_{ij}, which is the cost per unit flow on this link. We assume that the total cost of using a link is proportional to the flow, z_{ij}, on it. For the multicast, suppose we have a source node $s \in N$ producing data at a positive rate R that it wishes to transmit to a non-empty set of terminal nodes T in N.

It is shown in [10] that a subgraph z is capable of supporting a multicast connection of rate R from source s to T if and only if the min-cut from s to any $t \in T$ is greater than or equal to R. Hence, the problem of finding the min-cost subgraph can be formulated into the following LP problem [1]:

$$\begin{aligned} \text{minimize} \quad & \sum_{(i,j)\in A} a_{ij} z_{ij} \\ \text{subject to} \quad & z_{ij} \geq x_{ij}^{(t)}, \qquad \forall (i,j) \in A, t \in T, \\ & \sum_{\{j|(i,j)\in A\}} x_{ij}^{(t)} - \sum_{\{j|(j,i)\in A\}} x_{ji}^{(t)} = \delta_i^{(t)} \\ & \qquad\qquad \forall i \in N, t \in T, \\ & x_{ij}^{(t)} \geq 0, \qquad \forall (i,j) \in A, t \in T, \end{aligned} \tag{1}$$

where

$$\delta_i^{(t)} = \begin{cases} R, & \text{if } \ i = s, \\ -R, & \text{if } \ i = t, \\ 0 & \text{otherwise.} \end{cases}$$

We use the distributed algorithm proposed in [8] to solve the above LP problem. This algorithm is completely decentralized and each node only has to know the cost of its incoming and outgoing links, and exchange information with neighboring nodes. It employs the subgradient method on the Lagrangian dual of (1), which is given by:

$$\begin{aligned} \text{maximize} \quad & \sum_{t\in T} q^{(t)}(p^{(t)}) \\ \text{subject to} \quad & \sum_{t\in T} p_{ij}^{(t)} = a_{ij}, \qquad \forall (i,j) \in A, \\ & p_{ij}^{(t)} \geq 0, \qquad \forall (i,j) \in A, t \in T, \end{aligned} \tag{2}$$

where

$$q^{(t)}(p^{(t)}) = \min_{x^{(t)} \in F^{(t)}} \sum_{(i,j)\in A} p_{ij}^{(t)} x_{ij}^{(t)}, \qquad \forall t \in T, \tag{3}$$

and $F^{(t)}$ is the bounded polyhedron of points $x^{(t)}$ satisfying the conservation of flow constraints

$$\begin{aligned} \sum_{\{j|(i,j)\in A\}} x_{ij}^{(t)} - \sum_{\{j|(j,i)\in A\}} x_{ji}^{(t)} = \delta_i^{(t)}, \quad & \forall i \in N, \\ x_{ij}^{(t)} \geq 0, \qquad & \forall (i,j) \in A. \end{aligned}$$

Note that the subproblem (3) is a standard shortest path problem with link costs $p_{ij}^{(t)}$, which can be solved using a multitude of distributed algorithms (e.g., distributed Bellman-Ford).

The distributed algorithm for finding the min-cost subgraph based on (1) and (2) is summarized below. For details of this algorithm and related proofs, please refer to [8].

1) Before the first iteration, each node initializes $p[0]$.
2) In the nth iteration, use $p[n]$ as link costs, and run a distributed shortest path algorithm to determine $x[n]$.
3) Update $p[n+1]$ using subgradient obtained through $x[n]$ values.
$$g_{ij}^{(t)}[n] = x_{ij}^{(t)}[n],$$
$$p[n+1] := \left[p[n] + \theta[n] \sum_{t\in T} g^{(t)}[n] \right]_P^+,$$
where $g[n]$ is the subgradient for $p[n]$, and $\theta[n]$ is the step size for the nth iteration.
4) At the end of each iteration, nodes recover a primal solution, $\tilde{x}[n]$, based on the dual computations. More details on the recovery of primal solutions are presented in Section IV.
5) Each node computes the $\tilde{z}_{ij}[n]$ values from the $\tilde{x}_{ij}[n]$ values.
6) Steps (2) to (5) are repeated until the primal solution has converged.

Since the intermediate $\{\tilde{z}[n], \tilde{x}[n]\}$ values after each iteration are always feasible solutions to the primal problem, we do not have to wait till the primal solution converges to start the multicast. Instead, the multicast can be started after the first iteration, and we can shift the flows gradually through the iterations to operate on a more cost effective subgraph.

