Distributed subgradient methods for saddle-point problems

David Mateos-Núñez Jorge Cortés

Abstract— **We present provably correct distributed subgradient methods for general min-max problems with agreement constraints on a subset of the arguments of both the convex and concave parts. Applications include separable constrained minimization problems where each constraint is a sum of convex functions of local variables for the agents. The proposed algorithm then reduces to primal-dual updates using local subgradients and Laplacian averaging on local copies of the multipliers associated to the global constraints. The framework also encodes minimization problems with semidefinite constraints, which results in novel distributed strategies that are scalable if the order of the matrix inequalities is independent of the network size. Our analysis establishes for the case of general convex-concave functions the convergence of the running time-averages of the local estimates to a saddle point under periodic connectivity of the communication digraphs. Specifically, choosing the gradient step-sizes in a suitable way, we show that the evaluation error is proportional to $1/\sqrt{t}$, where t is the iteration step.**

I. INTRODUCTION

Saddle-point problems arise in constrained optimization via the Lagrangian formulation and also in min-max problems in game-theoretic models. Currently, these fields of research find applications in cooperative control of multi-agent systems and in large-scale machine learning, motivating the study of distributed strategies that scale well with the number of agents, are provably correct, and are robust against a variety of failures and uncertainties.

Literature review: We build on three related areas: iterative methods for saddle-point problems [1], [2], dual decompositions for constrained optimization [3], and consensus-based distributed optimization algorithms, see [4], [5], [6], [7], [8] and references therein. Historically, these fields have been driven by the need of solving constrained optimization problems and by an effort of parallelizing the computations [9], [10], leading to the use of consensus approaches that allow different processors with local memories to update the same components of a vector by averaging their estimates [11]. Saddle-point problems, or min-max problems, arise in optimization contexts such as worst-case design, exact penalty functions, duality theory, and zero-sum games, see e.g. [12]. The work [1] studies iterative subgradient methods to find a saddle point of Lagrangians and establishes convergence to an arbitrarily small neighborhood depending on the gradient stepsize. Along these lines, [2] presents an analysis for general convex-concave functions and studies the evaluation error of the running time-averages, showing convergence to an arbitrarily small neighborhood assuming boundedness of the estimates. In [2], [13], the boundedness of the estimates in the case of Lagrangians is achieved using a truncated projection onto a closed set that preserves the dual set, which [14] shows to be bounded when the strong Slater condition holds. The bound on the Lagrange multipliers depends on global information and hence must be known beforehand for its use in distributed implementations.

In distributed constrained optimization, an important distinction that determines the technical analysis and the applications arises depending of whether constraints fall in either of two categories: the first type concerns the *global* decision vector, in which agents need to agree (see e.g., [15], [6], where all the agents know the constraint, or [16], [17], [6], where the constraint is given by the intersection of (abstract) closed convex sets). The second type *couples* the *local* decision vectors across a network (see e.g., [18], [19], where the inequality constraint is a sum of convex functions and each one is only known to the corresponding agent, or [20], where in the case of linear equality constraints there is a distinction between constraint graph and network graph). Here, we address a combination of the two types of constraints, allowing the agreement constraint to play an independent role on a subset of both the primal and dual variables. This opens the way to the design of novel distributed coordination strategies for a suite of constrained convex optimization problems. The recent work [18] considers convex-concave functions arising from Lagrangians and uses primal-dual perturbed methods, which require the extra updates of the perturbation points to guarantee asymptotic convergence to a saddle point (as apposed to convergence of a subsequence). These computations require subgradient methods or proximal methods that add to the computation and the communication complexity.

Statement of contributions: We address the design and analysis of distributed algorithms for constrained optimization problems under a variety of saddle-point formulations. The explicit agreement constraints on a subset of the arguments of both the convex and concave parts allows to distribute both primal and dual variables independently. For instance, separable constraints can be decomposed using agreement on dual variables, while a subset of the primal variables can still be subject to agreement or eliminated through Fenchel conjugation; local constraints can be handled through projections; and part of the objective can be expressed as a maximization problem in extra variables. We present projected subgradient methods with Laplacian averaging, which naturally lend themselves to distributed implementation, and characterize their asymptotic convergence properties. The technical analysis entails computing bounds on the evaluation error with respect to a saddle point in terms of the disagreement, the size of the subgradients, the size of the estimates, and the gradient stepsizes. Finally, under assumptions on the boundedness of the estimates and the subgradients, we further bound the cumulative disagreement under joint connectivity of the communication graphs, regardless of the interleaved projections, and make a choice of decreasing stepsizes that guarantees convergence of the evaluation error as $1/\sqrt{t}$.

II. Preliminaries

Here we introduce basic notation and notions from graph theory and optimization used throughout the paper.

A. Notational conventions

We denote by $\mathbb{R}^n$ the n-dimensional Euclidean space, by $\mathrm{I}_n \in \mathbb{R}^{n\times n}$ the identity matrix in $\mathbb{R}^n$, and by $\mathbb{1}_N \in \mathbb{R}^n$ the vector of all ones. Given two vectors, $u,\ v \in \mathbb{R}^n$, we denote by $u \geq v$ the entry-wise set of inequalities $u_i \geq v_i$, for each $i = 1, \dots, n$. Given a vector $v \in \mathbb{R}^n$, we denote its Euclidean norm, or two-norm, by $\|v\|_2 = \sqrt{\sum_{i=1}^n v_i^2}$ and the one-norm by $\|v\|_1 = \sum_{i=1}^n |v_i|$. For a closed convex set $\mathcal{S}$, the orthogonal projection $\mathcal{P}_{\mathcal{S}}(\cdot)$ onto $\mathcal{S}$ is defined as

$$\mathcal{P}_{\mathcal{S}}(x) \in \arg\min_{x' \in \mathcal{S}} \|x - x'\|_2. \tag{1}$$

Given a convex set $\mathcal{S} \subseteq \mathbb{R}^n$, a function $f : \mathcal{S} \to \mathbb{R}$ is convex if $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$ for all $\alpha \in [0,1]$ and $x, y \in \mathcal{S}$. A vector $\xi_x \in \mathbb{R}^n$ is a subgradient of f at $x \in \mathcal{S}$ if $f(y) - f(x) \geq \xi_x^\top (y - x)$, for all $y \in \mathcal{S}$. We denote by $\partial f(x)$ the set of all such subgradients. The function f is concave if $-f$ is convex. A vector $\xi_x \in \mathbb{R}^n$ is a subgradient of f at $x \in \mathcal{S}$ if $-\xi_x \in \partial(-f)(x)$. For a symmetric matrix $A \in \mathbb{R}^{n\times n}$, we let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote its minimum and maximum eigenvalues. $A \succeq 0$ denotes a symmetric positive semidefinite matrix. We let $\mathbb{S}^n_{\succeq 0}$ denote the set of all symmetric positive semidefinite matrices.

