MULTIUSER OPTIMIZATION: DISTRIBUTED ALGORITHMS AND ERROR ANALYSIS

JAYASH KOSHAL, ANGELIA NEDIĆ AND UDAY V. SHANBHAG*

Abstract. Traditionally, a *multiuser problem* is a constrained optimization problem characterized by a set of users, an objective given by a sum of user-specific utility functions, and a collection of linear constraints that couple the user decisions. The users do not share the information about their utilities, but do communicate values of their decision variables. The multiuser problem is to maximize the sum of the users-specific utility functions subject to the coupling constraints, while abiding by the informational requirements of each user. In this paper, we focus on generalizations of convex multiuser optimization problems where the objective and constraints are not separable by user and instead consider instances where user decisions are coupled, both in the objective and through nonlinear coupling constraints. To solve this problem, we consider the application of gradient-based distributed algorithms on an approximation of the multiuser problem. Such an approximation is obtained through a Tikhonov regularization and is equipped with estimates of the difference between the optimal function values of the original problem and its regularized counterpart. In the algorithmic development, we consider constant steplength primal-dual and dual schemes in which the iterate computations are distributed naturally across the users, i.e., each user updates its own decision only. Convergence in the primal-dual space is provided in limited coordination settings, which allows for differing steplengths across users as well as across the primal and dual space. We observe that a generalization of this result is also available when users choose their regularization parameters independently from a prescribed range. An alternative to primal-dual schemes can be found in dual schemes which are analyzed in regimes where approximate primal solutions are obtained through a fixed number of gradient steps. Per-iteration error bounds are provided in such regimes and extensions are provided to regimes where users independently choose their regularization parameters. Our results are supported by a case-study in which the proposed algorithms are applied to a multi-user problem arising in a congested traffic network.

1. Introduction.

This paper deals with generic forms of multiuser problems arising often in network resource management, such as rate allocation in communication networks [7, 9, 13, 22, 23, 25]. A multiuser problem is a constrained optimization problem associated with a finite set of N users (or players). Each user i has a convex cost function $f_i(x_i)$ that depends only on its decision vector x_i. The decision vectors x_i, $i = 1, \ldots, N$ are typically subject to a finite system of linear inequalities $a_j^T(x_1, \ldots, x_N) \le b_j$ for $j = 1, \ldots, m$, which couple the user decision variables. The multiuser problem is formulated as a convex minimization of the form

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{N} f_i(x_i) \\
\text{subject to} \quad & a_j^T(x_1, \ldots, x_N) \le b_j, \quad && j = 1, \ldots, m \\
& x_i \in X_i, && i = 1, \ldots, N,
\end{aligned}
\tag{1.1}
$$

where X_i is the set constraint on user i decision x_i (often X_i is a box constraint). In many applications, users are characterized by their payoff functions rather than cost functions, in which case the multiuser problem is a concave maximization problem. In multiuser optimization, the problem information is distributed. In particular, it is assumed that user i knows only its function $f_i(x_i)$ and the constraint set X_i. Furthermore, user i can modify only its own decision x_i but may observe the decisions

* 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 awards CMMI 0948905 and CCF-0728863. The authors are grateful to the two reviewers and the associate editor Prof. S. Zhang for their comments and suggestions, all of which have greatly improved the paper.

$(x_j)_{j\neq i}$ of the other users. In effect, every user can see the entire vector x. Finally, it is often desirable that the algorithmic parameters (such as regularization parameters and steplengths) be chosen with relative independence across users since it is often challenging to both mandate choices and enforce consistency across users.

The goal in multiuser optimization is to solve problem (1.1) in compliance with the distributed information structure of the problem. More specifically, our focus is on developing distributed algorithms that satisfy several properties: (1) Limited informational requirements: Any given user does not have access to the utilities or the constraints of other users; (2) Single timescale: Two-timescale schemes for solving monotone variational problems require updating regularization parameters at a slow timescale and obtaining regularized solutions at a fast timescale. Coordinating across timescales is challenging and our goal lies in developing *single timescale* schemes; and (3) Limited coordination of algorithm parameters: In truly large-scale networks, enforcing consistency across algorithm parameters is often challenging and ideally, one would like minimal coordination across users in specifying algorithm parameters.

Our interest is in first-order methods, as these methods have small overhead per iteration and they exhibit stable behavior in the presence of various sources of noise in the computations, as well as in the information exchange due to possibly noisy links in the underlying communication network over which the users communicate.

Prior work [7, 9, 13, 22, 23, 25] has largely focused on multiuser problem (1.1). Both primal, primal-dual and dual schemes are discussed typically in a continuous-time setting (except for [13] where dual discrete-time schemes are investigated). Both dual and primal-dual discrete-time (approximate) schemes, combined with simple averaging, have been recently studied in [14–16] for a general convex constrained formulation. All of the aforementioned work establishes the convergence properties of therein proposed algorithms under the assumption that the users coordinate their steplengths, i.e., the steplength values are equal across all users.

This paper generalizes the standard multiuser optimization problem, defined in (1.1), in two distinct ways: (1) The user objectives are coupled by a congestion metric (as opposed to being separable). Specifically, the objective in (1.1) is replaced by a system cost given by $\sum_{i=1}^{N} f_i(x_i) + c(x_1, \ldots, x_N)$, with a convex coupling cost $c(x_1, \ldots, x_N)$; and (2) The linear inequalities in (1.1) are replaced with general convex inequalities. In effect, the constraints are nonlinear and not necessarily separable by user decisions.

To handle these generalizations of the multiuser problem, we propose approximating the problems with their regularized counterparts and, then, solving the regularized problems in a distributed fashion in compliance with the user specific information (user functions and decision variables). We provide an error estimate for the difference between the optimal function values of the original and the regularized problems. For solving the regularized problems, we consider distributed primal-dual and dual approaches, including those requiring inexact solutions of Lagrangian subproblems. We investigate the convergence properties and provide error bounds for these algorithms using two different assumptions on the steplengths, namely that the steplengths are the same across all users and the steplengths differ across different users. These results are extended to regimes where the users may select their regularization parameters from a broadcasted range.

The work in this paper is closely related to the distributed algorithms in [5, 27] and the more recent work on shared-constraint games [29, 30], where several classes of problems with the structures admitting decentralized computations are addressed.

However, the algorithms in the aforementioned work hinge on equal steplengths for all users and exact solutions for their success. In most networked settings, these requirements fail to hold, thus complicating the application of these schemes. Furthermore, due to the computational complexity of obtaining exact solutions for large scale problems, one is often more interested in a good approximate solution (with a provable error bound) rather than an exact solution.

Related is also the literature on centralized projection-based methods for optimization (see for example books [3,8,21]) and variational inequalities [8,10–12,20,24]. Recently, efficient projection-based algorithms have been developed in [1,2,18,26] for optimization, and in [17,19] for variational inequalities. The algorithms therein are all well suited for distributed implementations subject to some minor restrictions such as choosing Bregman functions that are separable across users' decision variables. The aforementioned algorithms will preserve their efficiency as long as the stepsize values are the same for all users. When the users are allowed to select their stepsizes within a certain range, there may be some efficiency loss. By viewing the stepsize variations as a source of noise, the work in this paper may be considered as an initial step into exploring the effects of "noisy" stepsizes on the performance of first-order algorithms, starting with simple first-order algorithms which are known to be stable under noisy data.

A final note is in order regarding certain terms that we use throughout the paper. The term "error analysis" pertains to the development of bounds on the difference between a given solution or function value and its optimal counterpart. The term "coordination" assumes relevance in distributed schemes where certain algorithmic parameters may need to satisfy a prescribed requirement across all users. Finally, it is worth accentuating why our work assumes relevance in implementing distributed algorithms in practical settings. In large-scale networks, the success of standard distributed implementations is often contingent on a series of factors. For instance, convergence often requires that steplengths match across users, exact/inexact solutions are available in bounded time intervals and finally, users have access to recent updates by the other network participants. In practice, algorithms may not subscribe to these restrictions and one may be unable to specify the choice of algorithm parameters, such as steplengths and regularization parameters, across users. Accordingly, we extend standard fixed-steplength gradient methods to allow for heterogeneous steplengths and diversity in regularization parameters.

The paper is organized as follows. In Section 2, we describe the problem of interest, motivate it through an example and recap the related fixed-point problem. In Section 3, we propose a regularized primal-dual method to allow for more general coupling among the constraints. Our analysis is equipped with error bounds when step-sizes and regularization parameters differ across users. Dual schemes are discussed in Section 4, where error bounds are provided for the case when inexact primal solutions are used. The behavior of the proposed methods is examined for a multiuser traffic problem in Section 5. We conclude in Section 6.

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., $\Pi_X(x) \triangleq \operatorname{argmin}_{z\in X} \|x - z\|$.

2. Problem Formulation. Traditionally, the multiuser problem (1.1) is characterized by a coupling of the user decision variables only through a separable objective function. In general, however, the objective need not be separable and the user decisions may be jointly constrained by a set of convex constraints. We consider a generalization to the canonical multiuser optimization problem of the following form:

$$\begin{aligned} \text{minimize} \quad & f(x) \triangleq \sum_{i=1}^{N} f_i(x_i) + c(x) \\ \text{subject to} \quad & d_j(x) \leq 0 \qquad \text{for all } j = 1, \ldots, m, \\ & x_i \in X_i \qquad \text{for all } i = 1, \ldots, N, \end{aligned} \tag{2.1}$$

where N is the number of users, $f_i(x_i)$ is user i cost function depending on a decision vector $x_i \in \mathbb{R}^{n_i}$ and $X_i \subseteq \mathbb{R}^{n_i}$ is the constraint set for user i. The function $c(x)$ is a joint cost that depends on the user decisions, i.e., $x = (x_1, \ldots, x_N) \in \mathbb{R}^n$, where $n = \sum_{i=1}^{N} n_i$. The functions $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}$ are *convex and continuously differentiable*. Further, we assume that $d_j : \mathbb{R}^n \to \mathbb{R}$ is a *continuously differentiable convex function* for every j. Often, when convenient, we will write the inequality constraints $d_j(x) \leq 0$, $j = 1, \ldots, m$, compactly as $d(x) \leq 0$ with $d(x) = (d_1(x), \ldots, d_m(x))^T$. Similarly, we use $\nabla d(x)$ to denote the vector of gradients $\nabla d_j(x)$, $j = 1, \ldots, m$, i.e., $\nabla d(x) = (\nabla d_1(x), \ldots, \nabla d_m(x))^T$. The user constraint sets X_i are assumed to be *nonempty, convex and closed*. We denote by f^* and X^*, respectively, the optimal value and the optimal solution set of this problem.

Before proceeding, we motivate the problem of interest via an example drawn from communication networks [9, 25], which can capture a host of other problems (such as in traffic or transportation networks).

Example 1. *Consider a network (see Fig 2.1) with a set of J link constraints and b_j being the finite capacity of link j, for $j \in J$. Let R be a set of user-specific routes, and let A be the associated link-route incidence matrix, i.e., $A_{jr} = 1$ if $j \in r$ implying that link j is traversed on route r, and $A_{jr} = 0$ otherwise.*

Suppose, the rth user has an associated route r and a rate allocation (flow) denoted by x_r. The corresponding utility of such a rate is given by $U_r(x_r)$. Assume further that utilities are additive implying that total utility is merely given by $\sum_{r \in R} U_r(x_r)$. Further, let $c(x)$ represent the congestion cost arising from using the same linkages in a route. Under this model the system optimal rates solve the following problem.

$$\begin{aligned} \textit{maximize} \quad & \sum_{r \in R} U_r(x_r) - c(x) \\ \textit{subject to} \quad & Ax \leq b, \quad x \geq 0. \end{aligned} \tag{2.2}$$

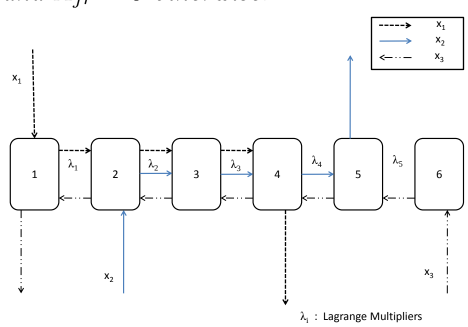


Fig. 2.1. *A network with 3 users and 5 links.*

At first glance, under suitable *concavity* assumptions on the utility functions and congestion cost, problem (2.2) is tractable from the standpoint of a centralized algorithm. However, if the utilities are not common knowledge, then such centralized schemes cannot be employed; instead our focus turns to developing distributed iterative schemes that respect the informational restrictions imposed by the application.

We are interested in algorithms aimed at solving system optimization problem (2.1) by each user executing computations only in the space of its own decision variables. Our approach is based on casting the system optimization problem as a fixed point problem through the variational inequality framework. Toward this goal, we let

$$\mathcal{L}(x,\lambda) = f(x) + \lambda^T d(x), \qquad X = X_1 \times X_2 \times \cdots \times X_N.$$

We also write $x = (x_i; x_{-i})$ with $x_{-i} = (x_j)_{j\neq i}$ to denote a vector where x_i is viewed as variable and x_{-i} is viewed as parameter. We let $\mathbb{R}^m_+$ denote the nonnegative orthant in $\mathbb{R}^m$. Under suitable strong duality conditions, from the first-order optimality conditions and the decomposable structure of X it can be seen that $(x^*,\lambda^*) \in X \times \mathbb{R}^m_+$ is a solution to (2.1) if and only x_i^* solves the parameterized variational inequalities $\mathrm{VI}(X_i, \nabla_{x_i}\mathcal{L}(x_i; x^*_{-i}, \lambda^*))$, $i = 1,\ldots,N$, and λ^* solves $\mathrm{VI}(\mathbb{R}^+_m, -\nabla_\lambda \mathcal{L}(x^*,\lambda))$. A vector (x^*,λ^*) solves $\mathrm{VI}(X_i, \nabla_{x_i}\mathcal{L}(x_i; x^*_{-i}, \lambda^*))$, $i = 1,\ldots,N$ and $\mathrm{VI}(\mathbb{R}^m_+, -\nabla_\lambda \mathcal{L}(x^*,\lambda))$ if and only if each x_i^* is a zero of the parameterized natural map[1] $\mathbf{F}^{\mathrm{nat}}_{X_i}(x_i; x^*_{-i}, \lambda^*) = 0$, for $i = 1,\ldots,N$, and λ^* is a zero of the parameterized natural map $\mathbf{F}^{\mathrm{nat}}_{\mathbb{R}^m_+}(\lambda; x^*) = 0$, i.e.,

$$\begin{aligned} \mathbf{F}^{\mathrm{nat}}_{X_i}(x_i; x^*_{-i}, \lambda^*) &\triangleq x_i - \Pi_{X_i}(x_i - \nabla_{x_i}\mathcal{L}(x_i; x^*_{-i}, \lambda^*)) \qquad \text{for } i = 1,\ldots,N, \\ \mathbf{F}^{\mathrm{nat}}_{\mathbb{R}^+_m}(\lambda; x^*) &\triangleq \lambda - \Pi_{\mathbb{R}^+_m}(\lambda + \nabla_\lambda \mathcal{L}(x^*,\lambda)), \end{aligned}$$

where $\nabla_x \mathcal{L}(x,\lambda)$ and $\nabla_\lambda \mathcal{L}(x,\lambda)$ are, respectively, the gradients of the Lagrangian function with respect to x and λ. Equivalently, letting $x^* = (x_1^*,\ldots,x_N^*) \in X$, a solution to the original problem is given by a solution to the following system of nonsmooth equations:

$$\begin{aligned} x^* &= \Pi_X(x^* - \nabla_x \mathcal{L}(x^*,\lambda^*)), \\ \lambda^* &= \Pi_{\mathbb{R}^m_+}(\lambda^* + \nabla_\lambda \mathcal{L}(x^*,\lambda^*)). \end{aligned} \tag{2.3}$$

Thus, x^* solves problem (2.1) if and only if it is a solution to the system (2.3) for some $\lambda^* \geq 0$. This particular relation motivates our algorithmic development.

We now discuss the conditions that we use in the subsequent development. Specifically, we assume that the Slater condition holds for problem (2.1).

Assumption 1. *(Slater Condition) There exists a Slater vector $\bar{x} \in X$ such that $d_j(\bar{x}) < 0$ for all $j = 1,\ldots,m$.*

Under the Slater condition, the primal problem (2.1) and its dual have the same optimal value, and a dual optimal solution λ^* exists. When X is compact for example, the primal problem also has a solution x^*. A primal-dual optimal pair (x^*,λ^*) is also a solution to the coupled fixed-point problems in (2.3). For a more compact notation, we introduce the mapping $\Phi(x,\lambda)$ as

$$\Phi(x,\lambda) \triangleq (\nabla_x \mathcal{L}(x,\lambda), -\nabla_\lambda \mathcal{L}(x,\lambda)) = (\nabla_x \mathcal{L}(x,\lambda), -d(x)), \tag{2.4}$$

and we let $z = (x,\lambda)$. In this notation, the preceding coupled fixed-point problems are equivalent to a variational inequality requiring a vector $z^* = (x^*,\lambda^*) \in X \times \mathbb{R}^m_+$ such that

$$(z - z^*)^T \Phi(z^*) \geq 0 \quad \text{for all } z = (x,\lambda) \in X \times \mathbb{R}^m_+. \tag{2.5}$$

In the remainder of the paper, in the product space $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_N}$, we use $\|x\|$ and $x^T y$ to denote the Euclidean norm and the inner product that are induced,

[1] See [8], volume 1, 1.5.8 Proposition, page 83.

respectively, by the Euclidean norms and the inner products in the component spaces. Specifically, for $x = (x_1, \ldots, x_N)$ with $x_i \in \mathbb{R}^{n_i}$ for all i, we have

$$x^T y = \sum_{i=1}^{N} x_i^T y_i \qquad \text{and} \qquad \|x\| = \sqrt{\sum_{i=1}^{N} \|x_i\|^2}.$$

We now state our basic assumptions on the functions and the constraint sets in problem (2.1).