In the following sections, we theoretically analyze the convergence rate of this algorithm. The convergence rate results on the dual problem is presented in Section III, and those for the primal problem are presented in Section IV.

III. Convergence rate analysis for the dual problem

With properly chosen step sizes, the standard subgradient method proposed in the previous section converges to dual optimal solutions eventually [11], but it is hard to analyze the convergence rate of the standard method. To this end, we consider the incremental subgradient method studied in [12]. The incremental subgradient method can be used here because the objective function in (2) is the sum of $|T|$ convex component functions, and the constraint set is non-empty, closed and convex (see Chapter 2 of [12]). At each iteration, p is changed incrementally through a sequence of $|T|$ steps. Each step is a subgradient iteration for a single component function $q^{(t)}$. Thus, an iteration can be viewed as a cycle of $|T|$ subiterations. Denote the terminal nodes by $\{1, 2, \ldots, N_T\}$, where $N_T = |T|$. The vector $p[n+1]$ is obtained from $p[n]$ as follows.

$$\begin{aligned} \psi_0[n] &:= p[n], \\ \psi_i[n] &:= \left[\psi_{i-1}[n] + \theta[n] g^{(i)}[n] \right]_P^+, \\ p[n+1] &:= \psi_{N_T}[n]. \end{aligned}$$

We first prove two propositions that are useful for the convergence rate analysis.

Proposition 1: Problem (2) satisfies the subgradient boundedness property, which means there exists a positive scalar C such that

$$||g|| \leq C, \quad \forall g \in \partial q^{(t)}(p[n]) \cup \partial q^{(t)}(\psi_{i-1,n}),$$
$$\forall i = 1, ..., N_T, \qquad \forall n.$$

Proof: This is true because $q^{(t)}$ is the pointwise minimum of a finite number of affine functions, and in this case, for every p, the set of subgradients $\partial q^{(t)}(p)$ is the convex hull of a finite number of vectors. Thus, the subgradients are bounded. ■

Proposition 2: Let the optimal solution set be P^*, there exists a positive scalar μ such that

$$q^* - q(p) \geq \mu(\text{dist}(p, P^*))^2, \qquad \forall p \in P.$$

Proof: Problem (2) can be reformulated into a linear programming problem as follows.

$$\begin{aligned} \text{maximize} \quad & q'(v) = \sum_{t\in T} q^{(t)} \\ \text{subject to} \quad & q^{(t)} \leq \sum_{(i,j)\in A} p_{ij}^{(t)} \hat{x}_{ij}^{(t)}, \quad && \forall\, t \in T, \\ & \sum_{t\in T} p_{ij}^{(t)} = a_{ij}, && \forall\, (i,j) \in A, \\ & p_{ij}^{(t)} \geq 0, && \forall\, (i,j) \in A, t \in T, \\ & \hat{x}^{(t)} \in F^{(t)}, && \forall\, t \in T. \end{aligned} \tag{4}$$

The decision vector, v, is a concatenation of vectors q, p, and x, and we denote the feasible set by V. For any feasible $p \in P$ from (2), there is a corresponding v in (4) with the same p-component and $q'(v) = q(p)$. Furthermore, for any feasible $v \in V$, we can extract a p vector from it that gives the same total cost in (2). Therefore, the two formulations (2) and (4) have the same optimal values, i.e., $q^* = q'^*$.

Since the set of solutions for a linear programming problem is a set of weak sharp minima [13], there exists a positive α such that

$$q'^* - q'(v) \geq \alpha(\text{dist}(v, V^*)), \qquad \forall v \in V.$$

So for any $p \in P$ in (2), we have

$$q^* - q(p) = q'^* - q'(v) \geq \alpha(\text{dist}(v, V^*)) \geq \alpha(\text{dist}(p, P^*)).$$

The last inequality comes from the fact that p/P^* is the projection of v/V^* on P, and the projection operation is non-expansive. Since P is a bounded polyhedron, the distance between any two points in P is bounded, i.e., $\text{dist}(p, p') \leq B$ for all $p, p' \in P$ for some positive B. Therefore,

$$q^* - q(p) \geq \frac{\alpha}{B}(\text{dist}(p, P^*))^2.$$

Let $\mu = \alpha/B$, and the proposition is proved. ■

With these propositions, we have the following result for constant step size.

