Distributed primal-dual method for multi-agent sharing problem with conic constraints

Necdet Serhat Aybat
Industrial & Manufacturing Engineering
Pennsylvania State University
University Park, PA 16802
Email: nsa10@psu.edu

Erfan Yazdandoost Hamedani
Industrial & Manufacturing Engineering
Pennsylvania State University
University Park, PA 16802
Email: evy5047@psu.edu

***Abstract*—We consider cooperative multi-agent resource sharing problems over an undirected network of agents, where only those agents connected by an edge can directly communicate. The objective is to minimize the sum of agent-specific composite convex functions subject to a conic constraint that couples agents' decisions. A distributed primal-dual algorithm is proposed to solve the saddle point formulation, which requires to compute a consensus dual price for the coupling constraint. We provide convergence rates in sub-optimality, infeasibility and consensus violation for agents' dual price assessments; examine the effect of underlying network topology on the convergence rates of the proposed decentralized algorithm; and compare our method with Prox-JADMM algorithm on the basis pursuit problem.**

I. Introduction

Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ denote a *connected* undirected graph of N computing nodes, where $\mathcal{N} \triangleq \{1, \ldots, N\}$ and $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ is the set of edges – without loss of generality, suppose $(i, j) \in \mathcal{E}$ implies $i < j$. Suppose nodes i and j can exchange information only if $(i, j) \in \mathcal{E}$, and each node $i \in \mathcal{N}$ has a *private* (local) cost function $\Phi_i : \mathbb{R}^{n_i} \to \mathbb{R} \cup \{+\infty\}$ such that

$$\Phi_i(\xi_i) \triangleq \rho_i(\xi_i) + f_i(\xi_i), \tag{1}$$

where $\rho_i : \mathbb{R}^{n_i} \to \mathbb{R} \cup \{+\infty\}$ is a proper, closed convex function (possibly *non-smooth*), and $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ is a *smooth* convex function. Assume that f_i is differentiable on an open set containing $\mathbf{dom}\,\rho_i$ with a Lipschitz continuous gradient ∇f_i, of which Lipschitz constant is L_i; the prox map of ρ_i,

$$\mathbf{prox}_{\rho_i}(\xi_i) \triangleq \underset{x_i \in \mathbb{R}^{n_i}}{\operatorname{argmin}} \left\{\rho_i(x_i) + \tfrac{1}{2}\|x_i - \xi_i\|^2\right\}, \tag{2}$$

is *efficiently* computable for $i \in \mathcal{N}$, where $\|.\|$ denotes the Euclidean norm. Let $\mathcal{N}_i \triangleq \{j \in \mathcal{N} : (i, j) \in \mathcal{E} \text{ or } (j, i) \in \mathcal{E}\}$ denote the set of neighboring nodes of $i \in \mathcal{N}$, and $d_i \triangleq |\mathcal{N}_i|$ is the degree of node $i \in \mathcal{N}$. Let $\mathbb{R}^n \ni \boldsymbol{\xi} = [\xi_i]_{i\in\mathcal{N}}$ such that $n \triangleq \sum_{i\in\mathcal{N}} n_i$. Consider the following minimization problem:

$$\min_{\boldsymbol{\xi}\in\mathbb{R}^n} \sum_{i\in\mathcal{N}} \Phi_i(\xi_i) \quad \text{s.t.} \quad \sum_{i\in\mathcal{N}} R_i\xi_i - r_i \in \mathcal{K}, \tag{3}$$

where closed, convex cone $\mathcal{K} \subseteq R^m$, $R_i \in \mathbb{R}^{m\times n_i}$ and $r_i \in \mathbb{R}^m$ are the problem data such that each node $i \in \mathcal{N}$ only have access to R_i, r_i and $\mathcal{K}$ along with its objective $\Phi_i(\xi_i)$ defined in (1). Our objective is to solve (3) in a *decentralized* fashion using the computing nodes $\mathcal{N}$ and exchanging information only along the edges $\mathcal{E}$. In this paper, we only consider the case when the topology of the connectivity graph is *static* – in a recent preprint [1], we extended our results to *time-varying* connectivity graphs.

In the remainder of this section, we briefly discuss some previous work related to ours, and give a specific implementation of the primal-dual algorithm (PDA) proposed in [2]. Next, in Section II, we consider the multi-agent sharing problem in (3), provide a distributed implementation of PDA for solving (3), and establish error bounds under the following assumption.

Assumption 1. *A primal-dual optimal solution to* (3) *exists and the duality gap is 0.*

A. Related Work

Now we briefly review some recent work on the distributed solution of a resource sharing problem among a set of agents, $\mathcal{N}$, communicating over the network $G = (\mathcal{N}, \mathcal{E})$ where the objective is to minimize the sum of local convex functions subject to some coupling constraints on local decisions. In [3] an *asynchronous* distributed method based on ADMM is proposed. This algorithm can handle coupling constraints with very particular structure on a static network: $\min_{\boldsymbol{\xi},\mathbf{z}} \sum_{i\in\mathcal{N}} \Phi_i(\boldsymbol{\xi}_i)$ subject to $D\boldsymbol{\xi} + H\mathbf{z} = 0$, $\mathbf{z} \in Z$, and $\boldsymbol{\xi}_i \in X_i$ for $i \in \mathcal{N}$, where $\boldsymbol{\xi} = [\xi_i]_{i\in\mathcal{N}}$ is the vector of local decision variables, $\mathbf{z}$ is the coupling variable, X_i's and Z are closed convex sets, H is diagonal and invertible, and each row of D has exactly one nonzero element while D has no columns of all zeros. Under the *compactness* assumption on $X = \Pi_{i\in\mathcal{N}} X_i$ and Z, it is shown that the method has $\mathcal{O}(1/k)$ convergence rate in terms of *expected* suboptimality and feasibility violation – in each iteration *exact* minimizations involving Φ_i is needed.

In [4], a method based on ADMM is proposed to reduce the computational work of ADMM due to exact minimizations in each iteration. First, a dual consensus ADMM is proposed for solving (3) in a distributed fashion when $\mathcal{K} = \{\mathbf{0}\}$, and $\Phi_i(\xi_i) = \rho_i(\xi_i) + f_i(A_i\xi_i)$ for ρ_i and f_i as in (1). Under strong duality assumption, it is shown that dual iterate sequence converges and every limit point of the primal sequence is optimal without giving a rate result. Next, to avoid exact minimizations in ADMM, an inexact variant taking proximal-gradient steps is analyzed. Convergence of primal-dual sequence is shown when each f_i is strongly convex – without a rate result; and a linear rate is given in the absence of the non-smooth ρ_i, i.e., $\Phi_i(\xi_i) = f_i(A_i\xi_i)$, and when each E_i has full row-rank and Φ_i is strongly convex. In [5] a proximal dual consensus ADMM method is proposed by Chang, under static network topology assumption. The objective is to minimize the sum of convex functions $\sum_{i\in\mathcal{N}}\Phi_i$ subject to coupling equality and agent-specific constraints. Each agent-specific set is assumed to be an intersection of a polyhedron and a "simple" set. The polyhedral constraints are handled using a penalty formulation without requiring projection onto them. Chang also proposed a randomized variant which can handle randomly on/off agents and imperfect communication links. It is shown that both algorithms have $\mathcal{O}(1/k)$ convergence rate; that said, in each iteration, costly *exact* minimizations involving Φ_i is needed.

In [6], a general setting for constrained distributed optimization in a time-varying network topology has been considered to minimize a composition of a global network function (smooth) with the summation of local objective functions (smooth), subject to inequality constraints on the summation of agent specific constrained functions and local compact sets. They propose a consensus-based distributed primal-dual perturbation (PDP) algorithm and show that the local primal-dual iterates converges to a global optimal primal-dual solution; however, no rate result was provided. The proposed PDP method can also handle non-smooth constraints with similar convergence guarantees. More recently, in [7], distributed continuous-time coordination algorithms are proposed to minimize convex separable objective subject to coupling equality and convex inequality constraints. Assuming the objective and constraint functions are Lipschitz, point-wise convergence is established without providing a rate result.