ASSUMPTION 2. *The set X is closed, convex, and bounded. The functions $f_i(x_i)$, $i = 1, \ldots, N$, and $c(x)$ are continuously differentiable and convex.*

Next, we define the gradient map

$$F(x) = \left(\nabla_{x_1}(f_1(x_1) + c(x))^T, \ldots, \nabla_{x_N}(f_N(x_N) + c(x))^T\right)^T,$$

for which we assume the following.

ASSUMPTION 3. *The gradient map $F(x)$ is Lipschitz continuous with constant L over the set X, i.e.,*

$$\|F(x) - F(y)\| \leq L\|x - y\| \qquad \text{for all } x, y \in X.$$

3. A Regularized Primal-Dual Method. In this section, we present a distributed gradient-based method that employs a fixed regularization in the primal and dual space. We begin by discussing the regularized problem in Section 3.1 and proceed to provide bounds on the error in Section 3.2. In Section 3.3, we examine the monotonicity and Lipschitzian properties of the regularized mapping and develop the main convergence result of this section in Section 3.4. Notably, the theoretical results in Section 3.4 prescribe a set from which users may independently select steplengths with no impact on the overall convergence of the scheme. Finally, in Section 3.5, we further weaken the informational restrictions of the scheme by allowing users to select regularization parameters from a broadcasted range, and we extend the Lipschitzian bounds and convergence rates to this regime.

3.1. Regularization. For approximately solving the variational inequality (2.5), we consider its regularized counterpart obtained by regularizing the Lagrangian in both primal and dual space. In particular, for $\nu > 0$ and $\epsilon > 0$, we let $\mathcal{L}_{\nu,\epsilon}$ denote the regularized Lagrangian, given by

$$\mathcal{L}_{\nu,\epsilon}(x, \lambda) = f(x) + \frac{\nu}{2}\|x\|^2 + \lambda^T d(x) - \frac{\epsilon}{2}\|\lambda\|^2. \tag{3.1}$$

The regularized variational inequality requires determining a vector $z^*_{\nu,\epsilon} = (x^*_{\nu,\epsilon}, \lambda^*_{\nu,\epsilon}) \in X \times \mathbb{R}^m_+$ such that

$$(z - z^*_{\nu,\epsilon})^T \Phi_{\nu,\epsilon}(z^*_{\nu,\epsilon}) \geq 0 \quad \text{for all } z = (x, \lambda) \in X \times \mathbb{R}^m_+, \tag{3.2}$$

where the regularized mapping $\Phi_{\nu,\epsilon}(x, \lambda)$ is given by

$$\Phi_{\nu,\epsilon}(x, \lambda) \triangleq (\nabla_x \mathcal{L}_{\nu,\epsilon}(x, \lambda), -\nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x, \lambda)) = (\nabla_x \mathcal{L}(x, \lambda) + \nu x, -d(x) + \epsilon\lambda). \tag{3.3}$$

The gradient map $\nabla_x \mathcal{L}_{\nu,\epsilon}(x, \lambda)$ is given by

$$\nabla_x \mathcal{L}_{\nu,\epsilon}(x, \lambda) \triangleq (\nabla_{x_1} \mathcal{L}_{\nu,\epsilon}(x, \lambda), \ldots, \nabla_{x_N} \mathcal{L}_{\nu,\epsilon}(x, \lambda))$$

where $\nabla_{x_i}\mathcal{L}_{\nu,\epsilon}(x,\lambda) = \nabla_{x_i}\left(f(x) + \lambda^T d(x)\right) + \nu x_i$. It is known that, under some conditions, the unique solutions $z^*_{\nu,\epsilon}$ of the variational inequality in (3.2) converge, as $\nu \to 0$ and $\epsilon \to 0$, to the smallest norm solution of the original variational inequality in (2.5) (see [8], Section 12.2). We, however, want to investigate approximate solutions and estimate the errors resulting from solving a regularized problem instead of the original problem, while the regularization parameters are kept fixed at some values.

To solve the variational inequality (3.2), one option lies in considering projection schemes for monotone variational inequalities (see Chapter 12 in [8]). However, the lack of Lipschitz continuity of the mapping precludes a direct application of these schemes. In fact, the Lipschitz continuity of $\Phi_{\nu,\epsilon}(z)$ cannot even be proved when the functions f and d_j have Lipschitz continuous gradients. In proving the Lipschitzian property, we observe that the boundedness of the multipliers cannot be assumed in general. However, the "bounding of multipliers λ" may be achieved under the Slater regularity condition. In particular, the Slater condition can be used to provide a compact convex region containing all the dual optimal solutions. Replacing $\mathbb{R}^m_+$ with such a compact convex set results in a variational inequality that is equivalent to (3.2),

Determining a compact set containing the dual optimal solutions can be accomplished by viewing the regularized Lagrangian $\mathcal{L}_{\nu,\epsilon}$ as a result of two-step regularization: we first regularize the original primal problem (2.1), and then we regularize its Lagrangian function. Specifically, for $\nu > 0$, the regularized problem (2.1) is given by

$$\begin{aligned} &\text{minimize} \;\; f_\nu(x) \triangleq \sum_{i=1}^{N}\left(f_i(x_i) + \frac{\nu}{2}\|x_i\|^2\right) + c(x) \\ &\text{subject to } d_j(x) \le 0 \quad \text{for all } j = 1,\dots,m, \\ &\qquad\qquad x_i \in X_i \quad \text{for all } i = 1,\dots,N. \end{aligned} \tag{3.4}$$

Its Lagrangian function is

$$\mathcal{L}_\nu(x,\lambda) = f(x) + \frac{\nu}{2}\|x\|^2 + \lambda^T d(x) \qquad \text{for all } x \in X,\; \lambda \ge 0, \tag{3.5}$$

and its corresponding dual problem is

$$\begin{aligned} &\text{maximize} \;\; v_\nu(\lambda) \triangleq \min_{x\in X}\mathcal{L}_\nu(x,\lambda) \\ &\text{subject to } \lambda \ge 0. \end{aligned}$$

We use v^*_ν to denote the optimal value of the dual problem, i.e., $v^*_\nu = \max_{\lambda\ge 0} v_\nu(\lambda)$, and we use Λ^*_ν to denote the set of optimal dual solutions. For $\nu = 0$, the value v^*_0 is the optimal dual value of the original problem (2.1) and Λ^*_0 is the set of the optimal dual solutions of its dual problem.

Under the Slater condition, for every $\nu > 0$, the solution x^*_ν to problem (3.4) exists and therefore strong duality holds [4]. In particular, the optimal values of problem (3.4) and its dual are equal, i.e., $f(x^*_\nu) = v^*_\nu$, and the dual optimal set Λ^*_ν is nonempty and bounded [28]. Specifically, we have

$$\Lambda^*_\nu \subseteq \left\{\lambda \in \mathbb{R}^m \;\Big|\; \sum_{j=1}^{m}\lambda_j \le \frac{f(\bar{x}) + \frac{\nu}{2}\|\bar{x}\|^2 - v^*_\nu}{\min_{1\le j\le m}\{-d_j(\bar{x})\}},\; \lambda \ge 0\right\} \qquad \text{for all } \nu > 0.$$

When the Slater condition holds and the optimal value f^* of the original problem (2.1) is finite, the strong duality holds for that problem as well, and therefore, the preceding relation also holds for $\nu = 0$, with v_0^* being the optimal value of the dual problem for (2.1). In this case, we have $f^* = v_0^*$, while for any $\nu > 0$, we have $v_\nu^* = f(x_\nu^*)$ for a solution x_ν^* of the regularized problem (3.4). Since $f(x_\nu^*) \geq f^*$, it follows that $v_\nu^* \geq v_0^*$ for all $\nu \geq 0$, and therefore,

$$\Lambda_\nu^* \subseteq \left\{ \lambda \in \mathbb{R}^m \;\Big|\; \sum_{j=1}^m \lambda_j \leq \frac{f(\bar{x}) + \frac{\nu}{2}\,\|\bar{x}\|^2 - v_0^*}{\min_{1\leq j\leq m}\{-d_j(\bar{x})\}},\ \lambda \geq 0 \right\} \qquad \text{for all } \nu \geq 0,$$

where the set Λ_0^* is the set of dual optimal solutions for the original problem (2.1). Noting that a larger set on the right hand side can be obtained by replacing v_0^* with any lower-bound estimate of v_0^* [i.e., $v(\bar{\lambda})$ for some $\bar{\lambda} \geq 0$], we can define a compact convex set $\mathcal{D}_\nu$ for every $\nu \geq 0$, as follows:

$$\mathcal{D}_\nu = \left\{ \lambda \in \mathbb{R}^m \;\Big|\; \sum_{j=1}^m \lambda_j \leq \frac{f(\bar{x}) + \frac{\nu}{2}\|\bar{x}\|^2 - v(\bar{\lambda})}{\min_{1\leq j\leq m}\{-d_j(\bar{x})\}},\ \lambda \geq 0 \right\} \qquad \text{for every } \nu \geq 0, \quad (3.6)$$

which satisfies

$$\Lambda_\nu^* \subset \mathcal{D}_\nu, \qquad \text{for every } \nu \geq 0. \qquad (3.7)$$

Observe that $v_0^* \leq v_\nu^* \leq v_{\nu'}$ for $0 \leq \nu \leq \nu'$, implying that $\mathcal{D}_0 \subseteq \mathcal{D}_\nu \subseteq \mathcal{D}_{\nu'}$. Therefore, the compact sets $\mathcal{D}_\nu$ are nested, and their intersection is a nonempty compact set $\mathcal{D}$ which contains the optimal dual solutions Λ_0^* of the original problem.

In the rest of the paper, we will assume that the Slater condition holds and the set X is compact (Assumption 2), so that the construction of such nested compact sets is possible. Specifically, we will *assume that a family of nested compact convex sets $\mathcal{D}_\nu \subset \mathbb{R}_+^m$, $\nu \geq 0$, satisfying relation (3.7) has already been determined.* In this case, the variational inequality of determining $z_{\nu,\epsilon} = (x_{\nu,\epsilon}, \lambda_{\nu,\epsilon}) \in X \times \mathcal{D}_\nu$ such that

$$(z - z_{\nu,\epsilon})^T \Phi_{\nu,\epsilon}(z_{\nu,\epsilon}) \geq 0 \qquad \text{for all } z = (x, \lambda) \in X \times \mathcal{D}_\nu, \qquad (3.8)$$

has the same solution set as the variational inequality in (3.2), where λ is constrained to lie in the nonnegative orthant.

3.2. Regularization error. We now provide an upper bound on the distances between $x_{\nu,\epsilon}$ and x_ν^*. Here, $x_{\nu,\epsilon}$ is the primal component of $z_{\nu,\epsilon}$, the solution of the variational inequality in (3.8) and x_ν^* is the solution of the regularized problem in (3.4) for given positive parameters ν and ϵ.

PROPOSITION 3.1. *Let Assumption 2 hold except for the boundedness of X. Also, let Assumption 1 hold. Then, for any $\nu > 0$ and $\epsilon > 0$, for the solution $z_{\nu,\epsilon} = (x_{\nu,\epsilon}, \lambda_{\nu,\epsilon})$ of variational inequality (3.8), we have*

$$\nu\|x_\nu^* - x_{\nu,\epsilon}\|^2 + \frac{\epsilon}{2}\|\lambda_{\nu,\epsilon}\|^2 \leq \frac{\epsilon}{2}\|\lambda_\nu^*\|^2 \qquad \textit{for all } \lambda_\nu^* \in \Lambda_\nu^*,$$

where x_ν^* is the optimal solution of the regularized problem (3.4) and Λ_ν^* is the set of optimal solutions of its corresponding dual problem.

Proof. The existence of a unique solution $x_\nu^* \in X$ of problem (3.4) follows from the continuity and strong convexity of f_ν. Also, by the Slater condition, the dual optimal set Λ_ν^* is nonempty. In what follows, let $\lambda_\nu^* \in \Lambda_\nu^*$ be an arbitrary but fixed dual optimal solution for problem (3.4). To make the notation simpler, we use ξ to denote the pair of regularization parameters (ν,ϵ), i.e., $\xi=(\nu,\epsilon)$. When the interplay between the parameters is relevant, we will write them explicitly.

From the definition of the mapping Φ_ξ it follows that the solution $z_\xi=(x_\xi,\lambda_\xi)\in X\times\mathcal{D}_\nu$ is a saddle-point for the regularized Lagrangian function $\mathcal{L}_\xi(x,\lambda)=\mathcal{L}(x,\lambda)+\frac{\nu}{2}\|x\|^2-\frac{\epsilon}{2}\|\lambda\|^2$, i.e.,

$$\mathcal{L}_\xi(x_\xi,\lambda)\le\mathcal{L}_\xi(x_\xi,\lambda_\xi)\le\mathcal{L}_\xi(x,\lambda_\xi)\qquad\text{for all } x\in X \text{ and } \lambda\in\mathcal{D}_\nu. \tag{3.9}$$

Recalling that $\Lambda_\nu^*\subseteq\mathcal{D}_\nu$, and by letting $\lambda=\lambda_\nu^*$ in the first inequality of the preceding relation, we obtain

$$0\le\mathcal{L}_\xi(x_\xi,\lambda_\xi)-\mathcal{L}_\xi(x_\xi,\lambda_\nu^*)=(\lambda_\xi-\lambda_\nu^*)^Td(x_\xi)-\frac{\epsilon}{2}\|\lambda_\xi\|^2+\frac{\epsilon}{2}\|\lambda_\nu^*\|^2. \tag{3.10}$$

We now estimate the term $(\lambda_\xi-\lambda_\nu^*)^Td(x_\xi)=\sum_{j=1}^m(\lambda_{\xi,j}-\lambda_{\nu,j}^*)d_j(x_\xi)$ by considering the individual terms, where $\lambda_{\nu,j}^*$ is the j-th component of λ_ν^*. By convexity of each d_j, we have

$$d_j(x_\xi)\le d_j(x_\nu^*)+\nabla d_j(x_\xi)^T(x_\xi-x_\nu^*)\le\nabla d_j(x_\xi)^T(x_\xi-x_\nu^*),$$

where the last inequality follows from x_ν^* being a solution to the primal regularized problem (hence, $d_j(x_\nu^*)\le 0$ for all j). By multiplying the preceding inequality with $\lambda_{\xi,j}$ (which is nonnegative) and by adding over all j, we obtain

$$\sum_{j=1}^m\lambda_{\xi,j}d_j(x_\xi)\le\sum_{j=1}^m\lambda_{\xi,j}\nabla d_j(x_\xi)^T(x_\xi-x_\nu^*).$$

By the definition of the regularized Lagrangian $\mathcal{L}_\xi(x,\lambda)$, we have

$$\begin{aligned}\sum_{j=1}^m\lambda_{\xi,j}\nabla d_j(x_\xi)^T(x_\xi-x_\nu^*)&=\nabla_x\mathcal{L}_\xi(x_\xi,\lambda_\xi)^T(x_\xi-x_\nu^*)-\left(\nabla f(x_\xi)+\nu x_\xi\right)^T(x_\xi-x_\nu^*)\\&\le-\left(\nabla f(x_\xi)+\nu x_\xi\right)^T(x_\xi-x_\nu^*),\end{aligned}$$

where the inequality follows from $\nabla_x\mathcal{L}_\xi(x_\xi,\lambda_\xi)^T(x_\xi-x_\nu^*)\le 0$, which holds in view of the second inequality in saddle-point relation (3.9) with $x=x_\nu^*\in X$. Therefore, by combining the preceding two relations, we obtain

$$\sum_{j=1}^m\lambda_{\xi,j}d_j(x_\xi)\le-\left(\nabla f(x_\xi)+\nu x_\xi\right)^T(x_\xi-x_\nu^*). \tag{3.11}$$

By convexity of each d_j, we have $d_j(x_\xi)\ge d_j(x_\nu^*)+\nabla d_j(x_\nu^*)^T(x_\xi-x_\nu^*)$ By multiplying the preceding inequality with $-\lambda_{\nu,j}^*$ (which is non-positive) and by adding over all j, we obtain

$$\begin{aligned}-\sum_{j=1}^m\lambda_{\nu,j}^*d_j(x_\xi)&\le-\sum_{j=1}^m\lambda_{\nu,j}^*d_j(x_\nu^*)-\sum_{j=1}^m\lambda_{\nu,j}^*\nabla d_j(x_\nu^*)^T(x_\xi-x_\nu^*)\\&=\sum_{j=1}^m\lambda_{\nu,j}^*\nabla d_j(x_\nu^*)^T(x_\nu^*-x_\xi),\end{aligned}$$

where the equality follows from $(\lambda_\nu^*)^T d(x_\nu^*) = 0$, which holds by the complementarity slackness of the primal-dual pair (x_ν^*, λ_ν^*) of the regularized problem (3.4). Using the definition of the Lagrangian function $\mathcal{L}_\nu$ in (3.5) for the problem (3.4), we have

$$\begin{aligned}\sum_{j=1}^m \lambda_{\nu,j}^* \nabla d_j(x_\nu^*)^T (x_\nu^* - x_\xi) &= \nabla_x \mathcal{L}_\nu(x_\nu^*, \lambda_\nu^*)^T (x_\nu^* - x_\xi) - (\nabla f(x_\nu^*) + \nu x_\nu^*)^T (x_\nu^* - x_\xi) \\ &\leq -(\nabla f(x_\nu^*) + \nu x_\nu^*)^T (x_\nu^* - x_\xi),\end{aligned}$$

where the inequality follows from relation $\nabla_x \mathcal{L}(x_\nu^*, \lambda_\nu^*)^T (x_\nu^* - x_\xi) \leq 0$, which in turn holds since (x_ν^*, λ_ν^*) is a saddle-point of the Lagrangian function $\mathcal{L}_\nu(x, \lambda)$ over $X \times \mathcal{D}_\nu$ and $x_\xi \in X$. Combining the preceding two relations, we obtain

$$-\sum_{j=1}^m \lambda_{\nu,j}^* d_j(x_\xi) \leq -(\nabla f(x_\nu^*) + \nu x_\nu^*)^T (x_\nu^* - x_\xi) = (\nabla f(x_\nu^*) + \nu x_\nu^*)^T (x_\xi - x_\nu^*).$$

