

Distributed Multiuser Optimization: Algorithms and Error Analysis

Jayash Koshal, Angelia Nedić and Uday V. Shanbhag

Abstract—We consider a class of multiuser optimization problems in which user interactions are seen through congestion cost functions or coupling constraints. Our primary emphasis lies on the convergence and error analysis of distributed algorithms in which users communicate through aggregate user information. Traditional implementations are reliant on strong convexity assumptions, require coordination across users in terms of consistent stepsizes, and often rule out early termination by a group of users. We consider how some of these assumptions can be weakened in the context of projection methods motivated by fixed-point formulations of the problem. Specifically, we focus on (approximate) primal and primal-dual projection algorithms. We analyze the convergence behavior of the methods and provide error bounds in settings with limited coordination across users and regimes where a group of users may prematurely terminate affecting the convergence point.

I. INTRODUCTION

This paper considers a multiuser optimization problem in a general setting, where users are coupled by a cost function and constraints. A problem of this kind may arise in network resource allocation such as rate allocation among multiple users, where the coupling cost may be due to congestion, while the coupling constraints may be due to the network link capacities. The goal of this paper is to study both primal and primal-dual algorithms, establish their convergence properties and provide error bounds.

This paper is related to the rate-control class of problems considered in [2], [4], [5], [12], [13]. The prior work is dealing with users with separable objectives, but coupled polyhedral constraints. Both primal and primal-dual schemes are discussed typically in a continuous-time setting (with exception for [5] investigating dual discrete-time schemes). In contrast, this paper deals with discrete-time schemes.

More generally, the work in this paper is related to the distributed algorithms in [1], [14], where several classes of problems with special structure admitting the decentralized computations are discussed including contractive mappings. Unlike the preceding work, in this paper we also consider noncontractive mappings for which we propose algorithms that through averaging admit error bounds per iteration. The use of averaging is motivated by the work in [9] and the more recent schemes for generating primal and primal-dual approximate solutions proposed in [6]–[8], as well as a merit function approach proposed in [11].

In this paper, we focus on a multiuser optimization problem where the system cost is given by $\sum_{i=1}^N f_i(x_i) + c(x)$,

with $f_i(x_i)$ being user-specific convex cost and $c(x)$ being a coupling cost. We cast such a problem as a Cartesian variational inequality [3], for which we consider primal and primal-dual projection algorithms. Our interest is in the convergence and error analysis for constant stepsize rules. We analyze the primal algorithm when the system cost is strongly convex and when the cost is only convex, but has bounded gradients. We also allow for coupling across user decisions through convex constraints and analyze a primal-dual algorithm for such a system. The contributions can be broadly categorized as lying in the development and analysis of *primal projection algorithms*, *average primal algorithms*, and *primal-dual algorithms*. The novelty is in our *error analysis and per-iteration error bounds* for the system behavior, including the cases when the users use different stepsize values, a class of users terminate prematurely, and different stepsize values are used within primal-dual scheme.

The paper is organized as follows. In section II, we describe the problem of interest and formulate it as a variational inequality. In section III, we propose a primal method, analyze its properties and provide error bounds under “strong monotonicity” condition. In section IV, under weaker conditions, we study the effects of averaging and the resulting error bounds. In section V, we extend our analysis to allow for more general coupling constraints and present a regularized primal-dual method. Our analysis is equipped with error bounds when primal and dual steplengths differ. We summarize our main contributions in Section VI.

Throughout this paper, we view vectors as columns. We write x^T to denote the transpose of a vector x , and $x^T y$ to denote the inner product of vectors x and y . We use $\|x\| = \sqrt{x^T x}$ to denote the Euclidean norm of a vector x . We use Π_X to denote the Euclidean projection operator onto a set X , i.e., $\Pi_X(x) \triangleq \operatorname{argmin}_{z \in X} \|x - z\|$.

II. PROBLEM FORMULATION

Consider a set of N users each having a specific cost $f_i(x_i)$ depending on a decision vector $x_i \in \mathbb{R}^{n_i}$. Let $c(x)$ be a coupling cost that depends on the user decision vectors, i.e., $x = (x_1, \dots, x_N) \in \mathbb{R}^n$, where $n = \sum_{i=1}^N n_i$. The functions $f_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R}$ and $c : \mathbb{R}^n \rightarrow \mathbb{R}$ are *convex and differentiable*¹. In addition, each user i has a constraint set $X_i \subseteq \mathbb{R}^{n_i}$, which is *non-empty, convex and closed*. The user-

Department of Industrial and Enterprise Systems Engineering, University of Illinois, Urbana IL 61801, Email: {koshal1,angelia,udaybag}@illinois.edu. This work has been supported by NSF award CCF-0728863.

¹It is not crucial to have the functions defined over the entire space. We make this assumption to simplify the discussion and notation. Also, some of our results are applicable to the case when f_i are not necessarily differentiable, by replacing the gradient with a subgradient.

system optimization problem is given by:

$$\begin{aligned} \text{minimize} \quad & f(x) \triangleq \sum_{i=1}^N f_i(x_i) + c(x) \\ \text{subject to} \quad & x_i \in X_i \text{ for all } i = 1, \dots, N. \end{aligned} \quad (1)$$

We note that the problem has ‘‘almost’’ separable structure in the sense that the user’s variables are coupled only through the cost $c(x)$. In the absence of such a cost, the problem would decompose into N independent user problems.

