Subgradient Methods in Network Resource Allocation: Rate Analysis

Angelia Nedić
Department of Industrial and
Enterprise Systems Engineering
University of Illinois
Urbana-Champaign, IL 61801
Email: angelia@uiuc.edu

Asuman Ozdaglar
Department of Electrical Engineering
and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02142
Email: asuman@mit.edu

***Abstract*— We consider dual subgradient methods for solving (nonsmooth) convex constrained optimization problems. Our focus is on generating approximate primal solutions with performance guarantees and providing convergence rate analysis. We propose and analyze methods that use averaging schemes to generate approximate primal optimal solutions. We provide estimates on the convergence rate of the generated primal solutions in terms of both the amount of feasibility violation and bounds on the primal function values. The feasibility violation and primal value estimates are given per iteration, thus providing practical stopping criteria. We provide a numerical example that illustrate the performance of the subgradient methods with averaging in a network resource allocation problem.**

I. Introduction

Lagrangian relaxation and duality have been effective tools for solving large-scale convex optimization problems and for systematically providing lower bounds on the optimal value of nonconvex (continuous and discrete) optimization problems. Subgradient methods have played a key role in this framework providing computationally efficient means to obtain near-optimal dual solutions and bounds on the optimal value of the primal optimization problem. Most remarkably, in networking applications, over the last few years, subgradient methods have been used with great success in developing decentralized cross-layer resource allocation mechanisms (see [10] and [22] for more on this subject).

The subgradient methods for solving dual problems have been extensively studied starting with Polyak [17] and Ermoliev [4]. Their convergence properties under various stepsize rules have been long established and can be found, for example, in [21], [18], [5], and [1]. Numerous extensions and implementations including parallel and incremental versions have been proposed and analyzed (see [7], [11], [12], [13]). Despite widespread use of the subgradient methods for solving dual (nondifferentiable) problems, there are limited results in the existing literature on the recovery of primal solutions and the convergence rate analysis in the primal space. In many network resource allocation problems, however, the main interest is in solving the primal problem. In this case, the question arises whether we can use the subgradient method in dual space and exploit the subgradient information to produce primal near-feasible and near-optimal solutions.

In this paper, we study generating approximate primal optimal solutions for general convex constrained optimization problems using dual subgradient methods. We consider a simple averaging scheme that constructs primal solutions by forming the running averages of the primal iterates generated when evaluating the subgradient of the dual function. We focus on methods that use a constant stepsize both in view of its simplicity and the potential to generate approximate solutions in a relatively small number of iterations.

We provide estimates on the convergence rate of the average primal sequences in terms of both the amount of feasibility violation and the primal objective function values. Our estimates depend on the norm of the generated dual iterates. Under the Slater condition, we show that the dual sequence is bounded and we provide an explicit bound on the norm of the dual iterates. Combining these results, we establish convergence rate estimates for the average primal sequence. Our estimates show that under the Slater condition, the amount of constraint violation goes to zero at the rate of $1/k$ with the number of subgradient iterations k. Moreover, the primal function values go to the optimal value within some error at the rate $1/k$. Our bounds explicitly highlight the dependence of the error terms on the constant stepsize and illustrate the tradeoffs between the solution accuracy and computational complexity in selecting the stepsize value.

Other than the papers cited above, our paper is also related to the literature on the recovery of primal solutions from subgradient methods (see for example [15], [21], [8], [20], [9], [6], [16], [19]). These works focus on the asymptotic behavior of the primal sequences, i.e., the convergence properties in the limit as the number of iterations increases to infinity. Since the focus is on the asymptotic behavior, the convergence analysis has been mostly limited to diminishing stepsize rules[1]. Moreover there is no convergence rate analysis on the generated primal sequences. In this paper, we focus on generating approximate primal solutions in finitely many iterations with convergence rate guarantees.

We thank Ali ParandehGheibi for his assistance with the numerical example.

This research was partially supported by the National Science Foundation under CAREER grants CMMI-0742538 and DMI-0545910.

[1]The exception is the paper [6] where a target-level based stepsize has been considered.

The paper is organized as follows: In Section II, we define the primal and dual problems, and provide an explicit bound on the level sets of the dual function under Slater condition. In Section III, we consider a subgradient method with a constant stepsize and study its properties under Slater. In Section IV, we introduce approximate primal solutions generated through averaging and provide bounds on their feasibility violation and primal cost values. Section V, we present a numerical example of network resource allocation that illustrates the performance of the dual subgradient method with averaging. Section VI contains our concluding remarks.