The preceding relation and inequality (3.11), yield

$$(\lambda_\xi - \lambda_\nu^*)^T d(x_\xi) = \sum_{j=1}^m (\lambda_{\xi,j} - \lambda_{\nu,j}^*) d_j(x_\xi) \leq (\nabla f(x_\nu^*) - \nabla f(x_\xi))^T (x_\xi - x_\nu^*) - \nu \|x_\xi - x_\nu^*\|^2.$$

From the monotonicity of ∇f, we have $(\nabla f(x_\nu^*) - \nabla f(x_\xi))^T (x_\xi - x_\nu^*) \leq 0$, thus implying $(\lambda_\xi - \lambda_\nu^*)^T d(x_\xi) \leq -\nu \|x_\xi - x_\nu^*\|^2$. Finally, by combining the preceding relation with (3.10), and recalling notation $\xi = (\nu, \epsilon)$, we obtain for any solution x_ν^*,

$$\nu \|x_{\nu,\epsilon} - x_\nu^*\|^2 + \frac{\epsilon}{2} \|\lambda_{\nu,\epsilon}\|^2 \leq \frac{\epsilon}{2} \|\lambda_\nu^*\|^2 \qquad \text{for all } \lambda^* \in \Lambda_\nu^*, \tag{3.12}$$

thus showing the desired relation. □

As an immediate consequence of Proposition 3.1, in view of $\Lambda_\nu^* \subset \mathcal{D}_\nu$, we have

$$\|x_{\nu,\epsilon} - x_\nu^*\| \leq \sqrt{\frac{\epsilon}{2\nu}} \max_{\lambda^* \in \mathcal{D}_\nu} \|\lambda^*\| \qquad \text{for all } \nu > 0 \text{ and } \epsilon > 0. \tag{3.13}$$

This relation provides a bound on the distances of the solutions x_ν^* of problem (3.4) and the component $x_{\nu,\epsilon}$ of the solution $z_{\nu,\epsilon}$ of the regularized variational inequality in (3.8). The relation suggests that a ν larger than ϵ would yield a better error bound. Note, however, that increasing ν would correspond to the enlargement of the set $\mathcal{D}_\nu$, and therefore, increasing value for $\max_{\lambda^* \in \mathcal{D}_\nu} \|\lambda^*\|$. When the specific structure of the sets $\mathcal{D}_\nu$ is available, one may try to optimize the term $\sqrt{\frac{\epsilon}{2\nu}}\ \max_{\lambda^* \in \mathcal{D}_\nu} \|\lambda^*\|$ with respect to ν, while ϵ is kept fixed. In fact, the following result provides a simple result when $\mathcal{D}_\nu$ is specified using the Slater point $\bar{x}$.

Lemma 3.2. *Under the assumptions of Proposition 3.1, for a fixed $\epsilon > 0$, the tightest bound for $\|x_{\nu,\epsilon} - x_\nu^*\|$ is given by*

$$\|x_{\nu,\epsilon} - x_\nu^*\| \leq \left(\frac{\sqrt{\epsilon \left(f(\bar{x}) - v(\bar{\lambda}) \right)}}{\min_{1 \leq j \leq m} \{-d_j(\bar{x})\}} \|\bar{x}\| \right).$$

Proof. Using $\|x\|_2 \leq \|x\|_1$, from relation (3.13) we have

$$\|x_{\nu,\epsilon} - x_\nu^*\| \leq \sqrt{\frac{\epsilon}{2\nu}} \left(\max_{\lambda \in \mathcal{D}_\nu} \|\lambda\|_2 \right) \leq \sqrt{\frac{\epsilon}{2\nu}} \left(\max_{\lambda \in \mathcal{D}_\nu} \|\lambda\|_1 \right).$$

But by the structure of the set D_ν, we have that

$$\begin{aligned}\sqrt{\frac{\epsilon}{2\nu}}\left(\max_{\lambda\in\mathcal{D}_\nu}\|\lambda\|_1\right) &= \sqrt{\frac{\epsilon}{2\nu}}\left(\frac{f(\bar{x})+\frac{\nu}{2}\|\bar{x}\|^2-v(\bar{\lambda})}{\min_{1\le j\le m}\{-d_j(\bar{x})\}}\right)\\ &= \sqrt{\frac{\epsilon}{2\nu}}\left(\frac{f(\bar{x})-v(\bar{\lambda})}{\min_{1\le j\le m}\{-d_j(\bar{x})\}}\right)+\sqrt{\frac{\nu\epsilon}{2}}\left(\frac{\frac{1}{2}\|\bar{x}\|^2}{\min_{1\le j\le m}\{-d_j(\bar{x})\}}\right)\\ &= \frac{\sqrt{\epsilon}}{\sqrt{2}\,\min_{1\le j\le m}\{-d_j(\bar{x})\}}\left(\frac{a}{\sqrt{\nu}}+b\sqrt{\nu}\right),\end{aligned}$$

where $a = f(\bar{x}) - v(\bar{\lambda})$ and $b = \frac{1}{2}\|\bar{x}\|^2$. It can be seen that the function $h(\nu) = a/\sqrt{\nu}+b\sqrt{\nu}$ has a unique minimum at $\nu^* = \frac{a}{b}$, with the minimum value $h(\nu^*) = 2\sqrt{ab}$. Thus, we have

$$\sqrt{\frac{\epsilon}{2\nu}}\left(\max_{\lambda\in\mathcal{D}_\nu}\|\lambda\|_1\right) \le \left(\frac{\sqrt{\epsilon\left(f(\bar{x})-v(\bar{\lambda})\right)}}{\min_{1\le j\le m}\{-d_j(\bar{x})\}}\right)\|\bar{x}\|,$$

implying the desired estimate. □

When the set X is bounded, as another consequence of Proposition 3.1, we may obtain the error bounds on the sub-optimality of the vector $x_{\nu,\epsilon}$ by using the preceding error bound. Specifically, we can provide bounds on the violation of the primal inequality constraints $d_j(x) \le 0$ at $x = x_{\nu,\epsilon}$. Also, we can estimate the difference in the values $f(x_{\nu,\epsilon})$ and the primal optimal value f^* of the original problem (2.1). This is done in the following lemma.

Lemma 3.3. *Let Assumptions 2 and 1 hold. For any $\nu,\epsilon > 0$, we have*

$$\begin{aligned}&\max\{0, d_j(x_{\nu,\epsilon})\} \le M_{d_j} M_\nu\sqrt{\frac{\epsilon}{2\nu}} \qquad \text{for all } j = 1,\ldots,m,\\ &|f(x_{\nu,\epsilon}) - f(x^*)| \le M_f M_\nu\sqrt{\frac{\epsilon}{2\nu}} + \frac{\nu}{2}D^2,\end{aligned}$$

with $M_{d_j} = \max_{x\in X}\|\nabla d_j(x)\|$ for each j, $M_f = \max_{x\in X}\|\nabla f(x)\|$, $M_\nu = \max_{\lambda\in\mathcal{D}_\nu}\|\lambda\|$ and $D = \max_{x\in X}\|x\|$.

Proof. Let $\nu > 0$ and $\epsilon > 0$ be given, and let $j \in \{1,\ldots,m\}$ be arbitrary. Since d_j is convex, we have

$$d_j(x_{\nu,\epsilon}) \le d_j(x_\nu^*) + \nabla d_j(x_\nu^*)^T(x_{\nu,\epsilon} - x_\nu^*) \le \|\nabla d_j(x_\nu^*)\|\,\|x_{\nu,\epsilon} - x_\nu^*\|,$$

where in the last inequality we use $d_j(x_\nu^*) \le 0$, which holds since x_ν^* is the solution to the regularized primal problem (3.4). Since X is compact, the gradient norm $\|\nabla d_j(x)\|$ is bounded by some constant, say M_{d_j}. From this and the estimate

$$\|x_{\nu,\epsilon} - x_\nu^*\| \le \sqrt{\frac{\epsilon}{2\nu}}\,\|\lambda_\nu^*\|, \tag{3.14}$$

which follows by Proposition 3.1, we obtain

$$d_j(x_{\nu,\epsilon}) \le M_{d_j}\sqrt{\frac{\epsilon}{2\nu}}\,\|\lambda_\nu^*\|,$$

where λ_ν^* is a dual optimal solution of the regularized problem. Since the set of dual optimal solutions is contained in the compact set $\mathcal{D}_\nu$, the dual solutions are bounded. Thus, for the violation of the constraint $d_j(x) \le 0$, we have

$$\max\{0, d_j(x_{\nu,\epsilon})\} \le M_{d_j} M_\nu \sqrt{\frac{\epsilon}{2\nu}},$$

where $M_\nu = \max_{\lambda\in\mathcal{D}_\nu} \|\lambda\|$. Next, we estimate the difference $|f(x_{\nu,\epsilon}) - f(x^*)|$. We can write

$$|f(x_{\nu,\epsilon}) - f(x^*)| \le |f(x_{\nu,\epsilon}) - f(x_\nu^*)| + f(x_\nu) - f^*, \tag{3.15}$$

where we use $0 \le f(x_\nu^*) - f^*$. By convexity of f, we have

$$\nabla f(x_\nu^*)^T(x_{\nu,\epsilon} - x_\nu^*) \le f(x_{\nu,\epsilon}) - f(x_\nu^*) \le \nabla f(x_{\nu,\epsilon})^T(x_{\nu,\epsilon} - x_\nu^*).$$

Since $x_{\nu,\epsilon}$, $x^* \in X$ and X is compact, by the continuity of the gradient $\|\nabla f(x)\|$, the gradient norm is bounded over the set X, say by a scalar M_f, so that

$$|f(x_{\nu,\epsilon}) - f(x_\nu^*)| \le M_f\,\|x_{\nu,\epsilon} - x_\nu^*\|.$$

Using the estimate (3.14) and the boundedness of the dual optimal multipliers, similar to the preceding analysis, we obtain the following bound

$$|f(x_{\nu,\epsilon}) - f(x_\nu^*)| \le M_f M_\nu \sqrt{\frac{\epsilon}{2\nu}}.$$

By substituting the preceding relation in inequality (3.15), we obtain

$$|f(x_{\nu,\epsilon}) - f(x^*)| \le M_f M_\nu \sqrt{\frac{\epsilon}{2\nu}} + f(x_\nu) - f^*.$$

Further, by using the estimate $f(x_\nu^*) - f^* \le \frac{\nu}{2}\max_{x\in X}\|x\|^2 = \frac{\nu}{2}D^2$ of Lemma 7.1 (see appendix), we obtain the desired relation. □

Next, we discuss how one may specify ν and ϵ. Given a threshold error δ on the deviation of the obtained function value from its optimal counterpart, we have that $|f(x_{\nu,\epsilon}) - f(x^*)| < \delta$, if the following holds

$$M_f M_\nu \sqrt{\frac{\epsilon}{2\nu}} + \frac{\nu}{2}D^2 < \delta.$$

But by the structure of the set D_ν, we have that

$$\begin{aligned}
M_\nu &= \sqrt{\frac{\epsilon}{2\nu}}\left(\max_{\lambda\in\mathcal{D}_\nu}\|\lambda\|_1\right)\\
&= \sqrt{\frac{\epsilon}{2\nu}}\left(\frac{f(\bar{x}) + \frac{\nu}{2}\|\bar{x}\|^2 - v(\bar{\lambda})}{\min_{1\le j\le m}\{-d_j(\bar{x})\}}\right)\\
&= \sqrt{\frac{\epsilon}{2\nu}}\left(\frac{f(\bar{x}) - v(\bar{\lambda})}{\min_{1\le j\le m}\{-d_j(\bar{x})\}}\right) + \sqrt{\frac{\nu\epsilon}{2}}\left(\frac{\frac{\|\bar{x}\|^2}{2}}{\min_{1\le j\le m}\{-d_j(\bar{x})\}}\right)\\
&= \frac{1}{\sqrt{2}\,\min_{1\le j\le m}\{-d_j(\bar{x})\}}\left(\frac{a\sqrt{\epsilon}}{\sqrt{\nu}} + b\sqrt{\epsilon\nu}\right),
\end{aligned}$$

where $a = f(\bar{x}) - v(\bar{\lambda})$ and $b = \frac{\|\bar{x}\|^2}{2}$. Thus, we have

$$M_f M_\nu \sqrt{\frac{\epsilon}{2\nu}} + \frac{\nu}{2} D^2 \leq \frac{M_f}{\sqrt{2}\, \min_{1\leq j\leq m}\{-d_j(\bar{x})\}} \left(\frac{a\sqrt{\epsilon}}{\sqrt{\nu}} + b\sqrt{\epsilon\nu} \right) + \frac{\nu}{2} D^2 < \delta.$$

Next, we may choose parameters ν and ϵ so that the above inequality is satisfied. The expression suggests that one must choose $\epsilon < \nu$ (as M_f could be large). Thus setting $\epsilon = \nu^3$, we will obtain a quadratic inequality in parameter ν which can subsequently allow for selecting ν and therefore ϵ.

Unfortunately, the preceding results do not provide a bound on $\|x_{\nu,\epsilon} - x^*\|$ and indeed for the optimal ν^* minimizing $\|x_{\nu,\epsilon} - x^*\|$, the error in $\|x_{\nu,\epsilon} - x^*_\nu\|$ can be large (due to error in $\|x^*_\nu - x^*\|$). The challenge in obtaining a bound on $\|x^*_{\nu,\epsilon} - x^*\|$ implicitly requires a bound on $\|x^*_\nu - x^*\|$ which we currently do not have access to. Note that by introducing a suitable growth property on the function, one may obtain a handle on $\|x^*_\nu - x^*\|$.

3.3. Properties of $\Phi_{\nu,\epsilon}$. We now focus on characterizing the mapping $\Phi_{\nu,\epsilon}$ under the following assumption on the constraint functions d_j for $j = 1, \ldots, m$.

ASSUMPTION 4. *For each j, the gradient $\nabla d_j(x)$ is Lipschitz continuous over X with a constant $L_j > 0$, i.e.,*

$$\|\nabla d_j(x) - \nabla d_j(y)\| \leq L_j \|x - y\| \qquad \text{for all } x, y \in X.$$

Under this and the Slater assumption, we prove and the strong monotonicity and the Lipschitzian nature of $\Phi_{\nu,\epsilon}(x, \lambda)$.

LEMMA 3.4. *Let Assumptions 2–4 hold and let $\nu, \epsilon \geq 0$. Then, the regularized mapping $\Phi_{\nu,\epsilon}$ is strongly monotone over $X \times \mathbb{R}^m_+$ with constant $\mu = \min\{\nu, \epsilon\}$ and Lipschitz over $X \times \mathcal{D}_\nu$ with constant $L_\Phi(\nu, \epsilon)$ given by*

$$L_\Phi(\nu, \epsilon) = \sqrt{(L + \nu + M_d + M_\nu L_d)^2 + (M_d + \epsilon)^2}, \quad L_d = \sqrt{\sum_{j=1}^m L_j^2},$$

where L is a Lipschitz constant for $\nabla f(x)$ over X, L_j is a Lipschitz constant for $\nabla d_j(x)$ over X, $M_d = \max_{x\in X} \|\nabla d(x)\|$, and $M_\nu = \max_{\lambda\in\mathcal{D}_\nu} \|\lambda\|$.

Proof. We use $\lambda_{1,j}$ and $\lambda_{2,j}$ to denote the jth component of vectors λ_1 and λ_2. For any two vectors $z_1 = (x_1, \lambda_1),\ z_2 = (x_2, \lambda_2) \in X \times \mathbb{R}^m_+$, we have

$$\begin{aligned}
&(\Phi_{\nu,\epsilon}(z_1) - \Phi_{\nu,\epsilon}(z_2))^T (z_1 - z_2) \\
&= \begin{pmatrix} \nabla_x \mathcal{L}(x_1, \lambda_1) - \nabla_x \mathcal{L}(x_2, \lambda_2) + \nu(x_1 - x_2) \\ -d(x_1) + \epsilon\lambda_1 + d(x_2) - \epsilon\lambda_2 \end{pmatrix}^T \begin{pmatrix} x_1 - x_2 \\ \lambda_1 - \lambda_2 \end{pmatrix} \\
&= (\nabla f(x_1) - \nabla f(x_2))^T (x_1 - x_2) + \nu\|x_1 - x_2\|^2 \\
&\quad + \sum_{j=1}^m (\lambda_{1,j} \nabla d_j(x_1) - \lambda_{2,j} \nabla d_j(x_2))^T (x_1 - x_2) \\
&\quad - \sum_{j=1}^m (d_j(x_1) - d_j(x_2))(\lambda_{1,j} - \lambda_{2,j}) + \epsilon\|\lambda_1 - \lambda_2\|^2.
\end{aligned}$$

By using the monotonicity of $\nabla f(x)$, and by grouping the terms with $\lambda_{1,j}$ and $\lambda_{2,j}$, separately, we obtain

$$
\begin{aligned}
&(\Phi_{\nu,\epsilon}(z_1)-\Phi_{\nu,\epsilon}(z_2))^T(z_1-z_2) \geq \nu\|x_1-x_2\|^2 \\
&+\sum_{j=1}^m \lambda_{1,j}\left(d_j(x_2)-d_j(x_1)+\nabla d_j(x_1)^T(x_1-x_2)\right) \\
&+\sum_{j=1}^m \lambda_{2,j}\left(d_j(x_1)-d_j(x_2)-\nabla d_j(x_2)^T(x_1-x_2)\right)+\epsilon\|\lambda_1-\lambda_2\|^2.
\end{aligned}
$$

Now, by non-negativity of $\lambda_{1,j}, \lambda_{2,j}$ and convexity of $d_j(x)$ for each j, we have

$$
\begin{aligned}
&\lambda_{1,j}\left(d_j(x_2)-d_j(x_1)+\nabla d_j(x_1)^T(x_1-x_2)\right) \geq 0, \\
&\lambda_{2,j}\left(d_j(x_1)-d(x_2)-\nabla d_j(x_2)^T(x_1-x_2)\right) \geq 0.
\end{aligned}
$$

Using the preceding relations, we get

$$
(\Phi_{\nu,\epsilon}(z_1)-\Phi_{\nu,\epsilon}(z_2))^T(z_1-z_2) \geq \nu\|x_1-x_2\|^2+\epsilon\|\lambda_1-\lambda_2\|^2 \geq \min\{\nu,\epsilon\}\,\|z_1-z_2\|^2,
$$

showing that $\Phi_{\nu,\epsilon}$ is strongly monotone with constant $\mu=\min\{\nu,\epsilon\}$.

