Distributed Optimization for Smart Cyber-Physical Networks

Other titles in Foundations and Trends® in Systems and Control

On the Control of Multi-Agent Systems: A Survey
Fei Chen and Wei Ren
ISBN: 978-1-68083-582-3

Analysis and Synthesis of Reset Control Systems
Christophe Prieur, Isabelle Queinnec, Sophie Tarbouriech and Luca Zaccarian
ISBN: 978-1-68083-522-9

Control and State Estimation for Max-Plus Linear Systems
Laurent Hardouin, Bertrand Cottenceau, Ying Shang and Jorg Raisch
ISBN: 978-1-68083-544-1

Distributed Optimization for Smart Cyber-Physical Networks

Giuseppe Notarstefano
Università di Bologna, Bologna, Italy
giuseppe.notarstefano@unibo.it

Ivano Notarnicola
Università di Bologna, Bologna, Italy
ivano.notarnicola@unibo.it

Andrea Camisa
Università di Bologna, Bologna, Italy
a.camisa@unibo.it

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Systems and Control

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

G. Notarstefano, I. Notarnicola and A. Camisa. *Distributed Optimization for Smart Cyber-Physical Networks.* Foundations and Trends® in Systems and Control, vol. 7, no. 3, pp. 253–383, 2020.

ISBN: 978-1-68083-619-6

Foundations and Trends® in Systems and Control

Volume 7, Issue 3, 2020

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Systems and Control publishes survey and tutorial articles in the following topics:

- Control of:
  - Hybrid and Discrete Event Systems
  - Nonlinear Systems
  - Network Systems
  - Stochastic Systems
  - Multi-agent Systems
  - Distributed Parameter Systems
  - Delay Systems
- Filtering, Estimation, Identification
- Optimal Control
- Systems Theory
- Control Applications

Information for Librarians

Foundations and Trends® in Systems and Control, 2020, Volume 7, 4 issues. ISSN paper version 2325-6818. ISSN online version 2325-6826. Also available as a combined paper and online subscription.

Contents

Distributed Optimization for Smart Cyber-Physical Networks

Giuseppe Notarstefano[1], Ivano Notarnicola[2] and Andrea Camisa[3]

[1] *Università di Bologna, Bologna (Italy); giuseppe.notarstefano@unibo.it*
[2] *Università di Bologna, Bologna (Italy); ivano.notarnicola@unibo.it*
[3] *Università di Bologna, Bologna (Italy); a.camisa@unibo.it*

ABSTRACT

The presence of embedded electronics and communication capabilities as well as sensing and control in smart devices has given rise to the novel concept of cyber-physical networks, in which agents aim at cooperatively solving complex tasks by local computation and communication. Numerous estimation, learning, decision and control tasks in smart networks involve the solution of large-scale, structured optimization problems in which network agents have only a partial knowledge of the whole problem. Distributed optimization aims at designing local computation and communication rules for the network processors allowing them to cooperatively solve the global optimization problem without relying on any central unit. The purpose of this survey is to provide an introduction to distributed optimization methodologies. Principal approaches, namely (primal) consensus-based, duality-based and constraint exchange methods, are formalized. An analysis of the basic schemes is supplied, and state-of-the-art extensions are reviewed.

Giuseppe Notarstefano, Ivano Notarnicola and Andrea Camisa (2020), Distributed Optimization for Smart Cyber-Physical Networks, Foundations and Trends® in Systems and Control: Vol. 7, No. 3, pp 253–383. DOI: 10.1561/2600000020.

Introduction

Motivation

In recent years, the breakthroughs in embedded electronics are giving the opportunity to include computation and communication capabilities in almost any device of several domains as factories, farms, buildings, grids and cities. Communication among devices has enabled a number of new challenges along the direction of turning smart devices into smart (cooperating) systems. The keyword "cyber-physical networks" is being adopted to refer to this permeating reality, whose distinctive feature is that a great advantage can be obtained if its interconnected, complex nature is exploited. A novel peer-to-peer *distributed* computational framework is emerging as a new opportunity in which peer processors, communicating over a network, cooperatively solve a task without resorting to a unique provider that knows and owns all the data.

Several challenges arising in cyber-physical networks can be stated as optimization problems. Examples are estimation, decision, learning and control applications. To solve optimization problems over cyber-physical networks, it is not possible to apply the classical optimization algorithms (that we call *centralized*), which require the data to be managed by a single entity. In fact, the problem data are spread over the network, and it is undesirable (or even impossible) to collect them at a unique node. To this end, parallel computing serves as a source of inspiration. In order to speed up the solution of large-scale optimization problems, several effort has been made in designing *parallel* algorithms by splitting the computational burden among several processors. However, for typical parallel optimization algorithms, a central coordinating node is required and the communication topology is designed ad hoc. In distributed computation the communication topology cannot be thought of as a design parameter. Rather, it is a given part of the problem. Thus, in cyber-physical networks, the goal is to design algorithms, based on the exchange of information among the processors, that take advantage of the aggregated computational power. All the agents must be treated as peers and each of them must perform the same tasks and no "master" node must be present. Moreover, information privacy is often a requirement (i.e., private problem data at each node must not be shared with the other nodes). These challenges call for tailored strategies and have given rise to a novel, growing research branch termed *distributed optimization*.

Scope of the Monograph

The purpose of this survey is to give a comprehensive overview of the most common approaches used to design distributed optimization algorithms, together with the theoretical analysis of the main schemes in their basic version. We identify and formalize classes of problem set-ups that arise in motivating application scenarios. For each set-up, in order to give the main tools for analysis, we review tailored distributed algorithms in simplified cases. Extensions and generalizations of the basic schemes are also discussed at the end of each chapter. The algorithms have been developed by combining mathematical tools from optimization theory (e.g., duality) and network control theory (e.g., average consensus). For some of the discussed algorithms, we will present also parallel algorithms that serve as a starting point for the development of distributed methods.

We focus on three main categories of distributed optimization approaches: *(i)* primal consensus-based methods, i.e., methods combining classical gradient or subgradient steps with local averaging schemes; *(ii)* dual methods, i.e., methods which employ the Lagrangian dual of suitable equivalent formulations of the target problem to obtain a distributed routine; *(iii)* constraint exchange methods, which are based on the exchange of (active) constraints among agents to compute a solution of the considered problem.

Survey papers on distributed optimization have been proposed in the literature. An early survey paper presenting a broad class of relevant optimization problems in control is [85]. It also discusses tailored, parallel and distributed optimization algorithms based on decomposition techniques and including also the distributed subgradient method. Recent surveys analyze thoroughly average consensus [87] and the distributed subgradient method [87, 88, 91], with a literature review on other distributed optimization techniques. The book [97] provides parallel and distributed asynchronous optimization algorithms, including gradient tracking techniques. Some latest advances in distributed optimization are collected in [45].

Organization

In Chapter 1, we introduce the relevant problem set-ups, that we call *cost-coupled*, *constraint-coupled* and *common cost*, along with several motivating applications of interest arising in estimation, learning, decision and control. In Chapter 2 we provide an overview of primal approaches to solve cost-coupled problems, namely the distributed subgradient algorithm and the gradient tracking algorithm. In Chapter 3, a discussion on relevant duality forms for distributed optimization is first provided, and then distributed algorithms relying on Lagrangian approaches are reviewed. Namely, for cost-coupled problems, distributed dual decomposition and distributed ADMM algorithms are considered, while for constraint-coupled problems, a distributed dual subgradient algorithm and a method based on relaxation and successive distributed decomposition are presented. In Chapter 4, we focus on constraint exchange methods. We introduce the Constraints Consensus algorithm applied to common-cost problems, along with its most relevant extensions.

We also provide illustrative numerical examples to highlight significant properties of the considered distributed optimization methods. Since the described algorithms are designed for different problem set-ups, different, relevant simulation scenarios are considered in each chapter.

Concluding Remarks

In this survey, we considered a distributed optimization framework arising in modern cyber-physical networks, in which computing units have only a partial knowledge of a global optimization problem and must solve it through local computation and communication without any central coordinator. First, we introduced main optimization set-ups addressed in distributed optimization (i.e., cost-coupled, common-cost, and constraint-coupled), and motivated them with relevant estimation, learning, decision and control applications arising in smart networks. Then, we reviewed three main approaches to design distributed optimization algorithms, namely (primal) consensus-based, duality-based and constraint-exchange methods, and provided a theoretical analysis under simplified communication assumptions and/or problem set-ups. To highlight the behavior of the presented algorithms, the theoretical results are also equipped with numerical examples.