Regarding notation, we view a vector as a column vector, and we denote by $x'y$ the inner product of two vectors x and y. We use $\|y\|$ to denote the standard Euclidean norm, $\|y\| = \sqrt{y'y}$. For a vector $u \in \mathbb{R}^m$, we write u^+ to denote the projection of u on the nonnegative orthant in $\mathbb{R}^m$, i.e.,

$$u^+ = (\max\{0, u_1\}, \cdots, \max\{0, u_m\})'.$$

For a concave function $q : \mathbb{R}^m \to [-\infty, \infty]$, we say that a vector $s_{\bar{\mu}} \in \mathbb{R}^m$ *is a subgradient of* $q(\mu)$ *at a given vector* $\bar{\mu} \in \mathrm{dom}(q)$ if

$$q(\bar{\mu}) + s'_{\bar{\mu}}(\mu - \bar{\mu}) \geq q(\mu) \qquad \text{for all } \mu \in \mathrm{dom}(q), \tag{1}$$

where $\mathrm{dom}(q) = \{\mu \in \mathbb{R}^m \mid q(\mu) > -\infty\}$. The set of all subgradients of q at $\bar{\mu}$ is denoted by $\partial q(\bar{\mu})$.

II. Primal and Dual Problems

We focus on the following constrained optimization problem:

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && g(x) \leq 0, \ x \in X, \end{aligned} \tag{2}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, $g = (g_1, \ldots, g_m)'$ and each $g_j : \mathbb{R}^n \to \mathbb{R}$ is a convex function, and $X \subset \mathbb{R}^n$ is a nonempty closed convex set. We refer to this as the *primal problem*. We denote the primal optimal value by f^*, and throughout this paper, *we assume that the value f^* is finite*.

To generate approximate solutions to the primal problem of Eq. (2), we consider approximate solutions to its dual problem. Here, the *dual problem* is the one arising from Lagrangian relaxation of the inequality constraints $g(x) \leq 0$, and it is given by

$$\begin{aligned} &\text{maximize} && q(\mu) \\ &\text{subject to} && \mu \geq 0, \ \mu \in \mathbb{R}^m, \end{aligned} \tag{3}$$

where q is the dual function defined by

$$q(\mu) = \inf_{x \in X} \{f(x) + \mu' g(x)\}. \tag{4}$$

We often refer to a vector $\mu \in \mathbb{R}^m$ with $\mu \geq 0$ as a *multiplier*. We denote the dual optimal value by q^* and the dual optimal set by M^*. We say that there is *zero duality gap* if the optimal values of the primal and the dual problems are equal, i.e., $f^* = q^*$.

We assume that the minimization problem associated with the evaluation of the dual function $q(\mu)$ has a solution for every $\mu \geq 0$. This is the case, for instance, when the set X is compact (since f and $g_j's$ are continuous due to being convex over $\mathbb{R}^n$). Furthermore, we assume that the minimization problem in Eq. (4) is simple enough so that it can be solved efficiently. For example, this is the case when the functions f and g_j's are affine or affine plus norm-square term [i.e., $c\|x\|^2 + a'x + b$], and the set X is the nonnegative orthant in $\mathbb{R}^n$. Many practical problems of interest, such as those arising in network resource allocation, often have this structure.

In our subsequent development, we consider subgradient methods as applied to the dual problem given by Eqs. (3) and (4). Due to the form of the dual function q, the subgradients of q at a vector μ are related to the primal vectors x_μ attaining the minimum in Eq. (4). Specifically, the set $\partial q(\mu)$ of subgradients of q at a given $\mu \geq 0$ is given by

$$\partial q(\mu) = \mathrm{conv}\left(\{g(x_\mu) \mid x_\mu \in X_\mu\}\right), \tag{5}$$

where $X_\mu = \{x_\mu \in X \mid q(\mu) = f(x_\mu) + \mu' g(x_\mu)\}$, and $\mathrm{conv}(Y)$ denotes the convex hull of a set Y. In the following, we omit some of the proofs due to space constraints and refer the interested reader to the longer version of our paper [14].

A. Slater Condition and Boundedness of the Multiplier Sets

In this section, we consider sets of the form $\{\mu \geq 0 \mid q(\mu) \geq q(\bar{\mu})\}$ for a fixed $\bar{\mu} \geq 0$, which are obtained by intersecting the nonnegative orthant in $\mathbb{R}^m$ and (upper) level sets of the concave dual function q. We show that these sets are bounded when the primal problem satisfies the standard Slater constraint qualification, formally given in the following.

Assumption 1: (Slater Condition) *There exists a vector $\bar{x} \in \mathbb{R}^n$ such that*

$$g_j(\bar{x}) < 0 \qquad \textit{for all } j = 1, \ldots, m.$$

We refer to a vector $\bar{x}$ satisfying the Slater condition as *a Slater vector*.

