for my family
ACKNOWLEDGEMENTS

This thesis presents the results of my research as visiting graduate student at the Mechanical & Aerospace Engineering Department (MAE), Jacobs School of Engineering, University of California, San Diego (UCSD).

First and foremost, I would like to thank my advisor Prof. Jorge Cortés for accepting me as visiting graduate student and for supervising my thesis. This work owes its very existence to him. His passion and enthusiasm for control theoretical and mathematical research as well as his unparalled dedication to the students make him an outstanding researcher and mentor. I want to thank him for making my research stay in San Diego, CA, USA so memorable and productive at the same time, for all the enlightening and fruitful discussions, and for pushing me in the right directions.

Second, I want to thank Prof. Dr.-Ing. Frank Allgöwer for all the opportunities he gave me during my studies in Engineering Cybernetics at the University of Stuttgart, Germany, and for inspiring me through his uncanny passion for control theory. As my mentor of a fellows group of the German National Academic Foundation (Studienstiftung des deutschen Volkes) I want to thank him for his permanent guidance and all helpful advices that have been most valuable to me.

I also want to thank Jan-Maximilian Montenbruck, M.Sc. from the Institute for Systems Theory and Automatic Control (IST) at the University of Stuttgart, Germany for supervising my work and for valuable comments and suggestions.

I would like to thank my colleagues from the Martínez & Cortés research group at UCSD for their contribution to a great working atmosphere. Special thanks to my labmate Ashish Cherukuri, for all the discussions and mentorship, both academically and personally.

Furthermore, I want to thank the German National Academic Foundation and the Dr. Jürgen & Irmgard Ulderup Foundation for financial support that made my research stay in San Diego, CA, USA possible.

Last but not least, I am eternally indebted to my wife Sophia and my son Jacob, and to my parents for all their support and love.

San Diego, July 2015
Simon K. Niederländer
# TABLE OF CONTENTS

List of Symbols ........................................................................................................... ix

Abstract .................................................................................................................. xi

Deutsche Kurzfassung .............................................................. xiii

Chapter 1 Introduction ................................................................. 1
  1.1 Motivation and Focus ................................................................. 2
  1.2 Literature Review ........................................................................ 2
    1.2.1 Distributed Optimization Algorithms ......................... 3
    1.2.2 Robust Optimization Algorithms ................................. 3
  1.3 Contributions ................................................................................. 4
    1.3.1 Distributed Coordination for Convex Optimization 4
    1.3.2 Robust Distributed Optimization ................................. 5
  1.4 Outline .......................................................................................... 5

Chapter 2 Preliminaries ............................................................... 7
  2.1 Nonsmooth Analysis ................................................................. 7
  2.2 Set-valued and Projected Dynamical Systems .................... 10
    2.2.1 Set-valued Dynamical Systems ................................. 10
    2.2.2 Projected Dynamical Systems and Viability Theory 12
  2.3 Convex and Robust Optimization ....................................... 14
    2.3.1 Nonsmooth Convex Optimization ....................... 14
    2.3.2 Introduction to $\infty/2$-Optimization .................. 17
  2.4 Summary ...................................................................................... 18

Chapter 3 Distributed Continuous-Time Coordination ....... 19
  3.1 Problem Statement ................................................................. 20
  3.2 Lagrangian Saddle-Point Characterization ..................... 21
  3.3 Distributed Coordination Algorithm ................................ 24
    3.3.1 Saddle-Point Dynamics ...................................... 24
    3.3.2 Convergence Analysis ...................................... 25
  3.4 Projected Saddle-Point-Like Dynamics ......................... 28
    3.4.1 Set-valued Projection Operator ...................... 29
    3.4.2 Stability Analysis ...................................... 30
LIST OF SYMBOLS

The following list contains symbols which are used throughout the thesis. Symbols which are defined locally are not listed below.

Sets

\( \mathbb{R}, \mathbb{R}_{\geq 0}, \mathbb{R}_>0 \)  
sets of real, nonnegative real, and positive numbers

\( \mathbb{Z}_{>0} \)  
set of positive integer numbers

\( \mathcal{B}(x, \delta) \)  
(open) ball of radius \( \delta \in \mathbb{R}_{>0} \) with center \( x \in \mathbb{R}^n \), i.e., \( \mathcal{B}(x, \delta) = \{ y \in \mathbb{R}^n \mid \| y - x \| < \delta \} \subset \mathbb{R}^n \)

\( \text{co}(C) \)  
convex hull of \( C \subset \mathbb{R}^n \)

\( \text{co}(C) \)  
closure of the convex hull of \( C \subset \mathbb{R}^n \)

\( \text{int}(C) \)  
interior of \( C \subset \mathbb{R}^n \)

\( \text{relint}(C) \)  
relative interior of \( C \subset \mathbb{R}^n \)

\( \text{bd}(C) \)  
boundary of \( C \subset \mathbb{R}^n \)

\( \text{cl}(C) \)  
closure of \( C \subset \mathbb{R}^n \), i.e., \( \text{cl}(C) = \text{int}(C) \cup \text{bd}(C) \)

\( \text{graph}(C) \)  
graph of \( C \subset \mathbb{R}^n \), i.e., \( \text{graph}(C) = \{(x, y) \in X \times Y \mid y \in C(x)\} \)

Vectors and Matrices

\( I \)  
identity matrix

\( 0 \)  
matrix of zeros

\( \mathbb{1}_n \)  
all-ones vector, i.e., \( \mathbb{1}_n = (1, \ldots, 1) \in \mathbb{R}^n \)

\( A^\top \)  
transpose of matrix \( A \)

\( A \otimes B \)  
Kronecker product of matrices \( A \) and \( B \)

\( A \oplus B \)  
direct sum of matrices \( A \) and \( B \)

\( \begin{pmatrix} A & B \\ \ast & C \end{pmatrix} \)  
\( \ast \) denotes the symmetric part of a (block-)matrix

\( A \succ 0, (A \succeq 0) \)  
matrix \( A \) is positive (semi-)definite

\( \lambda_{\text{min}}(A) \)  
smallest eigenvalue of matrix \( A \)

\( \lambda_{\text{max}}(A) \)  
largest eigenvalue of matrix \( A \)

Functions

\( \langle \cdot, \cdot \rangle \)  
Euclidean inner product

\( \| \cdot \|, \| \cdot \|_\infty \)  
\( \ell_2 \)- and \( \ell_\infty \)-norms in \( \mathbb{R}^n \)

\( \| x \|_Q^2 \)  
\( \| x \|_Q^2 = \langle x, Qx \rangle \) for \( x \in \mathbb{R}^n \), \( Q \in \mathbb{R}^{n \times n} \)

\( [x]^+ \)  
max operator, i.e., \( [x]^+ = \{ \max\{0, x_1\}, \ldots, \max\{0, x_n\} \} \in \mathbb{R}_{\geq 0}^n \)

\( d_C(x) \)  
distance function, i.e., \( d_C(x) = \inf_{y \in C} \| x - y \| \)

\( \text{proj}_C(x) \)  
projection operator, i.e., \( \text{proj}_C(x) = \arg\min_{y \in C} \| x - y \| \)

\( C^1 \)  
class of continuously differentiable functions
$C^{1,1}$ class of $C^1$ functions with Lipschitz continuous gradients
$C^2$ class of twice continuously differentiable functions

Acronyms

a.a. almost all
a.e. almost everywhere
cf. confer (compare)
e.g. exempli gratia (for example)
i.e. id est (that is)
iff if and only if
o.w. otherwise
s.t. such that
w.l.o.g. without loss of generality
ABSTRACT

NETWORKED systems in the context of modern engineering problems have provoked an extensive research activity on distributed control and optimization. Under this paradigm, individual agents in a network – described by a weighted undirected graph topology – seek to cooperatively achieve a common goal using only local information communicated by their immediate neighbors. In this thesis, we focus on network objectives that give rise to general nonsmooth convex and robust (i.e., semi-infinite) optimization problems with an inherent distributed structure. Our objective is to synthesize distributed continuous-time coordination algorithms that allow each agent in a network to find its component of the optimal solution vector.

The contributions of this thesis are twofold. First, we develop distributed coordination algorithms that solve general nonsmooth convex programs. In particular, we design saddle-point(-like) dynamics based on an augmented Lagrangian function associated with a nonsmooth convex program. We show that the proposed set-valued dynamics asymptotically converge to a point in the solution set. Moreover, the algorithms proposed are amenable for distributed implementation over a network of agents and are scalable with the dimension of the optimal solution vector. Second, we extend our previous results of distributed algorithmic nature to obtain robust (i.e., worst-case) optimal solutions of semi-infinite optimization problems. Specifically, we establish a framework that allows us to reformulate a robust optimization problem as a convex program with an intrinsic distributed structure preserving information from the underlying uncertainty sets. We then synthesize specifically tailored continuous-time algorithms for uncertain optimization problems and characterize their convergence towards a point in the robust solution set.

Throughout the thesis, we emphasize the development of provable correct coordination algorithms using relevant tools from nonsmooth analysis, set-valued and projected dynamical systems, viability theory, and convex as well as robust programming. Simulations in a linear model predictive control problem and a static semi-infinite optimization problem illustrate our results.
DEUTSCHE KURZFASSUNG


Chapter 1

Introduction

Networked systems in the context of modern engineering problems are understood as the interconnection of multiple components\(^1\) with the ability to communicate among each other concerning to a prescribed network topology\(^2\). Under this paradigm, a wide variety of applications such as social, biological, economic and engineering processes can be considered. Traditionally, the study of networked systems is divided into the parts modeling, design, and control. From the modeling aspect, it is of particular interest to predict the future network behavior and to obtain further insights into the underlying mechanisms that affect a network. With regards to the design aspect of networks, considerations on which components have to be installed in order to maximize profit respecting resource limitations are in the foreground. Finally the control aspect of a network refers to the development and implementation of distributed algorithms that steer the networked system towards a desired state or behavior.