We are interested in distributed algorithms aimed at solving the system problem, while each user executes computations in the space of its own decision variables. Our approach is based on casting the system optimization problem as a variational inequality, which is decomposed by exploiting the separable structure of the user’s constraint sets and ‘‘almost separable’’ structure of the objective function in (1). We let X be the Cartesian product of the sets X_i , i.e., $X = X_1 \times \dots \times X_N$, $X \subseteq \mathbb{R}^n$. By the first-order optimality conditions, it can be seen that $x^* \in X$ is a solution to problem (1) if and only if x^* solves the following variational inequality, denoted by $\text{VI}(X, F)$,

$$(y - x)^T F(x) \geq 0 \quad \text{for all } y \in X,$$

where the map $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is given by

$$F(x) \triangleq (F_1(x), \dots, F_N(x))^T,$$

$$F_i(x) = \nabla_{x_i}(f_i(x_i) + c(x)) \quad \text{for all } i = 1, \dots, N. \quad (2)$$

Being the Cartesian product of convex sets, the set X is convex. Thus, a vector $x^* \in X$ solves $\text{VI}(X, F)$ if and only if x^* is a fixed point of the natural map² $\mathbf{F}_X^{\text{nat}}(x) \triangleq x - \Pi_X(x - F(x))$, i.e., $\mathbf{F}_X^{\text{nat}}(x^*) = 0$. In view of the special structure of the map F , it can be seen that the preceding fixed point equation is decomposable per user, and it is equivalent to the following system of equations:

$$x_i - \Pi_{X_i}(x_i - \alpha F_i(x)) \quad \text{for } i = 1, \dots, N, \quad (3)$$

with any scalar $\alpha > 0$. Thus, $x^* = (x_1^*, \dots, x_N^*) \in X$ solves problem (1) if and only if it is a solution to the system (3).

In the sequel, we use $\|x\|$ and $x^T y$ to denote respectively the Euclidean norm and the inner product in the product space $\mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_N}$, that are induced by the Euclidean norms and the inner products in the component spaces.

In addition to assuming that each X_i is closed convex set, and each f_i and c are differentiable convex functions in problem (1), we also use the following assumptions:

- (A1) The gradient map $F(x)$ is strongly monotone and Lipschitz over the set X with constants ν and L .
- (A2) The set X is compact.

The assumption (A1) is satisfied, for example, when $c(x)$ is strongly convex with Lipschitz gradient over X and each $\nabla_{x_i} f_i$ is Lipschitz over X_i . The assumption (A2) is satisfied, for example, when each X_i is compact. In this case, the gradient map $F(x)$ is uniformly bounded over the set X , i.e., there is $M > 0$ such that $\|F(x)\| \leq M$ for all $x \in X$.

III. PRIMAL PROJECTION METHOD

Here, we consider a projection method for solving the user optimization problem (1). Let x_i^k denote the estimate of user i at iteration k . At each iteration, user i receives estimates from the other users and then updates according to:

$$x_i^{k+1} = \Pi_{X_i}(x_i^k - \tau_k F_i(x^k)) \quad \text{for all } k \geq 0, \quad (4)$$

where $\tau_k > 0$ is the stepsize, $F_i(x)$ is given by (2), and $x_i^0 \in X_i$ is the initial estimate for user i . In the following sections, we study the properties of the sequence $\{x^k\}$ whose components x_i^k are generated by (4).

A. Convergence

For a strongly monotone gradient map, the convergence of the method is geometric, as seen in the following lemma.

Lemma 1: Let (A1) hold, and let $x^* \in X$ be the optimal solution for problem (1). Let $\{x_i^k\}, i = 1, \dots, N$, be generated by (4) using a constant stepsize, i.e., $\tau_k = \tau$ for all $k \geq 0$, with $\tau < 2/L$. Then, we have

$$\|x^{k+1} - x^*\| \leq \sqrt{q} \|x^k - x^*\| \quad \text{for all } k \geq 0,$$

where $q = 1 - \nu(2\tau - \tau^2 L)$.

The proof is similar to that of Lemma 2 in [10], page 24.

B. Error Analysis

In this section, we provide error results associated with a finite termination of all users or a group of users.

1) *Finite Termination:* We provide an upper bound for the error resulting from a finite termination of the algorithm (4) using a constant stepsize τ . Let ϵ_i^k be the residual error for user i at iteration k , i.e., for all k and $i = 1, \dots, N$,

$$\epsilon_i^k \triangleq \|x_i^k - \Pi_{X_i}(x_i^k - \tau F_i(x^k))\|. \quad (5)$$

We have the following result.

Lemma 2: Under assumptions of Lemma 1, we have

$$\sum_{i=1}^N (\epsilon_i^k)^2 \leq (1 + \sqrt{q})^2 q^k \|x^0 - x^*\|^2.$$

Proof: Since $x_i^{k+1} = \Pi_{X_i}(x_i^k - \tau F_i(x^k))$, we have $(\epsilon_i^k)^2 = \|x_i^k - x_i^{k+1}\|^2$. By summing over all users, we obtain

$$\sum_{i=1}^N (\epsilon_i^k)^2 = \sum_{i=1}^N \|x_i^k - x_i^{k+1}\|^2 = \|x^k - x^{k+1}\|^2.$$

Therefore, $\sum_{i=1}^N (\epsilon_i^k)^2 \leq (\|x^k - x^*\| + \|x^{k+1} - x^*\|)^2$, and by using Lemma 1, we further have

$$\begin{aligned} \sum_{i=1}^N (\epsilon_i^k)^2 &\leq (1 + \sqrt{q})^2 \|x^k - x^*\|^2 \\ &\leq (1 + \sqrt{q})^2 q^k \|x^0 - x^*\|^2. \end{aligned}$$

The preceding bound can be used to determine the minimal number of iterations K that is required to guarantee the user errors are all below some prespecified error level $\bar{\epsilon}$. ■

²See [3], volume 1, 1.5.8 Proposition, page 83.