Under the assumption that f^* is finite, it is well-known that the Slater condition is sufficient for a zero duality gap as well as for the existence of a dual optimal solution (see e.g. [1]). Furthermore, the dual optimal set is bounded (see [5]). This property of the dual optimal set under the Slater condition, has been observed and used as early as in Uzawa's analysis of Arrow-Hurwicz gradient method in [23]. Interestingly, most work on subgradient methods has not made use of this powerful result, which is a key in our analysis.

The following proposition extends the result on the optimal dual set boundedness under the Slater condition. In particular, it shows that the Slater condition also guarantees the boundedness of the (level) sets $\{\mu \geq 0 \mid q(\mu) \geq q(\bar{\mu})\}$.

Lemma 1: Let the Slater condition hold [Assumption 1]. Let $\bar{\mu} \geq 0$ be a vector such that the set $Q_{\bar{\mu}} = \{\mu \geq 0 \mid q(\mu) \geq q(\bar{\mu})\}$ is nonempty. Then, the set $Q_{\bar{\mu}}$ is bounded and, in particular, we have

$$\max_{\mu \in Q_{\bar{\mu}}} \|\mu\| \leq \frac{1}{\gamma} \left(f(\bar{x}) - q(\bar{\mu})\right),$$

where $\gamma = \min_{1 \leq j \leq m}\{-g_j(\bar{x})\}$ and $\bar{x}$ is a Slater vector.

It follows from the preceding Lemma that under the Slater condition, the dual optimal set M^* is nonempty. In particular, by noting that $M^* = \{\mu \geq 0 \mid q(\mu) \geq q^*\}$ and by using Lemma 1, we see that

$$\max_{\mu^* \in M^*} \|\mu^*\| \leq \frac{1}{\gamma}\left(f(\bar{x}) - q^*\right), \tag{6}$$

with $\gamma = \min_{1\leq j\leq m}\{-g_j(\bar{x})\}$.

In practice, the dual optimal value q^* is not readily available. However, having a dual function value $q(\tilde{\mu})$ for some $\tilde{\mu} \geq 0$, we can still provide a bound on the norm of the dual optimal solutions. Furthermore, given any multiplier sequence $\{\mu_k\}$, we can use the relation $q^* \geq \max_{0\leq i\leq k} q(\mu_i)$ to generate a sequence of (possibly improving) upper bounds on the dual optimal solution norms $\|\mu^*\|$. In the following, we use these bounds to provide error estimates for our approximate solutions.

III. Subgradient Method

To solve the dual problem, we consider the classical subgradient algorithm with a constant stepsize:

$$\mu_{k+1} = [\mu_k + \alpha g_k]^+ \qquad \text{for } k = 0, 1, \ldots, \tag{7}$$

where the vector $\mu_0 \geq 0$ is an initial iterate and the scalar $\alpha > 0$ is a stepsize. The vector g_k is a subgradient of q at μ_k given by

$$g_k = g(x_k), \quad x_k \in \operatorname*{argmin}_{x\in X}\{f(x) + \mu_k' g(x)\} \quad \text{for all } k \geq 0 \tag{8}$$

[see Eq. (5)].

One may consider other stepsize rules for the subgradient method. Our choice of the constant stepsize is primarily motivated by its practical importance and in particular, because in practice the stepsize typically stays bounded away from zero. Furthermore, the convergence rate estimates for this stepsize can be explicitly written in terms of the problem parameters that are often available. Also, when implementing a subgradient method with a constant stepsize rule, the stepsize length α is the only parameter that a user has to select, which is often preferred to more complex stepsize choices involving several stepsize parameters without a good guidance on their selection.

A. Basic Relations

In this section, we establish some basic relations that hold for a sequence $\{\mu_k\}$ obtained by the subgradient algorithm of Eq. (7). These properties are important in our construction of approximate primal solutions, and in particular, in our analysis of the error estimates of these solutions.

We start with a lemma providing some basic relations that hold under minimal assumptions. The relations given in this lemma have been known and used in various ways to analyze subgradient approaches (see [21], [18], [3], [2], [11], [12]).

Lemma 2: (Basic Iterate Relation) *Let the sequence $\{\mu_k\}$ be generated by the subgradient algorithm (7). We then have:*

(a) *For any $\mu \geq 0$ and all $k \geq 0$,*

$$\|\mu_{k+1}-\mu\|^2 \leq \|\mu_k-\mu\|^2 - 2\alpha\left(q(\mu) - q(\mu_k)\right) + \alpha^2\|g_k\|^2.$$

(b) *When the optimal solution set M^* is nonempty, for all $k \geq 0$, there holds*

$$\begin{aligned} dist^2(\mu_{k+1}, M^*) \quad \leq \quad & dist^2(\mu_k, M^*) \\ & -2\alpha\left(q^* - q(\mu_k)\right) + \alpha^2\|g_k\|^2, \end{aligned}$$

where $dist(y, Y)$ denotes the Euclidean distance from a vector y to a set Y.