B. Graph theory

We review basic notions from graph theory following [21]. A (weighted) digraph $\mathcal{G} := (\mathcal{I}, \mathcal{E}, \mathsf{A})$ is a triplet where $\mathcal{I} := \{1, \dots, N\}$ is the vertex set, $\mathcal{E} \subseteq \mathcal{I} \times \mathcal{I}$ is the edge set, and $\mathsf{A} \in \mathbb{R}^{N\times N}_{\geq 0}$ is the weighted adjacency matrix with the property that $\mathsf{a}_{ij} := A_{ij} > 0$ if and only if $(i,j) \in \mathcal{E}$. The complete graph is the digraph with edge set $\mathcal{I} \times \mathcal{I}$. Given $\mathcal{G}_1 = (\mathcal{I}, \mathcal{E}_1, \mathsf{A}_1)$ and $\mathcal{G}_2 = (\mathcal{I}, \mathcal{E}_2, \mathsf{A}_2)$, their union is the digraph $\mathcal{G}_1 \cup \mathcal{G}_2 = (\mathcal{I}, \mathcal{E}_1 \cup \mathcal{E}_2, \mathsf{A}_1 + \mathsf{A}_2)$. A path is an ordered sequence of vertices such that any pair of vertices appearing consecutively is an edge. A digraph is strongly connected if there is a path between any pair of distinct vertices. A sequence of digraphs $\left\{\mathcal{G}_t := (\mathcal{I}, \mathcal{E}_t, \mathsf{A}_t)\right\}_{t\geq 1}$ is δ-nondegenerate, for $\delta \in \mathbb{R}_{>0}$, if the weights are uniformly bounded away from zero by δ whenever positive, i.e., for each $t \in \mathbb{Z}_{\geq 1}$, $\mathsf{a}_{ij,t} := (\mathsf{A}_t)_{ij} > \delta$ whenever $\mathsf{a}_{ij,t} > 0$. A sequence $\{\mathcal{G}_t\}_{t\geq 1}$ is B-jointly connected, for $B \in \mathbb{Z}_{\geq 1}$, if for each $k \in \mathbb{Z}_{\geq 1}$, the digraph $\mathcal{G}_{kB} \cup \dots \cup \mathcal{G}_{(k+1)B-1}$ is strongly connected. The Laplacian matrix $\mathsf{L} \in \mathbb{R}^{N\times N}$ of a digraph $\mathcal{G}$ is $\mathsf{L} := \mathrm{diag}(\mathsf{A}\mathbb{1}_N) - \mathsf{A}$. Note that $\mathsf{L}\mathbb{1}_N = 0_N$. The weighted out-degree and in-degree of $i \in \mathcal{I}$ are, respectively, $d_{\text{out}}(i) := \sum_{j=1}^N \mathsf{a}_{ij}$ and $d_{\text{in}}(i) := \sum_{j=1}^N \mathsf{a}_{ji}$. A digraph is weight-balanced if $d_{\text{out}}(i) = d_{\text{in}}(i)$ for all $i \in \mathcal{I}$, that is, $\mathbb{1}_N^\top \mathsf{L} = 0_N$. For convenience, we let $\mathsf{L}_{\mathcal{K}} := \mathrm{I}_N - \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^\top$ denote the Laplacian of the complete graph with edge weights $1/N$. Note that $\mathsf{L}_{\mathcal{K}}$ is idempotent, i.e., $\mathsf{L}_{\mathcal{K}}^2 = \mathsf{L}_{\mathcal{K}}$. For the sake of the reader, Table I collects some shorthand notation.

$\mathrm{M} = \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^\top$	$\mathsf{L}_{\mathcal{K}} = \mathrm{I}_N - \mathrm{M}$	$\mathsf{L}_t = \mathrm{diag}(\mathsf{A}_t\mathbb{1}_N) - \mathsf{A}_t$
$\mathbf{M} = \mathrm{M} \otimes \mathrm{I}_d$	$\mathbf{L}_{\mathcal{K}} = \mathsf{L}_{\mathcal{K}} \otimes \mathrm{I}_d$	$\mathbf{L}_t = \mathsf{L}_t \otimes \mathrm{I}_d$

TABLE I: Notation for graph matrices employed along the paper, where the dimension d depends on the context.

C. Optimization and saddle points

For any function $\mathcal{L} : \mathcal{W} \times \mathcal{M} \to \mathbb{R}$, the *max-min inequality* [22, Sec 5.4.1] states that

$$\inf_{w\in\mathcal{W}} \sup_{\mu\in\mathcal{M}} \mathcal{L}(w,\mu) \geq \sup_{\mu\in\mathcal{M}} \inf_{w\in\mathcal{W}} \mathcal{L}(w,\mu). \tag{2}$$

When equality holds, we say that $\mathcal{L}$ satisfies the *strong max-min property* or the *saddle-point* property. A point $(w^*, \mu^*) \in \mathcal{W} \times \mathcal{M}$ is called a *saddle point* if

$$w^* = \inf_{w\in\mathcal{W}} \mathcal{L}(w, \mu^*) \text{ and } \mu^* = \sup_{\mu\in\mathcal{M}} \mathcal{L}(w^*, \mu).$$

[12, Sec. 2.6] discusses sufficient conditions to guarantee the existence of saddle points. Note that the existence of saddle points implies the strong max-min property. Given functions $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^m \to \mathbb{R}$ and $h : \mathbb{R}^p \to \mathbb{R}$, the *Lagrangian* for the problem

$$\min_{w\in\mathbb{R}^n} f(w) \quad \text{s.t.} \quad g(w) \leq 0,\ h(w) = 0,$$

is defined as $\mathcal{L}(w,\mu,\lambda) = f(w) + \mu^\top g(w) + \lambda^\top h(w)$ for $(\mu, \lambda) \in \mathbb{R}^m_{\geq 0} \times \mathbb{R}^p$. In this case, inequality (2) is called *weak-duality*, and if the equality is satisfied, then we say that *strong-duality* (or Lagrangian duality) holds. The point (w^*, μ^*, λ^*) is a saddle point for the Lagrangian if and only if w^* solves the constrained minimization problem and (μ^*, λ^*) solves the *dual problem*, which is maximizing the *dual function* $q(\mu,\lambda) := \inf_{w\in\mathbb{R}^n} \mathcal{L}(w,\mu,\lambda)$ over $\mathbb{R}^m_{\geq 0} \times \mathbb{R}^p$. The vectors (μ^*, λ^*) are called *Lagrange multipliers* or *optimal dual vectors*.