Next, we show that $\Phi_{\nu,\epsilon}$ is Lipschitz over $X\times\mathcal{D}_\nu$. Thus, given $\nu,\epsilon\geq 0$, and any two vectors $z_1=(x_1,\lambda_1),\ z_2=(x_2,\lambda_2)\in X\times\mathcal{D}_\nu$, we have

$$
\begin{aligned}
&\|\Phi_{\nu,\epsilon}(z_1)-\Phi_{\nu,\epsilon}(z_2)\| \\
&=\left\|\begin{pmatrix}\nabla f(x_1)-\nabla f(x_2)+\nu(x_1-x_2)+\sum_{j=1}^m\left(\lambda_{1,j}\nabla d_j(x_1)-\lambda_{2,j}\nabla d_j(x_2)\right) \\ -d(x_1)+d(x_2)+\epsilon(\lambda_1-\lambda_2)\end{pmatrix}\right\| \\
&\leq \|\nabla f(x_1)-\nabla f(x_2)\|+\nu\|x_1-x_2\|+\left\|\sum_{j=1}^m(\lambda_{1,j}\nabla d_j(x_1)-\lambda_{2,j}\nabla d_j(x_2))\right\| \\
&\quad+\|d(x_1)-d(x_2)\|+\epsilon\|\lambda_1-\lambda_2\|. \qquad (3.16)
\end{aligned}
$$

By the compactness of X (Assumption 2) and the continuity of $\nabla d_j(x)$ for each j, the boundedness of $\nabla d(x)=(\nabla d_1(x),\ldots,\nabla d_m(x))^T$ follows, i.e.,

$$
\|\nabla d(x)\|\leq M_d \qquad \text{for all } x\in X \text{ and some } M_d>0. \qquad (3.17)
$$

Furthermore, by using the mean value theorem (see for example [3], page 682, Prop. A.22), we can see that $d(x)$ is Lipschitz continuous over the set X with the same constant M_d. Specifically, for all $x,y\in X$, there exists a $\theta\in[0,1]$ such that

$$
\|d(x)-d(y)\|=\|\nabla d(x+\theta(y-x))(x-y)\|\leq M_d\|x-y\|.
$$

By using the Lipschitz property of $\nabla f(x)$ and $d(x)$, and by adding and subtracting the term $\sum_{j=1}^m\lambda_{1,j}\nabla d_j(x_2)$, from relation (3.16) we have

$$
\begin{aligned}
\|\Phi_{\nu,\epsilon}(z_1)-\Phi_{\nu,\epsilon}(z_2)\| \leq L\|x_1-x_2\|+\nu\|x_1-x_2\|+\sum_{j=1}^m\lambda_{1,j}\,\|\nabla d_j(x_1)-\nabla d_j(x_2)\| \\
+\sum_{j=1}^m|\lambda_{1,j}-\lambda_{2,j}|\,\|\nabla d_j(x_2)\|+M_d\|x_1-x_2\|+\epsilon\|\lambda_1-\lambda_2\|,
\end{aligned}
$$

where we also use $\lambda_{1,j} \geq 0$ for all j. By using Hölder's inequality and the boundedness of the dual variables $\lambda_1, \lambda_2 \in \mathcal{D}_\nu$, we get

$$\begin{aligned}\sum_{j=1}^{m} \lambda_{1,j} \|\nabla d_j(x_1) - \nabla d_j(x_2)\| \leq & \|\lambda_1\| \sqrt{\sum_{j=1}^{m} \|\nabla d_j(x_1) - \nabla d_j(x_2)\|^2} \\ \leq & M_\nu \sqrt{\sum_{j=1}^{m} L_j^2} \, \|x_1 - x_2\|,\end{aligned}$$

where in the last inequality we also use the Lipschitz property of $\nabla d_j(x)$ for each j. Similarly, by Hölder's inequality and the boundedness of $\nabla d(x)$ [see (3.17)], we have

$$\sum_{j=1}^{m} |\lambda_{1,j} - \lambda_{2,j}| \, \|\nabla d_j(x_2)\| \leq M_d \|\lambda_1 - \lambda_2\|.$$

By combining the preceding three relations and letting $L_d = \sqrt{\sum_{j=1}^{m} L_j^2}$, we obtain

$$\|\Phi_{\nu,\epsilon}(z_1) - \Phi_{\nu,\epsilon}(z_2)\| \leq (L + \nu + M_d + M_\nu L_d)\|x_1 - x_2\| + (M_d + \epsilon)\|\lambda_1 - \lambda_2\|.$$

Further, by Hölder's inequality, we have

$$\begin{aligned}\|\Phi_{\nu,\epsilon}(z_1) - \Phi_{\nu,\epsilon}(z_2)\| &\leq \sqrt{(L + \nu + M_d + M_\nu L_d)^2 + (M_d + \epsilon)^2} \sqrt{\|x_1 - x_2\|^2 + \|\lambda_1 - \lambda_2\|^2} \\ &= L_\Phi(\nu, \epsilon)\|z_1 - z_2\|,\end{aligned}$$

thus showing the Lipschitz property of $\Phi_{\nu,\epsilon}$. □

3.4. Primal-dual method. The strong monotonicity and Lipschitzian nature of the regularized mapping $\Phi_{\nu,\epsilon}$ for given $\nu > 0$ and $\epsilon > 0$, imply that standard projection algorithms can be effectively applied. Our goal is to generalize these schemes to accommodate the requirements of *limited* coordination. While in theory, convergence of projection schemes relies on consistency of primal and dual step-lengths, in practice, this requirement is difficult to enforce. In this section, we allow for different step-lengths and show that such a scheme does indeed result in a contraction.

Now, we consider solving the variational inequality in (3.8) by using a primal-dual method in which the users can choose their primal steplengths independently with possibly differing dual steplengths. In particular, we consider the following algorithm:

$$\begin{aligned} x_i^{k+1} &= \Pi_{X_i}(x_i^k - \alpha_i \nabla_{x_i} \mathcal{L}_{\nu,\epsilon}(x^k, \lambda^k)), \\ \lambda^{k+1} &= \Pi_{\mathcal{D}_\nu}(\lambda^k + \tau \nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x^k, \lambda^k)), \end{aligned} \tag{3.18}$$

where $\alpha_i > 0$ is the primal steplength for user i and $\tau > 0$ is the dual steplength. Next, we present our main convergence result for the sequence $\{z^k\}$ with $z^k = (x^k, \lambda^k)$ generated using (3.18).

THEOREM 3.5. *Let Assumptions 2–4 hold. Let $\{z^k\}$ be a sequence generated by (3.18). Then, we have*

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

where $q_{\nu,\epsilon}$ *is given by*

$$q_{\nu,\epsilon} = \begin{cases} 1+\alpha_{\max}^2 L_\Phi^2(\nu,\epsilon) - 2\mu\tau + (\alpha_{\min}-\tau)\max\{1-2\nu,\, M_d^2\} & \\ +2(\alpha_{\max}-\alpha_{\min})L_\Phi(\nu,\epsilon), & \text{for } \tau < \alpha_{\min} \le \alpha_{\max}; \\ 1+\alpha_{\max}^2 L_\Phi^2(\nu,\epsilon) - 2\alpha_{\min}\mu & \\ +2(\alpha_{\max}-\alpha_{\min})L_\Phi(\nu,\epsilon), & \text{for } \alpha_{\min} \le \tau < \alpha_{\max}; \\ 1+\tau^2 L_\Phi^2(\nu,\epsilon) - 2\mu\alpha_{\min} + (\tau-\alpha_{\min})\max\{1-2\epsilon,\, M_d^2\} & \\ +2(\alpha_{\max}-\alpha_{\min})L_\Phi(\nu,\epsilon), & \text{for } \alpha_{\min} \le \alpha_{\max} \le \tau, \end{cases}$$

where $\alpha_{\min} = \min_{1\le i\le N}\{\alpha_i\}$, $\alpha_{\max} = \max_{1\le i\le N}\{\alpha_i\}$, $M_d = \max_{x\in X}\|\nabla d(x)\|$, $\mu = \min\{\nu,\epsilon\}$ *and* $L_\Phi(\nu,\epsilon)$ *is as defined in Lemma 3.4.*

Proof. Let $\{\alpha_i\}_{i=1}^N$ be the user dependent steplengths of the primal iterations and let $\alpha_{\min} = \min_{1\le i\le N}\{\alpha_i\}$ and $\alpha_{\max} = \max_{1\le i\le N}\{\alpha_i\}$ denote the minimum and maximum of the user steplengths. Using $x_{i,\nu,\epsilon} = \Pi_{X_i}(x_{i,\nu,\epsilon} - \alpha_i\nabla_{x_i}\mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon}))$, non-expansive property of projection operator and Cauchy-Schwartz inequality, it can be verified that

$$\begin{aligned} \|x^{k+1}-x_{\nu,\epsilon}\|^2 &\le \|x^k-x_{\nu,\epsilon}\|^2 + \alpha_{\max}^2\|\nabla_x\mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x\mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\|^2 \\ &- 2\alpha_{\min}(\nabla_x\mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x\mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon}))^T(x^k-x_{\nu,\epsilon}) \\ &+ 2(\alpha_{\max}-\alpha_{\min})\|\nabla_x\mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x\mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\|\|x^k-x_{\nu,\epsilon}\|, \end{aligned}$$

and

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

Summing the preceding two relations, we obtain

$$\begin{aligned} \|z^{k+1}-z_{\nu,\epsilon}\|^2 &\le \|z^k-z_{\nu,\epsilon}\|^2 + \max\{\alpha_{\max}^2,\tau^2\}\|\Phi(z^k)-\Phi(z_{\nu,\epsilon})\|^2 \\ &- 2\alpha_{\min}(\nabla_x\mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x\mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon}))^T(x^k-x_{\nu,\epsilon}) \\ &+ 2(\alpha_{\max}-\alpha_{\min})\|\nabla_x\mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x\mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\|\,\|x^k-x_{\nu,\epsilon}\| \\ &- 2\tau(\nabla_\lambda\mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_\lambda\mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon}))^T(\lambda^k-\lambda_{\nu,\epsilon}). \end{aligned} \tag{3.19}$$

We now consider three cases:

Case 1 ($\tau < \alpha_{\min} \le \alpha_{\max}$)*:* By adding and subtracting

$$2\tau(\nabla_x\mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x\mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon}))^T(x^k-x_{\nu,\epsilon}),$$

we see that relation in (3.19) can be written as

$$\begin{aligned} \|z^{k+1}-z_{\nu,\epsilon}\|^2 &\le \|z^k-z_{\nu,\epsilon}\|^2 + \alpha_{\max}^2\|\Phi_{\nu,\epsilon}(z^k)-\Phi_{\nu,\epsilon}(z_{\nu,\epsilon})\|^2 \\ &- 2\tau(\Phi_{\nu,\epsilon}(z^k)-\Phi_{\nu,\epsilon}(z_{\nu,\epsilon}))^T(z^k-z_{\nu,\epsilon}) \\ &- 2(\alpha_{\min}-\tau)\left(\nabla_x\mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x\mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\right)^T(x^k-x_{\nu,\epsilon}) \\ &+ 2(\alpha_{\max}-\alpha_{\min})\|\nabla_x\mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x\mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\|\|x^k-x_{\nu,\epsilon}\|. \end{aligned}$$

By Lemma 3.4, the mapping $\Phi_{\nu,\epsilon}$ is strongly monotone and Lipschitz with constants $\mu = \min\{\nu,\epsilon\}$ and $L_\Phi(\nu,\epsilon)$, respectively. Hence, from the preceding relation we obtain

$$\begin{aligned} \|z^{k+1}-z_{\nu,\epsilon}\|^2 &\le (1+\alpha_{\max}^2 L_\Phi^2(\nu,\epsilon) - 2\tau\mu)\|z^k-z_{\nu,\epsilon}\|^2 \\ &- 2(\alpha_{\min}-\tau)\left(\nabla_x\mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x\mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\right)^T(x^k-x_{\nu,\epsilon}) \\ &+ 2(\alpha_{\max}-\alpha_{\min})\|\nabla_x\mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x\mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\|\|x^k-x_{\nu,\epsilon}\|. \end{aligned}$$

Now $\|\nabla_x \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\|\|x^k - x_{\nu,\epsilon}\| \leq \|\Phi(z^k) - \Phi(z_{\nu,\epsilon})\|\|z^k - z_{\nu,\epsilon}\| \leq L_\Phi(\nu,\epsilon)\|z^k - z_{\nu,\epsilon}\|^2$ and thus we get

$$\begin{aligned}\|z^{k+1} - z_{\nu,\epsilon}\|^2 \leq (1 + \alpha_{\max}^2 L_\Phi^2(\nu,\epsilon) - 2\tau\mu)\|z^k - z_{\nu,\epsilon}\|^2 + 2(\alpha_{\max} - \alpha_{\min})L_\Phi(\nu,\epsilon)\|z^k - z_{\nu,\epsilon}\|^2 \\ - 2(\alpha_{\min} - \tau)\left(\nabla_x \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\right)^T (x^k - x_{\nu,\epsilon}).\end{aligned} \tag{3.20}$$

We next estimate the last term in the preceding relation. By adding and subtracting $\nabla_x \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda^k)$, we have

$$\begin{aligned}\left(\nabla_x \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k)\right. &- \left.\nabla_x \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\right)^T (x^k - x_{\nu,\epsilon}) \\ &= \left(\nabla_x \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda^k)\right)^T (x^k - x_{\nu,\epsilon}) \\ &\quad + \left(\nabla_x \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda^k) - \nabla_x \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\right)^T (x^k - x_{\nu,\epsilon}).\end{aligned}$$

Using the strong monotonicity of $\nabla_x \mathcal{L}_{\nu,\epsilon}$, and writing the second term on the right hand side explicitly, we get

$$\begin{aligned}\left(\nabla_x \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k)\right. &- \left.\nabla_x \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\right)^T (x^k - x_{\nu,\epsilon}) \geq \nu\|x^k - x_{\nu,\epsilon}\|^2 \\ &\quad + \sum_{j=1}^m \left(\nabla d_j(x_{\nu,\epsilon})(\lambda_j^k - \lambda_{\nu,\epsilon,j})\right)^T (x^k - x_{\nu,\epsilon}) \\ &\geq \nu\|x^k - x_{\nu,\epsilon}\|^2 \\ &\quad - \frac{1}{2}\left(\left\|\sum_{j=1}^m \nabla d_j(x_{\nu,\epsilon})(\lambda_j^k - \lambda_{\nu,\epsilon,j})\right\|^2 + \|x^k - x_{\nu,\epsilon}\|^2\right),\end{aligned}$$

where the last step follows by noting that $ab \geq -\frac{1}{2}(a^2 + b^2)$. Using Cauchy-Schwartz and Hölder's inequality, we have

$$\begin{aligned}\left\|\sum_{j=1}^m \nabla d_j(x_{\nu,\epsilon})(\lambda_j^k - \lambda_{\nu,\epsilon,j})\right\|^2 &\leq \left(\sum_{j=1}^m \|\nabla d_j(x_{\nu,\epsilon})\|\,|\lambda_j^k - \lambda_{\nu,\epsilon,j}|\right)^2 \\ &\leq \left(\sum_{j=1}^m \|\nabla d_j(x_{\nu,\epsilon})\|^2\right)\|\lambda^k - \lambda_{\nu,\epsilon}\|^2 \\ &\leq M_d^2\,\|\lambda^k - \lambda_{\nu,\epsilon}\|^2,\end{aligned}$$

where in the last step, the boundedness of $\nabla d(x)$ over X was employed ($\|\nabla d(x)\| \leq M_d$). By combining the preceding relations, we obtain

$$\begin{aligned}\left(\nabla_x \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k)\right. &- \left.\nabla_x \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\right)^T (x^k - x_{\nu,\epsilon}) \\ &\geq -\frac{1}{2}\left((1-2\nu)\|x^k - x_{\nu,\epsilon}\|^2 + M_d^2\|\lambda^k - \lambda_{\nu,\epsilon}\|^2\right) \\ &\geq -\frac{1}{2}\max\{1-2\nu,\ M_d^2\}\|z^k - z_{\nu,\epsilon}\|^2.\end{aligned}$$

If the above estimate is substituted in (3.20), we obtain

$$\|z^{k+1} - z_{\nu,\epsilon}\|^2 \leq q_{\nu,\epsilon}\|z^k - z_{\nu,\epsilon}\|^2,$$

where $q_{\nu,\epsilon} = 1 + \alpha_{\max}^2 L_\Phi^2(\nu,\epsilon) - 2\mu\tau + (\alpha_{\min} - \tau)\max\{1-2\nu,\ M_d^2\} + 2(\alpha_{\max} - \alpha_{\min})L_\Phi(\nu,\epsilon)$, thus showing the desired relation.