Proposition 3: For the sequence $\{p[n]\}$ generated by the incremental subgradient method with the step size $\theta[n]$ fixed to some positive constant θ, where $\theta \leq \frac{1}{2\mu}$, we have

$$\begin{aligned} (\text{dist}(p[n+1], P^*))^2 \quad \leq \quad & (1-2\theta\mu)^{n+1}(\text{dist}(p[0], P^*))^2 \\ & + \frac{\theta |T|^2 C^2}{2\mu}, \qquad \forall n. \end{aligned}$$

Proof: The proof for this proposition follows from Proposition 1.2 and the proof of Proposition 2.3 in [12]. ■

In summary, we have shown that the convergence rate for the incremental subgradient method on (2) is linear for a sufficiently small step size. However, only convergence to a neighborhood of the optimal solution set can be guaranteed.

IV. Convergence rate analysis for the primal problem

With the above results on the dual convergence rate, we can now move on to analyze the primal convergence rate, which is of main interest here. We first present some more details on the recovery of primal solutions from the dual iterations, which is Step (4) in the distributed algorithm described in Section II. Let $\{\mu_l[n]\}_{l=1,...,n}$ be a sequence of convex combination weights for each non-negative integer n, i.e., $\sum_{l=1}^n \mu_l[n] = 1$ and $\mu_l[n] \geq 0$ for all $l = 1, ..., n$. Further, let us define

$$\gamma_{ln} = \frac{\mu_l[n]}{\theta[l]}, \qquad l = 1, ...n, n = 0, 1, ...,$$

and

$$\Delta\gamma_n^{\max} = \max_{l=2,...n} \{\gamma_{ln} - \gamma_{(l-1)n}\}.$$

According to [16], if the step sizes $\{\theta[n]\}$ and the convex combination weights $\{\mu_l[n]\}$ are chosen such that

1) $\gamma_{ln} \geq \gamma_{(l-1)n}$ for all $l = 2, ...n$, and $n = 0, 1, ...$,
2) $\Delta\gamma_n^{\max} \to 0$ as $n \to \infty$, and
3) $\gamma_{1n} \to 0$ as $n \to \infty$ and $\gamma_{nn} \leq \delta$ for all $n = 0, 1, ...$, for some $\delta > 0$,

then we obtain an optimal solution to the primal problem (1) from any accumulation point of the sequence of primal iterates $\{\tilde{x}[n]\}$ given by

$$\tilde{x}_{ij}^{(t)}[n] = \sum_{l=1}^{n} \mu_l[n] x_{ij}^{(t)}[l], \qquad n = 0, 1, ...$$

The flow on each link, $\tilde{z}_{ij}[n]$, is then computed based on the recovered $\tilde{x}_{ij}^{(t)}[n]$ values.

We consider the scenario where on the dual side, simple constant step size is used in the subgradient method, i.e., $\theta[n] = \theta$ for all n, and on the primal side, the convex combination weights $\{\mu_l[n]\}$ are given by $\mu_l[n] = 1/n$ for all $l = 1, ...n$, and $n = 0, 1, ...$ The convergence rate of the primal solutions under this scenario is given in the following proposition.

Proposition 4: If constant step size is used in the dual subgradient method and simple averaging is used in primal solution recovery, the primal cost of the nth iteration converges to a neighborhood of the optimal solution with rate $O(1/n)$.

We use two different methods to prove this proposition, and the first one is shown below.

Proof: The first proof for this proposition makes use of Proposition 2 in [15]. We first show that the following two conditions are satisfied by our optimization problem.

1) **The Slater condition**, i.e., there exist a vector $\{\bar{z}, \bar{x}\}$, $\{\bar{z}, \bar{x}\} \in F$ such that

$$z_{ij} > x_{ij}^{(t)} \quad \forall (i,j) \in A, t \in T.,$$

where F is the feasible set for the conservation of flow constraints.
This condition clearly holds in our primal problem (1), because the variables z_{ij} do not appear in the constraints of F, and we can always find z_{ij} values that make the inequality constraints, $z_{ij} \geq x_{ij}^{(t)}$, strict for any set of feasible x values.

2) **Bounded subgradients**, i.e., there exists a scalar $L > 0$ such that

$$\|g[n]\| \leq L \quad \forall n > 0.$$

This follows directly from Proposition 1 in the previous section.