Appendices

A

Centralized Optimization Methods

A.1 Gradient Method

Consider the following unconstrained optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x}), \tag{A.1}$$

where $f : \mathbb{R}^d \to \mathbb{R}$. The gradient method is an iterative algorithm given by

$$\mathbf{x}^{t+1} = \mathbf{x}^t - \gamma^t \nabla f(\mathbf{x}^t), \tag{A.2}$$

where $t \geq 0$ denotes the iteration counter and γ^t is the step-size. The following result states the convergence of the gradient method for constant step-size.

Proposition A.1 ([9, Proposition 1.2.3])**.** Assume that f is a $\mathcal{C}^1$ function with Lipschitz continuous gradient ∇f with constant L. Let the step-size be constant, i.e., $\gamma^t = \gamma$, for all $t \geq 0$, and such that $0 < \gamma < 2/L$. Then, every limit point of the sequence $\{\mathbf{x}^t\}_{t\geq 0}$ generated by the gradient method (A.2), is a stationary point of problem (A.1), i.e., there exists a subset of indices $\mathcal{K} \subseteq \mathbb{N}$ such that

$$\lim_{\mathcal{K}\ni t\to\infty} \|\mathbf{x}^t - \bar{\mathbf{x}}\| = 0,$$

where $\bar{\mathbf{x}}$ is a stationary point of (A.1). □

The previous result can be extended in several ways, e.g., with different step-size rules and adapted to constrained problems. We refer the interested reader to [9] and references therein.

A.2 Subgradient Method

Consider the following constrained optimization problem

$$\min_{\mathbf{x}\in X} \ f(\mathbf{x}), \tag{A.3}$$

with $f : \mathbb{R}^d \to \mathbb{R}$ a convex function and $X \subseteq \mathbb{R}^d$ a closed, convex set.

A vector $\widetilde{\nabla} f(\mathbf{x}) \in \mathbb{R}^d$ is called a subgradient of the convex function f at $\mathbf{x} \in \mathbb{R}^d$ if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \widetilde{\nabla} f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x})$$

for all $\mathbf{y} \in \mathbb{R}^d$. The (projected) subgradient method is the iterative algorithm given by

$$\mathbf{x}^{t+1} = \mathcal{P}_X\left(\mathbf{x}^t - \gamma^t \widetilde{\nabla} f(\mathbf{x}^t)\right), \tag{A.4}$$

where $t \geq 0$ denotes the iteration counter, γ^t is the step-size, $\widetilde{\nabla} f(\mathbf{x}^t)$ denotes a subgradient of f at $\mathbf{x}^t$, and $\mathcal{P}_X(\,\cdot\,)$ is the Euclidean projection onto X.

Assumption A.1 (Diminishing Step-size). The step-size sequence $\{\gamma^t\}_{t\geq 0}$ is such that $\gamma^t \geq 0$ and satisfies

$$\lim_{t\to\infty} \gamma^t = 0, \qquad \sum_{t=0}^{\infty} \gamma^t = \infty, \qquad \sum_{t=0}^{\infty} (\gamma^t)^2 < \infty. \qquad \square$$

The following proposition formally states the convergence of the subgradient method (A.4).

Proposition A.2 ([10, Proposition 3.2.6]). Assume that all the subgradients of f are bounded at each $\mathbf{x} \in X$. Moreover, assume the optimal solution set of problem (A.3) is not empty. Let the step-size γ^t satisfy Assumption A.1. Then, the sequence $\{\mathbf{x}^t\}_{t\geq 0}$ generated by the subgradient method (A.4) converges to an optimal solution $\mathbf{x}^\star$ of problem (A.3), i.e.,

$$\lim_{t\to\infty} \|\mathbf{x}^t - \mathbf{x}^\star\| = 0, \qquad \lim_{t\to\infty} \|f(\mathbf{x}^t) - f^\star\| = 0. \qquad \square$$

A.3 Lagrangian Duality and Dual Subgradient Method

Consider a constrained optimization problem, addressed as primal problem, having the form

$$\begin{aligned} \min_{\mathbf{x} \in X} \; & f(\mathbf{x}) \\ \text{subj. to } & \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \end{aligned} \tag{A.5}$$

where $X \subseteq \mathbb{R}^d$ is a convex, compact set, $f : \mathbb{R}^d \to \mathbb{R}$ is a convex function and $\mathbf{g} : \mathbb{R}^d \to \mathbb{R}^S$ is such that each component $\mathbf{g}_s : \mathbb{R}^d \to \mathbb{R}$, $s \in \{1, \ldots, S\}$, is a convex (scalar) function.

The following optimization problem

$$\begin{aligned} \max_{\boldsymbol{\mu}} \; & q(\boldsymbol{\mu}) \\ \text{subj. to } & \boldsymbol{\mu} \geq \mathbf{0} \end{aligned} \tag{A.6}$$

is called the dual of problem (A.5), where $q : \mathbb{R}^S \to \mathbb{R}$ is obtained by minimizing with respect to $\mathbf{x} \in X$ the Lagrangian function $\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) + \boldsymbol{\mu}^\top \mathbf{g}(\mathbf{x})$, i.e., $q(\boldsymbol{\mu}) = \min_{\mathbf{x} \in X} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu})$. It can be shown that the domain of q (i.e., the set of $\boldsymbol{\mu}$ such that $q(\boldsymbol{\mu}) > -\infty$) is convex and that q is concave on its domain. A vector $\bar{\boldsymbol{\mu}} \in \mathbb{R}^S$ is said to be a Lagrange multiplier if it holds $\bar{\boldsymbol{\mu}} \geq \mathbf{0}$ and

$$\inf_{\mathbf{x} \in X} \mathcal{L}(\mathbf{x}, \bar{\boldsymbol{\mu}}) = \inf_{\mathbf{x} \in X \,:\, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}} f(\mathbf{x}).$$

It can be shown that the following inequality holds [9]

$$\inf_{\mathbf{x} \in X} \sup_{\boldsymbol{\mu} \geq \mathbf{0}} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) \geq \sup_{\boldsymbol{\mu} \geq \mathbf{0}} \inf_{\mathbf{x} \in X} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}), \tag{A.7}$$

which is called weak duality. When in (A.7) the equality holds, then we say that strong duality holds and, thus, solving the primal problem (A.5) is equivalent to solving its dual formulation (A.6). In this case the right-hand-side problem in (A.7) is referred to as *saddle-point problem* of (A.5).

Definition A.1. A pair $(\mathbf{x}^\star, \boldsymbol{\mu}^\star)$ is called a primal-dual optimal solution of problem (A.5) if $\mathbf{x}^\star \in X$ and $\boldsymbol{\mu}^\star \geq \mathbf{0}$, and $(\mathbf{x}^\star, \boldsymbol{\mu}^\star)$ is a saddle point of the Lagrangian, i.e.,

$$\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\mu}) \leq \mathcal{L}(\mathbf{x}^\star, \boldsymbol{\mu}^\star) \leq \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}^\star)$$

for all $\mathbf{x} \in X$ and $\boldsymbol{\mu} \geq \mathbf{0}$. □

Given the dual function q, an important property is as follows. A subgradient of $-q$ at a given $\bar{\boldsymbol{\mu}}$ can be efficiently computed as $g(\bar{\mathbf{x}})$, where $\bar{\mathbf{x}} = \text{argmin}_{\mathbf{x}\in X}\ f(\mathbf{x}) + \bar{\boldsymbol{\mu}}^\top g(\mathbf{x})$ (see [9, Section 6] for further details). Then, a subgradient method to solve the dual problem (A.6) reads

$$\mathbf{x}^{t+1} = \underset{\mathbf{x}\in X}{\text{argmin}}\ f(\mathbf{x}) + (\boldsymbol{\mu}^t)^\top g(\mathbf{x})$$
$$\boldsymbol{\mu}^{t+1} = \mathcal{P}_{\boldsymbol{\mu}\geq 0}\Big(\boldsymbol{\mu}^t + \gamma^t g(\mathbf{x}^{t+1})\Big),$$

where γ^t is a suitable step-size and $\boldsymbol{\mu}^0 \geq 0$ is arbitrary.

A.4 ADMM Algorithm

In this section, we review the Alternating Direction Method of Multipliers (ADMM) following [12, Section 3.4]. Consider the following optimization problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^d}\ & G_1(\mathbf{x}) + G_2(A\mathbf{x}) \\ \text{subj. to}\ & \mathbf{x} \in C_1,\ A\mathbf{x} \in C_2, \end{aligned} \tag{A.8}$$

where $G_1 : \mathbb{R}^d \to \mathbb{R}$ and $G_2 : \mathbb{R}^S \to \mathbb{R}$ are convex functions, A is a $S \times d$ matrix, and $C_1 \subseteq \mathbb{R}^d$ and $C_2 \subseteq \mathbb{R}^S$ are nonempty, closed convex sets. We assume that the optimal solution set $X^\star$ of problem (A.8) is nonempty. Furthermore, either C_1 is bounded or else $A^\top A$ is invertible.