This thesis predominantly reflects the control perspective of networked systems, and thus focusses on the design of provably correct coordination algorithms. In general, a network control algorithm can be understood as a policy that incorporates only local information to dictate the behavior of the subsystems. To this end, there exist two main architectures to control, i.e., influence networked systems. On the one hand, centralized approaches identify a single entity to generate and communicate the control signal for all other agents in the network. Consequently, in order to implement a centralized controller, full network information has to be accessible to the controller which in turn requires direct communication links between all subsystems in the network. Yet, centralized control strategies that may ignore the inherent network structure of the underlying application are often favored because of the simplicity of designing coordination algorithms. On the other hand, decentralized approaches, as we consider in this thesis, allow each subsystem to compute and implement its own control signal based on locally available information. Decentralized control strategies often yield advantages over centralized solvers when the problem under consideration requires inexpensive and low-performance computations, robustness against malfunctions, or the ability to quickly react to changes in the network topology. Moreover,

\(^1\)In this thesis, we use the terms component, subsystem, agent, and node interchangeably.
\(^2\)The network topologies considered throughout the thesis are described by means of connected, weighted, and undirected graphs [GR01].
distr** cours control algorithms enjoy the property of being naturally scalable with respect to the network size. However, the main challenge in the synthesis of distributed coordination algorithms is the fact that the control signal for each agent may only be generated using locally available information provided by its immediate neighbors. Interestingly, there is rarely a straightforward or intuitive mapping between the local behavior of a single component and the emergent global behavior of the network.

The remainder of the introduction focusses on the specific problem statements we consider in this thesis, arranges the problems within the context of the body on current literature, and summarizes the contributions of this work.

### 1.1 Motivation and Focus

This thesis focusses on the systematic development of distributed implementable coordination algorithms to achieve a desired network objective. For this purpose, a particularly powerful and widespread framework is mathematical programming. In general, the optimization approach allows to effectively capture the network objective in terms of a performance measure while enforcing the physical network constraints.

Motivated by network objectives that give rise to general nonsmooth convex optimization problems with an inherent distributed structure, our objective is to develop distributed continuous-time coordination algorithms that allow each agent to find its component of the optimal solution vector of the underlying convex program. This setup substantially differs from consensus-based distributed optimization strategies where the agents in a network agree on the entire optimal solution vector. We are particularly interested in establishing conditions on the convex programs that guarantee correctness and further induce strong convergence properties of the distributed coordination algorithms.

We then turn our attention to network objectives that can be described by optimization problems that explicitly involve a notion of variability (which models uncertainty) in the problem data. The basic idea of robust optimization is to seek an optimal solution which remains feasible and near-optimal under the perturbation of parameters in the optimization problem data. Here, our objective is to design specifically tailored distributed algorithms for robust computationally tractable optimization problems that allow subsystems to cooperatively compute reliable solutions which are immune against all realizations of uncertain data prescribed by some uncertainty sets. By definition, this problem setup is worst-case oriented in the sense that agents cooperatively compute minimum cost solutions satisfying a continuum of constraints.

### 1.2 Literature Review

In this section, we review the current state of the art regarding to the development of distributed continuous-time coordination algorithms for convex and robust optimization problems.
1.2. LITERATURE REVIEW

1.2.1 Distributed Optimization Algorithms

Mathematical programming plays an important role in a wide variety of networked systems, see e.g., [Ber98, SS07] and references therein. In particular, convex optimization problems model a broad array of engineering and economic scenarios and find numerous applications in diverse areas such as operations research, network flow optimization, control systems and signal processing.

Motivated by large scale problems and systems with data naturally partitioned over a network [BT97], there has been vast interest on distributed convex optimization, including the works [NOP10, ZM12, WE11, JKJJ08] and references therein. These works particularly build on consensus-based dynamics [OSFM07, BCM09, ME10, GC14]. However, a direct transcription of those methods to the present setup would result in cooperative strategies where agents interact with their neighbors but operate over a vector that represents the whole network state, and are therefore not scalable.

Considering scenarios closer to our approach, various distributed algorithms based on saddle-point dynamics [AHU58] have been proposed. The work [LCZL14] studies partial primal-dual dynamics for separable convex optimization problems with equality constraints in the context of power networks. More generally, the works [AHU58, FP10] introduce saddle-point dynamics of a smooth Lagrangian function associated with convex programs that only possess inequality constraints. The resulting dynamics are discontinuous in both the primal and dual variables because of the projections taken to keep the evolution within the feasible set. Both works establish convergence in the primal variables under the assumption that the solution of the optimization problem is unique [AHU58] or that the Slater constraint qualification certificate is satisfied [FP10], but do not characterize the properties of the final convergence point in the dual variables, which might indeed not be a dual solution. The work [DE11] proposes smooth saddle-point dynamics for general convex optimization problems that do not rely on projection operators, but converge to the set of primal-dual minimizer. However, the inequality constraints are only feasible asymptotically. More generally, the work [CC15] provides conditions under which asymptotic stability of saddle points under saddle-point dynamics that can be established. Our algorithm design builds on [RC15], which develops set-valued and discontinuous saddle-point dynamics specifically tailored for linear programs.

1.2.2 Robust Optimization Algorithms

Robust optimization techniques (as a general reference, see e.g., [BTGN09, BBC11]) are commonly used in diverse areas such as operations research and finance. In general, if the nominal problem data is slightly perturbed, the optimal solution of the underlying optimization problem can be completely meaningless.

A first attempt to incorporate hard uncertain constraints in mathematical programming goes back to [Soy73]. From there on, the framework of robust optimization has mainly been developed in the works [BTN98, BTN99, BTN00]. However, most existing algorithms to solve robust optimization problems are centralized, and thus not suitable for networked systems that demand distributed solutions. More recently,
the work [YHW+12] develops distributed discrete-time algorithms for robust optimization problems under linear uncertain constraints. To our knowledge, a rigorous investigation into distributed continuous-time coordination algorithms for semi-infinite optimization problems does not exist in the literature.

1.3 Contributions

This section summarizes our contributions to the body of research on distributed convex and robust optimization using continuous-time coordination algorithms. In what follows, we categorize our contributions with respect to distributed convex optimization and distributed robust mathematical programming.

1.3.1 Distributed Coordination for Convex Optimization

In Chapter 3, we consider general nonsmooth convex optimization scenarios defined by separable convex objective functions with coupling equality (affine) and inequality (convex) constraints. Both the objective function and the inequality constraints are Lipschitz continuous. As a result of this generality in the problem statement, we face various technical challenges in both the design and analysis of distributed coordination algorithms, particularly in what concerns the lack of differentiability of the problem data. Our first main contribution is the development of continuous-time saddle-point dynamics (that is, gradient descent in one variable and gradient ascent in the other) associated with an augmented Lagrangian function. It should be noted that, in general, saddle points are only guaranteed to be stable (but not necessarily asymptotically stable) under the corresponding saddle-point dynamics. We establish global convergence of trajectories of the dynamics to a point in the set of primal-dual solutions of general nonsmooth convex programs. Our proof strategy incorporates knowledge of a global parameter that in turn renders the algorithm to be not fully distributed. However, we propose alternative discontinuous saddle-point-like dynamics that rely on set-valued projection operations and are fully distributed over the underlying network topology. We show that the discontinuous dynamics share the same convergence properties as the saddle-point algorithm by establishing that, for sufficiently large values of the global parameter, the trajectories of the former are also trajectories of the latter. Our second main contribution is the development of modified saddle-point dynamics whose trajectories converge to the primal-dual solution exponentially fast, provided that mild convexity and regularity conditions on the aggregate objective function of the convex program are satisfied. Specifically, we characterize a performance bound on the trajectories of the modified dynamics in the absence of inequality constraints.

Chapter 3, in part, is a reprint of the material [NC15] as it appears in “Distributed Coordination for Separable Convex Optimization with Coupling Constraints” by S. K. Niederländer and J. Cortés in the proceedings of the 2015 IEEE Conference on Decision and Control. The author of this thesis is the primary investigator and author of the paper.
1.3.2 Robust Distributed Optimization

In Chapter 4, we consider robust optimization scenarios defined by a Lipschitz continuous convex objective function with linear uncertain inequality constraints, where the uncertainties are modeled by means of bounded polyhedral uncertainty sets. Our first contribution is the development of a framework that allows to reformulate the semi-infinite optimization problem as a convex optimization problem with linear equality and affine inequality coupling constraints, comprising the information of the uncertainty sets. The second contribution is the development of provably correct distributed continuous-time coordination algorithms customized for solving robust programs. As in Chapter 3, our algorithmic design builds on an augmented Lagrangian function. Also, we establish global convergence of trajectories of the derived saddle-point dynamics to a point in the set of primal-dual solutions of robust optimization problems. However, the proposed algorithm is, in general, not fully distributed over the network topology associated with the robust program. This motivates our next contribution on the design of an alternative algorithm that is amenable for distributed implementation using an explicit projection operator. We finally show that the projected dynamics share the same convergence properties as the saddle-point algorithm.

1.4 Outline

The thesis is organized as follows. Chapter 2 introduces basic notions and relevant tools from nonsmooth analysis, set-valued and projected dynamical systems, viability theory, and convex as well as robust optimization. In Chapter 3, we present the problem of solving a nonsmooth convex program with an inherent distributed structure over a network of agents. Here, we develop distributed continuous-time coordination algorithms that enable each agent to find its component of the optimal solution vector using only local information provided by its neighbors. Next, Chapter 4 develops distributed continuous-time coordination algorithms that are capable of solving convex optimization problems that explicitly include some notion of variability in the problem formulation. Finally, Chapter 5 summarizes our contributions and discusses ideas for future work.
Chapter 2

Preliminaries

This chapter introduces basic notions and relevant tools from nonsmooth analysis, set-valued and projected dynamical systems, viability theory, and convex as well as robust optimization.