B. Bounded Multipliers

Here, we show that the multiplier sequence $\{\mu_k\}$ produced by the subgradient algorithm is bounded under the Slater condition and the bounded subgradient assumption. We formally state the latter requirement in the following.

Assumption 2: (Bounded Subgradients) *The subgradient sequence $\{g_k\}$ is bounded, i.e., there exists a scalar $L > 0$ such that*

$$\|g_k\| \leq L \qquad \text{for all } k \geq 0.$$

This assumption is satisfied, for example, when the primal constraint set X is compact. In this case, due to the convexity of the constraint functions g_j over $\mathbb{R}^n$, each g_j is continuous over $\mathbb{R}^n$. Thus, $\max_{x\in X}\|g(x)\|$ is finite and provides an upper bound on the norms of the subgradients g_k, and hence, we can let

$$L = \max_{x\in X}\|g(x)\| \qquad \text{or} \qquad L = \max_{1\leq j\leq m}\max_{x\in X}|g_j(x)|.$$

In the following lemma, we establish the boundedness of the multiplier sequence. In this, we use the boundedness of the dual sets $\{\mu \geq 0 \mid q(\mu) \geq q(\bar{\mu})\}$ [cf. Lemma 1] and the basic relation for the sequence $\{\mu_k\}$ of Lemma 2(a).

Lemma 3: (Bounded Multipliers) *Let the multiplier sequence $\{\mu_k\}$ be generated by the subgradient algorithm of Eq. (7). Also, let the Slater condition and the bounded subgradient assumption hold [Assumptions 1 and 2]. Then, the sequence $\{\mu_k\}$ is bounded and, in particular, we have*

$$\begin{aligned} \|\mu_k\| \quad \leq \quad & \frac{2}{\gamma}\left(f(\bar{x}) - q^*\right) \\ & + \max\left\{\|\mu_0\|, \frac{1}{\gamma}\left(f(\bar{x}) - q^*\right) + \frac{\alpha L^2}{2\gamma} + \alpha L\right\}, \end{aligned}$$

where $\gamma = \min_{1\leq j\leq m}\{-g_j(\bar{x})\}$, L is the subgradient norm bound of Assumption 2, $\bar{x}$ is a Slater vector, and $\alpha > 0$ is the stepsize.

The bound of Lemma 3 depends explicitly on the dual optimal value q^*. In practice, the value q^* is not readily available. However, since $q^* \geq q(\mu_0)$, we can replace q^* with $q(\mu_0)$ in the bound of Lemma 3. Note that this bound depends on the algorithm parameters and problem data, i.e., it only involves the initial iterate μ_0 of the subgradient method, the stepsize α, the vector $\bar{x}$ satisfying the Slater condition, and the subgradient norm bound L. In some practical applications, such as those in network optimization, such data is readily available. One may think of optimizing this bound with respect to the Slater vector $\bar{x}$. This might be an interesting and challenging problem on its own. However, this is outside the scope of our paper.

IV. Approximate Primal Solutions

In this section, we provide approximate primal solutions by considering the running averages of the primal sequence $\{x_k\}$ generated as a by-product of the subgradient method [cf. Eq. (8)]. Intuitively, one would expect that, by averaging, the primal cost and the amount of constraint violation of primal infeasible vectors can be reduced due to the convexity of the cost and the constraint functions. It turns out that the benefits of averaging are far more reaching than merely cost and infeasibility reduction. We show here that under the Slater condition, we can also provide upper bounds for the number of subgradient iterations needed to generate a primal solution within a given level of constraint violation. We also derive upper and lower bounds on the gap from the optimal primal value. These bounds depend on some prior information such as a Slater vector and a bound on subgradient norms.

We now introduce the notation that we use in our averaging scheme throughout the rest of the paper. We consider the multiplier sequence $\{\mu_k\}$ generated by the subgradient algorithm of Eq. (7), and the corresponding sequence of primal vectors $\{x_k\} \subset X$ that provide the subgradients g_k in the algorithm of Eq. (7) [i.e., x_k as given by Eq. (8)]. We define $\hat{x}_k$ as the average of the vectors $x_0, \ldots, x_{k-1}$, i.e.,

$$\hat{x}_k = \frac{1}{k} \sum_{i=0}^{k-1} x_i \qquad \text{for all } k \geq 1. \tag{9}$$

The average vectors $\hat{x}_k$ lie in the set X because $x_i \in X$ for all i and the set X is convex. However, these vectors need not satisfy the primal inequality constraints $g_j(x) \leq 0$, $j = 0, \ldots, m$, and therefore, they can be primal infeasible.