Case 2 ($\alpha_{\min} \leq \tau < \alpha_{\max}$)*:* By adding and subtracting

$$2\alpha_{\min}(\nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon}))^T(\lambda^k - \lambda_{\nu,\epsilon}),$$

for $\tau < \alpha_{\max}$ relation (3.19) reduces to

$$\begin{aligned}\|z^{k+1} - z_{\nu,\epsilon}\|^2 \leq{}& \|z^k - z_{\nu,\epsilon}\|^2 + \alpha_{\max}^2\|\Phi(z^k) - \Phi(z_{\nu,\epsilon})\|^2\\ &- 2\alpha_{\min}(\Phi(z^k) - \Phi(z_{\nu,\epsilon}))^T(z^k - z_{\nu,\epsilon})\\ &- 2(\tau - \alpha_{\min})(\nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon}))^T(\lambda^k - \lambda_{\nu,\epsilon})\\ &+ 2(\alpha_{\max} - \alpha_{\min})\|\nabla_x \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\|\|x^k - x_{\nu,\epsilon}\|,\end{aligned}$$

which by Lipschitz continuity and strong monotonicity of Φ implies,

$$\begin{aligned}\|z^{k+1} - z_{\nu,\epsilon}\|^2 \leq{}& (1 + \alpha_{\max}^2 L_\Phi^2(\nu,\epsilon) - 2\alpha_{\min}\mu)\|z^k - z_{\nu,\epsilon}\|^2\\ &+ 2(\alpha_{\max} - \alpha_{\min})\|\nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\|\|\lambda^k - \lambda_{\nu,\epsilon}\|\\ &+ 2(\alpha_{\max} - \alpha_{\min})\|\nabla_x \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\|\|x^k - x_{\nu,\epsilon}\|.\end{aligned}$$

Using Hölder's inequality, we get

$$\begin{aligned}\|z^{k+1} - z_{\nu,\epsilon}\|^2 \leq{}& (1 + \alpha_{\max}^2 L_\Phi^2(\nu,\epsilon) - 2\alpha_{\min}\mu)\|z^k - z_{\nu,\epsilon}\|^2\\ &+ 2(\alpha_{\max} - \alpha_{\min})\|\Phi(z^k) - \Phi(z_{\nu,\epsilon})\|\|z^k - z_{\nu,\epsilon}\|.\end{aligned}$$

Finally using Lipschitz continuity of Φ we get

$$\|z^{k+1} - z_{\nu,\epsilon}\|^2 \leq q\|z^k - z_{\nu,\epsilon}\|^2,$$

where $q_{\nu,\epsilon} = 1 + \alpha_{\max}^2 L_\Phi^2(\nu,\epsilon) - 2\alpha_{\min}\mu + 2(\alpha_{\max} - \alpha_{\min})L_\Phi(\nu,\epsilon)$.

Case 3 ($\alpha_{\min} \leq \alpha_{\max} \leq \tau$)*:* Note that

$$\begin{aligned}&(\nabla_x \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon}))^T(x^k - x_{\nu,\epsilon})\\ &= \left(\Phi_{\nu,\epsilon}(z^k) - \Phi_{\nu,\epsilon}(z_{\nu,\epsilon})\right)^T(z^k - z_{\nu,\epsilon})\\ &-(-d(x^k) + \epsilon\lambda^k + d(x_{\nu,\epsilon}) - \epsilon\lambda_{\nu,\epsilon})^T(\lambda^k - \lambda_{\nu,\epsilon}).\end{aligned}$$

Thus, from the preceding equality and relation (3.19), where $\alpha_{\max} < \tau$, we have

$$\begin{aligned}\|z^{k+1} - z_{\nu,\epsilon}\|^2 \leq{}& \|z^k - z_{\nu,\epsilon}\|^2 + \tau^2\|\Phi(z^k) - \Phi(z_{\nu,\epsilon})\|^2\\ &- 2\alpha_{\min}(\Phi(z^k) - \Phi(z_{\nu,\epsilon}))^T(z^k - z_{\nu,\epsilon})\\ &- 2(\tau - \alpha_{\min})(\nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon}))^T(\lambda^k - \lambda_{\nu,\epsilon})\\ &+ 2(\alpha_{\max} - \alpha_{\min})\|\nabla_x \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\|\|x^k - x_{\nu,\epsilon}\|.\end{aligned}$$

By Lemma 3.4, the mapping $\Phi_{\nu,\epsilon}$ is strongly monotone and Lipschitz with constants $\mu = \min\{\nu,\epsilon\}$ and $L_\Phi(\nu,\epsilon)$, respectively. Hence, it follows

$$\begin{aligned}\|z^{k+1} - z_{\nu,\epsilon}\|^2 \leq{}& (1 + \tau^2 L_\Phi^2(\nu,\epsilon) - 2\alpha_{\min}\mu)\|z^k - z_{\nu,\epsilon}\|^2\\ &- 2(\tau - \alpha_{\min})(\nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon}))^T(\lambda^k - \lambda_{\nu,\epsilon})\\ &+ 2(\alpha_{\max} - \alpha_{\min})\|\nabla_x \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_x \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon})\|\|x^k - x_{\nu,\epsilon}\|,\end{aligned}$$

which can be further estimated as

$$\|z^{k+1} - z_{\nu,\epsilon}\|^2 \leq (1 + \tau^2 L_\Phi^2(\nu,\epsilon) - 2\alpha_{\min}\mu)\|z^k - z_{\nu,\epsilon}\|^2 + 2(\alpha_{\max} - \alpha_{\min})L_\Phi(\nu,\epsilon)\|z^k - z_{\nu,\epsilon}\|^2$$
$$- 2(\tau - \alpha_{\min})(\nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k) - \nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x_{\nu,\epsilon},\lambda_{\nu,\epsilon}))^T(\lambda^k - \lambda_{\nu,\epsilon}). \tag{3.21}$$

Next, we estimate the last term on the right hand side of the preceding relation. Through the use of Cauchy-Schwartz inequality, we have

$$\left(d(x^k) - d(x_{\nu,\epsilon})\right)^T (\lambda^k - \lambda_{\nu,\epsilon}) \leq \frac{1}{2}\left\|d(x^k) - d(x_{\nu,\epsilon})\right\|^2 + \frac{1}{2}\|\lambda^k - \lambda_{\nu,\epsilon}\|^2.$$

By the continuity of the gradient mapping of $d(x) = (d_1(x), \ldots, d_m(x))^T$ and its boundedness ($\|\nabla d(x)\| \leq M_d$), using the Mean-value Theorem we further have

$$\left\|d(x^k) - d(x_{\nu,\epsilon})\right\|^2 \leq M_d^2 \|x^k - x_{\nu,\epsilon}\|^2.$$

From the preceding two relations we have

$$\left(d(x^k) - d(x_{\nu,\epsilon})\right)^T (\lambda^k - \lambda_{\nu,\epsilon}) \leq \frac{M_d^2}{2}\|x^k - x_{\nu,\epsilon}\|^2 + \frac{1}{2}\|\lambda^k - \lambda_{\nu,\epsilon}\|^2,$$

which when substituted in inequality (3.21) yields

$$\|z^{k+1} - z_{\nu,\epsilon}\|^2 \leq \left(1 + \tau^2 L_\Phi^2(\nu,\epsilon) - 2\mu\alpha_{\min} + 2(\alpha_{\max} - \alpha_{\min})L_\Phi(\nu,\epsilon)\right)\|z^k - z_{\nu,\epsilon}\|^2$$
$$+(\tau - \alpha_{\min})(1 - 2\epsilon)\|\lambda^k - \lambda_{\nu,\epsilon}\|^2 + (\tau - \alpha_{\min})M_d^2\|x^k - x_{\nu,\epsilon}\|^2.$$

The desired relation follows by observing that

$$(1 - 2\epsilon)\|\lambda^k - \lambda_{\nu,\epsilon}\|^2 + M_d^2\|x^k - x_{\nu,\epsilon}\|^2 \leq \max\{1 - 2\epsilon,\, M_d^2\}\,\|z^k - z_{\nu,\epsilon}\|^2.$$

□

An immediate corollary of Theorem 3.5 is obtained when all users have the same steplength. More precisely, we have the following algorithm:

$$\begin{aligned} x^{k+1} &= \Pi_X(x^k - \alpha\nabla_x \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k)), \\ \lambda^{k+1} &= \Pi_{\mathcal{D}_\nu}(\lambda^k + \tau\nabla_\lambda \mathcal{L}_{\nu,\epsilon}(x^k,\lambda^k)), \end{aligned} \tag{3.22}$$

where $\alpha > 0$ and $\tau > 0$ are, respectively, primal and dual steplengths. We present the convergence of the sequence $\{z^k\}$ with $z^k = (x^k, \lambda^k)$ in the next corollary.

Corollary 3.6. *Let Assumptions 2–4 hold. Let $\{z^k\}$ be a sequence generated by (3.22) with the primal and dual step-sizes chosen independently. Then, we have*

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

where $q_{\nu,\epsilon}$ is given by

$$q_{\nu,\epsilon} = 1 - 2\mu\min\{\alpha,\tau\} + \max\{\alpha^2,\tau^2\}L_\Phi^2(\nu,\epsilon) + \theta(\alpha,\tau),$$

$$\text{and } \theta(\alpha,\tau) \triangleq \begin{cases} (\alpha - \tau)\max\{1 - 2\nu,\, M_d^2\} & \text{for } \tau \leq \alpha, \\ (\tau - \alpha)\max\{1 - 2\epsilon,\, M_d^2\} & \text{for } \alpha < \tau, \end{cases}$$

$\mu = \min\{\nu, \epsilon\}$ *and* $L_\Phi^2(\nu,\epsilon)$ *is as given in Lemma 3.4.*

Note that when $\alpha_{\min} = \alpha_{\max} = \tau$ and $\tau < 2\mu/L_\Phi^2(\nu,\epsilon)$, Theorem 3.5 implies the standard contraction result for a strongly monotone and Lipschitz mapping. However, Theorem 3.5 does not guarantee the existence of a tuple $(\alpha_{\min}, \alpha_{\max}, \tau)$ resulting in a contraction in general, i.e., does not ensure that $q_{\nu,\epsilon} \in (0,1)$. This is done in the following lemma.

LEMMA 3.7. *Let $q_{\nu,\epsilon}$ be as given in Theorem 3.6. Then, there exists a tuple $(\alpha_{\min}, \alpha_{\max}, \tau)$ such that $q_{\nu,\epsilon} \in (0,1)$.*

Proof. It suffices to show that there exists a tuple $(\alpha_{\min}, \alpha_{\max}, \tau)$ such that

$$\begin{aligned}
&0 < 1 + \alpha_{\max}^2 L_\Phi^2(\nu,\epsilon) - 2\mu\tau + (\alpha_{\min} - \tau)\max\{1-2\nu,\ M_d^2\} \\
&+2(\alpha_{\max} - \alpha_{\min})L_\Phi(\nu,\epsilon) < 1 && \tau < \alpha_{\min} \le \alpha_{\max} \\
&0 < 1 + \alpha_{\max}^2 L_\Phi^2(\nu,\epsilon) - 2\alpha_{\min}\mu \\
&+2(\alpha_{\max} - \alpha_{\min})L_\Phi(\nu,\epsilon) < 1 && \alpha_{\min} \le \tau < \alpha_{\max} \\
&0 < 1 + \tau^2 L_\Phi^2(\nu,\epsilon) - 2\mu\alpha_{\min} + (\tau - \alpha_{\min})\max\{1-2\epsilon,\ M_d^2\} \\
&+2(\alpha_{\max} - \alpha_{\min})L_\Phi(\nu,\epsilon) < 1 && \alpha_{\min} \le \alpha_{\max} \le \tau.
\end{aligned}$$

Also, it suffices to prove only one of the cases since the other cases follow by interchanging the roles of τ and $\alpha_{\min}$ or τ and $\alpha_{\max}$. We consider the case where $\tau < \alpha_{\min} \le \alpha_{\max}$. Here, if $\alpha_{\max} < 2\mu/L_\Phi^2(\nu,\epsilon)$ then there is $\beta < 1$ such that setting $\tau = \beta\alpha_{\max}$ we have $q < 1$. To see this let $\alpha_{\min} = \beta_1\alpha_{\max}$ such that $\beta < \beta_1 \le 1$ and $\max\{1-2\nu, M_d^2\} = M_d^2$. Consider

$$q_{\nu,\epsilon} - 1 = -2\mu\tau + \alpha_{\max}^2 L_\Phi^2(\nu,\epsilon) + (\alpha_{\min} - \tau)M_d^2 + 2(\alpha_{\max} - \alpha_{\min})L_\Phi(\nu,\epsilon).$$

Setting $\tau = \beta\alpha_{\max}$, $\alpha_{\min} = \beta_1\alpha_{\max}$, the preceding relation reduces to

$$q_{\nu,\epsilon} - 1 = -2\mu\beta\alpha_{\max} + \alpha_{\max}^2 L_\Phi^2(\nu,\epsilon) + \alpha_{\max}(\beta_1 - \beta)M_d^2 + 2\alpha_{\max}(1-\beta_1)L_\Phi(\nu,\epsilon).$$

Using $\beta < \beta_1 \le 1$ we obtain

$$q_{\nu,\epsilon} - 1 \le -2\mu\beta\alpha_{\max} + \alpha_{\max}^2 L_\Phi^2(\nu,\epsilon) + \alpha_{\max}(1-\beta)M_d^2 + 2\alpha_{\max}(1-\beta)L_\Phi(\nu,\epsilon).$$

We are done if we show that the expression on the right hand side of the preceding relation is negative for some β i.e.,

$$-2\mu\beta\alpha_{\max} + \alpha_{\max}^2 L_\Phi^2(\nu,\epsilon) + \alpha_{\max}(1-\beta)M_d^2 + 2\alpha_{\max}(1-\beta)L_\Phi(\nu,\epsilon) < 0.$$

Following some rearrangement it can be verified that

$$\beta > \frac{\alpha_{\max}L_\Phi^2(\nu,\epsilon) + M_d^2 + 2L_\Phi(\nu,\epsilon)}{2\mu + M_d^2 + 2L_\Phi(\nu,\epsilon)}.$$

Since we have $\alpha_{\max}L_\Phi^2(\nu,\epsilon) < 2\mu$ it follows that the expression on right hand side of the preceding relation is strictly less than 1 and we have

$$\beta \in \left(\frac{\alpha_{\max}L_\Phi^2(\nu,\epsilon) + M_d^2 + 2L_\Phi(\nu,\epsilon)}{2\mu + M_d^2 + 2L_\Phi(\nu,\epsilon)}, 1\right),$$

implying that we have $\beta \in (0,1)$. □

The previous result is motivated by several issues arising in practical settings. First there may be errors due to noisy links in the communication network, which may cause inconsistencies across steplengths. Often, it may be difficult to even enforce this consistency. As a consequence, we examine the extent to which the convergence theory is affected by a lack of consistency. A related question is whether one can, in a distributed setting, impose alternative requirements that weaken consistency. This can be achieved by setting bounds on the primal and dual steplengths which are independent. For instance, if $\alpha_{\max} < \frac{2\mu}{L_\Phi^2(\nu,\epsilon)}$, then it suffices to choose τ independently as $\tau \leq \beta\alpha_{\max} \leq \beta\frac{2\mu}{L_\Phi^2(\nu,\epsilon)}$, where β is chosen independently. Importantly, Lemma 3.7 provides a characterization of the relationship between $\alpha_{\min}$, $\alpha_{\max}$ and τ using the values of problem parameters, to ensure convergence of the scheme. Expectedly, as the numerical results testify, the performance does deteriorate when there α_i's and τ do not match.

Finally, we remark briefly on the relevance of allowing for differing steplengths. In distributed settings, communication of steplengths may be corrupted via error due to noisy communication links. A majority of past work on such problems (cf. [13,14]) requires that steplengths be consistent across users. Furthermore, in constrained regimes, there is a necessity to introduce both primal (user) steplengths and dual (link) steplengths. We show that there may be limited diversity across all of these parameters while requiring that these parameters together satisfy some relationship. One may question if satisfying this requirement itself requires some coordination. In fact, we show that this constraint is implied by a set of private user-specific and dual requirements on their associated steplengths, allowing for ease of implementation.

3.5. Extension to independently chosen regularization parameters. In this subsection, we extend the results of the preceding section to a regime where the ith user selects its own regularization parameter ν_i. Before proceeding, we provide a brief motivation of such a line of questioning. In networked settings specifying steplengths and regularization parameters for the users at every instant is generally challenging. Enforcing consistent choices across these users is also difficult. An alternative lies in *broadcasting* a range of choices for steplengths, as done in the previous subsection. In this section, we show that an analogous approach can be leveraged for specifying regularization parameters, with limited impacts on the final results. Importantly, the benefit of these results lies in the fact that enforcement of consistency of regularization parameters is no longer necessary. We start with definition of the regularized Lagrangian function with user specific regularization terms. In particular, we let

$$\mathcal{L}_{V,\epsilon}(x,\lambda) = f(x) + \frac{1}{2}x^TVx + \lambda^Td(x) - \frac{\epsilon}{2}\|\lambda\|^2 \tag{3.23}$$

where V is a diagonal matrix with diagonal entries $\nu_1,\ldots,\nu_N$. In this case, letting $\nu_{\max} \triangleq \max_{i\in\{1,\ldots,N\}}\{\nu_i\}$, for some $\bar{x} \in X$ and $\bar{\lambda} \geq 0$, we define the set $\mathcal{D}_{\nu_{\max}}$ given by:

$$\mathcal{D}_{\nu_{\max}} = \left\{\lambda \in \mathbb{R}^m \;\Big|\; \sum_{j=1}^m \lambda_j \leq \frac{f(\bar{x}) + \frac{\nu_{\max}}{2}\|\bar{x}\|^2 - v(\bar{\lambda})}{\min_{1\leq j\leq m}\{-d_j(\bar{x})\}},\ \lambda \geq 0\right\}. \tag{3.24}$$

We consider the regularized primal problem (2.1) with the regularization term $\frac{1}{2}x^TVx$. We let Λ_V^* be the set of dual optimal solutions of such regularized primal problem. Then, relation (3.6) holds for Λ_V^* and $\mathcal{D}_{\nu_{\max}}$, i.e., $\Lambda_V^* \subseteq \mathcal{D}_{\nu_{\max}}$ and, therefore, the development in the preceding two sections extends to this case as well. We let x_V^* and λ_V^* denote primal-dual of the regularized primal problem with the regularization term $\frac{1}{2}x^TVx$. Analogously, we let $x_{V,\epsilon}^*$ and $\lambda_{V,\epsilon}^*$ denote respectively the primal and the dual part of the saddle point solution for $\mathcal{L}_{V,\epsilon}(x,\lambda)$ over $X \times \mathbb{R}^m_+$. We present the modified results in the form of remarks and omit the details of proofs.