A set \( X \subset \mathbb{R}^n \) is said to be \( \text{convex} \) if for any \( x, y \in X \) and any \( \theta \in [0,1] \), the point \( \theta x + (1-\theta)y \) also belongs to \( X \). Given \( X, Y \subset \mathbb{R}^n \), the \( \text{Minkowski sum} \) and the \( \text{Pontryagin set difference} \) of \( X \) and \( Y \) are defined by

\[
X + Y = \{ x + y \mid x \in X, y \in Y \}, \quad X - Y = \{ x - y \mid x \in X, y \in Y \}.
\]

A function \( f : \mathbb{R}^n \supset X \rightarrow \mathbb{R} \) is \( \text{positive definite} \) if \( f(x) = 0 \) for \( x = 0 \) and \( f(x) > 0 \) for all \( x \in X \setminus \{0\} \). A function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is called \( \text{radially unbounded} \) if \( f(x) \rightarrow \infty \) when \( \|x\| \rightarrow \infty \). A function \( f \) is \( \text{proper} \) if it is both positive definite and radially unbounded. Given \( \gamma \in \mathbb{R}_{>0} \), the \( \gamma \)-\( \text{sublevel set} \) of \( f \) is defined by

\[
f^{-1}(\leq \gamma) = \{ x \in X \mid f(x) \leq \gamma \}.
\]

If \( f \) is radially unbounded, then the set \( f^{-1}(\leq \gamma) \) is compact. A function \( f : X \rightarrow \mathbb{R} \) defined on a convex set \( X \subset \mathbb{R}^n \) is said to be \( \text{convex} \) if

\[
\forall x, y \in X, \forall \theta \in [0,1] : \quad f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y).
\]

Note that \( f \) is \( \text{concave} \) if \( -f \) is convex. A function \( f : \mathbb{R}^n \times \mathbb{R}^m \supset X \times Y \rightarrow \mathbb{R} \) is called \( \text{convex-concave} \) if \( x \mapsto f(x,y) \) is convex and \( y \mapsto f(x,y) \) is concave. A \( \text{set-valued map} \) \( F : \mathbb{R}^n \supset X \Rightarrow \mathbb{R}^n \) assigns to each element in \( x \in X \) the set \( F(x) \subset \mathbb{R}^n \). Finally, let \( \text{Ln} : \mathbb{R}^n \supset X \Rightarrow \mathbb{R}^n \) be a set-valued map that associates to each subset of \( F \) the set of least-norm elements of its closure \( \text{cl}(F) \). If \( F \) is convex and closed, then \( \text{Ln}(F) = \text{proj}_F(0) \) is singleton valued.

2.1 Nonsmooth Analysis

This section presents some facts on nonsmooth analysis [Cla83] that will play a central role in both the derivation and analysis of distributed coordination algorithms. In particular, extensions on notions of derivatives offer new prospects regarding to the
synthesis of gradient(-like) algorithmic strategies based on functions that are otherwise not differentiable, i.e., nonsmooth.

**Definition 1.1** (Lipschitz continuity). A function \( f : \mathbb{R}^n \to \mathbb{R} \) is said to be **locally Lipschitz** at \( x \in \mathbb{R}^n \) if
\[
\exists \delta_x > 0, \exists L_x \geq 0 : |f(y) - f(z)| \leq L_x \|y - z\|, \quad \forall y, z \in B(x, \delta_x).
\]
A function \( f \) is **locally Lipschitz** on \( X \subset \mathbb{R}^n \) if it is locally Lipschitz at \( x \), for all \( x \in X \). If \( f \) is locally Lipschitz on \( \mathbb{R}^n \), then it is said to be **locally Lipschitz**.

A function that is locally Lipschitz at \( x \in \mathbb{R}^n \) is continuous at \( x \in \mathbb{R}^n \). Note that continuously differentiable functions at \( x \in \mathbb{R}^n \) are locally Lipschitz at \( x \in \mathbb{R}^n \). Rademacher’s Theorem [Cla83] states that locally Lipschitz functions are differentiable a.e. (in the sense of Lebesgue measure). If \( f \) is convex (or concave), then it is locally Lipschitz. The right directional derivative and generalized directional derivative of \( f \) at \( x \in \mathbb{R}^n \) in the direction of \( v \in \mathbb{R}^n \) are, respectively,
\[
f'(x; v) = \lim_{\delta \to 0^+} \frac{f(x + \delta v) - f(x)}{\delta}, \quad f^0(x; v) = \limsup_{\delta \to 0^+} \frac{f(y + \delta v) - f(y)}{\delta}.
\]
The advantage of the generalized directional derivative compared to the right directional derivative is that the limit always exists. This motivates the following definition.

**Definition 1.2** (Regular functions). A function \( f : \mathbb{R}^n \to \mathbb{R} \) is said to be **regular** at \( x \in \mathbb{R}^n \) if for all \( v \in \mathbb{R}^n \), \( f'(x; v) \) exists and \( f^0(x; v) = f'(x; v) \).

A continuously differentiable function \( f \) at \( x \in \mathbb{R}^n \) is regular at \( x \in \mathbb{R}^n \). Also, a locally Lipschitz function at \( x \in \mathbb{R}^n \) which is convex (or concave) is regular (cf. Proposition 2.3.6 in [Cla83]). The following notion of a gradient is of particular importance throughout the work.

**Definition 1.3** (Generalized gradient). The **generalized gradient** \( \partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \) of \( f \) at \( x \in \mathbb{R}^n \) is defined by
\[
\partial f(x) = \text{co} \left\{ \lim_{i \to \infty} \nabla f(x_i) \mid x_i \to x, \ x_i \notin S \cup \Omega_f \right\},
\]
where \( \Omega_f \) denotes the set of points in \( \mathbb{R}^n \) at which \( f \) fails to be differentiable, and \( S \) is any other (arbitrarily chosen) set of measure zero (in the sense of Lebesgue).

Note that the resulting set \( \partial f(x) \) is blind under the choice of \( S \). From the definition, the generalized gradient of \( f \) at \( x \in \mathbb{R}^n \) consists of all convex combinations of all of the possible limits of the gradient at neighboring points where \( f \) is differentiable. Note that if \( f \) is continuously differentiable at \( x \in \mathbb{R}^n \), then \( \partial f(x) \) reduces to the singleton set \( \{ \nabla f(x) \} \).

A point \( x \in \mathbb{R}^n \) which satisfies \( 0 \in \partial f(x) \) is called **critical point** of \( f \). For a convex function \( f \), the **1st-order convexity condition** [Roc97] states that
\[
f(y) \geq f(x) + \langle \pi, y - x \rangle, \quad \forall \pi \in \partial f(x), \forall x, y \in \mathbb{R}^n.
\]
The following proposition combines some important results adopted from [Cor08].

**Proposition 1.4** (Calculus of generalized gradients). Let \( \{ f_i : \mathbb{R}^n \to \mathbb{R} \mid i \in \{1, \ldots, m\} \} \) be a finite collection of locally Lipschitz functions at \( x \in \mathbb{R}^n \). The following statements hold:

(i) (Dilation rule) If \( \kappa \in \mathbb{R}_{\geq 0} \), then \( \partial(\kappa f)(x) = \kappa \partial f(x) \).

(ii) (Sum rule) Given \( \{ f_i : \mathbb{R}^n \to \mathbb{R} \mid i \in \{1, \ldots, m\} \} \). Then \( \sum_{i=1}^{m} f_i \) is locally Lipschitz at \( x \in \mathbb{R}^n \), and

\[
\partial \left( \sum_{i=1}^{m} f_i \right)(x) \subseteq \sum_{i=1}^{m} \partial f_i(x).
\]

Moreover, if each \( f_i \) is regular at \( x \in \mathbb{R}^n \) for \( i \in \{1, \ldots, m\} \), then equality holds and \( \sum_{i=1}^{m} f_i \) is regular at \( x \in \mathbb{R}^n \).

(iii) (Max rule) Let \( \mathbb{R}^n \ni x \mapsto f(x) = \max_{i \in \{1, \ldots, m\}} f_i(x) \). Then, \( f \) is locally Lipschitz at \( x \in \mathbb{R}^n \), and

\[
\partial f(x) \subseteq \text{co}\bigcup \{ \partial f_i(x) \mid f_i(x) = f(x) \},
\]

where \( I(x) = \{ i \in \{1, \ldots, m\} \mid f_i(x) = f(x) \} \). Furthermore, if \( f_i \) is regular at \( x \in \mathbb{R}^n \) for all \( i \in \{1, \ldots, m\} \), then equality holds and \( f \) is regular at \( x \in \mathbb{R}^n \).

The following definitions and concepts related to boundedness and continuity of set-valued maps are motivated from the analysis needed to establish convergence results of the distributed algorithms developed in the forthcoming sections.

**Definition 1.5** (Local boundedness of set-valued maps). A set-valued map \( F : \mathbb{R}^n \supset X \rightrightarrows \mathbb{R}^n \) is said to be locally bounded, if

\[
\forall x \in X, \exists \delta, M > 0 : \|z\| \leq M, \quad \forall z \in \partial F(y), \ y \in B(x, \delta).
\]

**Definition 1.6** (Semi-continuity of set-valued maps). A set-valued map \( F : \mathbb{R}^n \supset X \rightrightarrows \mathbb{R}^n \) is upper semi-continuous, if

\[
\forall x \in X, \forall \varepsilon > 0, \exists \delta_x > 0 : F(y) \subseteq F(x) + B(0, \varepsilon), \quad \forall y \in B(x, \delta_x).
\]

Conversely, a set-valued map \( F : \mathbb{R}^n \supset X \rightrightarrows \mathbb{R}^n \) is lower semi-continuous, if

\[
\forall x \in X, \forall \varepsilon > 0, \exists \delta_x > 0 : F(x) \subseteq F(y) + B(0, \varepsilon), \quad \forall y \in B(x, \delta_x).
\]

The set-valued map \( F : \mathbb{R}^n \supset X \rightrightarrows \mathbb{R}^n \) is continuous if it is both upper and lower semi-continuous. The notion of upper semi-continuity of a map \( f : \mathbb{R}^n \supset X \to \mathbb{R} \) is weaker than the notion of upper semi-continuity of \( f \) when viewed as (singleton-valued) set-valued map from \( X \) to subset of \( \mathbb{R}^n \). Indeed, the latter is equivalent to the condition that \( f : \mathbb{R}^n \supset X \to \mathbb{R} \) is continuous. The following result states some important properties of the generalized gradient (cf. [Cla83]).