Finally, while we were preparing this paper, we became aware of a recent work [8], which also use dual consensus formulation to decompose separable constraints. A distributed algorithm on time-varying directed communication networks is proposed for solving saddle-point problems subject to consensus constraints. The algorithm can also be applied to solve consensus optimization problems with inequality constraints that can be written as summation of local convex functions of local and global variables. Assuming each agents local iterates and subgradient sets are uniformly bounded, it is shown that the ergodic average of primal-dual sequence converges with $\mathcal{O}(1/\sqrt{k})$ rate in terms of saddle-point evaluation error; however, when applied to constrained optimization problems, no rate in terms of suboptimality and infeasibility is provided.

B. Preliminary

There has been active research on developing efficient algorithms for convex-concave saddle point problems $\min_{\mathbf{x}}\max_{\mathbf{y}}\mathcal{L}(\mathbf{x},\mathbf{y})$, e.g., [9], [10], [11], [12]. Recently, a primal-dual algorithm (PDA) is proposed in [2] for the following composite convex-concave saddle-point problem:

$$\min_{\mathbf{x}\in\mathcal{X}}\max_{\mathbf{y}\in\mathcal{Y}}\mathcal{L}(\mathbf{x},\mathbf{y}) \triangleq \Phi(\mathbf{x}) + \langle T\mathbf{x},\mathbf{y}\rangle - h(\mathbf{y}), \tag{4}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are finite-dimensional vector spaces, $\Phi(\mathbf{x}) \triangleq \rho(\mathbf{x}) + f(\mathbf{x})$, ρ and h are possibly non-smooth convex functions, f is a convex function and has a Lipschitz continuous gradient defined on $\mathbf{dom}\,\rho$ with constant L. Briefly, given $\mathbf{x}^0,\mathbf{y}^0$ and algorithm parameters $\nu_x,\nu_y>0$, PDA consists of two proximal-gradient steps:

$$\mathbf{x}^{k+1} \leftarrow \operatorname*{argmin}_{\mathbf{x}}\ \rho(\mathbf{x}) + f(\mathbf{x}^k) + \langle\nabla f(\mathbf{x}^k),\ \mathbf{x}-\mathbf{x}^k\rangle + \langle T\mathbf{x},\mathbf{y}^k\rangle + \tfrac{1}{\nu_x}D_x(\mathbf{x},\mathbf{x}^k) \tag{5a}$$

$$\mathbf{y}^{k+1} \leftarrow \operatorname*{argmin}_{\mathbf{y}}\ h(\mathbf{y}) - \langle T(2\mathbf{x}^{k+1}-\mathbf{x}^k),\mathbf{y}\rangle + \tfrac{1}{\nu_y}D_y(\mathbf{y},\mathbf{y}^k), \tag{5b}$$

where D_x and D_y are Bregman distance functions corresponding to some continuously differentiable strongly convex functions ψ_x and ψ_y such that $\mathbf{dom}\,\psi_x \supset \mathbf{dom}\,\rho$ and $\mathbf{dom}\,\psi_y \supset \mathbf{dom}\,h$. In particular, $D_x(\mathbf{x},\bar{\mathbf{x}}) \triangleq \psi_x(\mathbf{x}) - \psi_x(\bar{\mathbf{x}}) - \langle\nabla\psi_x(\bar{\mathbf{x}}),\ \mathbf{x}-\bar{\mathbf{x}}\rangle$, and D_y is defined similarly. In [2], a simple proof for the ergodic convergence is provided; indeed, it is shown that, when the convexity modulus for ψ_x and ψ_y is 1, if $\nu_x,\nu_y>0$ are chosen such that $(\frac{1}{\nu_x}-L)\frac{1}{\nu_y} \geq \sigma^2_{\max}(T)$, then for any $\mathbf{x},\mathbf{y}\in\mathcal{X}\times\mathcal{Y}$, the following holds for all $K\geq 1$:

$$\mathcal{L}(\bar{\mathbf{x}}^K,\mathbf{y}) - \mathcal{L}(\mathbf{x},\bar{\mathbf{y}}^K) \leq \frac{1}{K}\Big(\frac{1}{\nu_x}D_x(\mathbf{x},\mathbf{x}^0) + \frac{1}{\nu_y}D_y(\mathbf{y},\mathbf{y}^0) - \langle T(\mathbf{x}-\mathbf{x}^0),\mathbf{y}-\mathbf{y}^0\rangle\Big) \tag{6}$$

where $\bar{\mathbf{x}}^K \triangleq \frac{1}{K}\sum_{k=1}^K\mathbf{x}^k$ and $\bar{\mathbf{y}}^K \triangleq \frac{1}{K}\sum_{k=1}^K\mathbf{y}^k$.

First, in Theorem 1, we discuss a special case of (4), which will help us develop a *decentralized* algorithm for the sharing problem in (3), and allow us to establish its convergence. The proposed algorithm can distribute the computation over the nodes such that each node's computation is based on the local topology of $\mathcal{G}$ and information only available to that node.

It is worth mentioning the connection between PDA and the alternating direction method of multipliers (ADMM). Indeed, under certain settings, (preconditioned) ADMM is equivalent to PDA [2], [10]. There is also a strong connection between the linearized ADMM algorithm, PG-ADMM, proposed by Aybat et al. [13] and PDA – for details of these relations, see [1].

Notation. Throughout the paper, $\|\cdot\|$ denotes the Euclidean or the spectral norm depending on its argument, i.e., for a matrix R, $\|R\| = \sigma_{\max}(R)$. Given a convex set $\mathcal{S}$, let $\mathbb{1}_{\mathcal{S}}(\cdot)$ denote the indicator function of $\mathcal{S}$, i.e., $\mathbb{1}_{\mathcal{S}}(w)=0$ for $w\in\mathcal{S}$ and equal to $+\infty$ otherwise, and let $\mathcal{P}_{\mathcal{S}}(w) \triangleq \operatorname{argmin}\{\|v-w\| : v\in\mathcal{S}\}$ denote the projection onto $\mathcal{S}$. For a closed convex set $\mathcal{S}$, we define the distance function as $d_{\mathcal{S}}(w) \triangleq \|\mathcal{P}_{\mathcal{S}}(w)-w\|$. Given a convex cone $\mathcal{K}\in\mathbb{R}^m$, let $\mathcal{K}^*$ denote its dual cone, i.e., $\mathcal{K}^* \triangleq \{\theta\in\mathbb{R}^m :\ \langle\theta,w\rangle\geq 0\ \ \forall w\in\mathcal{K}\}$, and $\mathcal{K}^\circ \triangleq -\mathcal{K}^*$ denote the polar cone of $\mathcal{K}$. Cone $\mathcal{K}$ is called *proper* if it is closed, convex, pointed, and it has a nonempty interior. Given a convex function $g:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$, its convex conjugate is $g^*(w) \triangleq \sup_{\theta\in\mathbb{R}^n}\langle w,\theta\rangle - g(\theta)$. $\otimes$ denotes the Kronecker product, and $\mathbf{I}_n$ is the $n\times n$ identity matrix.

Definition 1. *Let* $\mathcal{X} \triangleq \Pi_{i\in\mathcal{N}}\mathbb{R}^{n_i}\times\mathbb{R}^{n_0}$ *and* $\mathcal{X}\ni\mathbf{x} = [\boldsymbol{\xi}^\top\mathbf{w}^\top]^\top$ *and* $\boldsymbol{\xi} = [\xi_i]_{i\in\mathcal{N}}$*; and* $\mathcal{Y} \triangleq \Pi_{i\in\mathcal{N}}\mathbb{R}^m$, $\mathcal{Y}\ni\mathbf{y} = [y_i]_{i\in\mathcal{N}}$*, and* Π *denotes the Cartesian product. Given parameters* $\gamma>0$*, and* $\tau_i,\kappa_i>0$ *for* $i\in\mathcal{N}$*, let* $\mathbf{D}_\gamma \triangleq \frac{1}{\gamma}\mathbf{I}_{n_0}$, $\mathbf{D}_\tau \triangleq \mathbf{diag}([\frac{1}{\tau_i}\mathbf{I}_{n_i}]_{i\in\mathcal{N}})$*, and* $\mathbf{D}_\kappa \triangleq \mathbf{diag}([\frac{1}{\kappa_i}\mathbf{I}_m]_{i\in\mathcal{N}})$*. Defining* $\psi_x(\mathbf{x}) \triangleq \frac{1}{2}\boldsymbol{\xi}^\top\mathbf{D}_\tau\boldsymbol{\xi} + \frac{1}{2}\mathbf{w}^\top\mathbf{D}_\gamma\mathbf{w}$ *and* $\psi_y(\mathbf{y}) \triangleq \frac{1}{2}\mathbf{y}^\top\mathbf{D}_\kappa\mathbf{y}$ *leads to the following Bregman distance functions:* $D_x(\mathbf{x},\bar{\mathbf{x}}) =$

$\frac{1}{2}\left\|\boldsymbol{\xi}-\bar{\boldsymbol{\xi}}\right\|_{\mathbf{D}_\tau}^2+\frac{1}{2}\left\|\mathbf{w}-\bar{\mathbf{w}}\right\|_{\mathbf{D}_\gamma}^2$, *and* $D_y(\mathbf{y},\bar{\mathbf{y}})=\frac{1}{2}\left\|\mathbf{y}-\bar{\mathbf{y}}\right\|_{\mathbf{D}_\kappa}^2$, *where the Q-norm is defined as* $\|z\|_Q \triangleq (z^\top Q z)^{\frac{1}{2}}$ *for* $Q \succ 0$.