A. Basic Properties of the Averaged Primal Sequence

In this section, we provide upper and lower bounds on the primal cost of the running averages $\hat{x}_k$. We also provide an upper and a lower bound on the amount of feasibility violation of these vectors. These bounds are given per iteration, as seen in the following.

Proposition 1: Let the multiplier sequence $\{\mu_k\}$ be generated by the subgradient method of Eq. (7). Let the vectors $\hat{x}_k$ for $k \geq 1$ be the averages given by Eq. (9). Then, for any $k \geq 1$, the following hold at the vector $\hat{x}_k$:

(a) *An upper bound on the amount of constraint violation given by*

$$\left\| g(\hat{x}_k)^+ \right\| \leq \frac{\|\mu_k\|}{k\alpha}.$$

(b) *An upper bound on the primal cost given by*

$$f(\hat{x}_k) \leq q^* + \frac{\|\mu_0\|^2}{2k\alpha} + \frac{\alpha}{2k} \sum_{i=0}^{k-1} \|g(x_i)\|^2.$$

(c) *A lower bound on the primal cost given by*

$$f(\hat{x}_k) \geq q^* - \|\mu^*\| \, \|g(\hat{x}_k)^+\|,$$

where μ^* is a dual optimal solution.

Proof: (a) By using the definition of the iterate μ_{k+1} in Eq. (7), we obtain

$$\mu_k + \alpha g_k \leq [\mu_k + \alpha g_k]^+ = \mu_{k+1} \qquad \text{for all } k \geq 0.$$

Since $g_k = g(x_k)$ with $x_k \in X$, it follows that

$$\alpha g(x_k) \leq \mu_{k+1} - \mu_k \qquad \text{for all } k \geq 0.$$

Therefore,

$$\sum_{i=0}^{k-1} \alpha g(x_i) \leq \mu_k - \mu_0 \leq \mu_k \qquad \text{for all } k \geq 1,$$

where the last inequality in the preceding relation follows from $\mu_0 \geq 0$. Since $x_k \in X$ for all k, by the convexity of X, we have $\hat{x}_k \in X$ for all k. Hence, by the convexity of each of the functions g_j, it follows that for all $k \geq 1$,

$$g(\hat{x}_k) \leq \frac{1}{k} \sum_{i=0}^{k-1} g(x_i) = \frac{1}{k\alpha} \sum_{i=0}^{k-1} \alpha g(x_i) \leq \frac{\mu_k}{k\alpha}.$$

Because $\mu_k \geq 0$ for all k, we have $g(\hat{x}_k)^+ \leq \mu_k/(k\alpha)$ for all $k \geq 1$ and, therefore,

$$\left\| g(\hat{x}_k)^+ \right\| \leq \frac{\|\mu_k\|}{k\alpha} \qquad \text{for all } k \geq 1.$$

(b) By the convexity of the primal cost $f(x)$ and the definition of x_k as a minimizer of the Lagrangian function $f(x) + \mu_k' g(x)$ over $x \in X$ [cf. Eq. (8)], we have

$$f(\hat{x}_k) \leq \frac{1}{k} \sum_{i=0}^{k-1} f(x_i) = \frac{1}{k} \sum_{i=0}^{k-1} \{f(x_i) + \mu_i' g(x_i)\} - \frac{1}{k} \sum_{i=0}^{k-1} \mu_i g(x_i).$$

Since $q(\mu_i) = f(x_i) + \mu_i' g(x_i)$ and $q(\mu_i) \leq q^*$ for all i, it follows that for all $k \geq 1$,

$$f(\hat{x}_k) \leq \frac{1}{k} \sum_{i=0}^{k-1} q(\mu_i) - \frac{1}{k} \sum_{i=0}^{k-1} \mu_i' g(x_i) \leq q^* - \frac{1}{k} \sum_{i=0}^{k-1} \mu_i' g(x_i). \tag{10}$$

From the definition of the algorithm in Eq. (7), by using the nonexpansive property of the projection, and the facts $0 \in \{\mu \in \mathbb{R}^m \mid \mu \geq 0\}$ and $g_i = g(x_i)$, we obtain

$$\|\mu_{i+1}\|^2 \leq \|\mu_i\|^2 + 2\alpha \mu_i' g(x_i) + \alpha^2 \|g(x_i)\|^2 \qquad \text{for all } i \geq 0,$$

implying that

$$-\mu_i' g(x_i) \leq \frac{\|\mu_i\|^2 - \|\mu_{i+1}\|^2 + \alpha^2 \|g(x_i)\|^2}{2\alpha} \qquad \text{for all } i \geq 0.$$