**Proposition 1.7** (Properties of the generalized gradient). Let \( f : \mathbb{R}^n \supset X \to \mathbb{R} \) be locally Lipschitz at \( x \in X \). Then, the set-valued mapping \( \partial f : \mathbb{R}^n \supset X \rightrightarrows \mathbb{R}^n \) is
upper semi-continuous and locally bounded at \( x \in X \). Moreover, the set \( \partial f(x) \) takes nonempty, convex, and compact values.

In subsequent sections, the result stated above will ensure existence of solutions of gradient(-like) algorithms based on locally Lipschitz functions.

**Definition 1.8 (Generalized Hessian).** Let \( f \in C^{1,1}(\mathbb{R}^n, \mathbb{R}) \). The generalized Hessian \( \partial(\nabla f) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \) of \( f \) at \( x \in \mathbb{R}^n \) is defined by

\[
\partial(\nabla f)(x) = \text{co} \left\{ \lim_{i \to \infty} \nabla^2 f(x_i) \mid x_i \to x, \, x_i \notin \Omega_f \right\}.
\]

Note that if \( f \) is twice continuously differentiable at \( x \in \mathbb{R}^n \), then \( \partial(\nabla f)(x) \) reduces to the singleton set \( \{ \nabla^2 f(x) \} \). The following result is a direct extension of Lebourg’s Mean-Value Theorem [Mor06] to vector-valued functions (cf. Proposition 2.6.5 in [Cla83]).

**Proposition 1.9 (Extended Mean-Value Theorem).** Let \( \nabla f : \mathbb{R}^n \to \mathbb{R}^n \) be Lipschitz continuous on \( X \subset \mathbb{R}^n \), and let \( x, y \in X \). Then,

\[
\nabla f(y) - \nabla f(x) \in \text{co} \{ \partial(\nabla f([x, y])) \} (y - x),
\]

where \([x, y] = x + \theta(y - x), \) for \( \theta \in [0, 1] \).

## 2.2 Set-valued and Projected Dynamical Systems

The exposition on set-valued and projected dynamical systems in this section is following [Cor08, AC84, NZ96, Smi01].

### 2.2.1 Set-valued Dynamical Systems

The dynamics proposed in this work are defined by means of differential inclusions, i.e., a generalization of differential equations. At each state, a differential inclusion specifies at set of possible evolutions, rather than a single one. The notion of so-called classical solutions to differential equations is not applicable to set-valued dynamics. This motivates the following definition.

**Definition 2.1 (Absolute continuity).** A function \( x : [a, b] \to \mathbb{R}^n \) is called absolutely continuous if, for all \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that, for each finite collection \( \{(a_1, b_1), \ldots, (a_n, b_n)\} \) of disjoint open intervals contained in \([a, b]\) with \( \sum_{i=1}^{n}(b_i - a_i) < \delta \), it follows that

\[
\sum_{i=1}^{n}\|x(b_i) - x(a_i)\| < \varepsilon.
\]

Every absolutely continuous function is continuous. Moreover, every continuously differentiable function is absolutely continuous. Note that every absolutely continuous
function is differentiable a.e. Finally, every locally Lipschitz function is absolutely continuous.

Let \( F : \mathbb{R}^n \supset X \rightrightarrows \mathbb{R}^n \) be a set-valued map. A time-invariant set-valued dynamical system is represented by the differential inclusion
\[
\dot{x}(t) \in F(x(t)), \quad x(t_0) = x_0 \in X.
\] (2.1)
A solution of (2.1) on an interval \([t_0, t_1] \subset \mathbb{R}\) is defined as an absolutely continuous mapping \( x : [t_0, t_1] \rightarrow X \) such that \( \dot{x}(t) \in F(x(t)) \) for a.a. \( t \in [t_0, t_1] \). Given \( x_0 \in X \), the existence of solutions of (2.1) with initial condition \( x_0 \) is guaranteed by the following lemma (cf. [Cor08]).

**Lemma 2.2** (Existence of solutions). Let the set-valued mapping \( F : \mathbb{R}^n \supset X \rightrightarrows \mathbb{R}^n \) be upper semi-continuous with nonempty, compact and convex values. Then, given \( x_0 \in X \), there exists a solution of (2.1) with initial condition \( x_0 \).

A solution is called maximal if it cannot be extended forward in time. The set of equilibria of \( F \) is defined by \( \text{eq}(F) = \{ x \in X \mid 0 \in F(x) \} \). A set \( \Omega \subset \mathbb{R}^n \) is weakly (respectively strongly) invariant for (2.1) if for each \( x_0 \in \Omega \), the set \( \Omega \) contains a maximal solution (respectively all maximal solutions) of (2.1).

A common theme in nonsmooth stability analysis is establishing the monotonic evolution of a real-valued function along the trajectories of a set-valued dynamical system. This motivates the following definition.

**Definition 2.3** (Lie-derivative). Given a locally Lipschitz function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) and a set-valued map \( F : \mathbb{R}^n \supset X \rightrightarrows \mathbb{R}^n \). The set-valued Lie-derivative \( L_F f : \mathbb{R}^n \supset X \rightrightarrows \mathbb{R} \) of \( f \) with respect to \( F \) at \( x \in X \) is defined as
\[
(L_F f)(x) = \{ \psi \in \mathbb{R} \mid \exists \xi \in F(x) : \langle \xi, \pi \rangle = \psi, \ \forall \pi \in \partial f(x) \}\.
\]
If \( F \) takes convex and compact values, then, for each \( x \in X \), \( (L_F f)(x) \) is a closed and bounded interval in \( \mathbb{R} \), possibly empty. If \( f \) is continuously differentiable at \( x \in X \), then
\[
(L_F f)(x) = \{ \langle \nabla f, \xi \rangle \mid \xi \in F(x) \}.
\]

The following results help to establish asymptotic convergence properties of set-valued dynamical systems (cf. [BC99]).

**Theorem 2.4** (Lyapunov stability for differential inclusions). Let \( F : \mathbb{R}^n \supset X \rightrightarrows \mathbb{R}^n \) be a set-valued map satisfying the hypotheses of Lemma 2.2, let \( x^* \in \text{eq}(F) \) be an equilibrium of the differential inclusion (2.1), and let \( X \subset \mathbb{R}^n \) be an open and connected set with \( x^* \in X \). Let \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) be such that the following conditions hold:

(i) \( f \) is locally Lipschitz and regular on \( X \).
(ii) \( f(x^*) = 0 \), and \( f(x) > 0 \) for all \( x \in X \setminus \{ x^* \} \).
(iii) \( (L_F f)(x) \subset (-\infty, 0] \) for all \( x \in X \).

Then, \( x^* \in \text{eq}(F) \) is a strongly stable equilibrium point of (2.1). Moreover, the set \( f^{-1}(\leq f(x_0)) \) is strongly invariant with respect to (2.1). If (iii) is replaced by
(iii)’ \((L_f f)(x) \subset (-\infty, 0)\) for all \(x \in X\),
then \(x^* \in \text{eq}(F)\) is a strongly asymptotically stable equilibrium of (2.1).

**Theorem 2.5** (LaSalle invariance principle for differential inclusions). Let \(f: \mathbb{R}^n \supset X \rightarrow \mathbb{R}\) be a locally Lipschitz and regular function. Let \(x_0 \in X\) and let \(f^{-1}(\leq f(x_0))\) be the connected component of the sublevel set \(\{x \in X \mid f(x) \leq f(x_0)\}\) containing \(x_0\). Assume the set \(f^{-1}(\leq f(x_0))\) is bounded and assume either \((L_f f)(x) \subset (-\infty, 0]\) or \((L_f f)(x) = \emptyset\) for all \(x \in f^{-1}(\leq f(x_0))\), where \(F\) is a locally bounded and upper semi-continuous set-valued map with nonempty, compact and convex values. Then, \(f^{-1}(\leq f(x_0))\) is strongly invariant with respect to (2.1) and any solution \(x: [t_0, +\infty) \rightarrow X\) of (2.1) starting from \(x_0 \in X\) converges to the largest weakly invariant set

\[\Omega \subset \text{cl}\{x \in X \mid 0 \in (L_f f)(x)\} \cap f^{-1}(\leq f(x_0))\].

Moreover, if the set \(\Omega\) consists of a finite number of points, then the limit of all solutions starting at \(x_0 \in X\) exists and is an element of \(\Omega\).

Differential inclusions are of great importance when dealing with differential equations that incorporate discontinuities. Let \(f: \mathbb{R}^n \supset X \rightarrow \mathbb{R}^n\) be piecewise continuous and consider the differential equation

\[\dot{x}(t) = f(x(t)), \quad x(t_0) = x_0 \in X\]. \hspace{1cm} (2.2)

for all \(t \in [t_0, t_1] \subset \mathbb{R}\). Note that the notion of a classical solution is not applicable to (2.2) due to its discontinuous right-hand side. This motivates the next definition.