Theorem 1. *Let $\mathcal{X}$, $\mathcal{Y}$, and Bregman functions D_x, D_y be as in Definition 1. Suppose $\Phi_i=\rho_i+f_i$ is composite convex function defined as in* (1) *for $i\in\mathcal{N}$; $\rho_0:\mathbb{R}^{n_0}\to\mathbb{R}\cup\{+\infty\}$ and $h_i:\mathbb{R}^m\to\mathbb{R}\cup\{+\infty\}$ for $i\in\mathcal{N}$ are proper, closed convex functions with simple prox-maps. Let $\Phi(\mathbf{x})\triangleq\rho(\mathbf{x})+f(\mathbf{x})$ and $h(\mathbf{y})\triangleq\sum_{i\in\mathcal{N}}h_i(y_i)$, where $\rho(\mathbf{x})\triangleq\rho_0(\mathbf{w})+\sum_{i\in\mathcal{N}}\rho_i(\xi_i)$ and $f(\mathbf{x})\triangleq\sum_{i\in\mathcal{N}}f_i(\xi_i)$. Given matrix $T\in\mathbb{R}^{m|\mathcal{N}|\times(n+n_0)}$ and the initial point $(\mathbf{x}^0,\mathbf{y}^0)$, the PDA iterate sequence $\{\mathbf{x}^k,\mathbf{y}^k\}_{k\geq1}$, generated according to* (5) *when $\nu_x=\nu_y=1$ satisfies* (6) *for all $K\geq1$ if $\bar{\mathbf{Q}}\triangleq\begin{bmatrix}\mathbf{D} & -T^\top\\ -T & \mathbf{D}_\kappa\end{bmatrix}\succeq0$, where $\mathbf{D}\triangleq\begin{bmatrix}\bar{\mathbf{D}}_\tau & \mathbf{0}\\ \mathbf{0} & \mathbf{D}_\gamma\end{bmatrix}$, and $\bar{\mathbf{D}}_\tau\triangleq\mathbf{diag}([(\frac{1}{\tau_i}-L_i)\mathbf{I}_{n_i}]_{i\in\mathcal{N}})$. Moreover, if a saddle point exists for* (4)*, and $\bar{\mathbf{Q}}\succ0$, then $\{\mathbf{x}^k,\mathbf{y}^k\}_{k\geq1}$ converges to a saddle point of* (4)*; hence, $\{\bar{\mathbf{x}}^k,\bar{\mathbf{y}}^k\}_{k\geq1}$ converges to the same point.*

Although the proof of Theorem 1 follows from the lines of [2], we provide the proof here for the sake of completeness. To this end, we prove the following key lemma first.

Lemma 1. *Let $\mathbf{z}=[\mathbf{x}^\top\mathbf{y}^\top]^\top$ for $\mathbf{x}\in\mathcal{X}$, $\mathbf{y}\in\mathcal{Y}$. For any $\mathbf{x}\in\mathcal{X}$, and $\mathbf{y}\in\mathcal{Y}$, the iterate sequence $\{\mathbf{z}^k\}_{k\geq1}$ defined as in the statement of Theorem 1 satisfies for all $k\geq0$*

$$\begin{aligned}&\mathcal{L}(\mathbf{x}^{k+1},\mathbf{y})-\mathcal{L}(\mathbf{x},\mathbf{y}^{k+1})\\ &\leq\Big[D_x(\mathbf{x},\mathbf{x}^k)+D_y(\mathbf{y},\mathbf{y}^k)-\Big\langle T(\mathbf{x}-\mathbf{x}^k),\ \mathbf{y}-\mathbf{y}^k\Big\rangle\Big]\\ &\quad-\Big[D_x(\mathbf{x},\mathbf{x}^{k+1})+D_y(\mathbf{y},\mathbf{y}^{k+1})-\Big\langle T(\mathbf{x}-\mathbf{x}^{k+1}),\ \mathbf{y}-\mathbf{y}^{k+1}\Big\rangle\Big]\\ &\quad-\frac{1}{2}(\mathbf{z}^{k+1}-\mathbf{z}^k)^\top\bar{\mathbf{Q}}(\mathbf{z}^{k+1}-\mathbf{z}^k).\end{aligned}\tag{7}$$

Proof. Fix $\nu_x=1$. Since ρ is a proper, closed, convex function and D_x is a Bregman function, Property 1 in [14] applied to (5a) implies that

$$\begin{aligned}&\rho(\mathbf{x})-\rho(\mathbf{x}^{k+1})+\Big\langle\nabla f(\mathbf{x}^k)+T^\top\mathbf{y}^k,\ \mathbf{x}-\mathbf{x}^{k+1}\Big\rangle\\ &\qquad\geq D_x(\mathbf{x},\mathbf{x}^{k+1})-D_x(\mathbf{x},\mathbf{x}^k)+D_x(\mathbf{x}^{k+1},\mathbf{x}^k).\end{aligned}\tag{8}$$

Convexity of f_i and Lipschitz continuity of ∇f_i implies that

$$\begin{aligned}f_i(\xi_i)&\geq f_i(\xi_i^k)+\Big\langle\nabla f_i(\xi_i^k),\ \xi_i-\xi_i^k\Big\rangle\\ &\geq f_i(\xi_i^{k+1})+\Big\langle\nabla f_i(\xi_i^k),\ \xi_i-\xi_i^{k+1}\Big\rangle-\frac{L_i}{2}\left\|\xi_i^{k+1}-\xi_i^k\right\|^2.\end{aligned}$$

Summing this inequality over $i\in\mathcal{N}$, combining the sum with (8) and from the definition of $\mathbf{D}$, we get

$$\begin{aligned}&\Phi(\mathbf{x})-\Phi(\mathbf{x}^{k+1})+\Big\langle T(\mathbf{x}-\mathbf{x}^{k+1}),\ \mathbf{y}^k\Big\rangle\\ &\qquad\geq D_x(\mathbf{x},\mathbf{x}^{k+1})-D_x(\mathbf{x},\mathbf{x}^k)+\tfrac{1}{2}\left\|\mathbf{x}^{k+1}-\mathbf{x}^k\right\|_{\mathbf{D}}^2.\end{aligned}\tag{9}$$

Fix $\nu_y=1$. Since h is a proper, closed, convex function and D_y is a Bregman function, Property 1 in [14] applied to (5b) implies that

$$\begin{aligned}&h(\mathbf{y})-h(\mathbf{y}^{k+1})+\Big\langle T(2\mathbf{x}^{k+1}-\mathbf{x}^k),\ \mathbf{y}^{k+1}-\mathbf{y}\Big\rangle\\ &\qquad\geq D_y(\mathbf{y},\mathbf{y}^{k+1})-D_y(\mathbf{y},\mathbf{y}^k)+\tfrac{1}{2}\left\|\mathbf{y}^{k+1}-\mathbf{y}^k\right\|_{\mathbf{D}_\kappa}^2.\end{aligned}\tag{10}$$

Summing (9) and (10) gives the desired result. □

Proof of Theorem 1: Since $\bar{\mathbf{Q}}\succeq0$, we can drop the last term in (7) for each $k\geq0$. Next, after summing it for $k=0,\ldots,K-1$, we divide the resulting inequality by K and use Jensen's inequality to get (6) for $\nu_x=\nu_y=1$ because we also have $D_x(\mathbf{x},\mathbf{x}^K)+D_y(\mathbf{y},\mathbf{y}^K)-\big\langle T(\mathbf{x}-\mathbf{x}^K),\ \mathbf{y}-\mathbf{y}^K\big\rangle\geq0$, which follows from $\bar{\mathbf{Q}}\succeq0$.