III. Distributed algorithms for saddle-point problems under agreement constraints

This section describes the problem of interest. Given closed convex sets $\boldsymbol{W} \subseteq \mathbb{R}^{d_{\boldsymbol{w}}}$, $\mathcal{D} \subseteq \mathbb{R}^{d_D}$, $\boldsymbol{M} \subseteq \mathbb{R}^{d_{\boldsymbol{\mu}}}$, $\mathcal{Z} \subseteq \mathbb{R}^{d_{\boldsymbol{z}}}$, and the function $\boldsymbol{\phi} : \boldsymbol{W} \times \mathcal{D}^N \times \boldsymbol{M} \times \mathcal{Z}^N \to \mathbb{R}$ which is jointly convex on the first two arguments and jointly concave on the last two arguments, we seek to solve the constrained saddle-point problem:

$$\min_{\substack{\boldsymbol{w}\in\boldsymbol{W},\, \boldsymbol{D}\in\mathcal{D}^N \\ D^i = D^j,\, \forall i,j}} \ \max_{\substack{\boldsymbol{\mu}\in\boldsymbol{M},\, \boldsymbol{z}\in\mathcal{Z}^N \\ z^i = z^j,\, \forall i,j}} \ \boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{D}, \boldsymbol{\mu}, \boldsymbol{z}), \tag{3}$$

where $\boldsymbol{D} := (D^1, \dots, D^N)$ and $\boldsymbol{z} := (z^1, \dots, z^N)$. The motivation for the consideration of explicit agreement constraints comes from various applications in network optimization and machine learning. In such scenarios, global decision variables that affect local objective functions and constraints can be duplicated into distinct ones so that each agent has its own local version to operate with, and agreement constraints are imposed to ensure the equivalence to the original optimization problem. We present examples of this procedure next.

A. Optimization problems with separable constraints

We illustrate here how optimization problems with constraints given by a sum of convex functions can be reformulated in the form (3) to make them amenable to distributed algorithmic solutions. Consider a group of agents $\{1, \ldots, N\}$, and let $f^i : \mathbb{R}^{n_i} \times \mathbb{R}^{d_D} \to \mathbb{R}$ and the components of $g^i : \mathbb{R}^{n_i} \times \mathbb{R}^{d_D} \to \mathbb{R}^m$ be convex functions associated to agent $i \in \{1, \ldots, N\}$. These functions depend on both a local decision vector $w^i \in \mathcal{W}_i$, with $\mathcal{W}_i \subseteq \mathbb{R}^{n_i}$ convex, and on a global decision vector $D \in \mathcal{D}$, with $\mathcal{D} \subseteq \mathbb{R}^{d_D}$ convex. The separable optimization problem then reads as

$$\begin{aligned} \min_{\substack{w^i \in \mathcal{W}_i, \forall i \\ D \in \mathcal{D}}} & \sum_{i=1}^N f^i(w^i, D) \\ \text{s.t.} \; & g^1(w^1, D) + \cdots + g^N(w^N, D) \leq 0. \end{aligned} \tag{4}$$

This problem can be reformulated as a constrained saddle-point problem as follows. One first constructs the corresponding Lagrangian function and introduces copies $\{z^i\}_{i=1}^N$ of the Lagrange multiplier z associated to the global constraint in (4), associates each z^i to g^i, and imposes the agreement constraint $z^i = z^j$ for all i, j. Similarly, we also introduce copies $\{D^i\}_{i=1}^N$ of the global decision vector D subject to agreement, $D^i = D^j$ for all i, j. The saddle point exists and gives a solution of the optimization (4) if strong duality holds. Formally, we have

$$\min_{\substack{w^i \in \mathcal{W}_i \\ D \in \mathcal{D}}} \; \max_{z \in \mathbb{R}^m_{\geq 0}} \sum_{i=1}^N f^i(w^i, D) + z^\top \sum_{i=1}^N g^i(w^i, D) \tag{5a}$$

$$= \min_{\substack{w^i \in \mathcal{W}_i \\ D \in \mathcal{D}}} \; \max_{\substack{z^i \in \mathbb{R}^m_{\geq 0} \\ z^i = z^j, \forall i,j}} \sum_{i=1}^N \big(f^i(w^i, D) + {z^i}^\top g^i(w^i, D) \big) \tag{5b}$$

$$= \min_{\substack{w^i \in \mathcal{W}_i \\ D^i \in \mathcal{D} \\ D^i = D^j, \forall i,j}} \; \max_{\substack{z^i \in \mathbb{R}^m_{\geq 0} \\ z^i = z^j, \forall i,j}} \sum_{i=1}^N \big(f^i(w^i, D^i) + {z^i}^\top g^i(w^i, D^i) \big). \tag{5c}$$

This formulation has its roots in the classical dual decompositions surveyed in [3, Ch. 2]. (See also [23, Sec. 1.2.3] and [24, Sec. 5.4] for the particular case of resource allocation.) While [3], [23] suggest to broadcast a centralized update of the multiplier, and the method in [24] has an implicit projection onto the probability simplex, this formulation has the multiplier associated to the global constraint estimated in a decentralized way. The recent work [18] implicitly rests on the above formulation.

B. Optimization problems with semidefinite constraints

Consider now an optimization problem analogous to (4) but with a semidefinite constraint,

$$\begin{aligned} \min_{w^i \in \mathcal{W}_i, \forall i} & \sum_{i=1}^N f^i(w^i) \\ \text{s.t.} \; & C - A^1 w^1 + \cdots + A^N w^N \preceq 0, \end{aligned} \tag{6}$$

where $\{A^i\}_{i=1}^N$, C are symmetric matrices in $\mathbb{R}^{d \times d}$ and $\mathcal{W}_i \subseteq \mathbb{R}$ for each $i \in \{1, \ldots, N\}$. Choosing symmetric matrices $\{C^i\}_{i=1}^N$ such that $\sum_{i=1}^N C^i = C$, we can formulate this problem as

$$\min_{w^i \in \mathcal{W}_i, \forall i} \; \max_{Z \in \mathbb{S}^d_{\succeq 0}} \sum_{i=1}^N f^i(w^i) + Z^\top \sum_{i=1}^N (C^i - A^i w^i) \tag{7a}$$

$$= \min_{w^i \in \mathcal{W}_i, \forall i} \; \max_{\substack{Z^i \in \mathbb{S}^d_{\succeq 0}, \forall i \\ Z^i = Z^j, \forall i,j}} \sum_{i=1}^N \big(f^i(w^i) + {Z^i}^\top (C^i - A^i w^i) \big). \tag{7b}$$

To motivate this dual decomposition in optimization problems with quasi-convex objectives, we examine the following example taken from [25, p. 51].