Problem (A.8) can be equivalently rewritten as

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^d, \mathbf{z}\in\mathbb{R}^S}\ & G_1(\mathbf{x}) + G_2(\mathbf{z}) \\ \text{subj. to}\ & A\mathbf{x} = \mathbf{z}, \\ & \mathbf{x} \in C_1,\ \mathbf{z} \in C_2. \end{aligned} \tag{A.9}$$

Let $\boldsymbol{\lambda} \in \mathbb{R}^S$ be a multiplier associated to the equality constraint $A\mathbf{x} = \mathbf{z}$ and introduce the *augmented* Lagrangian of problem (A.9)

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) = G_1(\mathbf{x}) + G_2(\mathbf{z}) + \boldsymbol{\lambda}^\top(A\mathbf{x} - \mathbf{z}) + \frac{\rho}{2}\|A\mathbf{x} - \mathbf{z}\|^2$$

where $\rho > 0$ is a penalty parameter. The ADMM algorithm is an iterative procedure in which at each iteration $t \geq 0$, the following steps are performed

$$\mathbf{x}^{t+1} = \underset{\mathbf{x} \in C_1}{\operatorname{argmin}} \ \mathcal{L}_\rho(\mathbf{x}, \mathbf{z}^t, \boldsymbol{\lambda}^t) \tag{A.10a}$$

$$\mathbf{z}^{t+1} = \underset{\mathbf{z} \in C_2}{\operatorname{argmin}} \ \mathcal{L}_\rho(\mathbf{x}^{t+1}, \mathbf{z}, \boldsymbol{\lambda}^t) \tag{A.10b}$$

$$\boldsymbol{\lambda}^{t+1} = \boldsymbol{\lambda}^t + \rho \, (A\mathbf{x}^{t+1} - \mathbf{z}^{t+1}), \tag{A.10c}$$

where the initialization of the variables $\mathbf{z}^0$ and $\boldsymbol{\lambda}^0$ can be arbitrary.

The ADMM algorithm is very similar to dual ascent and to the Method of Multipliers (MM): it consists of an $\mathbf{x}$-minimization, a $\mathbf{z}$-minimization, and a dual variable update. As in the method of multipliers, the dual variable update uses a step-size equal to the augmented Lagrangian parameter ρ. In the MM, the augmented Lagrangian $\mathcal{L}_\rho$ is minimized jointly with respect to the two primal variables. In ADMM, on the other hand, $\mathbf{x}$ and $\mathbf{z}$ are updated in an alternating or sequential fashion, which accounts for the term *alternating direction*.

Proposition A.3 ([12, Proposition 4.2])**.** Consider a sequence

$$\{\mathbf{x}^t, \mathbf{z}^t, \boldsymbol{\lambda}^t\}_{t \geq 0}$$

generated by the ADMM algorithm (A.10). Then, the generated sequence is bounded and every limit point of $\{\mathbf{x}^t\}_{t \geq 0}$ is an optimal solution of problem (A.8). Furthermore, the sequence $\{\boldsymbol{\lambda}^t\}_{t \geq 0}$ converges to an optimal solution of the dual of problem (A.8). □

In [18] a more general problem set-up for ADMM is considered. Specifically, let us consider a two-variable problem defined as

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^d, \mathbf{z} \in \mathbb{R}^S} \ & G_1(\mathbf{x}) + G_2(\mathbf{z}) \\ \text{subj. to } \ & A\mathbf{x} + B\mathbf{z} + c = 0 \\ & \mathbf{x} \in C_1, \ \mathbf{z} \in C_2. \end{aligned} \tag{A.11}$$

with $A \in \mathbb{R}^{p \times d}$, $B \in \mathbb{R}^{p \times S}$ and $c \in \mathbb{R}^{p \times 1}$. Then, the ADMM algorithm applied to problem (A.11) reads as follows

$$\mathbf{x}^{t+1} = \underset{\mathbf{x} \in C_1}{\operatorname{argmin}} \ \mathcal{L}_\rho(\mathbf{x}, \mathbf{z}^t, \boldsymbol{\lambda}^t) \tag{A.12a}$$

$$\mathbf{z}^{t+1} = \underset{\mathbf{z}\in C_2}{\operatorname{argmin}}\ \mathcal{L}_\rho(\mathbf{x}^{t+1}, \mathbf{z}, \boldsymbol{\lambda}^t) \tag{A.12b}$$

$$\boldsymbol{\lambda}^{t+1} = \boldsymbol{\lambda}^t + \rho\,(A\mathbf{x}^{t+1} + B\mathbf{z}^{t+1} + c), \tag{A.12c}$$

where the augmented Lagrangian is defined as

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) = G_1(\mathbf{x}) + G_2(\mathbf{z}) + \boldsymbol{\lambda}^\top(A\mathbf{x} + B\mathbf{z} + c) + \frac{\rho}{2}\|A\mathbf{x} + B\mathbf{z} + c\|^2.$$

B

Consensus Over Networks

Consensus and distributed averaging are fundamental building blocks in distributed optimization.

We introduce the consensus problem for a group of N agents that considers conditions under which, using a certain message-passing protocol, the local variables of each agent converge to the same value. There exist several results related to the convergence of local variables to a common value using various information exchange protocols among agents.

B.1 Average Consensus over Static Networks

One of the most used models for consensus is based on the following discrete-time iteration: to generate an estimate at iteration $t+1$, agent i forms a convex combination of its current estimate $\mathbf{z}_i^t$ with the estimates received from other agents as

$$\mathbf{z}_i^{t+1} = \sum_{j \in \mathcal{N}_i} a_{ij}\, \mathbf{z}_j^t, \tag{B.1}$$

where a_{ij} denotes a (positive) weight that agent i assigns to each neighbor j, and we recall that $\mathcal{N}_i$ is the set of neighbors of agent i in the (static) undirected communication graph. The weights a_{ij} are set to zero if i and j are not neighbors in the communication graph $\mathcal{G}$ and are doubly stochastic, i.e., they satisfy $\sum_{j=1}^N a_{ij} = 1$, for all $i \in \{1, \ldots, N\}$, and $\sum_{i=1}^N a_{ij} = 1$, for all $j \in \{1, \ldots, N\}$.

The consensus algorithm can be written in an aggregate form by stacking all the agents' estimates in a single variable which evolves according to

$$\mathbf{z}^{t+1} = \begin{bmatrix} \mathbf{z}_1^{t+1} \\ \vdots \\ \mathbf{z}_N^{t+1} \end{bmatrix} = A\mathbf{z}^t, \tag{B.2}$$

where A is a matrix whose (i,j)-th entry is a_{ij} for all $i,j \in \{1,\dots,N\}$.

A useful property of doubly stochastic matrices is the following. Given A doubly stochastic, it holds

$$\|A\mathbf{z} - \bar{\mathbf{z}}\| \le \sigma_A \|\mathbf{z} - \bar{\mathbf{z}}\|,$$

where $\bar{\mathbf{z}} \triangleq \frac{1}{N}\sum_{i=1}^N \mathbf{z}_i$ and σ_A is the spectral radius of $A - \mathbf{1}\mathbf{1}^\top/N$. It can be proven (see [148]) that if the graph is connected and A is doubly stochastic, then $\sigma_A \in (0,1)$, and specifically $\sigma_A = \max\{|\lambda_2|, |\lambda_N|\}$, where λ_h denotes the h-th largest eigenvalue of A.

Theorem B.1. Let $\mathcal{G}$ be a connected graph and let a_{ij}, $i,j \in \{1,\dots,N\}$ be doubly stochastic weights matching the graph. Then, the sequences $\{\mathbf{z}_i^t\}_{t\ge 0}$, $i \in \{1,\dots,N\}$, generated by (B.1) satisfy

$$\lim_{t\to\infty} \|\mathbf{z}_i^t - \bar{\mathbf{z}}^0\| = 0,$$

for all $i \in \{1,\dots,N\}$, where $\bar{\mathbf{z}}^0 = \frac{1}{N}\sum_{i=1}^N \mathbf{z}_i^0$. □

Several extensions of the basic consensus algorithm (B.1) exist. For instance, one can consider time-varying networks that have some long-term connectivity properties. The consensus algorithm needs to be adapted to accommodate the time-varying network by considering time-varying weights a_{ij}^t. Also, it is possible to design a consensus algorithm that works under delays and is robust to packet losses. See [46] for a recent survey on this topic. Next, we describe another extension in which the consensus algorithm is tailored for directed networks.

B.2 Push-sum Consensus over Directed Networks

In this section we describe how the average consensus algorithm can be adapted to work on directed networks. This algorithm is known as push-sum algorithm and has been introduced in [7].