Now suppose $\bar{\mathbf{Q}}\succ0$, and let $\mathbf{z}^*=[\mathbf{x}^{*\top}\mathbf{y}^{*\top}]^\top$ be a saddle point for (4). From the definition of $\bar{\mathbf{Q}}$, for all $\mathbf{z},\mathbf{z}'$, we have

$$D_x(\mathbf{x},\mathbf{x}')+D_y(\mathbf{y},\mathbf{y}')-\big\langle T(\mathbf{x}-\mathbf{x}'),\ \mathbf{y}-\mathbf{y}'\big\rangle\geq\tfrac{1}{2}\left\|\mathbf{z}-\mathbf{z}'\right\|_{\bar{\mathbf{Q}}}^2.\tag{11}$$

Evaluating (7) at $\mathbf{z}=\mathbf{z}^*$, we get $k\geq0$

$$\begin{aligned}&0\leq\mathcal{L}(\mathbf{x}^{k+1},\mathbf{y}^*)-\mathcal{L}(\mathbf{x}^*,\mathbf{y}^{k+1})\leq\\ &\big[D_x(\mathbf{x}^*,\mathbf{x}^k)+D_y(\mathbf{y}^*,\mathbf{y}^k)-\langle T(\mathbf{x}^*-\mathbf{x}^k),\ \mathbf{y}^*-\mathbf{y}^k\rangle\big]\\ &-\big[D_x(\mathbf{x}^*,\mathbf{x}^{k+1})+D_y(\mathbf{y}^*,\mathbf{y}^{k+1})-\langle T(\mathbf{x}^*-\mathbf{x}^{k+1}),\mathbf{y}^*-\mathbf{y}^{k+1}\rangle\big]\\ &-\tfrac{1}{2}\left\|\mathbf{z}^{k+1}-\mathbf{z}^k\right\|_{\bar{\mathbf{Q}}}^2;\end{aligned}\tag{12}$$

hence, we get for all $k\geq0$

$$\begin{aligned}&\tfrac{1}{2}\left\|\mathbf{z}^{k+1}-\mathbf{z}^*\right\|_{\bar{\mathbf{Q}}}^2\\ &\qquad\leq D_x(\mathbf{x}^*,\mathbf{x}^0)+D_y(\mathbf{y}^*,\mathbf{y}^0)-\big\langle T(\mathbf{x}^*-\mathbf{x}^0),\ \mathbf{y}^*-\mathbf{y}^0\big\rangle.\end{aligned}$$

Therefore, both $\{\mathbf{z}^k\}$ and $\{\bar{\mathbf{z}}^k\}$ are bounded sequences. Hence, there is a subsequence $\{\mathbf{z}^{k_n}\}_{n\geq1}$ converging to a limit point $\hat{\mathbf{z}}$. From (12), it follows that $\sum_{k=0}^{\infty}\left\|\mathbf{z}^{k+1}-\mathbf{z}^k\right\|_{\bar{\mathbf{Q}}}^2<\infty$. Since $\bar{\mathbf{Q}}\succ\mathbf{0}$, for any $\epsilon>0$, there exists N_1 such that for all $n\geq N_1$, we have $\left\|\mathbf{z}^{k_n+1}-\mathbf{z}^{k_n}\right\|<\frac{\epsilon}{2}$. From the fact that $\mathbf{z}^{k_n}\to\hat{\mathbf{z}}$, there exists N_2 such that for all $n\geq N_2$, we have $\left\|\mathbf{z}^{k_n}-\hat{\mathbf{z}}\right\|<\frac{\epsilon}{2}$. Therefore, by letting $N=\max\{N_1,N_2\}$ we get $\left\|\mathbf{z}^{k_n+1}-\hat{\mathbf{z}}\right\|$, i.e., $\mathbf{z}^{k_n+1}\to\hat{\mathbf{z}}$.

The optimality conditions for (5) imply that for all $n\in\mathbb{Z}_+$, we have $\mathbf{q}^n\in\partial\rho(\mathbf{x}^{k_n+1})$ and $\mathbf{p}^n\in\partial h(\mathbf{y}^{k_n+1})$, where

$$\begin{aligned}\mathbf{q}^n&\triangleq\nabla\psi_x(\mathbf{x}^{k_n})-\nabla\psi_x(\mathbf{x}^{k_n+1})-\big(\nabla f(\mathbf{x}^{k_n})+T^\top\mathbf{y}^{k_n}\big),\\ \mathbf{p}^n&\triangleq\nabla\psi_y(\mathbf{y}^{k_n})-\nabla\psi_y(\mathbf{y}^{k_n+1})+T^\top\big(2\mathbf{x}^{k_n+1}-\mathbf{x}^{k_n}\big).\end{aligned}$$

Since $\nabla\psi_x$ and $\nabla\psi_y$ are continuously differentiable on $\mathbf{dom}\,\rho$ and $\mathbf{dom}\,h$, respectively, and since ρ and h are proper, closed convex functions, it follows from Theorem 24.4 in [15] that

$$\partial\rho(\hat{\mathbf{x}})\ni\lim_n\mathbf{q}^n=-\nabla f(\hat{\mathbf{x}})-T^\top\hat{\mathbf{y}},\ \text{and}\ \partial h(\hat{\mathbf{y}})\ni\lim_n\mathbf{p}^n=T\hat{\mathbf{x}},$$

which also implies that $\hat{\mathbf{z}}$ is a saddle point of (4).

Since $\bar{\mathbf{Q}}\succ\mathbf{0}$, and (12) holds for any saddle point $\mathbf{z}^*$, setting $\mathbf{z}^*=\hat{\mathbf{z}}$ gives us a *nonincreasing* sequence $\{s^k\}_{k\geq0}$, where

$$0\leq s^k\triangleq D_x(\hat{\mathbf{x}},\mathbf{x}^k)+D_y(\hat{\mathbf{y}},\mathbf{y}^k)-\Big\langle T(\hat{\mathbf{x}}-\mathbf{x}^k),\hat{\mathbf{y}}-\mathbf{y}^k\Big\rangle.\tag{13}$$

Note $s\triangleq\lim_k s^k\geq0$ exists. Thus, $s=\lim_n s^{k_n}$; and since $\lim_n\big\langle T(\hat{\mathbf{x}}-\mathbf{x}^{k_n}),\hat{\mathbf{y}}-\mathbf{y}^{k_n}\big\rangle=0$ (from $\mathbf{z}^{k_n}\to\hat{\mathbf{z}}$),

$$s=\lim_{n\to\infty}\ D_x(\hat{\mathbf{x}},\mathbf{x}^{k_n})+D_y(\mathbf{y}^*,\mathbf{y}^{k_n})=0.$$

Therefore, $\mathbf{z}^k\to\hat{\mathbf{z}}$ follows from (13) and (11).

II. Distributed method for resource sharing

Suppose $\mathcal{K}$ in (3) is a proper cone. Let $\xi_i \in \mathbb{R}^{n_i}$ denote the *local* decision vector of node $i \in \mathcal{N}$. We can reformulate (3) as the following saddle point problem:

$$\min_{\boldsymbol{\xi}} \max_{y\in\mathcal{K}^\circ} \left\{\sum_{i\in\mathcal{N}} \Phi_i(\xi_i) + \langle \sum_{i\in\mathcal{N}} R_i\xi_i - r_i,\ y\rangle\right\}, \qquad (14)$$

where $y \in \mathbb{R}^m$ denotes the dual variable for (3). Next, (14) can be written as a dual consensus formation problem:

$$\min_{\boldsymbol{\xi}} \max_{\substack{y_i\in\mathcal{K}^\circ \\ y_i=y_j\ (i,j)\in\mathcal{E}}} \left\{\sum_{i\in\mathcal{N}} \Big(\Phi_i(\xi_i) + \langle R_i\xi_i - r_i,\ y_i\rangle\Big)\right\}. \qquad (15)$$

The consensus constraints $y_i = y_j$ for $(i,j) \in \mathcal{E}$ can be formulated as $M\mathbf{y} = 0$, where $M \triangleq H \otimes \mathbf{I}_m \in \mathbb{R}^{m|\mathcal{E}|\times m|\mathcal{N}|}$ and H is the oriented edge-node incidence matrix, i.e., the entry $H_{(i,j),l}$, corresponding to edge $(i,j) \in \mathcal{E}$ and $l \in \mathcal{N}$, is equal to 1 if $l = i$, -1 if $l = j$, and 0 otherwise. Note that $M^\top M = H^\top H \otimes \mathbf{I}_m = \Omega \otimes \mathbf{I}_m$, where $\Omega \in \mathbb{R}^{|\mathcal{N}|\times|\mathcal{N}|}$ denotes the graph Laplacian of $\mathcal{G}$, i.e., $\Omega_{ii} = d_i$, $\Omega_{ij} = -1$ if $(i,j) \in \mathcal{E}$ or $(j,i) \in \mathcal{E}$, and equal to 0 otherwise.