The bound of Proposition 3.1 when user i uses its own regularization parameter ν_i will reduce to:

$$(x_V^* - x_{V,\epsilon})^T V(x_V^* - x_{V,\epsilon}) + \frac{\epsilon}{2}\|\lambda_{V,\epsilon}\|^2 \le \frac{\epsilon}{2}\|\lambda_{\mathbf{V}}^*\|^2 \qquad \text{for all } \lambda_V^* \in \Lambda_V^*,$$

and thus we have the following bound

$$\|x_V^* - x_{V,\epsilon}\| \le \sqrt{\frac{\epsilon}{2\nu_{\min}}} \max_{\lambda^* \in \mathcal{D}_{\nu_{\max}}} \|\lambda^*\|,$$

where $\nu_{\min} \triangleq \min_{i\in\{1,\ldots,N\}} \{\nu_i\}$.

The result in Lemma 3.3 is replaced by the following one.

Lemma 3.8. *Let Assumptions 1 and 2 hold. For any $\nu_i > 0$, $i = 1,\ldots,N$, and $\epsilon > 0$, we have*

$$\max\{0, d_j(x_{V,\epsilon})\} \le M_{d_j} M_{\nu_{\max}} \sqrt{\frac{\epsilon}{2\nu_{\min}}} \qquad \textit{for all } j = 1,\ldots,m,$$

$$|f(x_{V,\epsilon}) - f(x^*)| \le M_f M_{\nu_{\max}} \sqrt{\frac{\epsilon}{2\nu_{\min}}} + \frac{\nu_{\max}}{2} D^2,$$

with $M_{d_j} = \max_{x\in X} \|\nabla d_j(x)\|$ for each $j = 1,\ldots,m$, $M_f = \max_{x\in X}\|\nabla f(x)\|$, $M_{\nu_{\max}} = \max_{\lambda^*\in\mathcal{D}_{\nu_{\max}}} \|\lambda^*\|$ and $D = \max_{x\in X}\|x\|$.

Our result following Lemma 3.3 where we describe how one may choose parameters ϵ and ν to get within a given threshold error on the deviation of the obtained function value from its optimal counterpart will have to be reconsidered using the appropriate parameters $\nu_{\min}$ and $\nu_{\max}$. More precisely, we will have $|f(x_{V,\epsilon}) - f(x^*)| < \delta$ if we have

$$M_f M_{\nu_{\max}} \sqrt{\frac{\epsilon}{\nu_{\min}}} + \frac{\nu_{\max}}{2} D^2 < \delta.$$

Following a similar analysis and using the structure of the set $D_{\nu_{\max}}$, we have

$$\begin{aligned} M_{\nu_{\max}} \sqrt{\frac{\epsilon}{2\nu_{\min}}} &= \sqrt{\frac{\epsilon}{2\nu_{\min}}} \left(\max_{\lambda\in\mathcal{D}_{\nu_{\max}}} \|\lambda\|_1 \right) \\ &= \sqrt{\frac{\epsilon}{2\nu_{\min}}} \left(\frac{f(\bar{x}) + \frac{\nu_{\max}}{2}\|\bar{x}\|^2 - v(\bar{\lambda})}{\min_{1\le j\le m}\{-d_j(\bar{x})\}} \right) \\ &= \sqrt{\frac{\epsilon}{2\nu_{\min}}} \left(\frac{f(\bar{x}) - v(\bar{\lambda})}{\min_{1\le j\le m}\{-d_j(\bar{x})\}} \right) + \sqrt{\frac{\epsilon}{2\nu_{\min}}} \left(\frac{\frac{\nu_{\max}}{2}\|\bar{x}\|^2}{\min_{1\le j\le m}\{-d_j(\bar{x})\}} \right), \end{aligned}$$

Letting $a = \frac{f(\bar{x}) - v(\bar{\lambda})}{\sqrt{2}\,\min_{1\le j\le m}\{-d_j(\bar{x})\}}$ and $b = \frac{\|\bar{x}\|^2}{2\sqrt{2}\,\min_{1\le j\le m}\{-d_j(\bar{x})\}}$, we have

$$M_f M_{\nu_{\max}} \sqrt{\frac{\epsilon}{\nu_{\min}}} + \frac{\nu_{\max}}{2} D^2 \le M_f \left(\frac{a\sqrt{\epsilon}}{\sqrt{\nu_{\min}}} + b\sqrt{\frac{\epsilon\nu_{\max}^2}{\nu_{\min}}} \right) + \frac{\nu_{\max}}{2} D^2 < \delta.$$

Next, we may choose parameters $\nu_{\min}$, $\nu_{\max}$ and ϵ so that the above inequality is satisfied. The expression suggests that one must choose $\epsilon < \nu_{\min}$ (as M_f could be large). Thus, setting $\epsilon = \nu_{\min}\nu_{\max}^2$, we will obtain a quadratic inequality in parameter $\nu_{\max}$ which can subsequently allow for selecting $\nu_{\max}$ and, therefore, selecting $\nu_{\min}$ and ϵ.

Analogous to the definition of the mapping $\Phi_{\nu,\epsilon}(x,\lambda)$ in (3.3), we define the regularized mapping corresponding to the Lagrangian in (3.23). Specifically, we have the regularized mapping $\Phi_{V,\epsilon}(x,\lambda)$ given by

$$\Phi_{V,\epsilon}(x,\lambda) \triangleq (\nabla_x \mathcal{L}_{V,\epsilon}(x,\lambda), -\nabla_\lambda \mathcal{L}_{V,\epsilon}(x,\lambda)) = (\nabla_x \mathcal{L}(x,\lambda) + Vx, -d(x) + \epsilon\lambda).$$

The properties of $\Phi_{V,\epsilon}$, namely, strong monotonicity and Lipschitz continuity remain. Specifically, $\Phi_{V,\epsilon}$ is strongly monotone with the same constant μ as before, i.e., $\mu = \min\{\nu_{\min}, \epsilon\}$. However, Lipschitz constant is not the same. Letting $L_\Phi(V,\epsilon)$ denote a Lipschitz constant for $\Phi_{V,\epsilon}$, we have

$$L_\Phi(V,\epsilon) = \sqrt{(L + \nu_{\max} + M_d + M_{\nu_{\max}} L_d)^2 + (M_d + \epsilon)^2}. \tag{3.25}$$

The result of Theorem 3.5 can be expressed as in the following corollary.

Corollary 3.9. *Let Assumptions 1-4 hold. Let $\{z^k\}$ be a sequence generated by* (3.18) *with each user using ν_i as its regularization parameter instead of ν. Then, we have*

$$\|z^{k+1} - z_{V,\epsilon}\| \le \sqrt{q_{V,\epsilon}}\|z^k - z_{V,\epsilon},\|$$

with $q_{V,\epsilon}$ as given in Theorem 3.5, where $L_\Phi(\nu,\epsilon)$ is replaced by $L_\Phi(V,\epsilon)$ from (3.25).

4. A Regularized Dual Method. The focus in Section 3 has been on primal-dual method dealing with problems where a set of convex constraints couples the user decisions. A key property of our primal-dual method is that both schemes have the same time-scales. In many practical settings, the primal and dual updates are carried out by very different entities so that the time-scales may be vastly different. For instance, the dual updates of the Lagrange multipliers could be controlled by the network operator and might be on a slower time-scale than the primal updates that are made by the users. Dual methods have proved useful in multiuser optimization problems and their convergence to the *optimal primal* solution has been studied for the case when the user objectives are strongly convex [13, 25].

In this section, we consider regularization to deal with the lack of strong convexity of Lagrangian subproblems and to also accommodate inexact solutions of the Lagrangian subproblems. For the inexact solutions, we develop error bounds. Inexactness is essential in constructing distributed online schemes that require primal solutions within a fixed amount of time. In the standard dual framework, for each $\lambda \in \mathbb{R}^m_+$, a solution $x(\lambda) \in X$ of a Lagrangian subproblem is given by a solution to $\mathrm{VI}(X, \nabla_x \mathcal{L}(x,\lambda))$, which satisfies the following inequality:

$$(x - x(\lambda))^T \nabla_x \mathcal{L}(x(\lambda),\lambda) \ge 0 \qquad \text{for all } x \in X,$$

where $\nabla_\lambda \mathcal{L}(x(\lambda),\lambda) = \nabla_x f(x(\lambda)) + \sum_{j=1}^m \lambda_j \nabla d_j(x(\lambda))$. An optimal dual variable λ is a solution of $\mathrm{VI}(\mathbb{R}^m_+, -\nabla_\lambda \mathcal{L}(x(\lambda),\lambda))$ given by

$$(\hat{\lambda} - \lambda)^T (-\nabla_\lambda \mathcal{L}(x(\lambda),\lambda)) \ge 0 \qquad \text{for all } \hat{\lambda} \in \mathbb{R}^m_+,$$

where $\nabla_\lambda \mathcal{L}(x,\lambda) = d(x)$. We consider a regularization in both primal and dual space as discussed in Section 3. In Section 4.1, we discuss the exact dual method and provide the contraction results in the primal and dual space as well as bounds on the infeasibility. These results are extended to allow for inexact solutions of Lagrangian subproblems in Section 4.2.

4.1. Regularized exact dual method. We begin by considering an *exact* dual scheme for the regularized problem given by

$$x^t = \Pi_X(x^t - \alpha\nabla_x\mathcal{L}_{\nu,\epsilon}(x^t,\lambda^t)), \tag{4.1}$$
$$\lambda^{t+1} = \Pi_{\mathcal{D}_\nu}(\lambda^t + \tau\nabla_\lambda\mathcal{L}_{\nu,\epsilon}(x^t,\lambda^t)) \qquad \text{for } t \geq 0, \tag{4.2}$$

where the set $\mathcal{D}_\nu$ is as defined in (3.6). In the primal step (4.1), the vector x^t denotes the solution $x(\lambda^t)$ of the fixed-point equation corresponding to the current Lagrange multiplier λ^t.

We now focus on the conditions ensuring that the sequence $\{\lambda^t\}$ converges to the optimal dual solution $\lambda^*_{\nu,\epsilon}$ and that the corresponding $\{x(\lambda^t)\}$ converges to the primal optimal $x^*_{\nu,\epsilon}$ of the regularized problem. We note that Proposition 3.1 combined with Lemma 3.3 provide bounds on the constraint violations, and a bound on the difference in the function values $f(x^*_{\nu,\epsilon})$ and the primal optimal value f^* of the original problem.

LEMMA 4.1. *Under Assumption 2, the function $-d(x(\lambda))$ is co-coercive in λ with constant $\frac{\nu}{M_d^2}$, where $M_d = \max_{x\in X}\|\nabla d(x)\|$.*

Proof. Let λ_1 and $\lambda_2 \in \mathbb{R}^m_+$. Let x_1 and x_2 denote the solutions to $\mathrm{VI}(X, \nabla_x\mathcal{L}_{\nu,\epsilon}(x,\lambda_1))$ and $\mathrm{VI}(X, \nabla_x\mathcal{L}_{\nu,\epsilon}(x,\lambda_2))$, respectively. Then, we have the following inequalities:

$$(x_2 - x_1)^T(\nabla f_\nu(x_1) + \sum_{j=1}^m \lambda_{1,j}\nabla d_j(x_1)) \geq 0,$$
$$(x_1 - x_2)^T(\nabla f_\nu(x_2) + \sum_{j=1}^m \lambda_{2,j}\nabla d_j(x_2)) \geq 0,$$

where $\lambda_{1,j}$ and $\lambda_{2,j}$ denote the jth component of vectors λ_1 and λ_2, respectively. Summing these inequalities, we get

$$\begin{aligned}&\sum_{j=1}^m \lambda_{1,j}(x_2 - x_1)^T\nabla d_j(x_1) + \sum_{j=1}^m \lambda_{2,j}(x_1 - x_2)^T\nabla d_j(x_2)\\ &\geq (x_2 - x_1)^T(\nabla f_\nu(x_2) - \nabla f_\nu(x_1)) \geq \nu\|x_2 - x_1\|^2.\end{aligned} \tag{4.3}$$

By using the convexity of the functions d_j and inequality (4.3), we obtain

$$\begin{aligned}(\lambda_2 - \lambda_1)^T(-d(x_2) + d(x_1)) &= \sum_{j=1}^m \lambda_{1,j}(d_j(x_2) - d_j(x_1)) + \sum_{j=1}^m \lambda_{2,j}(d_j(x_1) - d_j(x_2))\\ &\geq \sum_{j=1}^m \lambda_{1,j}(x_2 - x_1)^T\nabla d_j(x_1) + \sum_{j=1}^m \lambda_{2,j}(x_1 - x_2)^T\nabla d_j(x_2)\\ &\geq \nu\|x_2 - x_1\|^2.\end{aligned} \tag{4.4}$$

Now, by using the Lipschitz continuity of $d(x)$, as implied by Assumption 2, we see that $\|x_2 - x_1\|^2 \geq \frac{\nu}{M_d^2}\|d(x_2) - d(x_1)\|^2$ with $M_d = \max_{x\in X}\nabla d(x)$, which when substituted in the preceding relation yields the result. $\square$

We now prove our convergence result for the dual method, relying on the exact solution of the corresponding Lagrangian subproblem.

PROPOSITION 4.2. *Let Assumptions 1 and 2 hold, and let the step size τ be such that*

$$\tau < \frac{2\nu}{M_d^2 + 2\epsilon\nu} \qquad \textit{with} \qquad M_d = \max_{x \in X} \nabla d(x).$$

Then, for the sequence $\{\lambda_t\}$ generated by the dual method in (4.2), we have

$$\|\lambda^{t+1} - \lambda^*_{\nu,\epsilon}\| \leq q\|\lambda^t - \lambda^*_{\nu,\epsilon}\| \qquad \textit{where } q = 1 - \tau\epsilon.$$

Proof. By using the definition of the dual method in (4.2) and the non-expansivity of the projection, we obtain the following set of inequalities:

$$\begin{aligned}\|\lambda^{t+1} - \lambda^*_{\nu,\epsilon}\|^2 &\leq \|\lambda^t + \tau(d(x^t) - \epsilon\lambda^t) - \left(\lambda^*_{\nu,\epsilon} + \tau(d(x^*_{\nu,\epsilon}) - \epsilon\lambda^*_{\nu,\epsilon})\right)\|^2 \\ &= \|(1-\tau\epsilon)(\lambda^t - \lambda^*_{\nu,\epsilon}) - \tau\left(d(x^*_{\nu,\epsilon}) - d(x^t)\right)\|^2 \\ &= (1-\tau\epsilon)^2\|\lambda^t - \lambda^*_{\nu,\epsilon}\|^2 + \tau^2\|d(x^*_{\nu,\epsilon}) - d(x^t)\|^2 \\ &\quad - 2\tau(1-\tau\epsilon)(\lambda^t - \lambda^*_{\nu,\epsilon})^T\left(d(x^*_{\nu,\epsilon}) - d(x^t)\right).\end{aligned}$$

By invoking the co-coercivity of $-d(x)$ from Lemma 4.1, we further obtain

$$\|\lambda^{t+1} - \lambda^*_{\nu,\epsilon}\|^2 \leq (1-\tau\epsilon)^2\|\lambda^t - \lambda^*_{\nu,\epsilon}\|^2 + \left(\tau^2 - 2\tau(1-\tau\epsilon)\frac{\nu}{M_d^2}\right)\|d(x^*_{\nu,\epsilon}) - d(x^t)\|^2.$$

A contraction may be obtained by choosing τ such that $(\tau^2 - 2\tau(1-\tau\epsilon)\frac{\nu}{M_d^2}) < 0$

$$\text{and } \tau < \frac{1}{\epsilon} \text{ as given by } \tau < \frac{2\nu/M_d^2}{1 + 2\epsilon\nu/M_d^2} < \frac{1}{\epsilon}.$$

We therefore conclude that $\|\lambda^{t+1} - \lambda^*_{\nu,\epsilon}\|^2 \leq (1-\tau\epsilon)^2\|\lambda^t - \lambda^*_{\nu,\epsilon}\|^2$ for all $t \geq 0$. ☐

Next, we examine two remaining concerns. First, can a bound on the norm $\|x^t - x^*_{\nu,\epsilon}\|$ be obtained, where $x^t = x(\lambda^t)$? Second, can one make a rigorous statement regarding the infeasibility of x^t, similar to that provided in the context of the primal-dual method in Section 3?