In particular, since $\max_{1 \leq i \leq N} \epsilon_i^k \leq \sqrt{\sum_{i=1}^N (\epsilon_i^k)^2}$, we can determine the minimal $K \geq 0$ satisfying

$$(1 + \sqrt{q})\sqrt{q^K} \|x^0 - x^*\| \leq \bar{\epsilon}.$$

This gives us $K = \left\lceil \frac{2 \ln((1 + \sqrt{q})\|x^0 - x^*\|) - 2 \ln \bar{\epsilon}}{(-\ln q)} \right\rceil$. Note that K increases as $\bar{\epsilon}$ decreases, which is expected since a higher accuracy would naturally require more iterations. Also, K decreases as q decreases which is consistent with intuition that ‘‘a more contractive’’ map would require fewer iterations.

2) *Group of Users Terminates:* We study the error properties when a certain group of users terminates updating as they have their error below a critical level, while the other users continue updating. We provide error estimate for the global system under assumptions (A1)-(A2).

Let \tilde{k} be the time when a group of users terminates computations, and let \mathcal{I} be the index set of the users who have terminated. For $j \in \mathcal{I}$, we have $\|x_j^{\tilde{k}} - x_j^*\| \leq \bar{\epsilon}$. Let $\{\tilde{x}^k\}$ be the resulting sequence generated by the users. This sequence is identical to the sequence $\{x_k\}$ obtained by (4) up to time \tilde{k} . In particular, we have

$$\begin{aligned} \tilde{x}_i^k &= x_i^k & \text{for all } k \leq \tilde{k} \text{ and all } i, \\ \|x_j^{\tilde{k}} - x_j^*\| &\leq \bar{\epsilon}, \quad \tilde{x}_j^k = \tilde{x}_j & \text{for } k \geq \tilde{k} \text{ and } j \in \mathcal{I}, \\ \tilde{x}_i^{k+1} &= \Pi_{X_i}(\tilde{x}_i^k - \tau F_i(\tilde{x}^k)) & \text{for } k \geq \tilde{k} \text{ and } i \in \mathcal{I}^c, \end{aligned} \quad (6)$$

where \mathcal{I}^c is the index set of the users that continue updating.

We next, provide an estimate for the difference between the resulting sequence $\{x^k\}$ and the optimal x^* .

Proposition 1: Let (A1) and (A2) hold, and let the constant stepsize satisfy $0 < \tau < \frac{2}{L}$. Then, the sequence $\{\tilde{x}^k\}$ is such that for $k \geq \tilde{k}$,

$$\|\tilde{x}^{k+1} - x^*\|^2 \leq q^{k+1} \|x^0 - x^*\|^2 + \frac{1 - q^{k+1-\tilde{k}}}{1 - q} 4\tau M |\mathcal{I}| \bar{\epsilon},$$

where $q = 1 - \nu(2\tau - \tau^2 L)$ and M is the upper bound on gradient norms $\|F(x)\|$ over the set X .

Proof: Since $\tilde{x}^k = x^k$ for $k \leq \tilde{k}$, by Lemma 1, we get

$$\|\tilde{x}^{\tilde{k}} - x^*\|^2 \leq q^{\tilde{k}} \|x^0 - x^*\|^2. \quad (8)$$

From relation (6) we have for $k \geq \tilde{k}$ and $i \in \mathcal{I}$,

$$\|\tilde{x}_i^{k+1} - x_i^*\| = \|\tilde{x}_i^k - x_i^*\|.$$

Using the nonexpansive property of the projection, from (7) we can see that for $k \geq \tilde{k}$ and $i \in \mathcal{I}^c$,

$$\begin{aligned} \|\tilde{x}_i^{k+1} - x_i^*\|^2 &\leq \|\tilde{x}_i^k - x_i^*\|^2 + \tau^2 \|F_i(\tilde{x}^k) - F_i(x^*)\|^2 \\ &\quad - 2\tau (F_i(\tilde{x}^k) - F_i(x^*))^T (\tilde{x}_i^k - x_i^*). \end{aligned}$$

Summing across all users we obtain

$$\begin{aligned} \|\tilde{x}^{k+1} - x^*\|^2 &\leq \|\tilde{x}^k - x^*\|^2 + \tau^2 \sum_{i \in \mathcal{I}^c} \|F_i(\tilde{x}^k) - F_i(x^*)\|^2 \\ &\quad - 2\tau \sum_{i \in \mathcal{I}^c} (F_i(\tilde{x}^k) - F_i(x^*))^T (\tilde{x}_i^k - x_i^*). \end{aligned}$$

Adding and subtracting the missing terms with $j \in \mathcal{I}$ and using the Lipschitz and strong monotonicity property of the gradient map F , we further obtain

$$\begin{aligned} \|\tilde{x}^{k+1} - x^*\|^2 &\leq \|\tilde{x}^k - x^*\|^2 + \tau^2 \|F(\tilde{x}^k) - F(x^*)\|^2 \\ &\quad - 2\tau (F(\tilde{x}^k) - F(x^*))^T (\tilde{x}^k - x^*) \\ &\quad + 2\tau \sum_{j \in \mathcal{I}} (F_j(\tilde{x}^k) - F_j(x^*))^T (\tilde{x}_j^k - x_j^*) \\ &\leq q \|\tilde{x}^k - x^*\|^2 + 2\tau \sum_{j \in \mathcal{I}} \|F_j(\tilde{x}^k) - F_j(x^*)\| \|\tilde{x}_j^k - x_j^*\| \\ &\leq q \|\tilde{x}^k - x^*\|^2 + 4\tau M |\mathcal{I}| \bar{\epsilon}, \end{aligned}$$

where in the last step we used the first relation in (6) and $\|F(x)\| \leq M$ for all $x \in X$, which implies that each F_i is also bounded by M over X . Therefore,

$$\begin{aligned} \|\tilde{x}^{k+1} - x^*\|^2 &\leq q^{k+1-\tilde{k}} \|\tilde{x}^{\tilde{k}} - x^*\|^2 + 4\tau M |\mathcal{I}| \bar{\epsilon} \sum_{t=0}^{k-\tilde{k}} q^t \\ &\leq q^{k+1} \|x^0 - x^*\|^2 + \frac{1 - q^{k+1-\tilde{k}}}{1 - q} 4\tau M |\mathcal{I}| \bar{\epsilon}, \end{aligned}$$

where in the second inequality we use (8). \blacksquare

Note that the results of Proposition 1 and Lemma 1 coincide when $|\mathcal{I}| = 0$.