Define Lagrangian function $\mathcal{L}$:

$$\mathcal{L}(\boldsymbol{\xi}, \mathbf{w}, \mathbf{y}) \triangleq \sum_{i\in\mathcal{N}} \left(\Phi_i(\xi_i) + \langle R_i\xi_i - r_i,\ y_i\rangle\right) - \langle \mathbf{w}, M\mathbf{y}\rangle\,; \quad (16)$$

hence, (15) can be equivalently written as follows

$$\min_{\boldsymbol{\xi}} \max_{\mathbf{y}\in\Pi_{i\in\mathcal{N}}\mathcal{K}^\circ} \min_{\mathbf{w}} \mathcal{L}(\boldsymbol{\xi}, \mathbf{w}, \mathbf{y}) = \min_{\boldsymbol{\xi},\mathbf{w}} \max_{\mathbf{y}\in\Pi_{i\in\mathcal{N}}\mathcal{K}^\circ} \mathcal{L}(\boldsymbol{\xi}, \mathbf{w}, \mathbf{y}) \quad (17)$$

where the last equality is justified since $\mathcal{K}$ is a pointed cone, hence $\mathbf{int}\,(\mathcal{K}^\circ) \neq \emptyset$; therefore, for each fixed $\boldsymbol{\xi}$, inner $\max$ and $\min$ can be interchanged.

Next, we study the distributed implementation of PDA in (5a)-(5b) to solve (17). Define the block-diagonal matrix $R \triangleq \mathbf{diag}([R_i]_{i\in\mathcal{N}}) \in \mathbb{R}^{m|\mathcal{N}|\times n}$ and $T = [R\ \ -M^\top]$. Therefore, given the initial iterates $\boldsymbol{\xi}^0, \mathbf{w}^0, \mathbf{y}^0$ and parameters $\gamma > 0$, $\tau_i, \kappa_i > 0$ for $i \in \mathcal{N}$, choosing D_x and D_y as defined in Definition 1, and setting $\nu_x = \nu_y = 1$, PDA iterations in (5a)-(5b) take the following form for $k \geq 0$:

$$\begin{aligned}
\xi_i^{k+1} \leftarrow & \operatorname*{argmin}_{\xi_i} \rho_i(\xi_i) + f_i(\xi_i^k) + \langle \nabla f(\xi_i^k), \xi_i - \xi_i^k\rangle && (18a)\\
& + \langle R_i\xi_i - r_i, y_i^k\rangle + \frac{1}{2\tau_i}\|\xi_i - \xi_i^k\|^2 \\
\mathbf{w}^{k+1} \leftarrow & \operatorname*{argmin}_{\mathbf{w}} \left\{-\langle M\mathbf{y}^k, \mathbf{w}\rangle + \frac{1}{2\gamma}\|\mathbf{w} - \mathbf{w}^k\|^2\right\} \\
& = \mathbf{w}^k + \gamma\ M\mathbf{y}^k && (18b)\\
\mathbf{y}^{k+1} \leftarrow & \operatorname*{argmin}_{\mathbf{y}\in\Pi_{i\in\mathcal{N}}\mathcal{K}^\circ} \langle 2\mathbf{w}^{k+1} - \mathbf{w}^k, M\mathbf{y}\rangle && (18c)\\
& + \sum_{i\in\mathcal{N}} \Big[-\langle R_i(2\xi_i^{k+1} - \xi_i^k) - r_i, y_i\rangle + \frac{1}{2\kappa_i}\|y_i - y_i^k\|^2\Big].
\end{aligned}$$

Using recursion in $\mathbf{w}$ update rule in (18), we can write $\mathbf{w}^{k+1}$ as a partial summation of dual iterates $\mathbf{y}^k$, i.e., $\mathbf{w}^k = \mathbf{w}^0 + \gamma\sum_{\ell=0}^{k-1} M\mathbf{y}^\ell$. Let $\mathbf{w}^0 \leftarrow \mathbf{0}$, and $\mathbf{s}^k \triangleq \mathbf{y}^k + \sum_{\ell=0}^{k}\mathbf{y}^\ell$ for $k \geq 0$; since $M^\top M = \Omega\otimes\mathbf{I}_m$ we obtain

$$\begin{aligned}
\langle M\mathbf{y},\ 2\mathbf{w}^{k+1} - \mathbf{w}^k\rangle &= \gamma\ \langle \mathbf{y},\ (\Omega\otimes\mathbf{I}_m)\mathbf{s}^k\rangle \\
&= \gamma \sum_{i\in\mathcal{N}} \langle y_i,\ \sum_{j\in\mathcal{N}_i} (s_i^k - s_j^k)\rangle.
\end{aligned}$$

Thus, PDA iterations given in (18) for the static graph $\mathcal{G}$ can be computed in *decentralized* way, via the node-specific computations as in Algorithm DPDA-S displayed in Fig. 1.

Algorithm DPDA-S ($\boldsymbol{\xi}^0, \mathbf{y}^0, \gamma, \{\tau_i, \kappa_i\}_{i\in\mathcal{N}}$)
Initialization: $s_i^0 \leftarrow 2y_i^0$, $\quad i \in \mathcal{N}$
Step k: ($k \geq 0$)
1. $\xi_i^{k+1} \leftarrow \mathbf{prox}_{\tau_i\rho_i}\Big(\xi_i^k - \tau_i\Big(\nabla f_i(\xi_i^k) + R_i^\top y_i^k\Big)\Big), \quad i \in \mathcal{N}$
2. $p_i^{k+1} \leftarrow \sum_{j\in\mathcal{N}_i}(s_j^k - s_i^k), \quad i \in \mathcal{N}$
3. $y_i^{k+1} \leftarrow \mathcal{P}_{\mathcal{K}^\circ}\Big[y_i^k + \kappa_i\Big(R_i(2\xi_i^{k+1} - \xi_i^k) - r_i + \gamma p_i^{k+1}\Big)\Big], \quad i \in \mathcal{N}$
4. $s_i^{k+1} \leftarrow y_i^{k+1} + \sum_{\ell=0}^{k+1} y_i^\ell, \quad i \in \mathcal{N}$

Fig. 1. Distributed Primal Dual Algorithm for Static $\mathcal{G}$ (DPDA-S)

The $\mathcal{O}(1/K)$ rate for DPDA-S, given in (6), follows from Theorem 1 with the help of following technical lemma.