Example III.1. (Convex fractional program): Let $\boldsymbol{c}$, $\boldsymbol{d} \in \mathbb{R}^N$, $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times N}$. Assuming that $\boldsymbol{d}^\top \boldsymbol{x} > 0$ whenever $A\boldsymbol{x} - b \geq 0$, we have

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^N} & \frac{(\boldsymbol{c}^\top \boldsymbol{x})^2}{\boldsymbol{d}^\top \boldsymbol{x}} \\ \text{s.t.} \; & A\boldsymbol{x} - b \geq 0 \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^N, \, t \in \mathbb{R}} \; & t \\ \text{s.t.} \; & A\boldsymbol{x} - b \geq 0 \\ & \begin{bmatrix} t & \boldsymbol{c}^\top \boldsymbol{x} \\ \boldsymbol{c}^\top \boldsymbol{x} & \boldsymbol{d}^\top \boldsymbol{x} \end{bmatrix} \succeq 0, \end{aligned} \tag{8}$$

where we think of $\boldsymbol{x} = (x_1, \ldots, x_N)$ as the aggregate of decision vectors across the network. Then we introduce copies $\{t_i\}_{i=1}^N$ of the resulting primal variable subject to agreement, and split the constraints accordingly,

$$\begin{aligned} \min_{\substack{\boldsymbol{x} \in \mathbb{R}^N, t_i \in \mathbb{R}, \\ t_i = t_j, \forall i,j}} & \sum \tfrac{t_i}{N} \\ \text{s.t.} \; & \sum_{i=1}^N \boldsymbol{a}^i \, x_i - \tfrac{b}{N} \geq 0 \\ & \sum_{i=1}^N \Big(\begin{bmatrix} \frac{1}{N} & 0 \\ 0 & 0 \end{bmatrix} t_i + \begin{bmatrix} 0 & c_i \\ c_i & d_i \end{bmatrix} x_i \Big) \succeq 0, \end{aligned}$$

where $\boldsymbol{a}^i$ is the ith column of A. The resulting min-max problem has the form in (3),

$$\begin{aligned} \min_{\substack{\boldsymbol{x} \in \mathbb{R}^N, t_i \in \mathbb{R}, \\ t_i = t_j, \forall i,j}} \; \max_{\substack{\boldsymbol{\lambda}^i \in \mathbb{R}^m_{\geq 0}, \, Z^i \in \mathbb{S}^2_{\succeq 0} \\ \boldsymbol{\lambda}^i = \boldsymbol{\lambda}^j, \, Z^i = Z^j, \forall i,j}} & \Big\{ \sum_{i=1}^N \tfrac{t_i}{N} + \sum_{i=1}^N {\boldsymbol{\lambda}^i}^\top \big(\tfrac{b}{N} - \boldsymbol{a}^i \, x_i \big) \\ & - \sum_{i=1}^N \operatorname{trace} \Big(Z^i \Big(\begin{bmatrix} \frac{1}{N} & 0 \\ 0 & 0 \end{bmatrix} t_i + \begin{bmatrix} 0 & c_i \\ c_i & d_i \end{bmatrix} x_i \Big) \Big) \Big\}. \end{aligned} \quad \bullet$$

C. Saddle-point dynamics with Laplacian averaging

We propose a projected subgradient method for constrained saddle-point problems of the form (3). The agreement constraints are addressed via Laplacian averaging, allowing the design of distributed algorithms when the convex-concave functions are separable as in the cases above.

$$\hat{\boldsymbol{w}}_{t+1} = \boldsymbol{w}_t - \eta_t g_{\boldsymbol{w}_t} \tag{9a}$$
$$\hat{\boldsymbol{D}}_{t+1} = \boldsymbol{D}_t - \sigma \mathbf{L}_t \boldsymbol{D}_t - \eta_t g_{\boldsymbol{D}_t} \tag{9b}$$
$$\hat{\boldsymbol{\mu}}_{t+1} = \boldsymbol{\mu}_t + \eta_t g_{\boldsymbol{\mu}_t} \tag{9c}$$
$$\hat{\boldsymbol{z}}_{t+1} = \boldsymbol{z}_t - \sigma \mathbf{L}_t \boldsymbol{z}_t + \eta_t g_{\boldsymbol{z}_t} \tag{9d}$$
$$(\boldsymbol{w}_{t+1}, \boldsymbol{D}_{t+1}, \boldsymbol{\mu}_{t+1}, \boldsymbol{z}_{t+1}) = \mathcal{P}_{\boldsymbol{S}}\big(\hat{\boldsymbol{w}}_{t+1}, \hat{\boldsymbol{D}}_{t+1}, \hat{\boldsymbol{\mu}}_{t+1}, \hat{\boldsymbol{z}}_{t+1}\big),$$

where $\mathbf{L}_t = \mathsf{L}_t \otimes \mathrm{I}_{d_D}$ or $\mathbf{L}_t = \mathsf{L}_t \otimes \mathrm{I}_{d_z}$, depending on the context, and L_t is the Laplacian matrix for the communication graph $\mathcal{G}_t$ that might change over time; σ is a design parameter, $\{\eta_t\}_{t\geq 1}$ are the learning rates;

$$\begin{aligned}
g_{\boldsymbol{w}_t} &\in \partial_{\boldsymbol{w}} \boldsymbol{\phi}(\boldsymbol{w}_t, \boldsymbol{D}_t, \boldsymbol{\mu}_t, \boldsymbol{z}_t),\\
g_{\boldsymbol{D}_t} &\in \partial_{\boldsymbol{D}} \boldsymbol{\phi}(\boldsymbol{w}_t, \boldsymbol{D}_t, \boldsymbol{\mu}_t, \boldsymbol{z}_t),\\
g_{\boldsymbol{\mu}_t} &\in \partial_{\boldsymbol{\mu}} \boldsymbol{\phi}(\boldsymbol{w}_t, \boldsymbol{D}_t, \boldsymbol{\mu}_t, \boldsymbol{z}_t),\\
g_{\boldsymbol{z}_t} &\in \partial_{\boldsymbol{z}} \boldsymbol{\phi}(\boldsymbol{w}_t, \boldsymbol{D}_t, \boldsymbol{\mu}_t, \boldsymbol{z}_t),
\end{aligned}$$

and $\mathcal{P}_{\boldsymbol{S}}$ represents the orthogonal projection onto the closed convex set $\boldsymbol{S} := \boldsymbol{W} \times \mathcal{D}^N \times \boldsymbol{M} \times \mathcal{Z}^N$. This class of algorithms particularize to a novel class of primal-dual consensus-based subgradient methods when the convex-concave function takes the Lagrangian form in (5c).