**Definition 2.6** (Krasovskii set-valued map, cf. [GST12]). Let \(f: \mathbb{R}^n \supset X \rightarrow \mathbb{R}^n\). The Krasovskii set-valued map \(\mathcal{K}[f]: \mathbb{R}^n \supset X \rightrightarrows \mathbb{R}^n\) of \(f\) at \(x \in X\) is defined by

\[\mathcal{K}[f](x) = \bigcap_{\delta > 0} \text{ccl}\{f(B(x, \delta))\}\].

An absolutely continuous map \(x: [t_0, t_1] \rightarrow X\) is said to be a Krasovskii solution of (2.2) if it is a solution of the differential inclusion

\[\dot{x}(t) \in \mathcal{K}[f](x(t)), \quad x(t_0) = x_0 \in X\], \hspace{1cm} (2.3)

for a.a. \(t \in [t_0, t_1]\). Note that if \(\mathcal{K}[f]\) is locally bounded, upper semi-continuous and takes nonempty, convex, and compact values, then the existence of solutions of (2.3) starting from \(x_0 \in X\) is guaranteed by Lemma 2.2. The solutions of (2.2) in the sense of Krasovskii are by definition the solutions of the differential inclusion (2.3).

### 2.2.2 Projected Dynamical Systems and Viability Theory

Projected dynamical systems are a particular class of discontinuous dynamical systems. The discontinuities usually arise when the dynamical system at hand incorporates (state) constraints that are underlying a particular application. The following definitions introduce concepts from viability theory [Aub91] that are most helpful when dealing with constrained set-valued dynamical systems.
Let $X \subset \mathbb{R}^n$ be a nonempty, closed, and convex set; called the viability set. Let $d_X(x) : \mathbb{R}^n \to \mathbb{R}$ denote the distance function. The Bouligand (contingent) cone and the normal cone of $X$ at $x \in X$ are
\[
T_X(x) = \left\{ v \in \mathbb{R}^n \mid \liminf_{\delta \downarrow 0} \frac{d_X(x + \delta v)}{\delta} = 0 \right\} = \text{cl} \bigcup_{\delta > 0} \frac{1}{\delta} (X - x),
\]
respectively,
\[
N_X(x) = \{ q \in \mathbb{R}^n \mid \langle q, v \rangle \leq 0, \forall v \in T_X(x) \} = \text{cl} \bigcup_{\lambda \geq 0} \lambda \partial d_X(x).
\]
Note that if $x \in \text{int}(X)$, then $T_X(x) = \mathbb{R}^n$ and $N_X(x) = \emptyset$.

**Definition 2.7** (Viable solutions). Let $F : \mathbb{R}^n \supset X \rightrightarrows \mathbb{R}^n$ be a set-valued map. A solution $x : [t_0, t_1] \to X$ of (2.1) is said to be viable in $X$ under $F$ if
\[
\forall t \in [t_0, t_1], \quad x(t) \in X \subset \mathbb{R}^n.
\]
Given $x_0 \in X$, a necessary and sufficient condition (cf. Proposition 3.4.1 and 3.4.2 in [Aub91]) for the existence of viable solutions of (2.1) is that
\[
\forall x \in X, \quad F(x) \cap T_X(x) \neq \emptyset.
\]
The orthogonal (set) projection $\Pi_{T_X} : \mathbb{R}^n \supset X \rightrightarrows T_X$ of a set-valued map $F : \mathbb{R}^n \supset X \rightrightarrows \mathbb{R}^n$ onto $T_X$ at $x \in X$ is defined by
\[
\Pi_{T_X}(x, F(x)) = \bigcup_{\xi \in F(x)} \lim_{\delta \downarrow 0} \text{proj}_X (x + \delta \xi) - x.
\]
Note that if $x \in \text{int}(X)$, then $\Pi_{T_X}(x, F(x))$ reduces to the set $F(x)$. By definition of the projection operator $\Pi_{T_X}$, the inclusions $F(x) \subset T_X(x)$ and $\Pi_{T_X}(x, F(x)) \subset F(x) - N_X(x)$ hold for all $x \in X$ (cf. [AC84]).

Consider the projected differential inclusion
\[
\dot{x}(t) \in \Pi_{T_X}(x(t), F(x(t))), \quad x(t_0) = x_0 \in X. \tag{2.4}
\]
In general, the projection operator $\Pi_{T_X}$ in (2.4) possesses no continuity properties and the values of $\Pi_{T_X}(x, F(x))$ are not necessarily convex. The following definition introduces a further notion of solution of the projected differential inclusion (2.4).

**Definition 2.8** (Slow solutions). Let $\Pi_{T_X} : \mathbb{R}^n \supset X \rightrightarrows T_X$ be a set-valued map, and let $L_n : T_X \rightrightarrows T_X$ denote the set of least-norm elements of $\text{cl}(\Pi_{T_X})$. A solution $x : [t_0, t_1] \to X$ of (2.4) is said to be slow in $X$ under $\Pi_{T_X}$ if it is a solution of
\[
\dot{x}(t) \in L_n (\Pi_{T_X}(x(t), F(x(t)))), \quad x(t_0) = x_0 \in X.
\]
Given $x_0 \in X$, the following lemma guarantees the existence of viable (respectively slow) solutions of (2.4) with initial condition $x_0 \in X$ (cf. [AC84]).

**Lemma 2.9** (Existence of viable and slow solutions). Let $X \subset \mathbb{R}^n$ be a nonempty, closed and convex set. Let the set-valued map $F : \mathbb{R}^n \supset X \rightrightarrows \mathbb{R}^n$ satisfy
\[
\]
(i) \( \text{graph}(F) \) is closed,
(ii) \( F(X) \) is bounded, and
(iii) \( \forall x \in X, F(x) \) is convex.

Then, for all \( x_0 \in X \), there exists a viable solution \( x : [t_0, t_1] \to X \) of
\[
\dot{x}(t) \in F(x(t)) - N_X(x(t)), \quad x(t_0) = x_0 \in X.
\]
(2.5)

Moreover, the set of viable solutions of (2.4) and (2.5) coincide. In addition, if (i) is replaced by

(i)' \( \text{graph}(F) \) is closed and \( F \) is continuous,

then there exists a slow solution \( x : [t_0, t_1] \to X \) of (2.5). Moreover, the set of slow solutions of (2.4) and (2.5) coincide.

2.3 Convex and Robust Optimization

This section introduces some basic notions and results on convex and robust optimization following [HUL93, DD12, BTGN09, BBC11].

2.3.1 Nonsmooth Convex Optimization

The term convex optimization refers to the minimization problem of a convex function over a convex set which is described by means of equality and inequality constraints. In this work, both the objective function and the inequality constraints are Lipschitz continuous. This motivates the following definitions and results on nonsmooth convex optimization and \( \ell_1 \)-exact-penalty methods.

Consider the mathematical optimization problem
\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \in X,
\end{align*}
\]
(2.6)
where \( f : \mathbb{R}^n \supset X \to \mathbb{R} \cup \{-\infty, +\infty\} \) is called the objective function and \( X \subset \mathbb{R}^n \) denotes the feasibility set. Any point \( x \in X \) is said to be a feasible point. Conversely, any point \( x \in \mathbb{R}^n \setminus X = \{ x \in \mathbb{R}^n \mid x \notin X \} \) is said to be infeasible.

**Definition 3.1** (Local and global minima). A point \( x^* \in X \) is said to be a local minimum of \( f \) on \( X \) if
\[
\exists \delta > 0 : f(x^*) \leq f(x), \forall x \in X \cap B(x^*, \delta).
\]
A point \( x^* \in X \) is said to be a global minimum of \( f \) on \( X \) if
\[
\exists \delta > 0 : f(x^*) \leq f(x), \forall x \in X.
\]
The value \( f(x^*) \) on \( X \) is said to be the *optimal value* of (2.6). Let \( X \) be a nonempty and convex set, and let \( f : \mathbb{R}^n \supset X \to \mathbb{R} \) be a convex function on \( X \), then (2.6) is said to be a *convex program*. The following proposition is a fundamental result in convex programming (cf. [HUL93]).

**Proposition 3.2** (Sufficient condition for global minima). Let \( x^* \in X \) be a local minimum of (2.6). Then, \( x^* \in X \) is also a global minimum. Moreover, the set

\[
X \cap \{ x^* \in \mathbb{R}^n \mid f(x^*) \leq f(x), \forall x \in X \}
\]

is a closed convex subset of \( X \).

The following result provides necessary and sufficient conditions (cf. [DD12]) for a point \( x^* \in X \) to be a global minimizer of (2.6).

**Theorem 3.3** (Abstract optimality conditions). Let \( X \subset \mathbb{R}^n \) be a nonempty, closed and convex set, and let \( f : \mathbb{R}^n \supset X \to \mathbb{R} \) be Lipschitz continuous. Then, \( x^* \in X \) is a global minimizer of (2.6) if and only if either one of the following conditions hold:

(i) \( f'(x^*; v) \geq 0, \forall v \in T_X(x^*) \), or

(ii) \( 0 \in \partial f(x^*) + N_X(x^*) \).

Note that condition (ii) appears as a dual formulation of (i). Equivalently, (ii) can be stated as \(- \partial f(x^*) \cap N_X(x^*) \neq \emptyset \). In what follows, suppose the set \( X \subset \mathbb{R}^n \) has a representation via equality and inequality constraints,

\[
X = \{ x \in \mathbb{R}^n \mid h(x) = 0, g(x) \leq 0 \},
\]

where \( g : \mathbb{R}^n \to \mathbb{R}^m \) is convex and Lipschitz continuous, and \( h : \mathbb{R}^n \to \mathbb{R}^p \) is affine, i.e., \( h(x) = Ax - b \), with \( A \in \mathbb{R}^{p \times n} \) and \( b \in \mathbb{R}^p \). The equality and inequality functions \( h \) and \( g \) are understood component-wise, i.e., \( h_\ell(x) = 0 \) for all \( \ell \in \{1, \ldots, p\} \), and \( g_k(x) \leq 0 \) for all \( k \in \{1, \ldots, m\} \).