IV. APPROXIMATE AVERAGE PRIMAL SOLUTIONS

In this section, we study the properties of the method (4) without the strong monotonicity assumption. In order to provide approximate solutions with error estimates, we consider the averages of user estimates. Specifically, each user i updates according to (4) and, in each iteration k , the user computes its average $\hat{x}_i^k = \frac{1}{k+1} \sum_{t=0}^k x_t^i$.

A. User Dependent Stepsize

We consider a situation when users have their individual stepsize in (4). In the following proposition, we provide a per-iteration bound on the system *optimal function value* f^* .

Proposition 2: Let (A2) hold. Let $\{x^k\}$ be sequence generated by (4) with the stepsize $\tau_k = \tau_i$ for user i . Let \hat{x}^k be the average vector with components \hat{x}_i^k (of user averages). We then have for any $k \geq 1$,

$$f(\hat{x}^k) - f(x^*) \leq \frac{D^2}{2(k+1)\tau_{\min}} + \frac{\tau_{\max}^2 M^2}{2\tau_{\min}} + \frac{\delta MD}{\tau_{\min}}, \quad (9)$$

where $D = \max_{u,v \in X} \|u - v\|$, $\tau_{\max} = \max_{1 \leq i \leq N} \tau_i$, $\tau_{\min} = \min_{1 \leq i \leq N} \tau_i$, and $\delta = \tau_{\max} - \tau_{\min}$.

Proof: When X is compact, by continuity of f , a solution $x^* \in X$ to problem (1) exists. By using the relation $x_i^* = \Pi_{X_i}(x_i^*)$ and the nonexpansive property of the projection operation, we can see that

$$\begin{aligned} \|x_i^{k+1} - x_i^*\|^2 &\leq \|x_i^k - x_i^*\|^2 + \tau_i^2 \|F_i(x^k)\|^2 \\ &\quad - 2\tau_i F_i(x^k)^T (x_i^k - x_i^*). \end{aligned}$$

Summing over all users i , and using $\tau_i \leq \tau_{\max}$, we get

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 + \tau_{\max}^2 \|F(x^k)\|^2 \\ &\quad - 2 \sum_{i=1}^N \tau_i F_i(x^k)^T (x_i^k - x_i^*). \end{aligned}$$

By adding and subtracting the terms $\tau_{\min} F_i(x^k)^T(x_i^k - x_i^*)$, and using $\|F(x^k)\| \leq M$, we further have

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 + \tau_{\max}^2 M^2 \\ &\quad - 2\tau_{\min} F(x^k)^T(x^k - x^*) + 2\delta \sum_{i=1}^N \|F_i(x^k)\| \|x_i^k - x_i^*\|, \end{aligned}$$

with $\delta = \tau_{\max} - \tau_{\min}$. By Hölders inequality, we have $\sum_{i=1}^N \|F_i(x^k)\| \|x_i^k - x_i^*\| \leq \|F(x^k)\| \|x^k - x^*\|$, which together with $\|F(x^k)\| \leq M$ and $\|x^k - x^*\| \leq D$, implies

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 + \tau_{\max}^2 M^2 \\ &\quad - 2\tau_{\min} F(x^k)^T(x^k - x^*) + 2\delta MD. \end{aligned}$$

By convexity of $f(x)$ we have $F(x^k)^T(x^k - x^*) \geq f(x^k) - f(x^*)$, and by rearranging the terms, we obtain

$$\begin{aligned} 2\tau_{\min}(f(x^k) - f(x^*)) &\leq \|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 \\ &\quad + \tau_{\max}^2 M^2 + 2\delta MD. \end{aligned}$$

Summing the preceding relations over $k = 0, \dots, K$ yields

$$\begin{aligned} 2\tau_{\min} \sum_{k=0}^K (f(x^k) - f(x^*)) &\leq \|x^0 - x^*\|^2 - \|x^{K+1} - x^*\|^2 \\ &\quad + (K+1)(\tau_{\max}^2 M^2 + 2\delta MD) \quad \text{for } K \geq 1. \end{aligned}$$

Dropping the nonpositive term, dividing by $2\tau_{\min}(K+1)$, and using $\|x^0 - x^*\| \leq D$ and the convexity of f , we obtain

$$f(\hat{x}^K) - f(x^*) \leq \frac{D^2}{2(K+1)\tau_{\min}} + \frac{\tau_{\max}^2 M^2}{2\tau_{\min}} + \frac{\delta MD}{\tau_{\min}}.$$

Observe that the constant error term $\frac{\tau_{\max}^2 M^2}{2\tau_{\min}} + \frac{\delta MD}{\tau_{\min}}$ depends on the stepsize and does not diminish even when all users have the same stepsize. When $\tau_{\max} = \tau_{\min} = \tau$, we have $\delta = 0$ yielding a known result for a convex function f :

$$f(\hat{x}^K) - f(x^*) \leq \frac{D^2}{2(K+1)\tau} + \frac{\tau M^2}{2}.$$

B. Error when a Group of Users Terminates

We next provide a bound for the case when a group of users ceases updating. In particular, we have the following result, which parallels the result of Proposition 1.