By summing over $i = 0, \ldots, k-1$ for $k \geq 1$, we have

$$-\frac{1}{k} \sum_{i=0}^{k-1} \mu_i' g(x_i) \leq \frac{\|\mu_0\|^2 - \|\mu_k\|^2}{2k\alpha} + \frac{\alpha}{2k} \sum_{i=0}^{k-1} \|g(x_i)\|^2.$$

Combining the preceding relation and Eq. (10), we further have for all $k \geq 1$,

$$f(\hat{x}_k) \leq q^* + \frac{\|\mu_0\|^2 - \|\mu_k\|^2}{2k\alpha} + \frac{\alpha}{2k} \sum_{i=0}^{k-1} \|g(x_i)\|^2,$$

implying the desired estimate.

(c) Given a dual optimal solution μ^*, we have

$$\begin{aligned} f(\hat{x}_k) &= f(\hat{x}_k) + (\mu^*)' g(\hat{x}_k) - (\mu^*)' g(\hat{x}_k) \\ &\geq q(\mu^*) - (\mu^*)' g(\hat{x}_k). \end{aligned}$$

Because $\mu^* \geq 0$ and $g(\hat{x}_k)^+ \geq g(\hat{x}_k)$, we further have

$$-(\mu^*)' g(\hat{x}_k) \geq -(\mu^*)' g(\hat{x}_k)^+ \geq -\|\mu^*\| \|g(\hat{x}_k)^+\|.$$

From the preceding two relations and the fact $q(\mu^*) = q^*$ it follows that

$$f(\hat{x}_k) \geq q^* - \|\mu^*\| \|g(\hat{x}_k)^+\|.$$

■

An immediate consequence of Proposition 1(a) is that the maximum violation $\|g(\hat{x}_k)^+\|_\infty$ of constraints $g_j(x)$, $j = 1, \ldots, m$, at $x = \hat{x}_k$ is bounded by the same bound. In particular, we have

$$\max_{1 \leq j \leq m} g_j(\hat{x}_k)^+ \leq \frac{\|\mu_k\|}{k\alpha} \qquad \text{for all } k \geq 1.$$

which follows from the proposition in view of the relation $\|y\|_\infty \leq \|y\|$ for any y.

We note that the results of Proposition 1 in parts (a) and (c) show how the amount of feasibility violation $\|g(\hat{x}_k)^+\|$ affects the lower estimate of $f(\hat{x}_k)$. Furthermore, we note that the results of Proposition 1 indicate that the bounds on the feasibility violation and the primal value $f(\hat{x}_k)$ are readily available provided that we have bounds on the multiplier norms $\|\mu_k\|$, optimal solution norms $\|\mu^*\|$, and subgradient norms $\|g(x_k)\|$. This is precisely what we use in the next section to establish our estimates.

B. Properties of the Averaged Primal Sequence under Slater

Here, we strengthen the relations of Proposition 1 under the Slater condition and the subgradient boundedness. Our main result is given in the following proposition.

Proposition 2: Let the sequence $\{\mu_k\}$ be generated by the subgradient algorithm (7). Let the Slater condition and the bounded subgradient assumption hold [Assumptions 1 and 2]. Also, let

$$\begin{aligned} B^* &= \frac{2}{\gamma}\left(f(\bar{x}) - q^*\right) \\ &\quad + \max\left\{\|\mu_0\|, \frac{1}{\gamma}\left(f(\bar{x}) - q^*\right) + \frac{\alpha L^2}{2\gamma} + \alpha L\right\}, \end{aligned}$$

where $\gamma = \min_{1 \leq j \leq m}\{-g_j(\bar{x})\}$, $\bar{x}$ is a Slater vector, and L is the subgradient norm bound. Let the vectors $\hat{x}_k$ for $k \geq 1$ be the averages given by Eq. (9). Then, the following hold at the vector $\hat{x}_k$ for all $k \geq 1$:

(a) *An upper bound on the amount of constraint violation given by*

$$\|g(\hat{x}_k)^+\| \leq \frac{B^*}{k\alpha}.$$

(b) *An upper bound on the primal cost given by*

$$f(\hat{x}_k) \leq f^* + \frac{\|\mu_0\|^2}{2k\alpha} + \frac{\alpha L^2}{2}.$$

(c) *A lower bound on the primal cost given by*

$$f(\hat{x}_k) \geq f^* - \frac{1}{\gamma}\left[\, f(\bar{x}) - q^* \,\right] \|g(\hat{x}_k)^+\|.$$

As indicated in Proposition 2(a), the amount of feasibility violation $\|g(\hat{x}_k)^+\|$ of the vector $\hat{x}_k$ diminishes to zero at the rate $1/k$ as the number of subgradient iterations k increases. By combining the results in (a)-(c), we see that the function values $f(\hat{x}_k)$ converge to f^* within error level $\alpha L^2/2$ with the same rate of $1/k$ as $k \to \infty$.