In directed networks is not always possible to construct a doubly stochastic matrix A, while a column stochastic matrix is always available. We use B to denote a column stochastic matrix, i.e., such that $\mathbf{1}^\top B = \mathbf{1}^\top$. Formally, the push-sum consensus reads

$$\phi_i^{t+1} = \sum_{j \in \mathcal{N}_i} b_{ij}\, \phi_j^t \tag{B.3a}$$

$$\mathbf{s}_i^{t+1} = \sum_{j \in \mathcal{N}_i} b_{ij}\, \mathbf{s}_j^t \tag{B.3b}$$

$$\mathbf{z}_i^{t+1} = \frac{\mathbf{s}_i^{t+1}}{\phi_i^{t+1}}, \tag{B.3c}$$

with the initial values $\phi_i^0 = 1$ for all $i \in \{1, \dots, N\}$.

The convergence of this scheme has been proven in [7], i.e., the sequences $\{\mathbf{z}_i^t\}_{t \geq 0}$, $i \in \{1, \dots, N\}$, generated by (B.3) satisfy

$$\lim_{t \to \infty} \|\mathbf{z}_i^t - \bar{\mathbf{z}}^0\| = 0,$$

for all $i \in \{1, \dots, N\}$, where $\bar{\mathbf{z}}^0 = \frac{1}{N} \sum_{i=1}^N \mathbf{z}_i^0$.

B.3 Dynamic Average Consensus Algorithm

In this section, we present a distributed algorithm to achieve dynamic average consensus that has been proposed in [162]. See also [59] for a very recent tutorial.

We consider a network of N agents in which each agent i is able to measure a local discrete-time signal $\{\mathbf{r}_i^t\}_{t \geq 0}$. The goal is to design a distributed algorithm that enables agents to eventually track the average of their signal $\mathbf{r}_i^t$, $i \in \{1, \dots, N\}$, by means of local communication only.

The dynamic consensus algorithm proposed in [162] consists in a consensus-based procedure in which each agent maintains a local estimate $\mathbf{z}_i^t$ of the average. The local estimate is iteratively updated according to

$$\mathbf{z}_i^{t+1} = \sum_{j \in \mathcal{N}_i} a_{ij}\, \mathbf{z}_j^t + \left(\mathbf{r}_i^{t+1} - \mathbf{r}_i^t\right), \tag{B.4}$$

where a_{ij} are entries of a doubly stochastic matrix.

If the input signals $\mathbf{r}_i^t$ asymptotically converge to a constant value, then the dynamic average consensus algorithm in (B.4) is guaranteed to converge, i.e., for all $i \in \{1, \dots, N\}$, it holds

$$\lim_{t \to \infty} \|\mathbf{z}_i^t - \bar{\mathbf{r}}^t\| = 0,$$

where $\bar{\mathbf{r}}^t = \frac{1}{N} \sum_{i=1}^{N} \mathbf{r}_i^t$ for all $t \geq 0$.

The interested reader can find a rigorous treatment and a more comprehensive discussion on this class of algorithms in [162, 59].

C

Linear Programming

A Linear Program (LP) is an optimization problem with linear cost function and linear constraints:

$$\begin{aligned} \min_{\mathbf{x}} \; & c^\top \mathbf{x} \\ \text{subj. to } & a_k^\top \mathbf{x} \le b_k, \quad k \in \{1, \ldots, K\}, \end{aligned} \tag{C.1}$$

where $c \in \mathbb{R}^d$ is the cost vector and $a_k \in \mathbb{R}^d$ and $b_k \in \mathbb{R}$ describe K inequality constraints. In the subsequent discussion, we assume that $d \le K$. The feasible set $\mathcal{X}$ of problem (C.1) is the set of vectors satisfying all the constraints, i.e.,

$$\mathcal{X} \triangleq \{\mathbf{x} \in \mathbb{R}^d \mid a_k^\top \mathbf{x} \le b_k \text{ for all } k \in \{1, \ldots, K\}\}.$$

Note that $\mathcal{X}$ is a polyhedron, for which the following definition of vertex can be given.

Definition C.1. A vector $\tilde{\mathbf{x}} \in \mathbb{R}^d$ is a vertex of $\mathcal{X}$ if there exists some $c \in \mathbb{R}^d$ such that $c^\top \tilde{\mathbf{x}} < c^\top \mathbf{x}$ for all $\mathbf{x} \in \mathcal{X}$ with $\mathbf{x} \neq \tilde{\mathbf{x}}$. □

If problem (C.1) admits an optimal solution, it can be shown that there exists an optimal vertex, i.e., a vertex which is an optimal solution of the problem (see, e.g., [13, Theorem 2.7]). Let $\mathbf{x}^\star$ be an optimal vertex of problem (C.1). Then, it is a standard result in linear programming theory that there exists an index set $\{\ell_1, \ldots, \ell_d\} \subset \{1, \ldots, K\}$, with cardinality d, such that $\mathbf{x}^\star$ is the unique optimal vertex of the problem

$$\begin{aligned} \min_{\mathbf{x}} \; & c^\top \mathbf{x} \\ \text{subj. to } & a_{\ell_h}^\top \mathbf{x} \le b_{\ell_h}, \quad h \in \{1, \ldots, d\}, \end{aligned}$$

which is a relaxed version of problem (C.1) in which only d constraints are considered. In addition, the vectors $a_{\ell_h}, h \in \{1, \ldots, d\}$ are linearly independent, so that they form a basis of $\mathbb{R}^d$. By analogy, the constraints

$a_{\ell_h}^\top \mathbf{x} \leq b_{\ell_h}, h \in \{1, \ldots, d\}$ are called a *basis* of the point $\mathbf{x}^\star$. Due to the optimality of $\mathbf{x}^\star$, we call it also a basis of problem (C.1). To compactly denote such basis, we introduce a matrix $P \in \mathbb{R}^{d \times d}$, obtained by stacking the row vectors $a_{\ell_h}^\top$, and a vector $q \in \mathbb{R}^d$, obtained by stacking the scalars b_{ℓ_h}, i.e.,

$$P = \begin{bmatrix} a_{\ell_1}^\top \\ \vdots \\ a_{\ell_d}^\top \end{bmatrix}, \quad q = \begin{bmatrix} b_{\ell_1} \\ \vdots \\ b_{\ell_d} \end{bmatrix}.$$

Then, $\mathbf{x}^\star = P^{-1} q$, and we say that the tuple (P, q) is a basis of (C.1).

If problem (C.1) has multiple optimal solutions, we say that the LP is *dual degenerate*. In presence of dual degeneracy, it is not trivial to guarantee convergence of distributed algorithms to the same optimal solution. In order to overcome this issue, it is possible to rely on the lexicographic ordering of vectors. We now give some definitions.

Definition C.2. A vector $\mathbf{v} \in \mathbb{R}^n$ is said to be *lexicographically positive* (or *lex-positive*) if $\mathbf{v} \neq \mathbf{0}$ and the first non-zero component of $\mathbf{v}$ is positive. In symbols:

$$\mathbf{u} \overset{L}{>} \mathbf{0}.$$

A vector $\mathbf{u} \in \mathbb{R}^n$ is said to be *lexicographically larger* (resp. *smaller*) than another vector $\mathbf{v} \in \mathbb{R}^n$ if $\mathbf{u} - \mathbf{v}$ is lex-positive (resp. $\mathbf{v} - \mathbf{u}$ is lex-positive), or, equivalently, if $\mathbf{u} \neq \mathbf{v}$ and the first nonzero component of $\mathbf{u} - \mathbf{v}$ is positive (resp., negative). In symbols:

$$\mathbf{u} \overset{L}{>} \mathbf{v} \quad \text{or} \quad \mathbf{u} \overset{L}{<} \mathbf{v}.$$

Given a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$, the lexicographic minimum is the element $\mathbf{v}_i$ such that $\mathbf{v}_j \overset{L}{>} \mathbf{v}_i$ for all $j \neq i$. In symbols:

$$\mathbf{v}_i = \text{lexmin}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}. \qquad \square$$

Now, consider the optimal solution set of problem (C.1), i.e., $\mathcal{X}^\star \triangleq \{\mathbf{x} \in \mathcal{X} \mid c^\top \mathbf{x} \leq c^\top \mathbf{x}' \text{ for all } \mathbf{x}' \in \mathcal{X}\} \subseteq \mathcal{X}$, where $\mathcal{X}$ is the feasible set of problem (C.1). Among all the optimal solutions in $\mathcal{X}^\star$, it is possible to compute the lexicographically minimal one, i.e., lexmin($\mathcal{S}^\star$). It turns out that finding lexmin($\mathcal{S}^\star$) is equivalent to finding the (unique) optimal solution to a modified (non dual-degenerate) version of the original problem (C.1), where the cost vector c is perturbed to $c' = c + \Delta$, with Δ a lexicographic perturbation vector:

$$\Delta^\top = [\Delta_0 \ \Delta_0^2 \ \dots \ \Delta_0^d],$$

for a sufficiently small $\Delta_0 > 0$ (see [56]). Therefore, the lex-optimal solution of problem (C.1) is the *unique* optimal solution of the problem with perturbed cost

$$\begin{aligned} \min_{\mathbf{x}} \ & (c + \Delta)^\top \mathbf{x} \\ \text{subj. to } & a_k^\top \mathbf{x} \le b_k, \quad k \in \{1, \dots, K\}. \end{aligned} \tag{C.2}$$

Thus, the lex-optimal solution of problem (C.1) exists if and only if problem (C.2) admits an optimal solution. Moreover, the optimal solution of (C.2) is attained at a vertex of (C.1), therefore it is an optimal vertex of problem (C.1).