Consider the nonsmooth convex optimization problem

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad h_\ell(x) = 0, \quad \ell \in \{1, \ldots, p\}, \\
& \quad g_k(x) \leq 0, \quad k \in \{1, \ldots, m\}.
\end{align*}
\]

Let the viability set be defined by \( G = \{ x \in \mathbb{R}^n \mid g(x) \leq 0 \} \). The following definition is of particular importance throughout the work.

**Definition 3.4** (Slater’s constraint qualification certificate). The *Slater constraint qualification certificate* is satisfied by (2.7) if

\[
\exists x \in \text{relint } X : \forall k \in \{1, \ldots, m\}, \ g_k(x) < 0 \land h(x) = 0.
\]

**Proposition 3.5** (Explicit normal cone formulation, cf. [DD12]). Let \( G \subset \mathbb{R}^n \) be a nonempty, closed and convex set, and suppose (2.7) satisfies Slater’s constraint
qualification certificate. Then, the normal cone of $G$ at $x \in G$ is
\[ N_G(x) = \left\{ \sum_{k \in K(x)} \nu_k \pi_{g_k} \in \mathbb{R}^n \mid \pi_{g_k} \in \partial g_k(x), \nu_k \geq 0, k \in K(x) \right\}, \]
where $K(x) = \{k \in \{1, \ldots, m\} \mid g_k(x) = 0\}$ is the active index set at $x \in G$ and $\nu_k \geq 0$ are so-called Lagrange multiplier.

Note that if $K(x)$ is nonempty and Slater’s constraint qualification certificate is satisfied by (2.7), then $N_G(x)$ is a closed and convex cone. The following theorem provides first order necessary and sufficient optimality conditions for (2.7), know as the Karush-Kuhn-Tucker (KKT) optimality conditions (cf. [DD12]).

**Theorem 3.6 (KKT optimality conditions).** Let the Slater constraint qualification certificate be satisfied by (2.7). A point $x^* \in X$ is a KKT point of (2.7) if and only if there exist Lagrange multiplier $\mu^* \in \mathbb{R}^p$ and $\nu^* \in \mathbb{R}^m_{\geq 0}$ such that
\[ 0 \in \partial f(x^*) + \sum_{\ell=1}^p \mu^*_\ell \partial h_\ell(x^*) + \sum_{k \in K(x^*)} \nu^*_k \partial g_k(x^*), \]
\[ h(x^*) = 0, \quad g(x^*) \preceq 0, \quad \langle \nu^*, g(x^*) \rangle = 0, \]
where $K(x^*) = \{k \in \{1, \ldots, m\} \mid g_k(x^*) = 0\}$.

The condition $\langle \nu^*, g(x^*) \rangle = 0$ is called transversality condition or complementary slackness condition; and strict complementary slackness holds if $g_k(x^*) = 0$ implies $\nu^*_k > 0$ for some $k \in K(x^*)$.

Let $\kappa \in \mathbb{R}_{\geq 0}$. Consider the intrinsically nonsmooth $\ell_1$-exact-penalty function $f^\kappa : \mathbb{R}^n \supset X \to \mathbb{R} \cup \{-\infty, +\infty\}$ defined by
\[ f^\kappa(x) = f(x) + \kappa \sum_{k \in K(x)} [g_k(x)]^+. \]

Let the $\ell_1$-exact-penalized convex optimization problem be defined by
\[
\begin{align*}
\text{minimize} & \quad f^\kappa(x) \\
\text{subject to} & \quad h_\ell(x) = 0, \quad \ell \in \{1, \ldots, p\}.
\end{align*}
\]

The following lemma states conditions under which the solutions of (2.7) and (2.8) coincide (cf. [Ber75]).

**Lemma 3.7 (Equivalence of solutions).** Suppose (2.7) is convex, has a nonempty and compact solution set, and satisfies Slater’s constraint qualification certificate. Then, (2.7) and (2.8) have identical solutions if $\kappa > \|\nu^*\|_\infty$ for some Lagrange multiplier $\nu^* \in \mathbb{R}^m_{\geq 0}$ of (2.7).

Note that the Lagrange multiplier $\mu^* \in \mathbb{R}^p$ and $\nu^* \in \mathbb{R}^m_{\geq 0}$ of (2.7) exist due to Slater’s constraint qualification certificate (cf. Definition 3.4).
2.3.2 Introduction to $\infty/2$-Optimization

Robust (or semi-infinite) optimization refers to a minimization problem in which some notion of variability is explicitly included in the problem formulation. The objective is to compute minimum cost solutions among all solutions that are feasible for all realizations of uncertainty.

Consider the semi-infinite optimization problem

$$\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad g_k(x, a_k) \leq 0, \quad \forall a_k \in A_k,
\end{align*}$$

for $k \in \{1, \ldots, m\}$, where $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$ is the convex and Lipschitz continuous objective function, and $g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ represents the inequality constraints (convex and Lipschitz continuous in $x$, affine in $a_k$). Let the uncertainty parameter $a_k \in \mathbb{R}^n$ take arbitrary values in the bounded uncertainty set $A_k \subset \mathbb{R}^n$.

Let $A = A_1 \times \cdots \times A_m$. The robust feasibility set of (2.9) is denoted by $X(A) = \{x \in \mathbb{R}^n \mid g_k(x, a_k) \leq 0, \; \forall a_k \in A_k, \; k \in \{1, \ldots, m\}\}$.

In general, (2.9) requires to evaluate infinitely many constraints. Alternatively, consider to evaluate the inequality constraints only at global uncertainty maximizers.

**Definition 3.8 (Robust inner problems).** The robust inner problems associated with (2.9) are defined by

$$\max_{a_k \in A_k} g_k(x, a_k) \leq 0, \quad k \in \{1, \ldots, m\}.$$

Note that computational tractability of the robust inner problems is tantamount for $A = A_1 \times \cdots \times A_m$ (respectively $X(A)$) to be convex.

**Definition 3.9 (Robust counterpart).** The robust counterpart associated with (2.9) is the bi-level (or min-max) structured optimization problem

$$\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad \max_{a_k \in A_k} g_k(x, a_k) \leq 0, \quad k \in \{1, \ldots, m\}.
\end{align*}$$

Let the inequality constraints in (2.9) be affine with respect to both $x$ and $a_k$

$$g_k(x, a_k) = \langle a_k, x \rangle - b_k \leq 0, \quad \forall a_k \in A_k, \quad k \in \{1, \ldots, m\},$$

where $a_k \in \mathbb{R}^n$ denotes the $k$th row of the matrix $A \in \mathbb{R}^{m \times n}$ with $m \leq n$, and $b_k \in \mathbb{R}$. In what follows, suppose that the unknown-but-bounded uncertainty set $A_k$ possesses a polyhedral representation defined by

$$A_k = \{a_k \in \mathbb{R}^n \mid D_k a_k \leq e_k\},$$

where $D_k \in \mathbb{R}^{p_k \times n}$, $e_k \in \mathbb{R}^{p_k}$, and $p_k \in \mathbb{Z}_{>0}$ denotes the number of halfspaces that enclose the uncertainty vector $a_k \in \mathbb{R}^n$. Note that under polyhedral uncertainty representation, the robust inner problem (cf. Definition 3.8) induces a linear structure. The
following proposition provides an explicit convex reformulation of the robust counter-
part of (2.9) under consideration of polyhedral uncertainty sets (cf. [BTGN09]).

**Proposition 3.10** (Explicit reformulation of the robust counterpart). Let \( \mathcal{A}_k = \{ a_k \in \mathbb{R}^n \mid D_k a_k \preceq e_k \} \) and suppose that the robust counterpart of (2.9) satisfies Slater’s constraint qualification certificate. Then, the robust counterpart of (2.9) explicitly admits the equivalent convex reformulation

\[
\begin{align*}
\text{minimize} \\ x \in \mathbb{R}^n, \{ \lambda_k \in \mathbb{R}^{p_k} \} \\
\text{subject to} & \langle e_k, \lambda_k \rangle \leq b_k, \quad k \in \{1, \ldots, m\}, \\
 & D_k^T \lambda_k = x, \quad k \in \{1, \ldots, m\}, \\
 & \lambda_k \succeq 0, \quad k \in \{1, \ldots, m\}.
\end{align*}
\]

Note that the size of the above optimization problem grows polynomially in the problem size of (2.9) and in the dimensions of the uncertainty sets \( \mathcal{A}_k \).

### 2.4 Summary

In this chapter, we have recalled basic notions and preliminaries relevant for the technical analysis provided in the remaining chapters of this thesis. In Section 2.1, we have introduced well-known concepts from nonsmooth analysis that will be most helpful in the development of coordination algorithms for optimization problems without differentiable data (cf. Chapter 3 and 4). In Section 2.2, we have discussed set-valued and projected dynamical systems that are of particular importance when the dynamics at hand are restricted to some viability set. Finally, we have considered convex and robust optimization problems in Section 2.3.
Chapter 3

Distributed Continuous-Time Coordination

Motivated by networks of agents that give rise to nonsmooth convex optimization problems with an inherent distributed structure, our objective in this chapter is to synthesize and analyze continuous-time coordination algorithms that allow each agent in a network to find its component of the optimal solution vector using only local information provided by its neighbors.

The algorithm proposed in this chapter builds on concepts of Lagrangian duality theory and strong duality that facilitate an alternative approach to characterize the solutions of a nonsmooth convex program in terms of saddle points of an augmented Lagrangian function. In particular, we show that the resulting continuous-time saddle-point algorithm is provably correct by relying on the LaSalle Invariance Principle for differential inclusions. In general, the saddle-point dynamics derived are not fully distributed because of a global parameter associated with the intrinsically nonsmooth $\ell_1$-exact-penalty function employed to encode the inequality constraints of the convex program. However, we further synthesize discontinuous saddle-point(-like) dynamics that, while enjoying the same convergence guarantees, are fully distributed over a network of agents and scalable with the dimension of the optimal solution vector.