Let us note that a more practical bound than the bound B^* of Proposition 2 can be obtained by using $\max_{0 \leq i \leq k} q(\mu_i)$ as an approximation of the dual optimal value q^*. These bounds can be used for developing practical stopping criteria for the subgradient algorithm with primal averaging. In particular, a user may specify a maximum tolerable infeasibility and/or a desired level of primal optimal accuracy. These specifications combined with the above estimates can be used to choose the stepsize value α, and to analyze the trade-off between the desired accuracy and the associated computational complexity (in terms of the number of subgradient iterations).

V. Numerical Example

In this section, we study a numerical example to illustrate the performance of the proposed subgradient method with primal averaging for a network resource allocation problem. Consider the network illustrated in Figure 1 with 2 serial links and 3 users each sending data at a rate x_i for $i = 1, 2, 3$. Link 1 has a capacity $c_1 = 1$ and link 2 has a capacity $c_2 = 2$. Assume that each user has an identical concave utility function $u_i(x_i) = \sqrt{x_i}$, which represents the utility gained from sending rate x_i. We consider allocating rates among the users as the optimal solution of the problem

$$\begin{aligned} &\text{maximize} && \textstyle\sum_{i=1}^3 \sqrt{x_i} \\ &\text{subject to} && x_1 + x_2 \leq 1, \quad x_1 + x_3 \leq 2, \\ &&& x_i \geq 0, \quad i = 1, 2, 3. \end{aligned}$$

The optimal solution of this problem is $x^* = [0.2686, 0.7314, 1.7314]$ and the optimal value is $f^* \approx 2.7$. We consider solving this problem using the dual subgradient method of Eq. (7) (with a constant stepsize $\alpha = 1$) combined with the primal averaging. In particular, when evaluating the subgradients of the dual function in Eq. (8), we obtain the primal sequence $\{x_k\}$. We generate the sequence $\{\hat{x}_k\}$ as the running average of the primal sequence [cf. Eq. (9)].

Figure 2 illustrates the behavior of the sequences $\{x_{ik}\}$ and $\{\hat{x}_{ik}\}$ for each user $i = 1, 2, 3$. As seen in this figure, for each user i, the sequences $\{x_{ik}\}$ exhibit oscillations whereas the average sequences $\{\hat{x}_{ik}\}$ converge smoothly to near-optimal solutions within 60 subgradient iterations.

Figure 3 illustrates the results for constraint violation and primal objective value for the sequences $\{x_k\}$ and $\{\hat{x}_k\}$. The plot to the left in Figure 3 shows the convergence behavior of the constraint violation $\|g(x)^+\|$ for the two sequences, i.e., $\|g(x_k)^+\|$ and $\|g(\hat{x}_k)^+\|$. Note that the constraint violation for the sequence $\{x_k\}$ oscillates within a large range while the constraint violation for the average

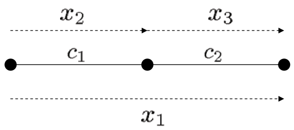

Fig. 1. A simple network with two links of capacities $c_1 = 1$ and $c_2 = 2$, and three users, each sending data at a rate x_i.

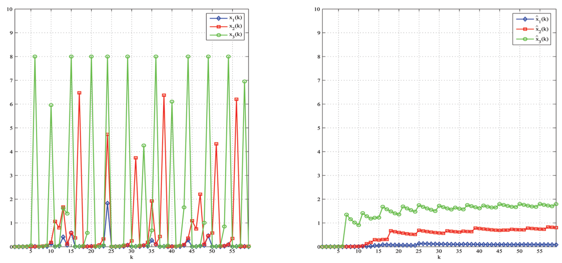

Fig. 2. The convergence behavior of the primal sequence $\{x_k\}$ (on the left) and $\{\hat{x}_k\}$ (on the right).

sequence $\{\hat{x}_k\}$ rapidly converges to 0. The plot to the right in Figure 3 shows a similar convergence behavior for the primal objective function values $f(x)$ along the sequences $\{x_k\}$ and $\{\hat{x}_k\}$.

VI. Conclusions

In this paper, we studied generating approximate primal solutions using dual subgradient method. The algorithm uses projected subgradient iterations to generate a dual sequence and an averaging scheme to produce approximate primal vectors. We have provided bounds on the amount of constraint violation at approximate primal solutions, as well as bounds on the primal function values. The subgradient algorithm produces primal vectors whose infeasibility diminishes to zero and whose function value converges to the primal optimal value within an error at a rate $1/k$ in the number of subgradient iterations k. We also illustrate the algorithm's behavior as applied to a rate allocation problem.