Acknowledgements

This work is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 638992 - OPT4SMART)

References

[1] Agarwal, P. K. and S. Sen. 2001. "Randomized algorithms for geometric optimization problems". *Combinatorial Optimization Dordrecht*. 9(1): 151–187.

[2] Alizadeh, M., X. Li, Z. Wang, A. Scaglione, and R. Melton. 2012. "Demand-side management in the smart grid: Information processing for the power switch". *IEEE Signal Processing Magazine*. 29(5): 55–67.

[3] Amenta, N. 1994. "Helly-type theorems and generalized linear programming". *Discrete & Computational Geometry*. 12(3): 241–261.

[4] Bastianello, N., R. Carli, L. Schenato, and M. Todescato. 2018. "A Partition-Based Implementation of the Relaxed ADMM for Distributed Convex Optimization over Lossy Networks". In: *IEEE Conference on Decision and Control (CDC)*. 3379–3384.

[5] Beck, A. 2017. *First-order methods in optimization*. Vol. 25. SIAM.

[6] Beck, A., A. Nedic, A. Ozdaglar, and M. Teboulle. 2014. "An $O(1/k)$ Gradient Method for Network Resource Allocation Problems". *IEEE Transactions on Control of Network Systems*. 1(1): 64–73.

[7] Bénézit, F., V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli. 2010. "Weighted gossip: Distributed averaging using non-doubly stochastic matrices". In: *IEEE International Symposium on Information Theory (ISIT)*. 1753–1757.

[8] Bertsekas, D. P. 1998. *Network optimization: continuous and discrete models*. Citeseer.

[9] Bertsekas, D. P. 1999. *Nonlinear programming*. Athena scientific.

[10] Bertsekas, D. P. 2015. *Convex Optimization Algorithms*. Athena Scientific.

[11] Bertsekas, D. P. and J. N. Tsitsiklis. 2000. "Gradient convergence in gradient methods with errors". *SIAM Journal on Optimization*. 10(3): 627–642.

[12] Bertsekas, D. P. and J. N. Tsitsiklis. 1989. *Parallel and distributed computation: numerical methods*. Vol. 23. Prentice Hall Englewood Cliffs, NJ.

[13] Bertsimas, D. and J. N. Tsitsiklis. 1997. *Introduction to linear optimization*. Vol. 6. Athena Scientific Belmont, MA.

[14] Bianchi, P., W. Hachem, and F. Iutzeler. 2016. "A coordinate descent primal-dual algorithm and application to distributed asynchronous optimization". *IEEE Transactions on Automatic Control*. 61(10): 2947–2957.

[15] Bianchi, P. and J. Jakubowicz. 2013. "Convergence of a multi-agent projected stochastic gradient algorithm for non-convex optimization". *IEEE Transactions on Automatic Control*. 58(2): 391–405.

[16] Bof, N., R. Carli, G. Notarstefano, L. Schenato, and D. Varagnolo. 2018. "Multi-Agent Newton-Raphson Optimization Over Lossy Networks". *IEEE Transactions on Automatic Control*.

[17] Boser, B. E., I. M. Guyon, and V. N. Vapnik. 1992. "A training algorithm for optimal margin classifiers". In: *Proceedings of the fifth annual workshop on Computational learning theory*. ACM. 144–152.

[18] Boyd, S., N. Parikh, E. Chu, B. Peleato, J. Eckstein, *et al.* 2011. "Distributed optimization and statistical learning via the alternating direction method of multipliers". *Foundations and Trends® in Machine learning*. 3(1): 1–122.

[19] Bürger, M., G. Notarstefano, and F. Allgöwer. 2014. "A Polyhedral Approximation Framework for Convex and Robust Distributed Optimization". *IEEE Transactions on Automatic Control*. 59(2): 384–395.

[20] Bürger, M., G. Notarstefano, F. Bullo, and F. Allgöwer. 2012. "A distributed simplex algorithm for degenerate linear programs and multi-agent assignments". *Automatica*. 48(9): 2298–2304.

[21] Camisa, A., I. Notarnicola, and G. Notarstefano. 2018. "A Primal Decomposition Method with Suboptimality Bounds for Distributed Mixed-Integer Linear Programming". In: *IEEE Conference on Decision and Control (CDC)*. 3391–3396.

[22] Carli, R. and G. Notarstefano. 2013. "Distributed partition-based optimization via dual decomposition". In: *IEEE Conference on Decision and Control (CDC)*. 2979–2984.

[23] Carli, R., G. Notarstefano, L. Schenato, and D. Varagnolo. 2015. "Analysis of Newton-Raphson Consensus for multi-agent convex optimization under asynchronous and lossy communications". In: *IEEE Conference on Decision and Control (CDC)*. 418–424.

[24] Cattivelli, F. S. and A. H. Sayed. 2010. "Diffusion LMS strategies for distributed estimation". *IEEE Transactions on Signal Processing*. 58(3): 1035–1048.

[25] Chamanbaz, M., G. Notarstefano, and R. Bouffanais. 2017. "Randomized Constraints Consensus for Distributed Robust Linear Programming". *IFAC-PapersOnLine*. 50(1): 4973–4978.

[26] Chang, T.-H. 2016. "A proximal dual consensus ADMM method for multi-agent constrained optimization". *IEEE Transactions on Signal Processing*. 64(14): 3719–3734.

[27] Chang, T.-H., M. Hong, and X. Wang. 2015. "Multi-agent distributed optimization via inexact consensus ADMM". *IEEE Transactions on Signal Processing*. 63(2): 482–497.

[28] Chang, T.-H., A. Nedić, and A. Scaglione. 2014. "Distributed constrained optimization by consensus-based primal-dual perturbation method". *IEEE Transactions on Automatic Control*. 59(6): 1524–1538.

[29] Chen, J. and A. H. Sayed. 2012. “Diffusion adaptation strategies for distributed optimization and learning over networks”. *IEEE Transactions on Signal Processing*. 60(8): 4289–4305.

[30] Cherukuri, A. and J. Cortés. 2015. “Distributed generator coordination for initialization and anytime optimization in economic dispatch”. *IEEE Transactions on Control of Network Systems*. 2(3): 226–237.

[31] Cherukuri, A. and J. Cortés. 2016. “Initialization-free distributed coordination for economic dispatch under varying loads and generator commitment”. *Automatica*. 74: 183–193.

[32] Di Lorenzo, P. and G. Scutari. 2015. “Distributed Nonconvex Optimization over Networks”. In: *IEEE Intern. Conf. on Comput. Advances in Multi-Sensor Adaptive Process. (CAMSAP)*. 229–232.

[33] Di Lorenzo, P. and G. Scutari. 2016. “Next: In-network nonconvex optimization”. *IEEE Transactions on Signal and Information Processing over Networks*. 2(2): 120–136.

[34] Dinh, Q. T., I. Necoara, and M. Diehl. 2013. “A dual decomposition algorithm for separable nonconvex optimization using the penalty function framework”. In: *IEEE Conference on Decision and Control (CDC)*. 2372–2377.

[35] Doan, M. D., M. Diehl, T. Keviczky, and B. De Schutter. 2017. “A Jacobi decomposition algorithm for distributed convex optimization in distributed model predictive control”. *IFAC-PapersOnLine*. 50(1): 4905–4911.

[36] Duchi, J. C., A. Agarwal, and M. J. Wainwright. 2012. “Dual averaging for distributed optimization: Convergence analysis and network scaling”. *IEEE Transactions on Automatic control*. 57(3): 592–606.

[37] Ebenbauer, C., S. Michalowsky, V. Grushkovskaya, and B. Gharesifard. 2017. “Distributed optimization over directed graphs with the help of Lie brackets”. *IFAC-PapersOnLine*. 50(1): 15343–15348.