Lemma 2. *Given $\{\tau_i, \kappa_i\}_{i\in\mathcal{N}}$ and γ such that $\gamma > 0$, and $\tau_i > 0$, $\kappa_i > 0$ for $i \in \mathcal{N}$, $\bar{\mathbf{Q}} \triangleq \begin{bmatrix}\bar{\mathbf{D}}_\tau & 0 & -R^\top\\ 0 & \mathbf{D}_\gamma & M\\ -R & M^\top & \mathbf{D}_\kappa\end{bmatrix} \succeq \mathbf{0}$ if $\{\tau_i, \kappa_i\}_{i\in\mathcal{N}}$ and γ are chosen such that $\frac{1}{\tau_i} > L_i$, and*

$$\Big(\frac{1}{\tau_i} - L_i\Big)\Big(\frac{1}{\kappa_i} - 2\gamma d_i\Big) \geq \|R_i\|^2, \quad \forall\, i \in \mathcal{N}. \qquad (19)$$

Moreover, $\bar{\mathbf{Q}} \succ \mathbf{0}$ if (19) *holds with strict inequality.*

Proof. Let $\mathbf{P} \triangleq \begin{bmatrix}\mathbf{I}_n & 0 & 0\\ 0 & 0 & \mathbf{I}_{m|\mathcal{N}|}\\ 0 & \mathbf{I}_{m|\mathcal{E}|} & 0\end{bmatrix}$ be permutation matrix. Hence, $\bar{\mathbf{Q}} \succeq 0$ is equivalent to $\mathbf{P}\bar{\mathbf{Q}}\mathbf{P}^{-1} \succeq 0$. Since $\mathbf{D}_\gamma \succ 0$, Schur complement condition implies that $\mathbf{P}\bar{\mathbf{Q}}\mathbf{P}^{-1} = \begin{bmatrix}\bar{\mathbf{D}}_\tau & -R^\top & 0\\ -R & \mathbf{D}_\kappa & M^\top\\ 0 & M & \mathbf{D}_\gamma\end{bmatrix} \succeq 0$ if and only if

$$B - \gamma\begin{bmatrix}0 & 0\\ 0 & M^\top M\end{bmatrix} \succeq 0 \quad \text{where} \quad B \triangleq \begin{bmatrix}\bar{\mathbf{D}}_\tau & -R^\top\\ -R & \mathbf{D}_\kappa\end{bmatrix}. \qquad (20)$$

Moreover, since $\bar{\mathbf{D}}_\tau \succ 0$, again using Schur complement and the fact that $M^\top M = \Omega\otimes\mathbf{I}_m$, one can conclude that (20) holds if and only if $\mathbf{D}_\kappa - \gamma(\Omega\otimes\mathbf{I}_n) - R\bar{\mathbf{D}}_\tau^{-1}R^\top \succeq 0$. By definition $\Omega = \mathbf{diag}([d_i]_{i\in\mathcal{N}}) - E$, where $E_{ii} = 0$ for all $i \in \mathcal{N}$ and $E_{ij} = E_{ji} = 1$ if $(i,j) \in \mathcal{E}$ or $(j,i) \in \mathcal{E}$. Note that $\mathbf{diag}([d_i]_{i\in\mathcal{N}}) + E \succeq 0$ since it is diagonally dominant. Therefore, $\Omega \preceq 2\,\mathbf{diag}([d_i]_{i\in\mathcal{N}})$. Hence, it is sufficient to have $(\frac{1}{\kappa_i} - 2\gamma d_i)\mathbf{I}_m - (\frac{1}{\tau_i} - L_i)^{-1}R_iR_i^\top \succeq 0$ for all $i \in \mathcal{N}$, and this condition holds if (20) is true. By the same argument, if (20) holds with strict inequality, then $\bar{\mathbf{Q}} \succ \mathbf{0}$. □

Remark II.1. *Note that for all $i \in \mathcal{N}$ by choosing $\tau_i = \frac{1}{c_i + L_i}$, $\kappa_i = \frac{c_i}{2c_i\gamma d_i + \|R_i\|^2}$ for any $c_i > 0$, the condition in Lemma 2 is satisfied.*

Next, we refine the error bound in (6), and quantify the suboptimality and infeasibility of the DPDA-S iterate sequence.

Theorem 2. *Suppose Assumption 1 holds. Let $\{\boldsymbol{\xi}^k, \mathbf{y}^k\}_{k\geq 0}$ be the DPDA-S iterate sequences generated as in Fig. 1, initialized from an arbitrary $\boldsymbol{\xi}^0$ and $\mathbf{y}^0 = \mathbf{0}$. Define $\mathbf{w}^k \triangleq \gamma\sum_{\ell=0}^{k-1} M\mathbf{y}^\ell$ for $k \geq 1$. Let primal-dual step-sizes*

$\{\tau_i,\kappa_i\}_{i\in\mathcal{N}}$ and γ be chosen such that the condition (19) holds with strict inequality. Then $\{(\boldsymbol{\xi}^k,\mathbf{w}^k,\mathbf{y}^k)\}$ converges to $(\boldsymbol{\xi}^*,\mathbf{w}^*,\mathbf{y}^*)$, a saddle point of (17) such that $\mathbf{y}_i^*=y^*$ for all $i\in\mathcal{N}$ and $(\boldsymbol{\xi}^*,y^*)$ is a primal-dual optimal solution to (3); and the following error bounds hold for all $K\geq 1$:

$$\|\mathbf{w}^*\|\left\|M\bar{\mathbf{y}}^K\right\|+\|y^*\|\,d_{\mathcal{K}}\big(\sum_{i\in\mathcal{N}}R_i\bar{\xi}_i^K-r_i\big)\leq\frac{\Theta_1}{K}$$

$$|\Phi(\bar{\boldsymbol{\xi}}^K)-\Phi(\boldsymbol{\xi}^*)|\leq\frac{\Theta_1}{K},$$

where $\Theta_1\triangleq\sum_{i\in\mathcal{N}}\left[\frac{1}{\tau_i}\|\xi_i^*-\xi_i^0\|^2+\frac{4}{\kappa_i}\|y^*\|^2\right]+\frac{2}{\gamma}\|\mathbf{w}^*\|^2$, $\bar{\boldsymbol{\xi}}^K=\frac{1}{K}\sum_{k=1}^K\boldsymbol{\xi}^k$ *and* $\bar{\mathbf{y}}^K=\frac{1}{K}\sum_{k=1}^K\mathbf{y}^k$.

Proof. Note the iterate sequence $\{\mathbf{x}^k,\mathbf{y}^k\}_{k\geq0}$ generated by Algorithm DPDA-S in Fig. 1 is the same as the PDA iterate sequence $\{\mathbf{x}^k,\mathbf{w}^k,\mathbf{y}^k\}_{k\geq0}$ computed according to (18) for solving (17) when $\mathbf{w}^0=\mathbf{0}$. Since the step-size parameters $\{\tau_i,\kappa_i\}_{i\in\mathcal{N}}$ and γ are chosen satisfying the condition (19) in Lemma 2 with strict inequality, the convergence condition, $\bar{\mathbf{Q}}\succ0$, in Theorem 1 is true, where $T=[R\ \ -M^\top]$ for problem in (17). Therefore, Theorem 1 implies that (6) holds for all $K\geq1$ with $\nu_x=\nu_y=1$ and Bregman functions D_x, D_y defined as in Definition 1. In particular, the result of Theorem 1 can be written more explicitly for (17) as follows: for any $\boldsymbol{\xi}\in\mathbb{R}^n$, $\mathbf{w}\in\mathbb{R}^{m|\mathcal{E}|}$, $\mathbf{y}\in\mathbb{R}^{m|\mathcal{N}|}$, and for all $K\geq1$,

$$\mathcal{L}(\bar{\boldsymbol{\xi}}^K,\bar{\mathbf{w}}^K,\mathbf{y})-\mathcal{L}(\boldsymbol{\xi},\mathbf{w},\bar{\mathbf{y}}^K)\leq\Theta(\boldsymbol{\xi},\mathbf{w},\mathbf{y})/K, \tag{21}$$

$$\Theta(\boldsymbol{\xi},\mathbf{w},\mathbf{y})\triangleq\frac{1}{2\gamma}\left\|\mathbf{w}-\mathbf{w}^0\right\|^2+\left\langle\mathbf{w}-\mathbf{w}^0,M(\mathbf{y}-\mathbf{y}^0)\right\rangle$$
$$+\sum_{i\in\mathcal{N}}\tfrac{1}{2\tau_i}\left\|\xi_i-\xi_i^0\right\|^2+\tfrac{1}{2\kappa_i}\left\|y_i-y_i^0\right\|^2-\left\langle R_i(\xi_i-\xi_i^0),y_i-y_i^0\right\rangle,$$

where $\bar{\mathbf{w}}^K\triangleq\frac{1}{K}\sum_{k=1}^K\mathbf{w}^k$. Note that under the assumption in (19), Schur complement condition guarantees that

$$\begin{bmatrix}\frac{1}{\tau_i}\mathbf{I}_n & -R_i^\top\\ -R_i & \frac{1}{\kappa_i}\mathbf{I}_{m_i}\end{bmatrix}\preceq\begin{bmatrix}\frac{2}{\tau_i}\mathbf{I}_n & \mathbf{0}^\top\\ \mathbf{0} & \frac{2}{\kappa_i}\mathbf{I}_{m_i}\end{bmatrix}.$$