IV. Convergence analysis

Here we present our technical analysis on the convergence properties of the dynamics (9). Our starting point is the assumption that a solution to (3) exists, namely, a saddle point $(\boldsymbol{w}^*, \boldsymbol{D}^*, \boldsymbol{\mu}^*, \boldsymbol{z}^*)$ of $\boldsymbol{\phi}$ on $\boldsymbol{S} := \boldsymbol{W} \times \mathcal{D}^N \times \boldsymbol{M} \times \mathcal{Z}^N$ under the agreement condition on $\mathcal{D}^N$ and $\mathcal{Z}^N$. That is, with $\boldsymbol{D}^* = D^* \otimes \mathbb{1}_N$ and $\boldsymbol{z}^* = z^* \otimes \mathbb{1}_N$ for some $(D^*, z^*) \in \mathcal{D} \times \mathcal{Z}$. We then study the evolution of the *running time-averages*

$$\boldsymbol{w}^{\text{av}}_{t+1} = \frac{1}{t}\sum_{s=1}^{t} \boldsymbol{w}_s, \quad \boldsymbol{D}^{\text{av}}_{t+1} = \frac{1}{t}\sum_{s=1}^{t} \boldsymbol{D}_s,$$
$$\boldsymbol{\mu}^{\text{av}}_{t+1} = \frac{1}{t}\sum_{s=1}^{t} \boldsymbol{\mu}_s, \qquad \boldsymbol{z}^{\text{av}}_{t} = \frac{1}{t}\sum_{s=1}^{t-1} \boldsymbol{z}_s,$$

towards the saddle point. Our strategy is the following. In Section IV-A, we bound the saddle-point evaluation error

$$t\boldsymbol{\phi}(\boldsymbol{w}^{\text{av}}_{t+1}, \boldsymbol{D}^{\text{av}}_{t+1}, \boldsymbol{\mu}^{\text{av}}_{t+1}, \boldsymbol{z}^{\text{av}}_{t+1}) - t\boldsymbol{\phi}(\boldsymbol{w}^*, \boldsymbol{D}^*, \boldsymbol{\mu}^*, \boldsymbol{z}^*). \tag{10}$$

in terms of the following quantities: the initial conditions, the size of the states of the dynamics, the size of the subgradients, and the cumulative disagreement of the running time-averages. Then, in Section IV-B we bound the cumulative disagreement in terms of the size of the subgradients and the learning rates. Finally, in Section IV-C we establish the saddle-point evaluation convergence result using the assumption that the estimates generated by the dynamics (9), as well as the subgradient sets, are uniformly bounded. (This assumption can be met in applications by designing projections that preserve the saddle points, particularly in the case of distributed constrained optimization.)

A. Saddle-point error in terms of the disagreement

Here, we bound the saddle-point evaluation error of the running time-averages in terms of the disagreement.

Lemma IV.1. (Evaluation error of the states in terms of the disagreement): *Let the sequence $\{(\boldsymbol{w}_t, \boldsymbol{D}_t, \boldsymbol{\mu}_t, \boldsymbol{z}_t)\}_{t\geq 1}$ be generated by the coordination algorithm (9) over a sequence of arbitrary weight-balanced digraphs $\{\mathcal{G}_t\}_{t\geq 1}$ such that $\sup_{t\geq 1} \lambda_{max}(\mathsf{L}_t) \leq \overline{\Lambda}$, and with*

$$\sigma \leq \big(\max\big\{\, d_{\text{out,t}}(k) \;:\; k \in \mathcal{I},\, t \in \mathbb{Z}_{\geq 1}\big\}\big)^{-1}. \tag{11}$$

Then, for any sequence of learning rates $\{\eta_t\}_{t\geq 1} \subset \mathbb{R}_{>0}$ and any $(\boldsymbol{w}_p, \boldsymbol{D}_p) \in \boldsymbol{W} \times \mathcal{D}^N$, the following holds:

$$\begin{aligned}
&2(\boldsymbol{\phi}(\boldsymbol{w}_t, \boldsymbol{D}_t, \boldsymbol{\mu}_t, \boldsymbol{z}_t) - \boldsymbol{\phi}(\boldsymbol{w}_p, \boldsymbol{D}_p, \boldsymbol{\mu}_t, \boldsymbol{z}_t))\\
&\leq \tfrac{1}{\eta_t}\big(\|\boldsymbol{w}_t - \boldsymbol{w}_p\|_2^2 - \|\boldsymbol{w}_{t+1} - \boldsymbol{w}_p\|_2^2\big)\\
&\quad + \tfrac{1}{\eta_t}\big(\|\mathbf{M}\boldsymbol{D}_t - \boldsymbol{D}_p\|_2^2 - \|\mathbf{M}\boldsymbol{D}_{t+1} - \boldsymbol{D}_p\|_2^2\big)\\
&\quad + 6\eta_t\|g_{\boldsymbol{w}_t}\|_2^2 + 6\eta_t\|g_{\boldsymbol{D}_t}\|_2^2\\
&\quad + 2\|g_{\boldsymbol{D}_t}\|_2(2+\sigma\overline{\Lambda})\|\mathbf{L}_{\mathcal{K}}\boldsymbol{D}_t\|_2 + 2\|g_{\boldsymbol{D}_t}\|_2\|\mathbf{L}_{\mathcal{K}}\boldsymbol{D}_p\|_2.
\end{aligned} \tag{12}$$

Also, for any $(\boldsymbol{\mu}_p, \boldsymbol{z}_p) \in \boldsymbol{M} \times \mathcal{Z}^N$, the analogous holds,

$$\begin{aligned}
&2(\boldsymbol{\phi}(\boldsymbol{w}_t, \boldsymbol{D}_t, \boldsymbol{\mu}_t, \boldsymbol{z}_t) - \boldsymbol{\phi}(\boldsymbol{w}_t, \boldsymbol{D}_t, \boldsymbol{\mu}_p, \boldsymbol{z}_p))\\
&\geq -\tfrac{1}{\eta_t}\big(\|\boldsymbol{\mu}_t - \boldsymbol{\mu}_p\|_2^2 - \|\boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}_p\|_2^2\big)\\
&\quad - \tfrac{1}{\eta_t}\big(\|\mathbf{M}\boldsymbol{z}_t - \boldsymbol{z}_p\|_2^2 - \|\mathbf{M}\boldsymbol{z}_{t+1} - \boldsymbol{z}_p\|_2^2\big)\\
&\quad - 6\eta_t\|g_{\boldsymbol{\mu}_t}\|_2^2 - 6\eta_t\|g_{\boldsymbol{z}_t}\|_2^2\\
&\quad - 2\|g_{\boldsymbol{z}_t}\|_2(2+\sigma\overline{\Lambda})\|\mathbf{L}_{\mathcal{K}}\boldsymbol{z}_t\|_2 - 2\|g_{\boldsymbol{z}_t}\|_2\|\mathbf{L}_{\mathcal{K}}\boldsymbol{z}_p\|_2.
\end{aligned} \tag{13}$$

In the previous result we have obtained inequalities on the evaluation error of the states of the dynamics with respect to a generic point in the variables of the convex and concave parts. Next, we obtain analogous bounds for the sum over time of these evaluation errors with respect to the same generic points and the running time-averages.