Proposition 3: Let (A2) hold, and let $\{\tilde{x}^k\}$ be generated by (6)–(7). Then, the average vectors $\hat{x}^k = \sum_{t=0}^k \tilde{x}^t / (k+1)$ are such that for all $k \geq \tilde{k}$,

$$f(\hat{x}^k) - f(x^*) \leq \frac{D^2}{2\tau(k+1)} + \frac{\tau M^2}{2} + \frac{(k - \tilde{k})M|\mathcal{I}|\bar{\epsilon}}{k+1}.$$

Proof: For users i continuing the updates, i.e., $i \in \mathcal{I}^c$, and all $k \geq \tilde{k}$, using $x_i^* = \Pi_{X_i}(x_i^*)$, it can be seen that

$$\begin{aligned} \|\tilde{x}_i^{k+1} - x_i^*\|^2 &\leq \|\tilde{x}_i^k - x_i^*\|^2 + \tau^2 \|F_i(\tilde{x}^k)\|^2 \\ &\quad - 2\tau F_i(\tilde{x}^k)^T(\tilde{x}_i^k - x_i^*). \end{aligned}$$

By (6), we have $x_i^k = \tilde{x}_i$ for all $i \in \mathcal{I}$ and $k \geq \tilde{k}$. By using this together with the preceding relation, and by summing

over all i , we obtain

$$\begin{aligned} \|\tilde{x}^{k+1} - x^*\|^2 &\leq \|\tilde{x}^k - x^*\|^2 + \tau^2 \sum_{i \in \mathcal{I}^c} \|F_i(\tilde{x}^k)\|^2 \\ &\quad - 2\tau \sum_{i \in \mathcal{I}^c} F_i(\tilde{x}^k)^T(\tilde{x}_i^k - x_i^*). \end{aligned}$$

Adding and subtracting the terms with $i \in \mathcal{I}$, we get

$$\begin{aligned} \|\tilde{x}^{k+1} - x^*\|^2 &\leq \|\tilde{x}^k - x^*\|^2 + \tau^2 \|F(\tilde{x}^k)\|^2 \\ &\quad - 2\tau F(\tilde{x}^k)^T(\tilde{x}^k - x^*) + 2\tau \sum_{j \in \mathcal{I}} F_j(\tilde{x}^k)^T(\tilde{x}_j^k - x_j^*). \end{aligned}$$

By rearranging the terms, and by invoking the boundedness of the gradients and relation (6), we obtain for $k \geq \tilde{k}$,

$$\begin{aligned} 2\tau F(\tilde{x}^k)^T(\tilde{x}^k - x^*) &\leq \|\tilde{x}^k - x^*\|^2 - \|\tilde{x}^{k+1} - x^*\|^2 \\ &\quad + \tau^2 M^2 + 2\tau M|\mathcal{I}|\bar{\epsilon}. \end{aligned}$$

By convexity of f , we have $F(\tilde{x}^k)^T(\tilde{x}^k - x^*) \geq f(\tilde{x}^k) - f(x^*)$ for all k , implying

$$\begin{aligned} 2\tau (f(\tilde{x}^k) - f(x^*)) &\leq \|\tilde{x}^k - x^*\|^2 - \|\tilde{x}^{k+1} - x^*\|^2 \\ &\quad + \tau^2 M^2 + 2\tau M|\mathcal{I}|\bar{\epsilon}. \end{aligned}$$

Using the analysis similar to that of the proof of Proposition 2, we can see that for all $k < \tilde{k}$,

$$2\tau (f(\tilde{x}^k) - f(x^*)) \leq \|\tilde{x}^k - x^*\|^2 - \|\tilde{x}^{k+1} - x^*\|^2 + \tau^2 M^2.$$

Thus, by summing the preceding relations over all k , and then dividing with $2\tau(k+1)$, we see that for $k \geq \tilde{k}$,

$$\frac{\sum_{t=0}^k f(\tilde{x}^t) - f(x^*)}{k+1} \leq \frac{\|x^0 - x^*\|^2}{2\tau(k+1)} + \frac{\tau M^2}{2} + \frac{(k - \tilde{k})M|\mathcal{I}|\bar{\epsilon}}{k+1}.$$

The desired estimate follows by the convexity of f and relation $\|x^0 - x^*\| \leq D$. \blacksquare

V. PRIMAL-DUAL PROJECTION METHOD

In many settings, an algebraic characterization of the constrained set may be essential for constructing convergent schemes. One such instance is a generalization of our canonical multiuser optimization problem given by

$$\begin{aligned} \text{minimize} \quad & f(x) \triangleq \sum_{i=1}^N f_i(x_i) \\ \text{subject to} \quad & x_i \in X_i \quad \text{for all } i = 1, \dots, N \quad (10) \\ & d(x) \leq 0, \end{aligned}$$

where $d(x) = (d_1(x), \dots, d_m(x))^T$ and each $d_j : \mathbb{R}^n \rightarrow \mathbb{R}$ is a *differentiable convex function*. We assume that there is no duality gap between this problem and its dual, and that a dual optimal solution exists.³ We use $\lambda \in \mathbb{R}^m$ to denote a Lagrange multiplier (dual variables). The optimal primal solution x^* and its corresponding optimal Lagrange multiplier λ^* are jointly referred to as a primal-dual optimal solution $z^* \triangleq (x^*, \lambda^*)$, which is also a solution to the coupled fixed-point problems:

$$x^* = \Pi_X(x^* - \phi_x(x^*, \lambda^*)),$$

³For this to hold, we may assume for example that Slater cognition holds, i.e., there is an $\bar{x} \in X = X_1 \times \dots \times X_N$ such that $d(\bar{x}) < 0$.