Therefore,

$$\Theta(\boldsymbol{\xi},\mathbf{w},\mathbf{y})\leq\frac{1}{2\gamma}\left\|\mathbf{w}-\mathbf{w}^0\right\|^2+\left\langle\mathbf{w}-\mathbf{w}^0,M(\mathbf{y}-\mathbf{y}^0)\right\rangle$$
$$+\sum_{i\in\mathcal{N}}\left(\frac{1}{\tau_i}\left\|\xi_i-\xi_i^0\right\|^2+\frac{1}{\kappa_i}\left\|y_i-y_i^0\right\|^2\right). \tag{22}$$

Under Assumption 1, a saddle point for (17) exists. For any saddle point $(\boldsymbol{\xi}^*,\mathbf{w}^*,\mathbf{y}^*)$, $\mathcal{L}(\boldsymbol{\xi}^*,\mathbf{w}^*,\mathbf{y}^*)=\Phi(\boldsymbol{\xi}^*)$, $\boldsymbol{\xi}^*$ is an optimal solution to (3) and $M\mathbf{y}^*=0$, i.e., for some $y^*\in\mathcal{K}^\circ$ we have $\mathbf{y}_i^*=y^*$ for all $i\in\mathcal{N}$. The bounds in the statement of the theorem are valid for an arbitrary saddle-point of $\mathcal{L}$ in (17); that said, without loss of generality we consider a specific saddle point $(\boldsymbol{\xi}^*,\mathbf{w}^*,\mathbf{y}^*)$ as defined next. According to Theorem 1, $(\boldsymbol{\xi}^k,\mathbf{w}^k,\mathbf{y}^k)$ converges to a saddle point of $\mathcal{L}$ in (17); let $(\boldsymbol{\xi}^*,\mathbf{w}^*,\mathbf{y}^*)$ be that point. Note that $(\bar{\boldsymbol{\xi}}^k,\bar{\mathbf{w}}^k,\bar{\mathbf{y}}^k)$ converges to $(\boldsymbol{\xi}^*,\mathbf{w}^*,\mathbf{y}^*)$ as well. Define $\tilde{z}\triangleq\sum_{i\in\mathcal{N}}R_i\bar{\xi}_i^K-r_i\in\mathbb{R}^m$. Since $\mathcal{K}$ is a closed convex cone, it induces a decomposition on $\mathbb{R}^m$, i.e., $\tilde{z}^1=\mathcal{P}_{\mathcal{K}}(\tilde{z})$ and $\tilde{z}^2=\mathcal{P}_{\mathcal{K}^\circ}(\tilde{z})$ satisfy $\tilde{z}=\tilde{z}^1+\tilde{z}^2$ and $\tilde{z}^1\perp\tilde{z}^2$. Note that since $\tilde{z}=\tilde{z}^1+\tilde{z}^2$, $\left\|\tilde{z}^2\right\|=\|\mathcal{P}_{\mathcal{K}}(\tilde{z})-\tilde{z}\|=d_{\mathcal{K}}(\tilde{z})$. Define $\widetilde{\mathbf{y}}=[\widetilde{y}_i]_{i\in\mathcal{N}}$ such that $\widetilde{y}_i\triangleq2\|y^*\|\frac{1}{\|\tilde{z}^2\|}\tilde{z}^2\in\mathcal{K}^\circ$, and note $M\widetilde{\mathbf{y}}=\mathbf{0}$. Since $\tilde{z}^1\perp\tilde{z}^2$,

$$\sum_{i\in\mathcal{N}}\left\langle R_i\bar{\xi}_i^K-r_i,\widetilde{y}_i\right\rangle=2\|y^*\|\,d_{\mathcal{K}}\Big(\sum_{i\in\mathcal{N}}R_i\bar{\xi}_i^K-r_i\Big). \tag{23}$$

Since $\mathbf{y}^*$ maximize $\mathcal{L}(\boldsymbol{\xi}^*,\mathbf{w}^*,\mathbf{y})$, and we set $\mathbf{y}^0=\mathbf{0}$ and $\mathbf{w}^0=\mathbf{0}$, the definitions of $\widetilde{\mathbf{y}}$ and (21), (22) together imply

$$\mathcal{L}(\bar{\boldsymbol{\xi}}^K,\bar{\mathbf{w}}^K,\widetilde{\mathbf{y}})-\mathcal{L}(\boldsymbol{\xi}^*,\mathbf{w}^*,\mathbf{y}^*)\leq\mathcal{L}(\bar{\boldsymbol{\xi}}^K,\bar{\mathbf{w}}^K,\widetilde{\mathbf{y}})-\mathcal{L}(\boldsymbol{\xi}^*,\mathbf{w}^*,\bar{\mathbf{y}}^K)$$
$$\leq\frac{1}{K}\Theta(\boldsymbol{\xi}^*,\mathbf{w}^*,\widetilde{\mathbf{y}})\leq\frac{\Theta_1}{K}. \tag{24}$$

Therefore, using (23) and (24), we can conclude that

$$\boldsymbol{\Phi}(\bar{\boldsymbol{\xi}}^K)-\boldsymbol{\Phi}(\boldsymbol{\xi}^*)+2\|y^*\|\,d_{\mathcal{K}}\Big(\sum_{i\in\mathcal{N}}R_i\bar{\xi}_i^K-r_i\Big)\leq\frac{\Theta_1}{K} \tag{25}$$

where we used the fact that $M\widetilde{\mathbf{y}}=\mathbf{0}$. Moreover, since $(\boldsymbol{\xi}^*,\mathbf{w}^*,\mathbf{y}^*)$ is a saddle-point for $\mathcal{L}$ in (17), we clearly have $\mathcal{L}(\bar{\boldsymbol{\xi}}^K,\mathbf{w}^*,\mathbf{y}^*)-\mathcal{L}(\boldsymbol{\xi}^*,\mathbf{w}^*,\mathbf{y}^*)\geq0$; therefore,

$$\boldsymbol{\Phi}(\bar{\boldsymbol{\xi}}^K)-\boldsymbol{\Phi}(\boldsymbol{\xi}^*)+\|y^*\|\,d_{\mathcal{K}}\Big(\sum_{i\in\mathcal{N}}R_i\bar{\xi}_i^K-r_i\Big)\geq0. \tag{26}$$

which follows from $y^*\in\mathcal{K}^\circ$, i.e., $\langle y^*,y\rangle=\langle y^*,\mathcal{P}_{\mathcal{K}^\circ}(y)\rangle\leq\|y^*\|\,d_{\mathcal{K}}(y)$ for all $y\in\mathbb{R}^m$. Thus, combining inequalities (25) and (26) immediately implies the *suboptimality* result.

Define $\widetilde{\mathbf{w}}\triangleq2\|\mathbf{w}^*\|\frac{M\bar{\mathbf{y}}^K}{\|M\bar{\mathbf{y}}^K\|}$. From (21) and (23), we have

$$\mathcal{L}(\bar{\boldsymbol{\xi}}^K,\bar{\mathbf{w}}^K,\widetilde{\mathbf{y}})-\mathcal{L}(\boldsymbol{\xi}^*,\widetilde{\mathbf{w}},\bar{\mathbf{y}}^K) \tag{27}$$
$$=\boldsymbol{\Phi}(\bar{\boldsymbol{\xi}}^K)-\boldsymbol{\Phi}(\boldsymbol{\xi}^*)+2\|y^*\|\,d_{\mathcal{K}}\Big(\sum_{i\in\mathcal{N}}R_i\bar{\xi}_i^K-r_i\Big)$$
$$-\sum_{i\in\mathcal{N}}\left\langle R_i\xi_i^*-r_i,\bar{y}_i^K\right\rangle+2\|\mathbf{w}^*\|\left\|M\bar{\mathbf{y}}^K\right\|\leq\frac{1}{K}\Theta(\boldsymbol{\xi}^*,\widetilde{\mathbf{w}},\widetilde{\mathbf{y}}).$$