Lemma IV.2. (Cumulative evaluation error of the states with respect to running time-averages in terms of disagreement): *Under the same assumptions of Lemma IV.1, for any $(\boldsymbol{w}_p, \boldsymbol{D}_p, \boldsymbol{\mu}_p, \boldsymbol{z}_p) \in \boldsymbol{W} \times \mathcal{D}^N \times \boldsymbol{M} \times \mathcal{Z}^N$, the difference*

$$\sum_{s=1}^{t} \boldsymbol{\phi}(\boldsymbol{w}_s, \boldsymbol{D}_s, \boldsymbol{\mu}_s, \boldsymbol{z}_s) - t\boldsymbol{\phi}(\boldsymbol{w}_p, \boldsymbol{D}_p, \boldsymbol{\mu}^{av}_{t+1}, \boldsymbol{z}^{av}_{t+1})$$

is upper-bounded by $\frac{\mathsf{u}(t, \boldsymbol{w}_p, \boldsymbol{D}_p)}{2}$, while the difference

$$\sum_{s=1}^{t} \boldsymbol{\phi}(\boldsymbol{w}_s, \boldsymbol{D}_s, \boldsymbol{\mu}_s, \boldsymbol{z}_s) - t\boldsymbol{\phi}(\boldsymbol{w}^{av}_{t+1}, \boldsymbol{D}^{av}_{t+1}, \boldsymbol{\mu}_p, \boldsymbol{z}_p)$$

is lower-bounded by $-\frac{\mathsf{u}(t,\boldsymbol{\mu}_p,\boldsymbol{z}_p)}{2}$, where

$$\begin{aligned}
&\mathsf{u}(t,\boldsymbol{w}_p,\boldsymbol{D}_p) \equiv \mathsf{u}\big(t,\boldsymbol{w}_p,\boldsymbol{D}_p,\{\boldsymbol{w}_s\}_{s=1}^t,\{\boldsymbol{D}_s\}_{s=1}^t\big) \quad (14)\\
&= \sum_{s=2}^{t}\big(\|\boldsymbol{w}_s-\boldsymbol{w}_p\|_2^2+\|\mathbf{M}\boldsymbol{D}_s-\boldsymbol{D}_p\|_2^2\big)\big(\tfrac{1}{\eta_s}-\tfrac{1}{\eta_{s-1}}\big)\\
&\quad+\tfrac{2}{\eta_1}\big(\|\boldsymbol{w}_1\|_2^2+\|\boldsymbol{w}_p\|_2^2+\|\boldsymbol{D}_1\|_2^2+\|\boldsymbol{D}_p\|_2^2\big)\\
&\quad+6\sum_{s=1}^{t}\eta_s(\|g_{\boldsymbol{w}_s}\|_2^2+\|g_{\boldsymbol{D}_s}\|_2^2)\\
&+2(2+\sigma\overline{\Lambda})\sum_{s=1}^{t}\|g_{\boldsymbol{D}_s}\|_2\|\mathbf{L}_{\mathcal{K}}\boldsymbol{D}_s\|_2+2\|\mathbf{L}_{\mathcal{K}}\boldsymbol{D}_p\|_2\sum_{s=1}^{t}\|g_{\boldsymbol{D}_s}\|_2, \quad (15)
\end{aligned}$$

and $\mathsf{u}(t,\boldsymbol{\mu}_p,\boldsymbol{z}_p) \equiv \mathsf{u}\big(t,\boldsymbol{\mu}_p,\boldsymbol{z}_p,\{\boldsymbol{\mu}_s\}_{s=1}^t,\{\boldsymbol{z}_s\}_{s=1}^t\big)$.

Next we combine the pair of inequalities obtained above to derive the saddle-point evaluation error of the running time-averages.

Proposition IV.3. (Saddle-point evaluation error of running time-averages): *Under the same hypotheses of Lemma IV.1, for any saddle point $(\boldsymbol{w}^*,\boldsymbol{D}^*,\boldsymbol{\mu}^*,\boldsymbol{z}^*)$ of $\boldsymbol{\phi}$ on $\boldsymbol{W}\times\mathcal{D}^N\times\boldsymbol{M}\times\mathcal{Z}^N$ with $\boldsymbol{D}^*=D^*\otimes\mathbb{1}_N$ and $\boldsymbol{z}^*=z^*\otimes\mathbb{1}_N$ for some $(D^*,z^*)\in\mathcal{D}\times\mathcal{Z}$, the following holds:*

$$\begin{aligned}
&-\mathsf{u}(t,\boldsymbol{\mu}^*,\boldsymbol{z}^*)-\mathsf{u}(t,\boldsymbol{w}_{t+1}^{av},\boldsymbol{D}_{t+1}^{av})\\
&\le 2t\boldsymbol{\phi}(\boldsymbol{w}_{t+1}^{av},\boldsymbol{D}_{t+1}^{av},\boldsymbol{\mu}_{t+1}^{av},\boldsymbol{z}_{t+1}^{av})-2t\boldsymbol{\phi}(\boldsymbol{w}^*,\boldsymbol{D}^*,\boldsymbol{\mu}^*,\boldsymbol{z}^*)\\
&\le \mathsf{u}(t,\boldsymbol{w}^*,\boldsymbol{D}^*)+\mathsf{u}(t,\boldsymbol{\mu}_{t+1}^{av},\boldsymbol{z}_{t+1}^{av}). \quad (16)
\end{aligned}$$