We then focus on the characterization of the convergence rate of the proposed saddle-point(-like) dynamics. Specifically, we identify a nonsmooth Lyapunov function for modified saddle-point dynamics that incorporates the augmented Lagrangian function of a nonsmooth convex program and a LaSalle function. This finding allows us to establish convergence properties without relying on the LaSalle Invariance Principle. In addition, under mild convexity and regularity conditions on the objective function of a convex program, we further characterize exponential convergence properties of the proposed saddle-point dynamics and provide a performance bound in the absence of inequality constraints.

Simulations in a linear model predictive control (MPC) example for a network of agents with coupling equality and inequality constraints, whose objective is to compute a control input sequence that simultaneously minimizes the actuation effort and the network state, illustrate our results and conclude this chapter.
3.1 Problem Statement

In this section, we introduce a general nonsmooth convex optimization problem which shall be solved over a network of agents using continuous-time coordination algorithms that only require local information provided through the communication links of neighboring agents.

Consider a network of \( n \in \mathbb{Z}_{>0} \) agents whose communication topology is represented by a weighted undirected graph \( \mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W}) \), where \( \mathcal{V} = \{1, \ldots, n\} \) is the set of vertices, \( \mathcal{E} \subset \mathcal{V} \times \mathcal{V} \) is the set of edges, and \( \mathcal{W} \in \mathbb{R}^{\mathcal{E}} \) is the vector of weights indexed by edges of \( \mathcal{G} \). The state of agent \( i \in \{1, \ldots, n\} \) is denoted by \( x_i \in \mathbb{R} \) and its set of neighbors is \( \mathcal{N}(i) = \{j \in \mathcal{V} | (i, j) \in \mathcal{E}\} \). Note that, w.l.o.g., each agents state \( x_i \) could belong to an arbitrary Euclidean space \( \mathbb{R}^d \) but, for simplicity, the following exposition is restricted to the case \( d = 1 \). The objective of the agents is to cooperatively solve the nonsmooth convex optimization problem

\[
\text{minimize} \quad \sum_{i=1}^{n} f_i(x_i) \\
\text{subject to} \quad h_\ell(x) = 0, \quad \ell \in \{1, \ldots, p\} \\
g_k(x) \leq 0, \quad k \in \{1, \ldots, m\} \tag{3.1}
\]

where \( f_i : \mathbb{R} \to \mathbb{R} \) is a convex and locally Lipschitz cost function associated to agent \( i \in \{1, \ldots, n\} \), \( g : \mathbb{R}^n \to \mathbb{R}^m \) is convex and locally Lipschitz, and \( h : \mathbb{R}^n \to \mathbb{R}^p \) is affine, i.e., \( h(x) = Ax - b \), with \( A \in \mathbb{R}^{p \times n} \) and \( b \in \mathbb{R}^p \), where \( p \leq n \). For convenience, the network state is denoted by \( x = (x_1, \ldots, x_n) \in \mathbb{R}^n \) and the aggregate objective function is \( f(x) = \sum_{i=1}^{n} f_i(x_i) \). The feasibility set of (3.1) is denoted by

\[ C = \{x \in \mathbb{R}^n | h(x) = 0, \quad g(x) \preceq 0\}. \]

Moreover, let the solution set of (3.1) be denoted by

\[ X = \{x^* \in \mathbb{R}^n | f(x^*) \leq f(x), \quad \forall x \in C\}. \]

Finally, the viability set associated with (3.1) is defined by

\[ G = \{x \in \mathbb{R}^n | g(x) \preceq 0\}. \]

We will make the following three assumptions throughout the remainder of this chapter:

**Assumption 1.1 (Feasibility).** The convex program (3.1) is feasible, i.e., \( C \neq \emptyset \), and possesses a finite optimal value \( f(x^*) \) for some \( x^* \in X \), i.e., \( |f(x^*)| < +\infty \). ●

**Assumption 1.2 (Constraint qualification).** The convex program (3.1) satisfies the Slater constraint qualification certificate (cf. Definition 3.4 in Ch. 2). ●

**Assumption 1.3 (Compatibility with network topology).** The constraint \( g_k(x) \leq 0 \) is compatible with \( \mathcal{G} \) if \( g_k \) can be expressed as a function of some components of the network state \( x \in \mathbb{R}^n \), say \( x_{S_k} \in \mathbb{R}^{S_k} \), and \( S_k \subset \mathcal{V} \) is a complete undirected subgraph of \( \mathcal{G} \). A similar definition can be stated for \( h_\ell \). ●
3.2 Lagrangian Saddle-Point Characterization

The central notions in this section are Lagrangian duality theory and the saddle-point optimality condition. Indeed, the *primal problem* (3.1) can be related to the so-called *Lagrangian dual problem*. Note that under Assumption 1.1 and 1.2, the optimal values of both the primal and dual (convex) programs coincide, i.e., strong duality holds. This is an immediate consequence of the Slater constraint qualification certificate (cf. Definition 3.4 in Ch. 2). In particular, the concept of strong duality facilitates an alternative approach to characterize primal-dual solutions of (3.1) in a more symmetric way, i.e., via saddle points.

**Definition 2.1** (Lagrangian function). The Lagrangian function associated with the convex program (3.1) is the function $L: \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m_{\geq 0} \to \mathbb{R}$ defined by

$$L(x, \mu, \nu) = \sum_{i=1}^{n} f_i(x_i) + \sum_{\ell=1}^{p} \mu_{\ell} h_{\ell}(x) + \sum_{k=1}^{m} \nu_k g_k(x),$$

(3.2)

or, in compact form

$$L(x, \mu, \nu) = f(x) + \langle \mu, h(x) \rangle + \langle \nu, g(x) \rangle,$$

where $\mu = (\mu_1, \ldots, \mu_p) \in \mathbb{R}^p$ and $\nu = (\nu_1, \ldots, \nu_m) \in \mathbb{R}^m_{\geq 0}$ are Lagrange multipliers.

Note that the primal problem (3.1) can be rewritten in terms of the Lagrangian function (3.2) as

$$\inf_{x \in \mathbb{R}^n} \sup_{(\mu, \nu) \in \mathbb{R}^p \times \mathbb{R}^m_{\geq 0}} L(x, \mu, \nu).$$

**Definition 2.2** (Lagrangian dual function). The Lagrangian dual function associated with (3.1) is the function $\phi: \mathbb{R}^p \times \mathbb{R}^m_{\geq 0} \to \mathbb{R} \cup \{-\infty, +\infty\}$ defined by

$$\phi(\mu, \nu) = \inf_{x \in \mathbb{R}^n} L(x, \mu, \nu) = \inf_{x \in \mathbb{R}^n} (f(x) + \langle \mu, h(x) \rangle + \langle \nu, g(x) \rangle).$$

Since the Lagrangian dual function is the pointwise infimum of a family of affine functions in $(\mu, \nu)$, it is always concave. The Lagrangian dual problem associated with the primal problem (3.1) is defined by

$$\maximize_{(\mu, \nu) \in \mathbb{R}^p \times \mathbb{R}^m} \phi(\mu, \nu)$$

subject to

$$-A^\top \mu \in \partial f(x) + \sum_{k \in K(x)} \nu_k \partial g_k(x),$$

$$\nu \succeq 0.$$  

(3.3)

Let the pair $(\mu^*, \nu^*)$ be an optimal solution of the dual problem (3.3) and let the set of solutions be denoted by $M \times N \subset \mathbb{R}^p \times \mathbb{R}^m_{\geq 0}$. Note that the dual problem (3.3) can be equivalently stated as

$$\sup_{(\mu, \nu) \in \mathbb{R}^p \times \mathbb{R}^m_{\geq 0}} \inf_{x \in \mathbb{R}^n} L(x, \mu, \nu).$$
Since Slater’s constraint qualification certificate (cf. Definition 3.4 in Ch. 2) is satisfied by assumption, strong duality holds, i.e., the strong min-max property

\[ \inf_{x \in \mathbb{R}^n} \sup_{(\mu, \nu) \in \mathbb{R}^p \times \mathbb{R}^m_{\geq 0}} \mathcal{L}(x, \mu, \nu) = \sup_{(\mu, \nu) \in \mathbb{R}^p \times \mathbb{R}^m_{\geq 0}} \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \mu, \nu) \]

is achieved. This motivates the notion of a saddle point associated with the Lagrangian function (3.2).

**Definition 2.3** (Saddle point). The triplet \((x^*, \mu^*, \nu^*) \in X \times M \times N\) is said to be a saddle point of \(\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m_{\geq 0} \to \mathbb{R}\) if

\[ \mathcal{L}(x^*, \mu, \nu) \leq \mathcal{L}(x^*, \mu^*, \nu^*) \leq \mathcal{L}(x, \mu^*, \nu^*), \quad \forall x \in \mathbb{R}^n, \forall \mu \in \mathbb{R}^p, \forall \nu \in \mathbb{R}^m_{\geq 0}. \quad \bullet \]

The following result (cf. [BNO03]) provides a powerful characterization of primal-dual solutions in terms of saddle points of the Lagrangian function (3.2), provided that strong duality holds, i.e., \(f(x^*) = \phi(\mu^*, \nu^*)\), where \(x^* \in X\) and \((\mu^*, \nu^*) \in M \times N\).

**Proposition 2.4** (Lagrangian saddle-point optimality condition). A point \(x^* \in X\) and the tuple \((\mu^*, \nu^*) \in M \times N\) are primal-dual solutions of (3.1) and (3.3) if and only if \((x^*, \mu^*, \nu^*)\) is a saddle point of the Lagrangian function (3.2).