[38] Eisen, M., A. Mokhtari, and A. Ribeiro. 2017. “Decentralized quasi-Newton methods”. *IEEE Transactions on Signal Processing*. 65(10): 2613–2628.

[39] Erseghe, T. 2012. "A Distributed and Scalable Processing Method Based Upon ADMM". *IEEE Signal Processing Letters*. 19(9): 563–566.

[40] Falsone, A., K. Margellos, S. Garatti, and M. Prandini. 2017. "Dual decomposition for multi-agent distributed optimization with coupling constraints". *Automatica*. 84: 149–158.

[41] Farina, F., A. Garulli, and A. Giannitrapani. 2018. "Distributed Interpolatory Algorithms for Set Membership Estimation". *IEEE Transactions on Automatic Control*.

[42] Farina, F., A. Garulli, A. Giannitrapani, and G. Notarstefano. 2019. "A Distributed Asynchronous Method of Multipliers for Constrained Nonconvex Optimization". *Automatica*. 103: 243–253.

[43] Gharesifard, B. and J. Cortés. 2014. "Distributed continuous-time convex optimization on weight-balanced digraphs". *IEEE Transactions on Automatic Control*. 59(3): 781–786.

[44] Giselsson, P., M. D. Doan, T. Keviczky, B. De Schutter, and A. Rantzer. 2013. "Accelerated gradient methods and dual decomposition in distributed model predictive control". *Automatica*. 49(3): 829–833.

[45] Giselsson, P. and A. Rantzer. 2018. *Large-scale and Distributed Optimization*. Vol. 2227. Springer.

[46] Hadjicostis, C. N., A. D. Domínguez-García, and T. Charalambous. 2018. "Distributed Averaging and Balancing in Network Systems: with Applications to Coordination and Control". *Foundations and Trends® in Systems and Control*. 5(2-3): 99–292.

[47] Hale, M. T., A. Nedić, and M. Egerstedt. 2017. "Asynchronous multiagent primal-dual optimization". *IEEE Transactions on Automatic Control*. 62(9): 4421–4435.

[48] Hatanaka, T., N. Chopra, T. Ishizaki, and N. Li. 2018. "Passivity-based distributed optimization with communication delays using PI consensus algorithm". *IEEE Transactions on Automatic Control*.

[49] Hochhaus, S. and M. T. Hale. 2018. "Asynchronous Distributed Optimization with Heterogeneous Regularizations and Normalizations". In: *IEEE Conference on Decision and Control (CDC)*. 4232–4237.

[50] Iutzeler, F., P. Bianchi, P. Ciblat, and W. Hachem. 2016. "Explicit convergence rate of a distributed alternating direction method of multipliers". *IEEE Transactions on Automatic Control*. 61(4): 892–904.

[51] Jakovetić, D., D. Bajovic, N. Krejic, and N. K. Jerinkic. 2016. "Distributed Gradient Methods with Variable Number of Working Nodes." *IEEE Transactions on Signal Processing*. 64(15): 4080–4095.

[52] Jakovetić, D., J. M. Moura, and J. Xavier. 2015. "Linear Convergence Rate of a Class of Distributed Augmented Lagrangian Algorithms". *IEEE Transactions on Automatic Control*. 60(4): 922–936.

[53] Jakovetić, D., J. Xavier, and J. M. Moura. 2011. "Cooperative convex optimization in networked systems: Augmented Lagrangian algorithms with directed gossip communication". *IEEE Transactions on Signal Processing*. 59(8): 3889–3902.

[54] Jakovetić, D., J. Xavier, and J. M. Moura. 2014. "Fast distributed gradient methods". *IEEE Transactions on Automatic Control*. 59(5): 1131–1146.

[55] Johansson, B., M. Rabi, and M. Johansson. 2009. "A randomized incremental subgradient method for distributed optimization in networked systems". *SIAM Journal on Optimization*. 20(3): 1157–1170.

[56] Jones, C. N., E. C. Kerrigan, and J. M. Maciejowski. 2007. "Lexicographic perturbation for multiparametric linear programming with applications to control". *Automatica*. 43(10): 1808–1816.

[57] Kia, S. S. 2017. "Distributed optimal in-network resource allocation algorithm design via a control theoretic approach". *Systems & Control Letters*. 107: 49–57.

[58] Kia, S. S., J. Cortés, and S. Martínez. 2015. "Distributed Convex Optimization via Continuous-time Coordination Algorithms with Discrete-time Communication". *Automatica*. 55: 254–264.

[59] Kia, S. S., B. Van Scoy, J. Cortés, R. A. Freeman, K. M. Lynch, and S. Martínez. 2019. "Tutorial on Dynamic Average Consensus: The Problem, Its Applications, and the Algorithms". *IEEE Control Systems Magazine*. 39(3): 40–72.

[60] Kumar, S., R. Jain, and K. Rajawat. 2017. "Asynchronous optimization over heterogeneous networks via consensus ADMM". *IEEE Transactions on Signal and Information Processing over Networks*. 3(1): 114–129.

[61] Latafat, P., L. Stella, and P. Patrinos. 2016. "New primal-dual proximal algorithm for distributed optimization". In: *IEEE Conference on Decision and Control (CDC)*. IEEE. 1959–1964.

[62] Lee, S. and A. Nedić. 2013. "Distributed random projection algorithm for convex optimization". *IEEE Journal of Selected Topics in Signal Processing*. 7(2): 221–229.

[63] Lee, S. and A. Nedić. 2016. "Asynchronous gossip-based random projection algorithms over networks". *IEEE Transactions on Automatic Control*. 61(4): 953–968.

[64] Lee, S., A. Nedić, and M. Raginsky. 2017. "Stochastic dual averaging for decentralized online optimization on time-varying communication graphs". *IEEE Transactions on Automatic Control*. 62(12): 6407–6414.

[65] Lee, S., A. Ribeiro, and M. M. Zavlanos. 2016. "Distributed continuous-time online optimization using saddle-point methods". In: *IEEE Conf. on Decision and Control (CDC)*. 4314–4319.

[66] Li, N. and J. R. Marden. 2013. "Designing games for distributed optimization". *IEEE Journal of Selected Topics in Signal Processing*. 7(2): 230–242.

[67] Lin, P., W. Ren, and Y. Song. 2016. "Distributed multi-agent optimization subject to nonidentical constraints and communication delays". *Automatica*. 65: 120–131.

[68] Ling, Q., Y. Liu, W. Shi, and Z. Tian. 2016. "Weighted ADMM for fast decentralized network optimization". *IEEE Transactions on Signal Processing*. 64(22): 5930–5942.

[69] Ling, Q. and A. Ribeiro. 2014. "Decentralized Dynamic Optimization Through the Alternating Direction Method of Multipliers". *IEEE Transactions on Signal Processing*. 5(62): 1185–1197.

[70] Ling, Q., W. Shi, G. Wu, and A. Ribeiro. 2015. "DLM: Decentralized linearized alternating direction method of multipliers". *IEEE Transactions on Signal Processing*. 63(15): 4051–4064.

[71] Liu, S., Z. Qiu, and L. Xie. 2017. "Convergence rate analysis of distributed optimization with projected subgradient algorithm". *Automatica*. 83: 162–169.

[72] Lobel, I. and A. Ozdaglar. 2011. "Distributed subgradient methods for convex optimization over random networks". *IEEE Transactions on Automatic Control*. 56(6): 1291–1306.

[73] Lu, J. and C. Y. Tang. 2012. "Zero-gradient-sum algorithms for distributed convex optimization: The continuous-time case". *IEEE Transactions on Automatic Control*. 57(9): 2348–2354.

[74] Makhdoumi, A. and A. Ozdaglar. 2017. "Convergence rate of distributed ADMM over networks". *IEEE Transactions on Automatic Control*. 62(10): 5082–5095.

[75] Margellos, K., A. Falsone, S. Garatti, and M. Prandini. 2018. "Distributed constrained optimization and consensus in uncertain networks via proximal minimization". *IEEE Transactions on Automatic Control*. 63(5): 1372–1387.

[76] Mateos-Núnez, D. and J. Cortés. 2017. "Distributed saddle-point subgradient algorithms with Laplacian averaging". *IEEE Transactions on Automatic Control*. 62(6): 2720–2735.

[77] Mateos, G., J. A. Bazerque, and G. B. Giannakis. 2010. "Distributed sparse linear regression". *IEEE Transactions on Signal Processing*. 58(10): 5262–5276.

[78] Michalowsky, S., B. Gharesifard, and C. Ebenbauer. 2018. "On the Lie bracket approximation approach to distributed optimization: Extensions and limitations". In: *IEEE European Control Conference (ECC)*. 119–124.

[79] Mokhtari, A., Q. Ling, and A. Ribeiro. 2017. "Network Newton distributed optimization methods". *IEEE Transactions on Signal Processing*. 65(1): 146–161.