PROPOSITION 4.3. *Let Assumptions 1 and 2 hold. Then, for the sequence $\{x^t\}$, with $x^t = x(\lambda^t)$, generated by the dual method* (4.1) *using a step-size τ such that $\tau < \frac{2\nu}{M_d^2 + 2\epsilon\nu}$, we have for all $t \geq 0$,*

$$\|x^t - x^*_{\nu,\epsilon}\| \leq \frac{M_d}{\nu}\|\lambda^t - \lambda^*_{\nu,\epsilon}\| \qquad \textit{and} \qquad \max\{0, d_j(x^t)\} \leq \frac{M_d^2}{\nu}\|\lambda^t - \lambda^*_{\nu,\epsilon}\|.$$

Proof. From relation (4.4) in the proof of Lemma 4.1, the Cauchy-Schwartz inequality and the boundedness of $\nabla d_j(x)$ for all $j = 1, \ldots, m$, we have

$$\begin{aligned}\|x^t - x^*_{\nu,\epsilon}\|^2 &= \|x(\lambda^t) - x(\lambda^*_{\nu,\epsilon})\|^2 \\ &\leq \frac{1}{\nu}\left((\lambda^t - \lambda^*_{\nu,\epsilon})^T(-d(x(\lambda^t)) + d(x(\lambda^*_{\nu,\epsilon})))\right) \\ &\leq \frac{M_d}{\nu}\|\lambda^t - \lambda^*_{\nu,\epsilon}\|\|x(\lambda^t) - x(\lambda^*_{\nu,\epsilon})\|,\end{aligned}$$

implying that $\|x^t - x^*_{\nu,\epsilon}\| \leq \frac{M_d}{\nu}\|\lambda^t - \lambda^*_{\nu,\epsilon}\|$. Furthermore, a bound on $\max\{0, d_j(x^t)\}$ can be obtained by invoking the convexity of each of the functions d_j and the boundedness of their gradients, as follows:

$$d_j(x^t) \leq d_j(x^*_{\nu,\epsilon}) + \nabla d_j(x^*_{\nu,\epsilon})^T(x^t - x^*_{\nu,\epsilon}) \leq M_d\|x^t - x^*_{\nu,\epsilon}\| \leq \frac{M_d^2}{\nu}\|\lambda^t - \lambda^*_{\nu,\epsilon}\|,$$

where in the second inequality we use $d_j(x^*_{\nu,\epsilon}) \leq 0$. Thus, a bound on the violation of constraints $d_j(x) \leq 0$ at $x = x^t$ is given by $\max\{0, d_j(x^t)\} \leq \frac{M_d^2}{\nu}\|\lambda^t - \lambda^*_{\nu,\epsilon}\|$. □

4.2. Regularized inexact dual method. The *exact* dual scheme requires solving the Lagrangian subproblem to optimality for a given value of the Lagrange multiplier. In practical settings, primal solutions are obtained via distributed iterative schemes and exact solutions are inordinately expensive from a computational standpoint. This motivates our study of the error properties resulting from solving the Lagrangian subproblem inexactly for every iteration in dual space. In particular, we consider a method executing a specified *fixed number of iterations*, say K, in the primal space for every iteration in the dual space. Our intent is to provide error bounds contingent on K. The inexact form of the dual method is given by the following:

$$x^{k+1}(\lambda^t) = \Pi_X(x^k(\lambda^t) - \alpha\nabla_x\mathcal{L}_{\nu,\epsilon}(x^k(\lambda^t), \lambda^t)) \qquad k = 0, \ldots, K-1,\ t \geq 0, \quad (4.5)$$

$$\lambda^{t+1} = \Pi_{\mathcal{D}_\nu}\left(\lambda^t + \tau\nabla_\lambda\mathcal{L}_{\nu,\epsilon}(x^K(\lambda^t), \lambda^t)\right) \qquad t \geq 0. \quad (4.6)$$

Throughout this section, we omit the explicit dependence of x on λ, by letting $x^k(t) \triangleq x^k(\lambda^t)$. We have the following result.

LEMMA 4.4. *Let Assumptions 1–4 hold. Let $\{x^k(t)\}, k = 1, \ldots, K, t \geq 0$ be generated by (4.5) using a step-size α, with $0 < \alpha < \frac{2}{L_f}$ where $L_f = L + \nu + M_\nu L_d$, L and L_d are Lipschitz constants for the gradient maps ∇f and ∇d respectively, while $M_\nu = \max_{\lambda\in\mathcal{D}_\nu}\|\lambda\|$. Then, we have for all t and all $k = 1, \ldots, K$,*

$$\|x^k(t) - x(t)\| \leq q_p^{k/2}\|x^0(t) - x(t)\|,$$

where $x(t) := x(\lambda^t)$ solves the Lagrangian subproblem corresponding to the multiplier λ^t and $q_p = 1 - \alpha\nu(2 - \alpha L_f)$.

Proof. We observe that for each λ_t the mapping $\nabla_x\mathcal{L}_{\nu,\epsilon}(x^k(\lambda^t), \lambda^t)$ of the Lagrangian subproblem is strongly monotone and Lipschitz continuous. The geometric convergence follows directly from [10], page 164, Theorem 13.1. □

Our next proposition provides a relation for $\|\lambda^{t+1} - \lambda^*_{\nu,\epsilon}\|$ in terms of $\|\lambda^t - \lambda^*_{\nu,\epsilon}\|^2$ with an error bound depending on K and t.

PROPOSITION 4.5. *Let Assumptions 2–4 hold. Let the sequence $\{\lambda^t\}$ be generated by (4.5)–(4.6) using a step-size α as in Lemma 4.4 and a step-size τ such that*

$$\tau < \min\left\{\frac{2\nu}{M_d^2 + 2\epsilon\nu}, \frac{2\epsilon}{1+\epsilon^2}\right\} \quad \text{with} \quad M_d = \max_{x\in X}\|\nabla d\|.$$

We then have for all $t \geq 0$,

$$\|\lambda^{t+1} - \lambda^*_{\nu,\epsilon}\|^2 \leq q_d^{t+1}\|\lambda^0 - \lambda^*_{\nu,\epsilon}\|^2 + \frac{1 - q_d^{t+1}}{1 - q_d}M_d^2\left(q_d q_p^K + 2\tau^2 M_x\, q_p^{K/2}\right),$$

where $q_p = 1 - \alpha\nu(2 - \alpha L_f)$, $q_d = (1 - \tau\epsilon)^2 + \tau^2$, and $M_x = \max_{x,y\in X}\|x - y\|$.

Proof. In view of (4.6) and the non-expansive property of the projection, we have

$$\begin{aligned}\|\lambda^{t+1}-\lambda^*_{\nu,\epsilon}\|^2 &\le \left\|(1-\tau\epsilon)(\lambda^t-\lambda^*_{\nu,\epsilon})-\tau\left(d(x^*_{\nu,\epsilon})-d(x^K(t))\right)\right\|^2\\ &= (1-\tau\epsilon)^2\|\lambda^t-\lambda^*_{\nu,\epsilon}\|^2+\tau^2\underbrace{\|d(x^*_{\nu,\epsilon})-d(x^K(t))\|^2}_{\textbf{Term 1}}\\ &\quad -2\tau(1-\tau\epsilon)\underbrace{(\lambda^t-\lambda^*_{\nu,\epsilon})^T\left(d(x^*_{\nu,\epsilon})-d(x^K(t))\right)}_{\textbf{Term 2}}. \end{aligned} \tag{4.7}$$

Next, we provide bounds on terms 1 and 2. For term 1 by adding and subtracting $d(x(t))$, we obtain

$$\begin{aligned}&\|d(x^*_{\nu,\epsilon})-d(x^K(t))\|^2=\|d(x^*_{\nu,\epsilon})-d(x(t))+d(x(t))-d(x^K(t))\|^2\\ &\le \|d(x^*_{\nu,\epsilon})-d(x(t))\|^2+\|d(x(t))-d(x^K(t))\|^2+2\|d(x^*_{\nu,\epsilon})-d(x(t))\|\|d(x(t))-d(x^K(t))\|.\end{aligned}$$

By using the Lipschitz continuity of $d(x)$ for $x\in X$, we further have for all $t\ge 0$,

$$\begin{aligned}\|d(x^*_{\nu,\epsilon})-d(x^K(t))\|^2 &\le \|d(x^*_{\nu,\epsilon})-d(x(t))\|^2+\|d(x(t))-d(x^K(t))\|^2\\ &\quad +2M_d^2\|x^*_{\nu,\epsilon}-x(t)\|\,\|x(t)-x^K(t)\|.\end{aligned} \tag{4.8}$$

Now, we consider term 2, for which by adding and subtracting $d(x(t))$, and by using the co-coercivity of $-d(x(\lambda))$ (see (4.4)), we obtain

$$\begin{aligned}&(\lambda^t-\lambda^*_{\nu,\epsilon})^T\left(d(x^*_{\nu,\epsilon})-d(x^K(t))\right)\\ &=(\lambda^t-\lambda^*_{\nu,\epsilon})^T\left(d(x^*_{\nu,\epsilon})-d(x(t))\right)+(\lambda^t-\lambda^*_{\nu,\epsilon})^T\left(d(x(t))-d(x^K(t))\right)\\ &\ge \frac{\nu}{M_d^2}\|d(x^*_{\nu,\epsilon})-d(x(t))\|^2+(\lambda^t-\lambda^*_{\nu,\epsilon})^T\left(d(x(t))-d(x^K(t))\right).\end{aligned}$$

Thus, we have

$$\begin{aligned}&-2\tau(1-\tau\epsilon)(\lambda^t-\lambda^*_{\nu,\epsilon})^T\left(d(x^*_{\nu,\epsilon})-d(x^K(t))\right)\le -2\tau(1-\tau\epsilon)\frac{\nu}{M_d^2}\|d(x^*_{\nu,\epsilon})-d(x(t))\|^2\\ &+\tau^2\|\lambda^t-\lambda^*_{\nu,\epsilon}\|^2+(1-\tau\epsilon)^2\left\|d(x(t))-d(x^K(t))\right\|^2.\end{aligned} \tag{4.9}$$

From relations (4.7), (4.8) and (4.9), by grouping the corresponding expressions accordingly, we obtain

$$\begin{aligned}\|\lambda^{t+1}-\lambda^*_{\nu,\epsilon}\|^2 &\le \left((1-\tau\epsilon)^2+\tau^2\right)\|\lambda^t-\lambda^*_{\nu,\epsilon}\|^2\\ &\quad+\left(\tau^2-2\tau(1-\tau\epsilon)\frac{\nu}{M_d^2}\right)\|d(x^*_{\nu,\epsilon})-d(x(t))\|^2\\ &\quad+\left((1-\tau\epsilon)^2+\tau^2\right)\|d(x(t))-d(x^K(t))\|^2\\ &\quad+2\tau^2M_d^2\|x^*_{\nu,\epsilon}-x(t)\|\,\|x(t)-x^K(t)\|.\end{aligned}$$

By Lemma 4.4, we have $\|x^K(t)-x(t)\|\le q_p^{K/2}\|x^0(t)-x(t)\|$ with $q_p=1-\alpha\nu(2-\alpha L_f)$. By using this, the Lipschitz continuity of $d(x)$ over X, and $\|x^*_{\nu,\epsilon}-x(t)\|\le M_x$ where $M_x=\max_{x,y\in X}\|x-y\|$, we obtain

$$\begin{aligned}\|\lambda^{t+1}-\lambda^*_{\nu,\epsilon}\|^2 &\le \left((1-\tau\epsilon)^2+\tau^2\right)\|\lambda^t-\lambda^*_{\nu,\epsilon}\|^2+\left(\tau^2-2\tau(1-\tau\epsilon)\frac{\nu}{M_d^2}\right)M_d^2M_x^2\\ &\quad+\left((1-\tau\epsilon)^2+\tau^2\right)M_d^2q_p^K+2\tau^2M_d^2M_x\,q_p^{K/2}.\end{aligned}$$

By choosing τ such that

$$\tau < \min\left\{\frac{2\nu}{M_d^2 + 2\epsilon\nu}, \frac{2\epsilon}{1+\epsilon^2}\right\},$$

we ensure that $(1-\tau\epsilon)^2 + \tau^2 < 1$ and $\tau^2 - 2\tau(1-\tau\epsilon)\frac{\nu}{M_d^2} < 0$. Therefore for such a τ, by letting $q_d = (1-\tau\epsilon)^2 + \tau^2$, we have

$$\|\lambda^{t+1} - \lambda^*_{\nu,\epsilon}\|^2 \le q_d\|\lambda^t - \lambda^*_{\nu,\epsilon}\|^2 + q_d M_d^2 q_p^K + 2\tau^2 M_d^2 M_x\, q_p^{K/2},$$

and by recursively using the preceding estimate, we obtain

$$\|\lambda^{t+1} - \lambda^*_{\nu,\epsilon}\|^2 \le q_d^{t+1}\|\lambda^0 - \lambda^*_{\nu,\epsilon}\|^2 + \frac{1-q_d^{t+1}}{1-q_d} M_d^2\left(q_d q_p^K + 2\tau^2 M_x\, q_p^{K/2}\right).$$

□

Note that by Proposition 4.5, we have $\lim_{K\to\infty} q_p^K = 0$ since $q_p < 1$ and, hence, the term $\frac{1-q_d^{t+1}}{1-q_d} M_d^2(q_d q_p^K + 2\tau^2 M_x\, q_p^{K/2})$ converges to zero. This is precisely what we expect: as the Lagrangian problem is solved to a greater degree of exactness, the method approaches the exact regularized counterpart of section 4.1. Also, note that when K is fixed the following limiting error holds

$$\lim_{t\to\infty} \|\lambda^{t+1} - \lambda^*_{\nu,\epsilon}\|^2 \le \frac{1}{1-q_d} M_d^2\left(q_d q_p^K + 2\tau^2 M_x\, q_p^{K/2}\right).$$

We now establish bounds on the norm $\|x^K(t) - x^*_{\nu,\epsilon}\|$ and the constraint violation $d_j(x)$ at $x = x^K(t)$ for all j.

Proposition 4.6. *Under assumptions of Proposition 4.5, for the sequence $\{x^K(t)\}$ generated by (4.5)–(4.6) we have for all $t \ge 0$,*

$$\|x^K(t) - x^*_{\nu,\epsilon}\| \le q_p^{K/2} M_x + \frac{M_d}{\nu}\|\lambda^t - \lambda^*_{\nu,\epsilon}\|,$$
$$\max\{0, d_j(x^K(t))\} \le M_d\left(q_p^{K/2} M_x + \frac{M_d}{\nu}\|\lambda^t - \lambda^*_{\nu,\epsilon}\|\right),$$

where q_p, M_x and M_d are as defined in Proposition 4.5.

Proof. Consider $\|x^K(t) - x^*_{\nu,\epsilon}\|$. By Lemma 4.4 we have $\|x^K(t) - x(t)\| \le q_p^{K/2}\|x^0(t) - x(t)\|$, while by co-coercivity of $-d(x)$, it can be seen that $\|x(t) - x^*_{\nu,\epsilon}\| \le \frac{M_d}{\nu}\|\lambda^t - \lambda^*_{\nu,\epsilon}\|$. Hence,

$$\|x^K(t) - x^*_{\nu,\epsilon}\| \le \|x^K(t) - x(t)\| + \|x(t) - x^*_{\nu,\epsilon}\| \le q_p^{K/2} M_x + \frac{M_d}{\nu}\|\lambda^t - \lambda^*_{\nu,\epsilon}\|,$$

where we also use $\|x^0(t) - x(t)\| \le M_x$. For the constraint d_j, by convexity of d_j and using $d_j(x^*_{\nu,\epsilon})$ we have for any $t \ge 0$,

$$\begin{aligned} d_j(x^K(t)) &\le d_j(x^*_{\nu,\epsilon}) + \nabla d(x^*_{\nu,\epsilon})^T(x^K(t) - x^*_{\nu,\epsilon}) \\ &\le \|\nabla d(x^*_{\nu,\epsilon})\|\,\|x^K(t) - x^*_{\nu,\epsilon}\| \le M_d\left(q_p^{K/2} M_x + \frac{M_d}{\nu}\|\lambda^t - \lambda^*_{\nu,\epsilon}\|\right), \end{aligned}$$

where in the last inequality we use the preceding estimate for $\|x^K(t) - x^*_{\nu,\epsilon}\|$. Thus, for the violation of $d_j(x)$ at $x = x^K(t)$ we have,

$$\max\{0, d_j(x^K(t))\} \leq M_d \left(q_p^{K/2} M_x + \frac{M_d}{\nu} \|\lambda^t - \lambda^*_{\nu,\epsilon}\| \right).$$

□

One may combine the result of Proposition 4.6 with the estimate for $\|\lambda^t - \lambda_{\nu,\epsilon}\|$ of Proposition 4.5 to bound the norm $\|x^K(t) - x(t)\|$ and the constraint violation $\max\{0, d_j(x^K(t))\}$ in terms of initial multiplier λ^0 and the optimal dual solution $\lambda^*_{\nu,\epsilon}$.

An obvious challenge in implementing such schemes is that convergence relies on exact primal solutions. Often, there is a fixed amount of time available for obtaining primal updates, leading us to consider whether one could construct error bounds for dual schemes where an approximate primal solution is obtained through a fixed number of gradient steps.

Finally, we discuss an extension of the preceding results to the case of independently chosen regularization parameters. Analogous to Section 3.5, we extend the results of dual method to the case when user i selects a regularization parameter ν_i for its own Lagrangian subproblem. As in Section 3.5, the results follow straightforwardly from the results developed so far in this section. We briefly discuss the modified results here for completeness.

As in Section 3.5, Lagrange multiplier λ belongs to set $\mathcal{D}_{\nu_{\max}}$ defined in (3.24). In this case, similar to the proof of Lemma 4.1, it can be seen that the function $-d(x(\lambda))$ is co-coercive in λ with constant $\frac{\nu_{\min}}{M_d^2}$. The result of Proposition 4.2 will require the dual steplength τ to satisfy the following relation:

$$\tau < \frac{2\nu_{\min}}{M_d^2 + 2\epsilon\nu_{\min}}.$$

Similarly, the result of Proposition 4.3 will hold with $\nu_{\min}$ replacing the regularization parameter ν i.e., for τ such that $\tau < \frac{2\nu_{\min}}{M_d^2 + 2\epsilon\nu_{\min}}$, we have for all $t \geq 0$,

$$\|x^t - x^*_{V,\epsilon}\| \leq \frac{M_d}{\nu_{\min}} \|\lambda^t - \lambda^*_{V,\epsilon}\| \qquad \text{and} \qquad \max\{0, d_j(x^t)\} \leq \frac{M_d^2}{\nu_{\min}} \|\lambda^t - \lambda^*_{V,\epsilon}\|.$$

Finally, Lemma 4.4 will hold with L_f defined by $L_f = L + \nu_{\max} + M_\nu L_d$ and $q_p = 1 - \alpha\nu_{\min}(2 - \alpha L_f)$. Also, for the result of Proposition 4.5 to hold, the dual steplength τ should be required to satisfy

$$\tau < \min\left\{ \frac{2\nu_{\min}}{M_d^2 + 2\epsilon\nu_{\min}}, \frac{2\epsilon}{1 + \epsilon^2} \right\}.$$

5. Case Study. In this section, we report some experimental results for the algorithms developed in preceding sections. We use the `knitro` solver [6] on Matlab 7 to compute a solution of the problem and examine the performance of our proposed methods on a multiuser optimization problem involving a serial network with multiple links. The problem captures traffic and communication networks where users are characterized by utility/cost functions and are coupled through a congestion cost. This case manifests itself through delay arising from the link capacity constraints. In Section 5.1, we describe the underlying network structure and the user objectives and we present the numerical results for the primal-dual and dual methods, respectively. In each instance, an emphasis will be laid on determining the impact of the extensions, specifically independent primal and dual step-lengths and independent primal regularization (primal-dual), and inexact solutions of the Lagrangian subproblems (dual).