With these two conditions satisfied, Proposition 2 in [15] shows that an upper bound on the primal cost of $\tilde{z}[n]$ is given by

$$f(\tilde{z}[n]) \leq f^* + \frac{\|p[0]\|^2}{2n\theta} + \frac{\theta L^2}{2}.$$

The above bound shows that, when constant step size is used, the primal solution converges to a neighborhood of the optimal solution with rate $O(1/n)$. The size of this neighborhood depends on the step size θ. In fact, there is a compromise between the rate of convergence and how close we get to the optimal solution. The smaller the value of θ, the closer we get to the optimal solution, but with a slower rate. ■

One thing we want to point out is that Proposition 2 in [15] also gives a bound on the amount of constraint violation of the primal solution recovered after each iterations. However, in our case, the constraint violation is always 0, i.e., after each iteration we recover a feasible primal solution. This is because the primal variables x_{ij} obtained in the dual subgradient methods always satisfy the flow conservation constraints, $x^{(t)} \in F^{(t)}$, and the z_{ij} values are recovered purely base on these feasible x values (they are not part of the dual computation). This makes sure that we always recover a feasible primal solution, and do not have to worry about constraint violations.

Next, we look at our second proof for Proposition 4.

Proof: In this method, we examine the complementary slackness between the primal the dual solutions after each iteration. In the following computations, we make use of some of the results presented in the proof of Theorem 1 in [16]. The complementary slackness after the nth iteration is bounded by

$$\begin{aligned}(\tilde{x}[n] - \tilde{z}[n])' \cdot p[n] \quad &\leq \quad \gamma_{nn}(p[n+1] - p^*)' \cdot p[n] \\ &+ \quad \gamma_{1n}(p^* - p[1])' \cdot p[n] \\ &+ \quad (\sum_{j=2}^{n}(\gamma_{jn} - \gamma_{(j-1)n})(p^* - p_j))' \\ &\quad\; \cdot p[n].\end{aligned} \tag{5}$$

Next, we take a close look at each of the three terms on the right hand side.

In the first term, both γ_{nn} and $\|p[n]\|$ are bounded. In the previous section, we have shown that using constant step size in the incremental subgradient method gives linear convergences rate on the dual side to a neighborhood of the optimal solution. Thus, $\|p[n+1] - p^*\|$ decreases to a non-zero term linearly with n, and the magnitude of the non-zero term depends on the size of the neighborhood on the dual side. Therefore, the first term decreases to some positive number linearly as n goes to infinity.

For the second term, $(p^* - p[1])$ is a constant with respect to n, and $\|p[n]\|$ is bounded. If we use constant step size in the subgradient method and simple averaging in primal recovery, we have

$$\gamma_{1n} = \frac{\mu_1[n]}{\theta[1]} = \frac{1/n}{\theta} = \frac{1}{n\theta},$$

and clearly, this term diminishes with rate $O(1/n)$ as n goes to inifinity.

Finally, we look at the third term. For any j in $\{2, ..., n\}$, we have

$$\gamma_{jn} - \gamma_{(j-1)n} = \frac{\mu_j[n]}{\theta[j]} - \frac{\mu_{j-1}[n]}{\theta[j-1]} = \frac{1}{n\theta} - \frac{1}{n\theta} = 0.$$

Therefore, the last term is always equal to 0.

Summing the three terms up, we can see that the value of $(\tilde{x}[n] - \tilde{z}[n])'p[n]$ decrease to a non-zero constant with rate $O(1/n)$ with n. Since, for a linear programming problem, ϵ-complementary slackness gives us a bound on the duality gap between the primal and dual solutions (see Theorem 4.5 in [17]), i.e.,

$$f(\tilde{z}[n]) - q(p[n]) < \epsilon,$$

and this gap is diminishing with rate $O(1/n)$, as shown above, the primal solution converges to a neighborhood of optimal one with rate $O(1/n)$.

■

From the above proof, we can see that if more complicated step sizes and convex combination coefficients are used, such that the dual solution converges to the optimal (rather than to a neighborhood of the optimal) with rate $O(1/n)$, then the first term in (5) also diminishes when n goes to infinity and the overall convergence rate would be $O(1/n)$ to the optimal solution.