[80] Mokhtari, A., W. Shi, Q. Ling, and A. Ribeiro. 2016. "DQM: Decentralized Quadratically Approximated Alternating Direction Method of Multipliers". *IEEE Transactions on Signal Processing*. 64(19): 5158–5173.

[81] Mota, J. F., J. M. Xavier, P. M. Aguiar, and M. Püschel. 2013. "D-ADMM: A communication-efficient distributed algorithm for separable optimization". *IEEE Transactions on Signal Processing*. 61(10): 2718–2723.

[82] Necoara, I. 2013. "Random coordinate descent algorithms for multi-agent convex optimization over networks". *IEEE Transactions on Automatic Control*. 58(8): 2001–2012.

[83] Necoara, I. and V. Nedelcu. 2015. "On linear convergence of a distributed dual gradient algorithm for linearly constrained separable convex problems". *Automatica*. 55: 209–216.

[84] Necoara, I., V. Nedelcu, D. Clipici, and L. Toma. 2017. "On fully distributed dual first order methods for convex network optimization". *IFAC-PapersOnLine*. 50(1): 2788–2793.

[85] Necoara, I., V. Nedelcu, and I. Dumitrache. 2011. "Parallel and distributed optimization methods for estimation and control in networks". *Journal of Process Control*. 21(5): 756–766.

[86] Nedić, A. 2011. "Asynchronous broadcast-based convex optimization over a network". *IEEE Transactions on Automatic Control*. 56(6): 1337–1351.

[87] Nedić, A. 2015. "Convergence rate of distributed averaging dynamics and optimization in networks". *Foundations and Trends® in Systems and Control*. 2(1): 1–100.

[88] Nedić, A. and J. Liu. 2018. "Distributed Optimization for Control". *Annual Review of Control, Robotics, and Autonomous Systems*. 1: 77–103.

[89] Nedić, A. and A. Olshevsky. 2015. "Distributed optimization over time-varying directed graphs". *IEEE Transactions on Automatic Control*. 60(3): 601–615.

[90] Nedić, A. and A. Olshevsky. 2016. "Stochastic gradient-push for strongly convex functions on time-varying directed graphs". *IEEE Transactions on Automatic Control*. 61(12): 3936–3947.

[91] Nedić, A., A. Olshevsky, and M. G. Rabbat. 2018a. "Network topology and communication-computation tradeoffs in decentralized optimization". *Proceedings of the IEEE*. 106(5): 953–976.

[92] Nedić, A., A. Olshevsky, and W. Shi. 2016. "A geometrically convergent method for distributed optimization over time-varying graphs". In: *IEEE Conf. on Decision and Control (CDC)*. 1023–1029.

[93] Nedić, A., A. Olshevsky, and W. Shi. 2017a. "Achieving geometric convergence for distributed optimization over time-varying graphs". *SIAM Journal on Optimization*. 27(4): 2597–2633.

[94] Nedić, A., A. Olshevsky, W. Shi, and C. A. Uribe. 2017b. "Geometrically convergent distributed optimization with uncoordinated step-sizes". In: *American Control Conference (ACC)*. IEEE. 3950–3955.

[95] Nedić, A. and A. Ozdaglar. 2009. "Distributed subgradient methods for multi-agent optimization". *IEEE Transactions on Automatic Control*. 54(1): 48–61.

[96] Nedić, A., A. Ozdaglar, and P. A. Parrilo. 2010. "Constrained Consensus and Optimization in Multi-Agent Networks". *IEEE Transactions on Automatic Control*. 55(4): 922–938.

[97] Nedić, A., J.-S. Pang, G. Scutari, and Y. Sun. 2018b. *Multi-agent Optimization*. Vol. 2224. Springer.

[98] Notarnicola, I., R. Carli, and G. Notarstefano. 2018a. "Distributed Partitioned Big-Data Optimization via Asynchronous Dual Decomposition". *IEEE Transactions on Control of Network Systems*. 5(4): 1910–1919.

[99] Notarnicola, I., M. Franceschelli, and G. Notarstefano. 2016. "A duality-based approach for distributed min-max optimization with application to demand side management". In: *IEEE Conference on Decision and Control (CDC)*. 1877–1882.

[100] Notarnicola, I., M. Franceschelli, and G. Notarstefano. 2019. "A duality-based approach for distributed min-max optimization". *IEEE Transactions on Automatic Control*. 64(6): 2559–2566.

[101] Notarnicola, I. and G. Notarstefano. 2017a. "A duality-based approach for distributed optimization with coupling constraints". In: *IFAC World Congress*. 14891–14896.

[102] Notarnicola, I. and G. Notarstefano. 2017b. "Asynchronous distributed optimization via randomized dual proximal gradient". *IEEE Transactions on Automatic Control*. 62(5): 2095–2106.

[103] Notarnicola, I. and G. Notarstefano. 2019. "Constraint-Coupled Distributed Optimization: a Relaxation and Duality Approach". *IEEE Transactions on Control of Network Systems*.

[104] Notarnicola, I., Y. Sun, G. Scutari, and G. Notarstefano. 2017a. "Distributed big-data optimization via block communications". In: *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. 1–5.

[105] Notarnicola, I., Y. Sun, G. Scutari, and G. Notarstefano. 2017b. "Distributed big-data optimization via block-iterative convexification and averaging". In: *IEEE Conference on Decision and Control (CDC)*. 2281–2288.

[106] Notarnicola, I., Y. Sun, G. Scutari, and G. Notarstefano. 2018b. "Distributed big-data optimization via block-iterative gradient tracking". *arXiv preprint arXiv:1808.07252*.

[107] Notarstefano, G. 2015. "A core-set approach for distributed quadratic programming in big-data classification". In: *IEEE Conference on Decision and Control (CDC)*. 1372–1377.

[108] Notarstefano, G. and F. Bullo. 2011. "Distributed Abstract Optimization via Constraints Consensus: Theory and Applications". *IEEE Transactions on Automatic Control*. 56(10): 2247–2261.

[109] Palomar, D. P. and M. Chiang. 2006. "A tutorial on decomposition methods for network utility maximization". *IEEE Journal on Selected Areas in Communications*. 24(8): 1439–1451.

[110] Paul, H., J. Fliege, and A. Dekorsy. 2013. "In-network-processing: Distributed consensus-based linear estimation". *IEEE Communications Letters*. 17(1): 59–62.

[111] Qu, G. and N. Li. 2016. "Harnessing smoothness to accelerate distributed optimization". In: *IEEE Conf. on Decision and Control (CDC)*. 159–166.

[112] Qu, G. and N. Li. 2017a. "Accelerated distributed Nesterov Gradient Descent for convex and smooth functions". In: *IEEE Conf. on Decision and Control (CDC)*. 2260–2267.

[113] Qu, G. and N. Li. 2017b. "Harnessing smoothness to accelerate distributed optimization". *IEEE Transactions on Control of Network Systems*. 5(3): 1245–1260.

[114] Ram, S. S., A. Nedić, and V. V. Veeravalli. 2010. "Distributed stochastic subgradient projection algorithms for convex optimization". *Journal of optimization theory and applications*. 147(3): 516–545.

[115] Rawat, A. and N. Elia. 2018. "Distributed Method of Multiplier for Coupled Lagrangian Problems: A Control Approach". In: *American Control Conference (ACC)*. 6475–6480.

[116] Rawlings, J. B. and D. Q. Mayne. 2009. "Model predictive control: Theory and design".

[117] Richert, D. and J. Cortés. 2015. "Robust distributed linear programming". *IEEE Transactions on Automatic Control*. 60(10): 2567–2582.

[118] Schizas, I. D., A. Ribeiro, and G. B. Giannakis. 2008. "Consensus in ad hoc WSNs with noisy links – Part I: Distributed estimation of deterministic signals". *IEEE Transactions on Signal Processing*. 56(1): 350–364.

[119] Scutari, G. and Y. Sun. 2019. "Distributed nonconvex constrained optimization over time-varying digraphs". *Mathematical Programming*. 176(1-2): 497–544.

[120] Shi, W., Q. Ling, G. Wu, and W. Yin. 2015a. "A proximal gradient algorithm for decentralized composite optimization". *IEEE Transactions on Signal Processing*. 63(22): 6013–6023.

[121] Shi, W., Q. Ling, G. Wu, and W. Yin. 2015b. "EXTRA: An Exact First-order Algorithm for Decentralized Consensus Optimization". *SIAM Journal on Optimization*. 25(2): 944–966.

[122] Shi, W., Q. Ling, K. Yuan, G. Wu, and W. Yin. 2014. "On the Linear Convergence of the ADMM in Decentralized Consensus Optimization". *IEEE Transactions on Signal Processing*. 62(7): 1750–1761.