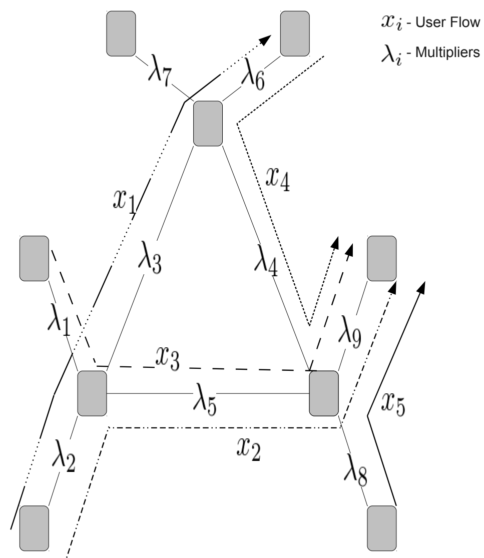


FIG. 5.1. *A network with 5 users and 9 links.*

5.1. Network and user data. The network comprises of a set of N users sharing a set $\mathscr{L}$ of links (see Fig. 5.1 for an illustration). A user $i \in N$ has a cost function $f_i(x_i)$ of its traffic rate x_i given by

$$f_i(x_i) = -k_i \log(1 + x_i) \qquad \text{for } i = 1, \ldots, N. \tag{5.1}$$

Each user selects an origin-destination pair of nodes on this network and faces congestion based on the links traversed along the prescribed path connecting the selected origin-destination nodes. We consider the congestion cost of the form:

$$c(x) = \sum_{i=1}^{N} \sum_{l \in \mathscr{L}} x_{li} \sum_{j=1}^{N} x_{lj}, \tag{5.2}$$

where, x_{lj} is the flow of user j on link l. The total cost of the network is given by

$$f(x) = \sum_{i=1}^{N} f_i(x) + c(x) = \sum_{i=1}^{N} -k_i \log(1 + x_i) + \sum_{i=1}^{N} \sum_{l \in \mathscr{L}} x_{li} \sum_{j=1}^{N} x_{lj}.$$

Let A denote the adjacency matrix that specifies the set of links traversed by the traffic generated by the users. More precisely, $A_{li} = 1$ if traffic of user i goes through link l and 0 otherwise. It can be seen that $\nabla c(x) = 2A^T Ax$ and thus the Lipschitz constant of the gradient map $\nabla f(x)$ is given by $L = \sqrt{\sum_i k_i^2} + 2\|A^T A\|$. Throughout this section, we consider a network with 9 links and 5 users. Table 5.1

User(i)	Links traversed	k_i
1	L2, L3, L6	10
2	L2, L5, L9	0
3	L1, L5, L9	10
4	L6, L4, L9	10
5	L8, L9	10

TABLE 5.1
Network and User Data

summarizes the traffic in the network as generated by the users and the parameters k_i of the user objective. The user traffic rates are coupled through the constraint of the form $\sum_{i=1}^{N} A_{li}x_i \leq C_l$ for all $l \in \mathscr{L}$, where C_l is the maximum aggregate traffic through link l. The constraint can be compactly written as $Ax \leq C$, where C is the link capacity vector and is given by $C = (10, 15, 20, 10, 15, 20, 20, 15, 25)$.

Regularized Primal-Dual Method. Figure 5.2 shows the number of iterations required to attain a desired error level for $\|z^k - z^*_{\nu,\epsilon}\|$ with $\{z^k\}$ generated by primal-dual algorithm (3.22) for different values of the step-size ratio $\beta = \alpha/\tau$ between the primal step-size α and dual stepsize τ. Note that in this case each user has the same step-size and the regularization parameter. Relations in Lemma 3.7 are used to obtain the theoretical range for the ratio parameter β and the corresponding step-lengths. The regularization parameters ν and ϵ were both set at 0.1, such that $\mu = \min\{\nu, \epsilon\} = 0.1$ and the algorithm was terminated when $\|z^k - z^*_{\nu,\epsilon}\| \leq 10^{-3}$. It can be observed that the number of iterations required for convergence decreases as the step-size ratio of approaches the value 1.

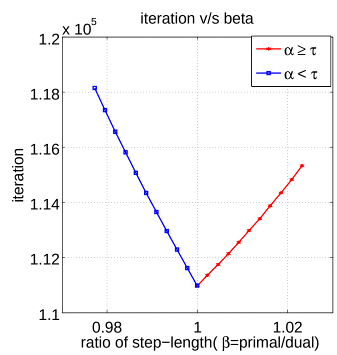


FIG. 5.2. *Performance of Primal-Dual Method for independent step-sizes in primal and dual space.*

Figure 5.3 illustrates the performance of the primal-dual algorithm in terms of the number of iterations required to attain $\|z^k - z^*_{\nu,\epsilon}\| < 10^{-3}$ as the steplength deviation in primal space $\alpha_{\max} - \alpha_{\min}$ increases. All users employ the same regularization parameter $\nu_i = \nu = 0.1$ and the dual regularization parameter ϵ is chosen to be 0.1. The plot demonstrates that, as the deviation between users' step-sizes increases, the number of iteration for a desired accuracy also increases.

Next, we let each user choose its own regularization parameter ν_i with uniform distribution over interval $(\nu_{\min}, 0.1)$ for a given $\nu_{\min} \leq 0.1$. Figure 5.4 shows the performance of the primal-dual algorithm in terms of the number of iterations required to attain the error $\|z^k - z^*_{V,\epsilon}\| < 10^{-3}$ as $\nu_{\min}$ is varied from 0.01 to 0.1. The dual

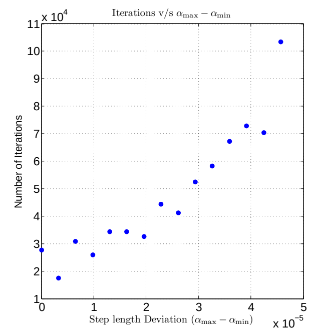


FIG. 5.3. *Performance of Primal-Dual method for deviation in user step-size.*

steplength was set at $\tau = 1.9\mu/L_\Phi^2$, where $\mu = \min\{\nu_{\min}, \epsilon\}$ with $\epsilon = 0.1$. The primal stepsizes that users employ are the same across the users and are given by $\alpha_i = \alpha = \beta\tau$, where β is as given in Lemma 3.7. As expected, the number of iterations increases when $\nu_{\min}$ decreases.

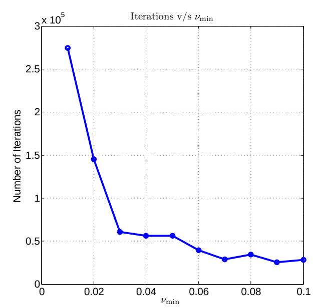
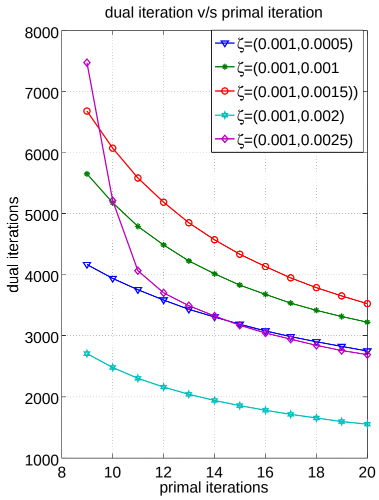


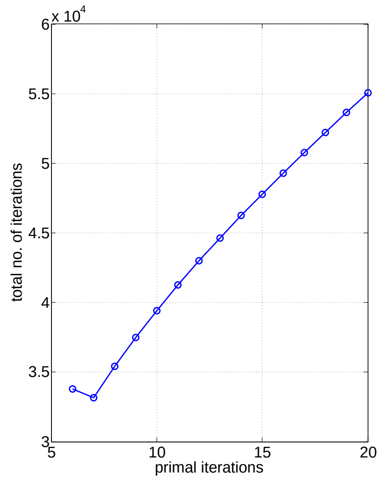
FIG. 5.4. *Performance of Primal-Dual method as user minimum regularization parameter* $\nu_{\min}$ *varies.*

Regularized Dual Method. Figure 5.5(a) compares dual iterations required to reach an accuracy level of $\|\lambda^k - \lambda^*_{\nu,\epsilon}\| \leq 10^{-6}$ for each K where $\{\lambda^k\}$ is generated using dual method (4.6) and K is the number of iterations in the primal space for each λ^k. The regularization parameter ϵ is varied from 0.0005 to 0.0025, while ν is fixed at 0.001. The primal step-size is set at $\alpha = 0.25/L_f$ and the dual step-size is taken as $\tau = 0.75\nu/M_d^2$ (see Section 4). Faster dual convergence was observed as K was increased for all ranges of parameters tested. For the case when $\nu = 0.001$ and $\epsilon = 0.001$, Figure 5.5(b) shows the dependency of total number of iterations required (primal $\times$ dual) for $\|\lambda^k - \lambda^*_{\nu,\epsilon}\| \leq 10^{-6}$ as the number K of primal iterations is varied. It can be observed that beyond a threshold level for K, the total number of iterations starts increasing. In effect, the extra effort in obtaining increasingly exact solutions to the Lagrangian subproblem is not met with faster convergence in the dual space.

6. Summary and conclusions. This paper focuses on a class of multiuser optimization problems in which user interactions are seen in the user objectives (through congestion or delay functions) and in the coupling constraints (as arising from shared resources). Traditional algorithms rely on a high degree of separability and cannot



(a)



(b)

FIG. 5.5. *Inexact Dual Method: (a) Comparison of dual iterations for a fixed number K of primal iterations; (b) Dependency of the total number of primal and dual iterations as the number K of primal iterations varies.*

be directly employed. They also rely on coordination in terms of uniform or equal step-sizes across users. The coordination requirements have been weakened to various degrees in the present work, which considers primal-dual and dual gradient algorithms, derived from the fixed-point formulations of the regularized problem. These schemes are analyzed in an effort to make rigorous statements regarding convergence behavior as well as provide error bounds in regularized settings that limited coordination across step-length choices and inexact solutions. Our main contributions are summarized next:

(1) Regularized primal-dual method: Under suitable convexity assumptions, we consider a regularized primal-dual projection scheme and provide error bounds for the regularized solution and optimal function value with respect to their optimal counterparts. In addition, we also obtain a bound on the infeasibility for the regularized solution. We also show that, under some conditions, the method can be extended to allow not only for independent selection of primal and dual stepsizes as well as independently chosen steplengths by every user but also when users choose their regularization parameter independently.

(2) Regularized dual method: In contrast with (1), applying dual schemes would require an optimal primal solution for every dual step. We show the contractive nature of a regularized dual scheme reliant on exact primal solutions. Furthermore, we develop asymptotic error bounds where for each dual iteration, the primal method for solving the Lagrangian subproblem terminates after a fixed number of steps. We also provide error bounds for the obtained solution and Lagrange multiplier as well as an upper bound on the infeasibility. Finally, we extend these results to the case when each user independently chooses its regularization parameter.

It is of future interest to consider the algorithms proposed in [17, 19] as applied to multiuser problem, whereby the users are allowed to implement step-sizes within a prescribed range of values. For this, at first, we would have to develop the error bounds for the algorithms in [17, 19] for the case when different users employ different steplengths.

7. Appendix. LEMMA 7.1. *Let Assumption 2 hold. Then, for each $\nu > 0$, we have*

$$0 \le f(x_\nu^*) - f^* \le \frac{\nu}{2}(D^2 - \|x_\nu^*\|^2) \qquad \text{where } D = \max_{x \in X} \|x\|.$$

Proof. Under Assumption 2, both the original problem and the regularized problem have solutions. Since the regularized problem is strongly convex, the solution $x_\nu^* \in X$ is unique for every $\nu > 0$. Furthermore, we have

$$f_\nu(x_\nu^*) - f_\nu(x) \le 0 \qquad \text{for all } x \in X.$$

Letting $x = x^*$ in the preceding relation, and using $f_\nu(x) = f(x) + \frac{\nu}{2}\|x\|^2$ and $f^* = f(x^*)$ we get

$$f(x_\nu^*) - f^* \le \frac{\nu}{2}\left(\|x^*\|^2 - \|x_\nu^*\|^2\right).$$

Since $x^* \in X$ solves the original problem and $x_\nu^* \in X$, we have $0 \le f(x_\nu^*) - f(x^*)$. Thus, from $f^* = f(x^*)$, using $D = \max_{x \in X}\|x\|$, it follows that $0 \le f(x_\nu^*) - f^* \le \frac{\nu}{2}\left(\|x^*\|^2 - \|x_\nu^*\|^2\right) \le \frac{\nu}{2}(D^2 - \|x_\nu^*\|^2)$. ∎

REFERENCES

[1] A. BECK AND M. TEBOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202.

[2] ———, *Gradient-based algorithms with applications to signal-recovery problems*, in Convex Optimization in Signal Processing and Communications, Y. Eldar and D. Palomar, eds., Cambridge University Press, 2010, pp. 42–88.

[3] D. P. BERTSEKAS, *Nonlinear Programming*, Athena Scientific, Bellmont, MA, 1999.

[4] D. P. BERTSEKAS, A. NEDIĆ, AND A. OZDAGLAR, *Convex Analysis and Optimization*, Athena Scientific, Belmont, MA, 2003.

[5] D. P. BERTSEKAS AND J. TSITSIKLIS, *Parallel and Distributed Computation: Numerical Methods*, Prentice Hall Inc., 1989.

[6] R. H. BYRD, M. E. HRIBAR, AND J. NOCEDAL, *An interior point algorithm for large-scale nonlinear programming*, SIAM Journal Optimization, 9 (1999), pp. 877–900 (electronic). Dedicated to John E. Dennis, Jr., on his 60th birthday.

[7] M. CHIANG, S. LOW, A. CALDERBANK, AND J. DOYLE, *Layering as optimization decomposition: A mathematical theory of network architectures*, Proceedings of the IEEE, 95 (2007), pp. 255–312.

[8] F. FACCHINEI AND J.-S. PANG, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, vol. 1 and 2, Springer-Verlag Inc., New York, 2003.

[9] F. KELLY, A. MAULLOO, AND D. TAN, *Rate control for communication networks: shadow prices, proportional fairness, and stability*, Journal of the Operational Research Society, 49 (1998), pp. 237–252.

[10] I. KONNOV, *Methods for solving multivalued variational inequalities*, in At the turn of the 20th–21st centuries (Russian), Kazan. Mat. Obs., Kazan, 2003, pp. 68–79.

[11] ———, *Equilibrium models and variational inequalities*, Elsevier, 2007.

[12] G. KORPELEVICH, *The extragradient method for finding saddle points and other problems*, Matekon, 12 (1976), pp. 747–756.

[13] S. LOW AND D. LAPSLEY, *Optimization flow control, I: Basic algorithm and convergence*, IEEE/ACM Transactions on Networking, 7 (1999), pp. 861–874.

[14] A. NEDIĆ AND A. OZDAGLAR, *Subgradient methods in network resource allocation: Rate analysis*, in 42nd Annual Conference on Information Sciences and Systems, CISS, Princeton, 2008, pp. 1189–1194.

[15] ———, *Approximate primal solutions and rate analysis for dual subgradient methods*, SIAM Journal on Optimization, 19 (2009), pp. 1757–1780.

[16] ———, *Subgradient methods for Saddle-Point problems*, Journal of Optimization Theory and Applications, 142 (2009), pp. 205–228.

[17] A. Nemirovski, *Prox-method with rate of convergence O(1/T) for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems*, SIAM Journal On Optimization, 15 (2004), pp. 229–251.

[18] Y. Nesterov, *Smooth minimization of non-smooth functions*, Mathematical Programming, Ser. A, 103 (2005), pp. 127–152.

[19] Y. Nesterov, *Dual extrapolation and its applications to solving variational inequalities and related problems*, Mathematical Programming, 109 (2007), pp. 319–344.

[20] J.-S. Pang and D. Chan, *Iterative methods for variational and complementarity problems*, Mathematical Programming, 24 (1982), pp. 284–313.

[21] B. T. Polyak, *Introduction to Optimization*, Optimization Software, May 1987.

[22] X. Qin and R. Berry, *Distributed power allocation and scheduling for parallel channel wireless networks*, in Third International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 2005. WIOPT, 2005, pp. 77–85.

[23] S. Shakkottai and R. Srikant, *Network optimization and control*, Foundations and Trends in Networking, 2 (2007), pp. 271–379.

[24] M. Solodov and P. Tseng, *Modified projection-type methods for monotone variational inequalities*, SIAM Journal on Control and Optimization, 34 (1996), pp. 1814–1830.

[25] R. Srikant, *Mathematics of Internet Congestion Control*, Birkhauser, 2004.

[26] P. Tseng, *On accelerated proximal gradient methods for convex-concave optimization*. submitted to SIAM Journal on Optimization, 2008.

[27] J. Tsitsiklis, *Problems in decentralized decision making and computation*, PhD thesis, Massachusetts Institute of Technology, 1984.

[28] H. Uzawa, *Iterative methods in concave programming*, Studies in Linear and Nonlinear Programming, (1958), pp. 154–165.

[29] H. Yin, U. Shanbhag, and P. Mehta, *Nash equilibrium problems with congestion costs and shared constraints*, in Proceedings of the 48th IEEE Conference on Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009., 2009, pp. 4649–4654.

[30] H. Yin, U. V. Shanbhag, and P. G. Mehta, *Nash equilibrium problems with congestion costs and shared constraints*, To appear in IEEE Transactions of Automatic Control (July 2011), (2009).