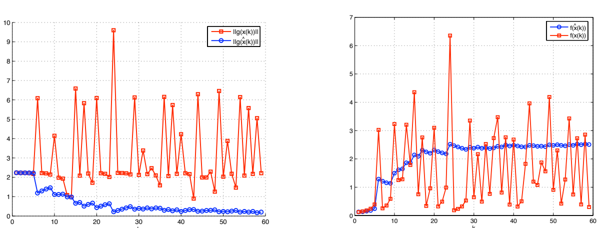

Fig. 3. The figure on the left shows the convergence behavior of the constraint violation for the two primal sequences, $\{x_k\}$ and $\{\hat{x}_k\}$. Similarly, the figure on the right shows the convergence of the corresponding primal objective function values.

References

[1] D.P. Bertsekas, A. Nedić, and A.E. Ozdaglar, *Convex analysis and optimization*, Athena Scientific, Cambridge, Massachusetts, 2003.

[2] R. Correa and C. Lemaréchal, *Convergence of some algorithms for convex minimization*, Mathematical Programming **62** (1993), 261–271.

[3] V.F. Demyanov and L.V. Vasilyev, *Nondifferentiable optimization*, Optimization Software, Inc., New York, 1985.

[4] Y.M. Ermoliev, *Methods for solving nonlinear extremal problems*, Kibernetika (1969), no. 4, 1–17.

[5] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms*, Springer-Verlag, Berlin, 1996.

[6] K.C. Kiwiel, T. Larsson, and P.O. Lindberg, *Lagrangian relaxation via ballstep subgradient methods*, Mathematics of Operations Research **32** (2007), no. 3, 669–686.

[7] K.C. Kiwiel and P.O. Lindberg, *Parallel subgradient methods for convex optimization*, Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications (D. Butnariu, Y. Censor, and S. Reich, eds.), Studies Comput. Math., Elsevier, 2001.

[8] T. Larsson and Z. Liu, *A Lagrangian relaxation scheme for structured linear programs with application to multicommodity network flows*, Optimization **40** (1997), 247–284.

[9] T. Larsson, M. Patriksson, and A. Strömberg, *Ergodic primal convergence in dual subgradient schemes for convex programming*, Mathematical Programming **86** (1999), 283–312.

[10] S. Low and D.E. Lapsley, *Optimization flow control, I: Basic algorithm and convergence*, IEEE/ACM Transactions on Networking **7** (1999), no. 6, 861–874.

[11] A. Nedić and D.P. Bertsekas, *Convergence rate of incremental subgradient algorithms*, Stochastic Optimization: Algorithms and Applications (S. Uryasev and P. Pardalos, eds.), Kluwer Academic Publishers, 2000.

[12] ———, *Incremental subgradient methods for nondifferentiable optimization*, SIAM Journal on Optimization **56** (2001), no. 1, 109–138.

[13] A. Nedić, D.P. Bertsekas, and V.S. Borkar, *Distributed asynchronous incremental subgradient methods*, Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications (D. Butnariu, Y. Censor, and S. Reich, eds.), Studies Comput. Math., Elsevier, 2001.

[14] A. Nedić and A. Ozdaglar, *Approximate primal solutions and rate analysis for dual subgradient methods*, Tech. Report LIDS 2753, Massachusetts Institute of Technology, 2007.

[15] A.S. Nemirovskii and D.B. Judin, *Cezare convergence of gradient method approximation of saddle points for convex-concave functions*, Doklady Akademii Nauk SSSR **239** (1978), 1056–1059.

[16] Y. Nesterov, *Primal-dual subgradient methods for convex problems*, Tech. Report 67, Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL), 2005.

[17] B.T. Polyak, *A general method for solving extremum problems*, Soviet Math. Doklady **8** (1967), no. 3, 593–597.

[18] ———, *Introduction to optimisation*, Optimization Software, Inc., New York, 1987.

[19] A. Ruszczyński, *A merit function approach to the subgradient method with averaging*, Optimization Methods and Software (2007), to appear.

[20] H.D. Sherali and G. Choi, *Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs*, Operations Research Letters **19** (1996), 105–113.

[21] N.Z. Shor, *Minimization methods for nondifferentiable functions*, Translated from Russian by K.C. Kiwiel and A. Ruszczynski, Springer, Berlin, 1985.

[22] R. Srikant, *Mathematics of Internet congestion control*, Birkhauser, 2004.

[23] H. Uzawa, *Iterative methods in concave programming*, Studies in Linear and Nonlinear Programming (K. Arrow, L. Hurwicz, and H. Uzawa, eds.), Stanford University Press, 1958, pp. 154–165.