$$\lambda^* = \Pi_{\mathbb{R}_+^m}(\lambda^* + d(x^*)), \quad (11)$$

where $\phi_x(x, \lambda)$ is defined as

$$\phi_x(x, \lambda) \triangleq (\phi_1(x, \lambda), \dots, \phi_N(x, \lambda))^T,$$

with $\phi_i(x, \lambda) = \nabla_{x_i}(f_i(x_i) + \lambda^T d(x))$ for all $i = 1, \dots, N$. Furthermore, we define the mapping $\phi(x, \lambda)$ as $\phi(x, \lambda) \triangleq (\phi_x(x, \lambda), -d(x))^T$, and we often use a more compact notation $z \triangleq (x, \lambda)$. Multiple approaches may be applied in the construction of primal-dual methods for solving such a class of problems. Given that our emphasis is on the construction of error estimates under a wide range of generalizations, we lay an accent on simple projection schemes. If we view problem (11) as a variational inequality of the form:

$$(z - z^*)^T \phi(z^*) \geq 0 \quad \text{for all } z = (x, \lambda) \in X \times \mathbb{R}_+^m,$$

then, the constant steplength algorithms discussed earlier may not converge. This is primarily because the mapping $\phi(z)$ is no longer strongly monotone but merely monotone. For this weaker set of mappings, convergent projection schemes may be constructed by using Tikhonov regularization and two-step projections, amongst others⁴.

We consider a regularization approach that allows for convergence to an ϵ -optimal solution. This requires the construction of a strongly monotone regularized mapping $\phi_\epsilon(x, \lambda)$, defined as $\phi_\epsilon(x, \lambda) \triangleq (\phi_x(x, \lambda), -d(x) + \epsilon\lambda)^T$. In this case, the algorithm for solving (11) is given by

$$\begin{aligned} x^{k+1} &= \Pi_X(x^k - \tau\phi_x(x^k, \lambda^k)), \\ \lambda^{k+1} &= \Pi_{\mathbb{R}_+^m}(\lambda^k - \tau(-d(x^k) + \epsilon\lambda)). \end{aligned} \quad (12)$$

For the mapping ϕ , we assume that:

(A3) The map $\phi(x, \lambda)$ is Lipschitz over the set $X \times \mathbb{R}_+^m$ with constant L_ϕ .

Under (A2), the boundedness of $\nabla d(x)$ follows, i.e., $\|\nabla d(x)\| \leq M_d$ for all $x \in X$ and some $M_d > 0$.

Next, we prove the Lipschitzian nature of $\phi(x, \lambda)$, its regularized counterpart $\phi_\epsilon(x, \lambda)$ as well as the strong monotonicity of the latter. This coupled with the stability of z^* allows us to rigorously relate z^* and z_ϵ .

Lemma 3: Let (A1)–(A3) hold. Then, the regularized mapping ϕ_ϵ is Lipschitz with constant $L(\epsilon) = L_\phi + \epsilon$ and strongly monotone with constant $\mu_\epsilon = \min\{\nu, \epsilon\}$.

Proof: We begin by proving the Lipschitzian nature of the mapping. We have

$$\begin{aligned} &\|\phi(z_1) - \phi(z_2)\| \\ &= \left\| \begin{pmatrix} \nabla f(x_1) - \nabla f(x_2) + \nabla d(x_1)^T \lambda_1 - \nabla d(x_2)^T \lambda_2 \\ -d(x_1) + d(x_2) + \epsilon(\lambda_1 - \lambda_2) \end{pmatrix} \right\| \\ &= \|\phi(z_1) - \phi(z_2) + \epsilon(\lambda_1 - \lambda_2)\| \\ &\leq L_\phi \|z_1 - z_2\| + \epsilon \|\lambda_1 - \lambda_2\| \\ &\leq L(\epsilon) \|z_1 - z_2\|. \end{aligned}$$

We now consider the strong monotonicity of ϕ_ϵ . We have

$$\nabla \phi_\epsilon(x, \lambda) = \begin{pmatrix} H(x, \lambda) & \nabla_x d(x)^T \\ -\nabla_x d(x) & \epsilon I \end{pmatrix},$$

⁴See for example [3], volume 2, Chapter 12.

where $H(x, \lambda) = \nabla_x \phi_x(x, \lambda)$, which is equal to $\nabla_x^2 f(x) + \sum_{j=1}^m \lambda_j \nabla_x^2 d_j(x)$.

Note that by (A1), $\nabla f(x)$ is strongly monotone. Since d_j are convex differentiable functions, it follows that ∇d_j are monotone for all j , implying that $u^T H(x, \lambda) u \geq \nu \|u\|^2$ for all $u \in \mathbb{R}^n$. Thus, for $u \in \mathbb{R}^n$ and $\gamma \in \mathbb{R}^m$,

$$\begin{aligned} \begin{pmatrix} u \\ \gamma \end{pmatrix}^T \nabla \phi_\epsilon(x, \lambda) \begin{pmatrix} u \\ \gamma \end{pmatrix} &= u^T H(x, \lambda) u + \epsilon \|\gamma\|^2 \\ &\geq \nu \|u\|^2 + \epsilon \|\gamma\|^2. \end{aligned}$$

Hence, ϕ_ϵ is strongly monotone with $\mu_\epsilon = \min\{\nu, \epsilon\}$. ■

We note that even when (A1) is relaxed we can obtain a similar result by regularizing in the primal and dual space.

A. Regularized Primal-Dual Method

We let z^k denote the primal-dual pair (x^k, λ^k) . We show that the regularized primal-dual method given by (12) converges to a point $z_\epsilon \in X \times \mathbb{R}_+^m$ with a geometric rate.