B. Cumulative disagreement

Here we bound the cumulative disagreement of the estimates. To study the disagreement over time on the estimates $\boldsymbol{D}_t$ and $\boldsymbol{z}_t$, we treat the subgradient terms as perturbations in the dynamics (9) and study its input-to-state stability properties. This approach is well suited for scenarios where the size of the subgradients can be uniformly bounded. Since the coupling in (9) with $\boldsymbol{w}_t$ and among the estimates $\boldsymbol{D}_t$ and $\boldsymbol{z}_t$ themselves takes place only through the subgradients, we focus on the following pair of decoupled dynamics,

$$\begin{aligned}
\hat{\boldsymbol{D}}_{t+1} &= \boldsymbol{D}_t-\sigma\mathbf{L}_t\boldsymbol{D}_t+\boldsymbol{u}_t^1 \quad (17a)\\
\hat{\boldsymbol{z}}_{t+1} &= \boldsymbol{z}_t-\sigma\mathbf{L}_t\boldsymbol{z}_t+\boldsymbol{u}_t^2 \quad (17b)\\
(\boldsymbol{D}_{t+1},\boldsymbol{z}_{t+1}) &= \mathcal{P}_{\mathcal{D}^N\times\mathcal{Z}^N}\big(\hat{\boldsymbol{D}}_{t+1},\hat{\boldsymbol{z}}_{t+1}\big),
\end{aligned}$$

where $\{\boldsymbol{u}_t^1\}_{t\ge1}\subset(\mathbb{R}^{d_D})^N$, $\{\boldsymbol{u}_t^2\}_{t\ge1}\subset(\mathbb{R}^{d_z})^N$ are arbitrary sequences of disturbances, and $\mathcal{P}_{\mathcal{D}^N\times\mathcal{Z}^N}$ is the orthogonal projection onto $\mathcal{D}^N\times\mathcal{Z}^N$.

The next result characterizes the input-to-state stability properties of (17) with respect to the agreement space.

Proposition IV.4. (Cumulative disagreement on (17) over jointly-connected weight-balanced digraphs): *Let $\{\mathcal{G}_s\}_{s\ge1}$ be a sequence of B-jointly connected, δ-nondegenerate, weight-balanced digraphs. For $\tilde{\delta}'\in(0,1)$, let*

$$\tilde{\delta} := \min\Big\{\,\tilde{\delta}',\ (1-\tilde{\delta}')\frac{\delta}{d_{\max}}\,\Big\}, \quad (18)$$

where

$$d_{\max} := \max\big\{\,d_{\text{out,t}}(k)\ :\ k\in\mathcal{I},\ t\in\mathbb{Z}_{\ge1}\,\big\}.$$

Then, for any choice

$$\sigma\in\Big[\,\frac{\tilde{\delta}}{\delta},\ \frac{1-\tilde{\delta}}{d_{\max}}\,\Big], \quad (19)$$

the dynamics (17a) over $\{\mathcal{G}_t\}_{t\ge1}$ is input-to-state stable with respect to the nullspace of the matrix $\hat{\mathbf{L}}_{\mathcal{K}}$. Specifically, for any $t\in\mathbb{Z}_{\ge1}$ and any $\{\boldsymbol{u}_s^1\}_{s=1}^{t-1}\subset(\mathbb{R}^{d_D})^N$,

$$\|\mathbf{L}_{\mathcal{K}}\boldsymbol{D}_t\|_2 \le \tfrac{2^4\|\boldsymbol{D}_1\|_2}{3^2}\Big(1-\frac{\tilde{\delta}}{4N^2}\Big)^{\lceil\frac{t-1}{B}\rceil}+C_u\max_{1\le s\le t-1}\|\boldsymbol{u}_s^1\|_2, \quad (20)$$

where

$$C_u := \frac{2^5/3^2}{1-\big(1-\frac{\tilde{\delta}}{4N^2}\big)^{1/B}} \quad (21)$$

and the cumulative disagreement satisfies

$$\sum_{t=1}^{t'}\|\mathbf{L}_{\mathcal{K}}\boldsymbol{D}_t\|_2 \le C_u\Big(\tfrac{\|\boldsymbol{D}_1\|_2}{2}+\sum_{t=1}^{t'-1}\|\boldsymbol{u}_t^1\|_2\Big). \quad (22)$$

Analogous bounds hold interchanging $\boldsymbol{D}_t$ by $\boldsymbol{z}_t$.

C. Convergence of the saddle-point dynamics with Laplacian averaging

We now state the saddle-point evaluation convergence of the dynamics (9). In words, under a mild connectivity assumption on the communication digraphs, a suitable choice of decreasing stepsizes, and assuming that the agents' estimates and the subgradient sets are uniformly bounded, then the saddle-point evaluation error decreases proportionally to $\frac{1}{\sqrt{t}}$.

Theorem IV.5. (Convergence of the saddle-point dynamics with Laplacian averaging): *Let the sequence $\{(\boldsymbol{w}_t,\boldsymbol{D}_t,\boldsymbol{\mu}_t,\boldsymbol{z}_t)\}_{t\ge1}$ be generated by the coordination algorithm (9) over a sequence $\{\mathcal{G}_t\}_{t\ge1}$ of B-jointly connected, δ-nondegenerate, weight-balanced digraphs satisfying $\sup_{t\ge1}\lambda_{max}(\mathsf{L}_t)\le\overline{\Lambda}$ with σ selected as in (19). Assume that the estimates are bounded as*

$$\|\boldsymbol{w}_t\|_2\le B_{\boldsymbol{w}},\ \ \|\boldsymbol{D}_t\|_2\le B_{\boldsymbol{D}},\ \ \|\boldsymbol{\mu}_t\|_2\le B_{\boldsymbol{\mu}},\ \ \|\boldsymbol{z}_t\|_2\le B_{\boldsymbol{z}},$$

for all $t\in\mathbb{Z}_{\ge1}$ whenever the sequence of learning rates $\{\eta_t\}_{t\ge1}\subset\mathbb{R}_{>0}$ is uniformly bounded. Assume also that the subgradients are bounded as

$$\|g_{\boldsymbol{w}_t}\|_2\le H_{\boldsymbol{w}},\ \|g_{\boldsymbol{D}_t}\|_2\le H_{\boldsymbol{D}},\ \|g_{\boldsymbol{\mu}_t}\|_2\le H_{\boldsymbol{\mu}},\ \|g_{\boldsymbol{z}_t}\|_2\le H_{\boldsymbol{z}}$$

for all $t\in\mathbb{Z}_{\ge1}$. Consider the following choice of learning rates called the Doubling Trick scheme*: for $m=0,1,2,\dots,\lceil\log_2 t\rceil$, we take $\eta_s=\frac{1}{\sqrt{2^m}}$ in each period of 2^m rounds $s=2^m,\dots,2^{m+1}-1$. Then, for any saddle point $(\boldsymbol{w}^*,\boldsymbol{D}^*,\boldsymbol{\mu}^*,\boldsymbol{z}^*)$ of $\boldsymbol{\phi}$ on $\boldsymbol{W}\times\mathcal{D}^N\times\boldsymbol{M}\times\mathcal{Z}^N$ with $\boldsymbol{D}^*=D^*\otimes\mathbb{1}_N$ and $\boldsymbol{z}^*=z^*\otimes\mathbb{1}_N$ for some $(D^*,z^*)\in\mathcal{D}\times\mathcal{Z}$, which is assumed to exist, the following holds:*