In addition, the second proof gives us a relationship between the closeness of the dual solution to its optimal to the quality of the primal solution recovered. Assume there is some inaccuracy in the link cost information, a_{ij}, and the real optimal cost, f^*_{real}, is not very far from the f^* computed with the wrong a_{ij} values. From the above proof, we can see that after sufficient number of iterations, the dual solutions and the corresponding recovered primal solutions are going to be pretty close to f^*, due to ϵ-complementary slackness. And by the triangle inequality, the resulting primal cost is close to the real optimal f^*_{real}. Therefore, even with slight inaccuracy in the link cost information, the dual or the primal solutions will not be very far from the optimal solution after some iterations.

V. CONCLUSIONS

In this paper, we look at the decentralized algorithm proposed in [8] for subgraph minimization for coded networks, and study the convergence rate of both the dual and the primal solutions. On the dual side, it is shown that using incremental subgradient method, the solutions can converge linearly to a neighborhood of the optimal solution. As for the primal solutions, with constant step size and averaging primal recovery method, the convergence rate is $O(1/n)$.

ACKNOWLEDGMENT

The authors would like to thank Professor Angelia Nedić for helpful discussions.

This work was supported by the National Science Foundation under grants "ITR: Network Coding - From Theory to Practice" (CCR-0325496), "XORs in the Air: Practical Wireless Network Coding" (CNS-0627021), by the Air Force Office of Scientific Research (AFOSR) under grant FA9550-06-1-0155, and by the BAE Systems National Security Solutions Inc. under grant "Control Based Mobile Ad-Hoc Networking Program (CBMANET)" (060786).

REFERENCES

[1] D. S. Lun, M. Médard, T. Ho, and R. Koetter, "Network coding with a cost criterion," in *Proc. 2004 International Symposium on Information Theory and its Applications (ISITA 2004)*, Oct. 2004, pp. 1232–1237.

[2] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "Toward a random operation of networks," submitted to *IEEE Trans. Inform. Theory*.

[3] T. Ho, M. Médard, J. Shi, M. Effros, and D. R. Karger, "On randomized network coding," in *Proc. 41th Annual Allterton Conference on Communication, Control, and Computing*, Oct. 2003.

[4] T. Ho, R. Koetter, M. Médard, D. R. Karger, and M. Effros, "The benefits of coding over routing in a randomized setting," in *Proc. 2003 IEEE International Symposium on Information Theory (ISIT 2003)*, 2003.

[5] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proc. 41th Annual Allterton Conference on Communication, Control, and Computing*, Oct. 2003.

[6] T. Ho, B. Leong, M. Médard, R. Koetter, Y.-H. Chang, and M. Effros, "On the utility of network coding in dynamic environments," in *Proc. 2004 International Workshop on Wireless Ad-hoc Networks (IWWAN 2004)*, 2004.

[7] Y. Wu, P. A. Chou, and S.-Y. Kung, "Minimum-energy multicast in mobile ad hoc networks using network coding," submitted to *IEEE Trans. Commun.*.

[8] D. S. Lun, N. Ratnakar, R. Koetter, M. Médard, E. Ahmed, and H. Lee, "Achieving minimum-cost multicast: A decentralized approach based on network coding," in *Proc. IEEE Infocom 2005*, vol. 3, Miami, FL, Mar. 2005, pp. 1608–1617.

[9] F. Zhao, D. S. Lun, M. Médard, and E. Ahmed, "Decentralized algorithms for operated coded wireless networks," in *Proc. IEEE Information Theory Workshop 2007*, Lake Tahoe, CA, Sept. 2007.

[10] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[11] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.

[12] A. Nedić, "Subgradient methods for convex minimization," Ph.D. dissertation, Massachusetts Institute of Technology, June 2002.

[13] J. V. Burke and M. C. Ferris, "Weak sharp minima in mathematical programming," *SIAM J. Control Optim.*, vol. 31, no. 5, pp. 1340–1359, Sept. 1993.

[14] S. Low and D. E. Lapsley, "Optimization flow control, I: Basic algorithm and convergence," *IEEE/ACM Transactions on Networking 7(6)*, pp. 861-874, Dec. 1999.

[15] A. Nedić, and A. Ozdaglar, "Approximate primal solutions and rate analysis for dual subgradient methods," *LIDS Report 2755*, 2007.

[16] H. D. Sherali and G. Choi, "Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs," *Oper. Res. Lett.*, vol. 19, pp. 105-113, 1996.

[17] D. P. Bertsekas, and J. N. Tsitsiklis, *Introduction to linear optimization*. Belmont, MA: Athena Scientific, 1997.