Lemma 4: Let (A1)–(A3) hold and let z_ϵ be the solution to the regularized problem. Let $\{z^k\}$ be generated by (12) using a constant stepsize $\tau_k = \tau$ for all $k \geq 0$, with $\tau < \frac{2\mu_\epsilon}{L(\epsilon)^2}$. Then

$$\|z^{k+1} - z_\epsilon\| \leq \sqrt{q_\epsilon} \|z^k - z_\epsilon\| \quad \text{for all } k \geq 0,$$

where $q_\epsilon = 1 + \tau^2 L(\epsilon)^2 - 2\tau\mu_\epsilon$, and $L(\epsilon)$ and μ_ϵ are as in Lemma 3.

Proof: Follows directly from Lemma 1. ■

A solution x^* of $\text{VI}(X, \phi)$ is stable if there exist positive scalars c and ϵ and a neighborhood \mathcal{D} such that for every $G \in B(\phi; \epsilon, X)^5$ and $x \in \text{SOL}(X, G) \cap \mathcal{D}$, $\|x - x^*\| \leq c \|\phi(v) - G(v)\|$ and $\text{SOL}(X, \phi) \cap \mathcal{D} = \{x^*\}$ (see e.g., [3], Definition 5.3.1).

We provide an error bound based on stability of z^* .

Proposition 4: Let (A1)–(A3) hold, and assume that z^* is a stable primal-dual solution. Then, there exists a $\beta > 0$ and a neighborhood \mathcal{D} of z^* such that for every $z_\epsilon \in \mathcal{D}$,

$$\frac{\|z_\epsilon - z^*\|}{\|z_\epsilon\|} \leq \beta \epsilon.$$

Proof: Based on assumption (A3), the mapping ϕ is Lipschitz. By Proposition 5.3.7 in [3], the Lipschitzian property coupled with the stability of the primal-dual equilibrium point z^* implies that there exists a $\beta > 0$ and a neighborhood \mathcal{D} such that for all $z \in \mathcal{D}$, we have

$$\|z^* - z_\epsilon\| \leq \beta \mathbf{F}_X^{\text{nat}}(z_\epsilon).$$

Let $\mathbf{F}_{X, \epsilon}^{\text{nat}}(z) \triangleq z - \Pi_X(z - \phi_\epsilon(z))$ denote the natural map of the perturbed map ϕ_ϵ . If z_ϵ is the solution of regularized problem, then $\mathbf{F}_{X, \epsilon}^{\text{nat}}(z_\epsilon) = 0$, implying that

$$\begin{aligned} \|\mathbf{F}_X^{\text{nat}}(z_\epsilon)\| &= \|\Pi_X(z_\epsilon - \phi(z_\epsilon)) - \Pi_X(z_\epsilon - \phi_\epsilon(z_\epsilon))\| \\ &= \|\phi(z_\epsilon) - \phi_\epsilon(z_\epsilon)\| \\ &\leq \epsilon \|z_\epsilon\|, \end{aligned}$$

which yields the desired bound. ■

⁵The set $B(g; \epsilon, X)$ is an ϵ -neighborhood of a given function g restricted to the set X , comprising of all continuous functions G such that $\sup_{y \in X} \|G(y) - g(y)\| < \epsilon$.

B. Independent Primal and Dual Stepsizes

In this section, we consider an algorithm in which there is less coordination between the primal and dual schemes. In particular, we consider a method which has independently chosen steplengths of the following form:

$$\begin{aligned} x^{k+1} &= \Pi_X(x^k - \alpha\phi_x(x^k, \lambda^k)), \\ \lambda^{k+1} &= \Pi_{\mathbb{R}_+^m}(\lambda^k - \tau(-d(x^k) + \epsilon\lambda^k)). \end{aligned} \quad (13)$$

We provide a relationship that holds for the sequence $\{z^k\}$ with $z^k = (x^k, \lambda^k)$ generated using (13) for the case when the primal stepsize α is greater than the dual stepsize τ .

Proposition 5: Let (A1)–(A3) hold. Let $\{z^k\}$ be a sequence generated by (13) with the primal and dual stepsizes such that $\alpha > \tau$ and $q_\epsilon < 1$, where $q_\epsilon = 1 + \alpha^2 L(\epsilon)^2 - 2\tau\mu_\epsilon + (\alpha - \tau)(1 - 2\nu + M_d^2)$. Then, we have

$$\|z^{k+1} - z_\epsilon\| \leq \sqrt{q_\epsilon} \|z^k - z_\epsilon\| \quad \text{for all } k \geq 0.$$

Proof: By the definition of the method we have,

$$\begin{aligned} \|x^{k+1} - x_\epsilon\|^2 &\leq \|x^k - x_\epsilon\|^2 + \alpha^2 \|\phi_x(x^k, \lambda^k) - \phi_x(x_\epsilon, \lambda_\epsilon)\|^2 \\ &\quad - 2\alpha(\phi_x(x^k, \lambda^k) - \phi_x(x_\epsilon, \lambda_\epsilon))^T(x^k - x_\epsilon), \end{aligned} \quad (14)$$

$$\begin{aligned} \|\lambda^{k+1} - \lambda_\epsilon\|^2 &\leq \|\lambda^k - \lambda_\epsilon\|^2 + \tau^2 \|-d(x^k) + \epsilon\lambda^k + d(x_\epsilon) - \epsilon\lambda_\epsilon\|^2 \\ &\quad - 2\tau(-d(x^k) + \epsilon\lambda^k + d(x_\epsilon) - \epsilon\lambda_\epsilon)^T(\lambda^k - \lambda_\epsilon). \end{aligned} \quad (15)$$