$$\begin{aligned}
-\frac{\alpha_{\boldsymbol{\mu},\boldsymbol{z}}+\alpha_{\boldsymbol{w},\boldsymbol{D}}}{2\sqrt{t-1}} &\le \boldsymbol{\phi}(\boldsymbol{w}_t^{av},\boldsymbol{D}_t^{av},\boldsymbol{z}_t^{av},\boldsymbol{\mu}_t^{av})-\boldsymbol{\phi}(\boldsymbol{w}^*,\boldsymbol{D}^*,\boldsymbol{z}^*,\boldsymbol{\mu}^*)\\
&\le \frac{\alpha_{\boldsymbol{w},\boldsymbol{D}}+\alpha_{\boldsymbol{\mu},\boldsymbol{z}}}{2\sqrt{t-1}}, \quad (23)
\end{aligned}$$

where $\alpha_{\boldsymbol{w},\boldsymbol{D}} := \frac{\sqrt{2}}{\sqrt{2}-1}\hat{\alpha}_{\boldsymbol{w},\boldsymbol{D}}$ *with*

$$\hat{\alpha}_{\boldsymbol{w},\boldsymbol{D}} := 4(B_{\boldsymbol{w}}^2 + B_{\boldsymbol{D}}^2) + 6(H_{\boldsymbol{w}}^2 + H_{\boldsymbol{D}}^2) + H_{\boldsymbol{D}}(3 + \sigma\overline{\Lambda})C_u\big(B_{\boldsymbol{D}} + 2H_{\boldsymbol{D}}\big)\,,$$

and $\alpha_{\boldsymbol{z},\boldsymbol{\mu}}$ *is analogously defined. The constant* C_u *given by* (21) *codifies the dependence on the network properties.*

We close this section commenting on the assumption in our main result about the boundedness of our dynamics' states, particularly in the application to constrained optimization.

Remark IV.6. (Saddle points of Lagrangians): Our general dynamics (9) can be particularized to a distributed strategy for constrained optimization problems of the form (4) via the saddle-point formulation with explicit agreement on the multipliers in (5c). The resulting algorithm consists of primal-dual subgradient updates followed by a projection step. The key is then the design of projections preserving the saddle points. While the Lagrange multipliers are naturally constrained to the positive orthant, under suitable conditions one can establish the boundedness of the optimal dual set [14], [13, Lemma 1]. One can then use these facts as in [13], [6], [18] to define a projection of the algorithms' estimates of the multipliers onto the intersection of the positive orthant and a sufficiently large ball, thereby preserving the set of Lagrange multipliers while at the same time ensuring the boundedness of the states of the dynamics. •

V. Conclusions and ideas for future work

We have proposed provably correct projected subgradient methods for saddle-point problems under explicit agreement constraints. We have shown that separable constrained optimization problems can be written in this form, where agreement plays a role in distributing the objectives (via agreement on a subset of the primal variables) as well as the constraints (via agreement on the dual variables). This approach enables the use of existing consensus-based ideas to tackle the algorithmic solution to these problems in a distributed fashion. Future extensions will include, first, a refined analysis for constrained optimization in terms of the cost error (instead of the saddle-point evaluation error). Second, the development of distributed strategies for bounding the optimal dual set necessary for the design of projections onto compact sets. Third, we envision applications to semidefinite programming, where chordal sparsity allows to tackle problems where the dimension of the matrices grows with the size of the network.

Acknowledgments

Research supported by NSF award CMMI-1300272.

References

[1] K. Arrow, L. Hurwitz, and H. Uzawa, *Studies in Linear and Non-Linear Programming*. Stanford, California: Stanford University Press, 1958.

[2] A. Nedić and A. Ozdaglar, "Subgradient methods for saddle-point problems," *Journal of Optimization Theory & Applications*, vol. 142, no. 1, pp. 205–228, 2009.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[4] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[5] B. Johansson, M. Rabi, and M. Johansson, "A randomized incremental subgradient method for distributed optimization in networked systems," *SIAM Journal on Control and Optimization*, vol. 20, no. 3, pp. 1157–1170, 2009.

[6] M. Zhu and S. Martínez, "On distributed convex optimization under inequality and equality constraints," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2012.

[7] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 781–786, 2014.

[8] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *IEEE Conf. on Decision and Control*, (Orlando, Florida), pp. 3800–3805, 2011.

[9] J. N. Tsitsiklis, *Problems in Decentralized Decision Making and Computation*. PhD thesis, Massachusetts Institute of Technology, Nov. 1984. Available at `http://web.mit.edu/jnt/www/Papers/PhD-84-jnt.pdf`.

[10] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.

[11] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Transactions on Automatic Control*, vol. 31, no. 9, pp. 803–812, 1986.

[12] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA: Athena Scientific, 1st ed., 2003.

[13] A. Nedić and A. Ozdaglar, "Approximate primal solutions and rate analysis for dual subgradient methods," *SIAM J. Optimization*, vol. 19, no. 4, pp. 1757–1780, 2010.

[14] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I*. Grundlehren Text Editions, New York: Springer, 1993.

[15] D. Yuan, S. Xu, and H. Zhao, "Distributed primal-dual subgradient method for multiagent optimization via consensus algorithms," *IEEE Trans. Systems, Man, and Cybernetics- Part B*, vol. 41, no. 6, pp. 1715–1724, 2011.

[16] A. Nedic, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.

[17] I. Necoara, I. Dumitrache, and J. A. K. Suykens, "Fast primal-dual projected linear iterations for distributed consensus in constrained convex optimization," in *IEEE Conf. on Decision and Control*, (Atlanta, GA), pp. 1366–1371, Dec. 2010.

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

[19] D. Richert and J. Cortés, "Robust distributed linear programming," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, 2015. To appear.

[20] D. Mosk-Aoyama, T. Roughgarden, and D. Shah, "Fully distributed algorithms for convex optimization problems," *SIAM J. Optimization*, vol. 20, no. 6, pp. 3260–3279, 2010.

[21] F. Bullo, J. Cortés, and S. Martínez, *Distributed Control of Robotic Networks*. Applied Mathematics Series, Princeton University Press, 2009. Electronically available at http://coordinationbook.info.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.

[23] A. Nedić and A. Ozdaglar, "Cooperative distributed multi-agent optimization," in *Convex Optimization in Signal Processing and Communications* (Y. Eldar and D. Palomar, eds.), Cambridge University Press, 2010.

[24] N. Parikh and S. Boyd, "Proximal algorithms," vol. 1, no. 3, pp. 123–231, 2013.

[25] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, pp. 49–95, Mar. 1996.