[123] Simonetto, A. and H. Jamali-Rad. 2016. "Primal recovery from consensus-based dual decomposition for distributed convex optimization". *Journal of Optimization Theory and Applications*. 168(1): 172–197.

[124] Simonetto, A., A. Koppel, A. Mokhtari, G. Leus, and A. Ribeiro. 2017. "Decentralized prediction-correction methods for networked time-varying convex optimization". *IEEE Transactions on Automatic Control*. 62(11): 5724–5738.

[125] Srivastava, K. and A. Nedic. 2011. "Distributed asynchronous constrained stochastic optimization". *IEEE Journal of Selected Topics in Signal Processing*. 5(4): 772–790.

[126] Stankovic, M. S., K. H. Johansson, and D. M. Stipanovic. 2011. "Distributed seeking of Nash equilibria with applications to mobile sensor networks". *IEEE Transactions on Automatic Control*. 57(4): 904–919.

[127] Sun, Y. and G. Scutari. 2017. "Distributed nonconvex optimization for sparse representation". In: *IEEE Intern. Conf. on Speech and Signal Process. (ICASSP)*. 4044–4048.

[128] Sun, Y., G. Scutari, and D. P. Palomar. 2016. "Distributed Nonconvex Multiagent Optimization over Time-varying Networks". In: *IEEE Asilomar Conf. on Signals, Systems, and Computers*. 788–794.

[129] Sundararajan, A., B. Hu, and L. Lessard. 2017. "Robust convergence analysis of distributed optimization algorithms". In: *Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE. 1206–1212.

[130] Tatarenko, T. and B. Touri. 2017. "Non-convex distributed optimization". *IEEE Transactions on Automatic Control*. 62(8): 3744–3757.

[131] Teixeira, A., E. Ghadimi, I. Shames, H. Sandberg, and M. Johansson. 2016. "The ADMM algorithm for distributed quadratic problems: Parameter selection and constraint preconditioning". *IEEE Transactions on Signal Processing*. 64(2): 290–305.

[132] Testa, A., I. Notarnicola, and G. Notarstefano. 2018. "Distributed Submodular Minimization over Networks: a Greedy Column Generation Approach". In: *IEEE Conference on Decision and Control (CDC)*. 4945–4950.

[133] Testa, A., A. Rucco, and G. Notarstefano. 2017. "A finite-time cutting plane algorithm for distributed mixed integer linear programming". In: *IEEE Conference on Decision and Control (CDC)*. 3847–3852.

[134] Testa, A., A. Rucco, and G. Notarstefano. 2019. "Distributed Mixed-Integer Linear Programming via Cut Generation and Constraint Exchange". *IEEE Transactions on Automatic Control*.

[135] Todescato, M., G. Cavraro, R. Carli, and L. Schenato. 2015. "A Robust Block-Jacobi Algorithm for Quadratic Programming under Lossy Communications". *IFAC-PapersOnLine*. 48(22): 126–131.

[136] Tsianos, K. I., S. Lawlor, and M. G. Rabbat. 2012. "Push-sum distributed dual averaging for convex optimization". In: *IEEE Conference on Decision and Control (CDC)*. 5453–5458.

[137] Varagnolo, D., F. Zanella, A. Cenedese, G. Pillonetto, and L. Schenato. 2016. "Newton-Raphson consensus for distributed convex optimization". *IEEE Transactions on Automatic Control*. 61(4): 994–1009.

[138] Wai, H.-T., J. Lafond, A. Scaglione, and E. Moulines. 2017. "Decentralized Frank–Wolfe Algorithm for Convex and Nonconvex Problems". *IEEE Transactions on Automatic Control*. 62(11): 5522–5537.

[139] Wang, J. and N. Elia. 2011. "A control perspective for centralized and distributed convex optimization". In: *IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*. 3800–3805.

[140] Wei, E. and A. Ozdaglar. 2013. "On the $O(1/k)$ convergence of asynchronous distributed alternating direction method of multipliers". In: *IEEE Global conference on signal and information processing (GlobalSIP)*. 551–554.

[141] Wei, E., A. Ozdaglar, and A. Jadbabaie. 2013a. "A distributed Newton method for network utility maximization–Part I: Algorithm". *IEEE Transactions on Automatic Control*. 58(9): 2162–2175.

[142] Wei, E., A. Ozdaglar, and A. Jadbabaie. 2013b. "A distributed Newton method for network utility maximization–Part II: Convergence". *IEEE Transactions on Automatic Control*. 58(9): 2176–2188.

[143] Wu, T., K. Yuan, Q. Ling, W. Yin, and A. H. Sayed. 2017. "Decentralized consensus optimization with asynchrony and delays". *IEEE Transactions on Signal and Information Processing over Networks*. 4(2): 293–307.

[144] Wu, X. and J. Lu. 2019. "Fenchel dual gradient methods for distributed convex optimization over time-varying networks". *IEEE Transactions on Automatic Control*.

[145] Xi, C. and U. A. Khan. 2017. "DEXTRA: A fast algorithm for optimization over directed graphs". *IEEE Transactions on Automatic Control*. 62(10): 4980–4993.

[146] Xi, C., V. S. Mai, R. Xin, E. H. Abed, and U. A. Khan. 2018a. "Linear convergence in optimization over directed graphs with row-stochastic matrices". *IEEE Transactions on Automatic Control*. 63(10): 3558–3565.

[147] Xi, C., R. Xin, and U. A. Khan. 2018b. "ADD-OPT: Accelerated distributed directed optimization". *IEEE Transactions on Automatic Control*. 63(5): 1329–1339.

[148] Xiao, L. and S. Boyd. 2004. "Fast linear iterations for distributed averaging". *Systems & Control Letters*. 53(1): 65–78.

[149] Xie, P., K. You, R. Tempo, S. Song, and C. Wu. 2018. "Distributed Convex Optimization with Inequality Constraints over Time-varying Unbalanced Digraphs". *IEEE Transactions on Automatic Control*. 63(12): 4331–4337.

[150] Xin, R. and U. A. Khan. 2018. "A linear algorithm for optimization over directed graphs with geometric convergence". *IEEE Control Systems Letters*. 2(3): 325–330.

[151] Xu, J., S. Zhu, Y. C. Soh, and L. Xie. 2015. "Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant stepsizes". In: *IEEE Conf. on Decision and Control (CDC)*. 2055–2060.

[152] Xu, J., S. Zhu, Y. C. Soh, and L. Xie. 2018a. "A Bregman splitting scheme for distributed optimization over networks". *IEEE Transactions on Automatic Control*. 63(11): 3809–3824.

[153] Xu, J., S. Zhu, Y. C. Soh, and L. Xie. 2018b. "Convergence of asynchronous distributed gradient methods over stochastic networks". *IEEE Transactions on Automatic Control*. 63(2): 434–448.

[154] Yang, B. and M. Johansson. 2010. "Distributed optimization and games: A tutorial overview". In: *Networked Control Systems*. Springer. 109–148.

[155] Yang, S., Q. Liu, and J. Wang. 2017. "A multi-agent system with a proportional-integral protocol for distributed constrained optimization". *IEEE Transactions on Automatic Control*. 62(7): 3461–3467.

[156] Yuan, K., Q. Ling, and W. Yin. 2016. "On the convergence of decentralized gradient descent". *SIAM Journal on Optimization*. 26(3): 1835–1854.

[157] Zanella, F., D. Varagnolo, A. Cenedese, G. Pillonetto, and L. Schenato. 2011. "Newton-Raphson consensus for distributed convex optimization". In: *IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*. 5917–5922.

[158] Zanella, F., D. Varagnolo, A. Cenedese, G. Pillonetto, and L. Schenato. 2012. "Asynchronous Newton-Raphson Consensus for Distributed Convex Optimization". In: *IFAC Workshop on Distributed Estimation and Control in Networked Systems*.

[159] Zeng, J. and W. Yin. 2018. "On nonconvex decentralized gradient descent". *IEEE Transactions on signal processing*. 66(11): 2834–2848.

[160] Zhang, Y. and M. M. Zavlanos. 2018. "A Consensus-Based Distributed Augmented Lagrangian Method". In: *IEEE Conference on Decision and Control (CDC)*. 1763–1768.

[161] Zhu, H., G. B. Giannakis, and A. Cano. 2009. "Distributed in-network channel decoding". *IEEE Transactions on Signal Processing*. 57(10): 3970–3983.

[162] Zhu, M. and S. Martínez. 2010. "Discrete-time dynamic average consensus". *Automatica*. 46(2): 322–329.

[163] Zhu, M. and S. Martínez. 2012. "On distributed convex optimization under inequality and equality constraints". *IEEE Transactions on Automatic Control*. 57(1): 151–164.