Adding (14) and (15), and using $z = (x, \lambda)$, $\|z\|^2 = \|x\|^2 + \|\lambda\|^2$, we obtain for $\alpha > \tau$,

$$\begin{aligned} \|z^{k+1} - z_\epsilon\|^2 &\leq \|z^k - z_\epsilon\|^2 + \alpha^2 \|\phi_\epsilon(z^k) - \phi_\epsilon(z_\epsilon)\|^2 \\ &\quad - 2\tau(\phi_\epsilon(z^k) - \phi_\epsilon(z_\epsilon))^T(z^k - z_\epsilon) \\ &\quad - 2(\alpha - \tau)(\phi_x(z^k) - \phi_x(z_\epsilon))^T(x^k - x_\epsilon), \end{aligned}$$

which through the use of Lipschitz and strong monotonicity property of ϕ_ϵ and adding and subtracting $\phi_x(x_\epsilon, \lambda^k)$ in the interaction term give us

$$\begin{aligned} \|z^{k+1} - z_\epsilon\|^2 &\leq (1 + \alpha^2 L^2(\epsilon) - 2\tau\mu_\epsilon) \|z^k - z_\epsilon\|^2 \\ &\quad - 2(\alpha - \tau)(\phi_x(x^k, \lambda^k) - \phi_x(x_\epsilon, \lambda^k))^T(x^k - x_\epsilon) \\ &\quad - 2(\alpha - \tau)(\phi_x(x_\epsilon, \lambda^k) - \phi_x(x_\epsilon, \lambda_\epsilon))^T(x^k - x_\epsilon). \end{aligned}$$

Using the strong monotonicity of ϕ_x , we obtain

$$\begin{aligned} \|z^{k+1} - z_\epsilon\|^2 &\leq (1 + \alpha^2 L^2(\epsilon) - 2\tau\mu_\epsilon) \|z^k - z_\epsilon\|^2 \\ &\quad - 2(\alpha - \tau)(\nu \|x^k - x_\epsilon\|^2 + (\nabla d(x_\epsilon))^T(\lambda^k - \lambda_\epsilon))^T(x^k - x_\epsilon). \end{aligned}$$

By using Cauchy-Schwartz's inequality, we have

$$\begin{aligned} \|z^{k+1} - z_\epsilon\|^2 &\leq (1 + \alpha^2 L^2(\epsilon) - 2\tau\mu_\epsilon) \|z^k - z_\epsilon\|^2 \\ &\quad + (\alpha - \tau)((1 - 2\nu) \|x^k - x_\epsilon\|^2 + \|\nabla d(x_\epsilon)\|^2 \|(\lambda^k - \lambda_\epsilon)\|^2). \end{aligned}$$

The result follows from $\|\nabla d(x)\| \leq M_d$ for all $x \in X$. ■
It can be seen that if $\alpha = \tau$, then the contraction result of Lemma 4 is obtained.

VI. SUMMARY AND CONCLUSIONS

This paper focuses on a class of multiuser optimization problems in which user interactions are seen either at the level of the objective or through a coupling constraint. Traditional algorithms rely on a high degree of separability and cannot be directly employed. Furthermore, much of the analysis is often contingent on strong convexity assumptions, requires coordination in terms of steplengths across users and rules out early termination. All of these are weakened to various degrees in the present work, which considers primal and primal-dual projection algorithms, derived from the fixed-point formulations of the solution. These schemes are analyzed in an effort to make rigorous statements regarding convergence behavior as well as provide error bounds in settings with limited coordination and premature termination.

VII. ACKNOWLEDGMENTS

The authors gratefully acknowledge the reviewers' comments.

REFERENCES

- [1] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and distributed computation: Numerical methods*, Prentice Hall Inc., 1989.
- [2] M. Chiang, S.H. Low, A.R. Calderbank, and J.C. Doyle, *Layering as optimization decomposition: A mathematical theory of network architectures*, Proceedings of the IEEE **95** (2007), no. 1, 255–312.
- [3] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, vol. 1 and 2, Springer-Verlag Inc., New York, 2003.
- [4] F.P. Kelly, A.K. Maulloo, and D.K. Tan, *Rate control for communication networks: shadow prices, proportional fairness, and stability*, Journal of the Operational Research Society **49** (1998), 237–252.
- [5] 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.
- [6] A. Nedić and A. Ozdaglar, *Subgradient methods in network resource allocation: Rate analysis*, 42nd Annual Conference on Information Sciences and Systems, CISS, Princeton, 2008, pp. 1189–1194.
- [7] ———, *Approximate primal solutions and rate analysis for dual subgradient methods*, SIAM Journal on Optimization **19** (2009), no. 4, 1757–1780.
- [8] ———, *Subgradient methods for saddle-point problems*, Journal of Optimization Theory and Applications **142** (2009), no. 1, 205–228.
- [9] A.S. Nemirovskii and D.B. Yudin, *Cezare convergence of gradient method approximation of saddle points for convex-concave functions*, Doklady Akademii Nauk SSSR **239** (1978), 1056–1059.
- [10] B.T. Polyak, *Introduction to optimisation*, Optimization Software, Inc., New York, 1987.
- [11] A. Ruszczyński, *A merit function approach to the subgradient method with averaging*, Optimization Methods and Software **23** (2008), no. 1, 161–172.
- [12] S. Shakkottai and R. Srikant, *Network optimization and control*, Foundations and Trends in Networking **2** (2007), no. 3, 271–379.
- [13] R. Srikant, *Mathematics of Internet congestion control*, Birkhauser, 2004.
- [14] J. N. Tsitsiklis, *Problems in decentralized decision making and computation*, Ph.D. thesis, Massachusetts Institute of Technology, 1984.