Hence, it is easy to see that $\left\langle\widetilde{\mathbf{w}},M\bar{\mathbf{y}}^K\right\rangle=2\|\mathbf{w}^*\|\left\|M\bar{\mathbf{y}}^K\right\|$. Note that $0\leq\mathcal{L}(\bar{\boldsymbol{\xi}}^K,\mathbf{w}^*,\mathbf{y}^*)-\mathcal{L}(\boldsymbol{\xi}^*,\mathbf{w}^*,\mathbf{y}^*)\leq\mathcal{L}(\bar{\boldsymbol{\xi}}^K,\mathbf{w}^*,\mathbf{y}^*)-\mathcal{L}(\boldsymbol{\xi}^*,\mathbf{w}^*,\bar{\mathbf{y}}^K)$. Hence, $y^*\in\mathcal{K}^\circ$ implies

$$0\leq\boldsymbol{\Phi}(\bar{\boldsymbol{\xi}}^K)-\boldsymbol{\Phi}(\boldsymbol{\xi}^*)+\|y^*\|\,d_{\mathcal{K}}\Big(\sum_{i\in\mathcal{N}}R_i\bar{\xi}_i^K-r_i\Big) \tag{28}$$
$$-\sum_{i\in\mathcal{N}}\left\langle R_i\xi_i^*-r_i,\bar{y}_i^K\right\rangle+\|\mathbf{w}^*\|\left\|M\bar{\mathbf{y}}^K\right\|.$$

Therefore, summing (27) and (28) leads to the desired *infeasibility* and *consensus* results. □

III. Numerical Experiments

In this section, we use the *Basis pursuit* problem to test DPDA-S and its variant, DPDA-D, which is proposed in a recent preprint [1] to extend DPDA-S to handle time-varying communication networks, and it requires each node $i\in\mathcal{N}$ to make $q_k>1$ communication rounds with the neighboring nodes at each iteration k (in contrast to $q_k=1$ for DPDA-S); and we compare them with Prox-JADMM algorithm proposed in [16]. For the *static* case, communication network $G=(\mathcal{N},\mathcal{E})$ is a connected graph that is generated by randomly adding edges to a spanning tree, generated uniformly at random, until a desired algebraic connectivity is achieved. For the *dynamic* case, for each consensus round $t\geq1$, $\mathcal{G}^t$ is generated as in the static case. Let $N\triangleq|\mathcal{N}|$. Basis pursuit problem has the following formulation:

$$\min_{\boldsymbol{\xi}}\ \ \|\boldsymbol{\xi}\|_1\quad\textbf{s.t.}\quad R\boldsymbol{\xi}=r, \tag{29}$$

where $R = [R_1, \ldots, R_N] \in \mathbb{R}^{m \times nN}$ and $r \in \mathbb{R}^m$ are the problem data; $\boldsymbol{\xi} \in \mathbb{R}^{nN}$ denotes the primal decision vector. Problem in (29) can be rewritten in the form of (3):

$$\min_{[\xi_i]_{i\in\mathcal{N}}} \sum_{i\in\mathcal{N}} \|\xi_i\|_1 \quad \textbf{s.t.} \quad \sum_{i\in\mathcal{N}} R_i\xi_i = r. \tag{30}$$

In a similar setting to [16], we set $N = 100$, $m = 300$, and $n = 10$. Matrix R is randomly generated with each entry sampled from Gaussian distribution and $r = R\boldsymbol{\xi}^*$, where $\boldsymbol{\xi}^*$ is randomly generated with 60 *nonzero* elements drawn from the standard Gaussian distribution.

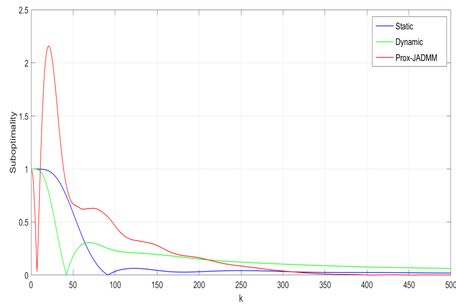

Fig. 2. Suboptimality

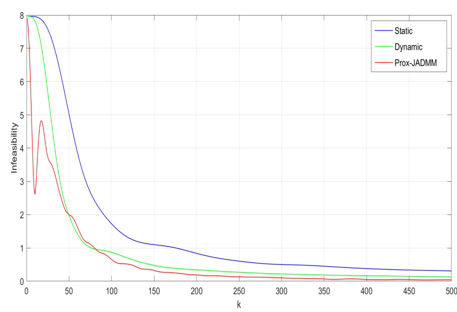

Fig. 3. Infeasibility

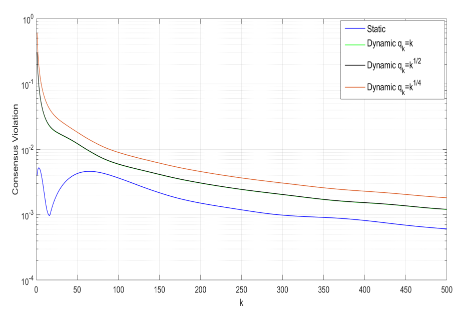

Fig. 4. Consensus violation

Fig. 2 and Fig. 3 illustrate the performance of algorithms in suboptimality and infeasibility versus number of iterations. Note that DPDA-S and DPDA-D use only local communication while this is not the case for Prox-JADMM which requires a central node for coordination. In terms of convergence behavior, our schemes can compete with Prox-JADMM. Fig. 4 illustrates consensus violation among dual variables, where the dual consensus violation is defined as $\max_{(i,j)\in\mathcal{E}} \|y_i - y_j\|$. As expected, the convergence rate for the static case is better than dynamic case. Moreover, in the dynamic case convergence of the consensus violation is slower when q_k, the rate of communication among nodes is lower.

References

[1] N. S. Aybat and E. Y. Hamedani, "A distributed ADMM-like method for resource sharing under conic constraints over time-varying networks," 2016, arXiv preprint arXiv:1611.07393.

[2] A. Chambolle and T. Pock, "On the ergodic convergence rates of a first-order primal–dual algorithm," *Mathematical Programming*, pp. 1–35, 2015.

[3] E. Wei and A. Ozdaglar, "On the $\mathcal{O}(1/k)$ convergence of asynchronous distributed alternating direction method of multipliers," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*. IEEE, 2013, pp. 551–554.

[4] T.-H. Chang, M. Hong, and X. Wang, "Multi-agent distributed optimization via inexact consensus admm," *IEEE Transactions on Signal Processing*, vol. 63, no. 2, pp. 482–497, 2015.

[5] T.-H. Chang, "A proximal dual consensus admm method for multi-agent constrained optimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3719–3734, 2014.

[6] T.-H. Chang, A. Nedic, and A. Scaglione, "Distributed constrained optimization by consensus-based primal-dual perturbation method," *Automatic Control, IEEE Transactions on*, vol. 59, no. 6, pp. 1524–1538, 2014.

[7] S. K. Niederländer and J. Cortés, "Distributed coordination for separable convex optimization with coupling constraints," in *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 694–699.

[8] D. Mateos-Núñez and J. Cortés, "Distributed subgradient methods for saddle-point problems," in *IEEE Conf. on Decision and Control. Osaka, Japan. Submitted*, 2015.

[9] A. Nedić and A. Ozdaglar, "Subgradient methods for saddle-point problems," *Journal of optimization theory and applications*, vol. 142, no. 1, pp. 205–228, 2009.

[10] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[11] B. He and X. Yuan, "Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 119–149, 2012.

[12] Y. Chen, G. Lan, and Y. Ouyang, "Optimal primal-dual methods for a class of saddle point problems," *SIAM Journal on Optimization*, vol. 24, no. 4, pp. 1779–1814, 2014.

[13] N. S. Aybat, Z. Wang, T. Lin, and S. Ma, "Distributed linearized alternating direction method of multipliers for composite convex consensus optimization," *arXiv preprint arXiv:1512.08122*, 2015.

[14] P. Tseng, "On accelerated proximal gradient methods for convex-concave optimization," 2008, available at http://www.mit.edu/ dimitrib/PTseng/papers/apgm.pdf.

[15] R. T. Rockafellar, *Convex analysis*. Princeton university press, 1997.

[16] W. Deng, M.-J. Lai, Z. Peng, and W. Yin, "Parallel multi-block admm with $\mathcal{O}(1/k)$ convergence," *arXiv preprint arXiv:1312.3040*, 2013.