The following proposition establishes a relationship between primal-dual solutions of (3.1) and (3.3), and the saddle points of an augmented Lagrangian function.

**Proposition 2.5** (Primal-dual solutions via saddle-point characterization). Given \(\kappa \in \mathbb{R}_{\geq 0}\) and \(\rho \in \mathbb{R}_{>0}\), let the augmented Lagrangian \(\mathcal{L}^\kappa : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}\) be defined by

\[ \mathcal{L}^\kappa(x, \mu) = f(x) + \frac{\rho}{2} \|Ax - b\|^2 + \langle \mu, Ax - b \rangle + \kappa \langle 1_m, [g(x)]^+ \rangle. \quad (3.4) \]

Then, the mapping \((x, \mu) \mapsto \mathcal{L}^\kappa(x, \mu)\) is convex-concave, and

(i) if \(x^* \in X\) and \((\mu^*, \nu^*) \in M \times N\) are primal-dual solutions of (3.1) and (3.3), then \((x^*, \mu^*) \in X \times M\) is a saddle point of \(\mathcal{L}^\kappa\) for any \(\kappa \geq \|\nu^*\|_\infty\),

(ii) if \((\tilde{x}, \tilde{\mu}) \in \mathbb{R}^n \times \mathbb{R}^p\) is a saddle point of \(\mathcal{L}^\kappa\) with \(\kappa > \|\nu^*\|_\infty\) for some \(\nu^* \in N\), then \(\tilde{x} \in \mathbb{R}^n\) is a primal solution of (3.1).

**Proof.** Since \(f, g\) are convex, and \(h\) is affine, the proposed augmented Lagrangian \(\mathcal{L}^\kappa\) in (3.4) is strictly convex in its primal variable \(x\) (because of the regularization term) and concave (in fact, linear) in its co-variable \(\mu\).

(i) Let \(x^* \in X\) and \((\mu^*, \nu^*) \in M \times N\) be primal-dual solutions of (3.1) and (3.3), respectively. By strong duality and the bound \(\kappa \geq \|\nu^*\|_\infty\), for any \(x \in \mathbb{R}^n\), it follows

\[ \mathcal{L}^\kappa(x, \mu^*) = f(x) + \frac{\rho}{2} \|Ax - b\|^2 + \langle \mu^*, Ax - b \rangle + \kappa \langle 1_m, [g(x)]^+ \rangle \]

\[ \geq f(x) + \langle \mu^*, Ax - b \rangle + \kappa \langle 1_m, [g(x)]^+ \rangle \]

\[ \geq f(x) + \langle \mu^*, Ax - b \rangle + \langle \nu^*, [g(x)]^+ \rangle \]

\[ \geq \inf_{x \in \mathbb{R}^n} (f(x) + \langle \mu^*, Ax - b \rangle + \langle \nu^*, g(x) \rangle) \]

\[ = \phi(\mu^*, \nu^*) = \mathcal{L}(x^*, \mu^*, \nu^*) = f(x^*) = \mathcal{L}^\kappa(x^*, \mu^*). \]
Due to the linearity of the augmented Lagrangian function $L^\kappa$ in $\mu$, the fact that $L^\kappa(x^*, \mu) = L^\kappa(x^*, \mu^*)$ for any $\mu \in \mathbb{R}^p$ is immediate.

(ii) Assume, to the contrary, that $(\tilde{x}, \tilde{\mu}) \in \mathbb{R}^n \times \mathbb{R}^p$ is a saddle point of (3.4) with $\kappa > \|\nu^*\|_\infty$ for some $\nu^* \in N$, but $\tilde{x} \in \mathbb{R}^n$ is not a primal solution. Let $x^* \in X$. For a fixed $x$, the mapping $\mu \mapsto L^\kappa(x, \mu)$ is concave and differentiable. A necessary condition for $(\tilde{x}, \tilde{\mu})$ to be a saddle point of $L^\kappa$ is that $A\tilde{x} - b = 0$. This fact yields $L^\kappa(x^*, \tilde{\mu}) \geq L^\kappa(\tilde{x}, \tilde{\mu})$, and hence

$$f(x^*) \geq f(\tilde{x}) + \kappa\langle 1_m, [g(\tilde{x})]^+ \rangle. \quad (3.5)$$

If $g(\tilde{x}) \preceq 0$, then $f(x^*) \geq f(\tilde{x})$, and $\tilde{x} \in \mathbb{R}^n$ would be a primal solution of (3.1). However, if $g(\tilde{x}) \not\preceq 0$, it follows

$$f(\tilde{x}) = f(\tilde{x}) + \langle \mu^*, A\tilde{x} - b \rangle - \langle \mu^*, A\tilde{x} - b \rangle + \langle \nu^*, g(\tilde{x}) \rangle - \langle \nu^*, g(\tilde{x}) \rangle \geq \inf_{\tilde{x} \in \mathbb{R}^n} (f(\tilde{x}) + \langle \nu^*, A\tilde{x} - b \rangle + \langle \nu^*, g(\tilde{x}) \rangle) - \langle \nu^*, A\tilde{x} - b \rangle - \langle \nu^*, g(\tilde{x}) \rangle = \phi(\mu^*, \nu^*) - \langle \nu^*, g(\tilde{x}) \rangle = f(x^*) - \langle \nu^*, g(\tilde{x}) \rangle > f(x^*) - \kappa\langle 1_m, [g(\tilde{x})]^+ \rangle,$$

which contradicts (3.5), concluding the proof.

Proposition 2.5 motivates to search for saddle points of the augmented Lagrangian function $L^\kappa$ rather than directly solving the constrained nonsmooth convex optimization problem (3.1). Since the augmented Lagrangian function $L^\kappa$ is convex-concave, a natural approach to find the saddle points is via its associated saddle-point dynamics. However, for an arbitrary Lagrangian function, such dynamics are known to render saddle points only stable, but not asymptotically stable. In fact, saddle-point dynamics derived using the standard Lagrangian function $L$ in (3.2) do not converge to solutions of the primal problem (3.1), see e.g., [AHU58, DSS58]. Interestingly, the convergence properties can be improved using regularization terms associated with the equality constraints of the nonsmooth convex optimization problem at hand.

Remark 2.6 (Structure of the augmented Lagrangian function). The standard Lagrangian function $L$ in (3.2) is augmented by an $\ell_1$-exact-penalty function in order to eliminate the inequality constraints of the nonsmooth convex program (3.1). Note that if the bound on $\kappa$ in Proposition 2.5 holds strictly, then exact equivalence between the saddle points of $L^\kappa$ and the primal-dual solutions of (3.1) and (3.3) is guaranteed (cf. Lemma 3.7 in Ch. 2). Moreover, a quadratic regularization term is added to $L^\kappa$ to improve the convergence properties of the underlying saddle-point dynamics. This results in the nonlinear and $\ell_1$-penalized nonsmooth constrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) + \frac{\rho}{2} \|Ax - b\|^2 + \kappa\langle 1_m, [g(x)]^+ \rangle \quad (3.6)$$

subject to $Ax = b$.

whose standard Lagrangian function is equivalent to $L^\kappa$ in (3.4).

Remark 2.7 (Bound on the parameter $\kappa$). Note that the lower bound on $\kappa$ in Proposition 2.5 is characterized by the dual solution $\nu^* \in N$ of (3.3) which is unknown a
priori. However, our forthcoming discussion proposes fully distributed dynamics that do not incorporate knowledge on the parameter $\kappa$.

### 3.3 Distributed Coordination Algorithm

This section presents a continuous-time coordination algorithm that asymptotically converges to the set of primal-dual solutions $X \times M$ of the nonsmooth convex program (3.6), given knowledge of a suitable lower bound on the parameter $\kappa$ (cf. Remark 2.7). The presentation proceeds by characterizing the stability properties of the algorithm and observing its limitations, leading up to the main contribution which is the introduction of saddle-point dynamics amenable for distributed implementation over a network of agents.

#### 3.3.1 Saddle-Point Dynamics

Based on the result in Proposition 2.5, consider the saddle-point dynamics (gradient descent in the primal variable $x$ and gradient ascent in the dual variable $\mu$) of the augmented Lagrangian function $L^\kappa$ in (3.4). The nonsmooth character of $L^\kappa : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ suggests that its associated saddle-point dynamics take the form of a differential inclusion,

\[
\begin{align*}
\dot{x}(t) &\in -\partial_x L^\kappa(x(t), \mu(t)), & x(0) = x_0 \in \mathbb{R}^n, \\
\dot{\mu}(t) &\in +\partial_\mu L^\kappa(x(t), \mu(t)), & \mu(0) = \mu_0 \in \mathbb{R}^p,
\end{align*}
\]

for a.a. $t \in [t_0, +\infty)$. The partial derivatives of $L^\kappa$ in (3.4) are, respectively,

\[
\begin{align*}
\partial_x L^\kappa(x, \mu) &\ni A^\top \left( \rho(Ax - b) + \mu \right) + \pi_f + \pi_g^+, \\
\partial_\mu L^\kappa(x, \mu) &\ni Ax - b,
\end{align*}
\]

where $\pi_f \in \partial f(x)$ and $\pi_g^+ \in \kappa \sum_{k \in K(x)} \partial [g_k(x)]^+$. In particular, the set-valued saddle-point dynamics (3.7) defined over $\mathbb{R}^n \times \mathbb{R}^p$ take the form

\[
\begin{align*}
\dot{x}(t) + A^\top \left( \rho(Ax(t) - b) + \mu(t) \right) &\in -\partial f(x(t)) \\
\dot{\mu}(t) &= \rho(Ax(t) - b),
\end{align*}
\]

for a.a. $t \in [t_0, +\infty)$ with initial condition $(x_0, \mu_0) \in \mathbb{R}^n \times \mathbb{R}^p$. Proposition 1.7 in Chapter 2 and Lemma 2.2 in Chapter 2 guarantee the existence of solutions $(x, \mu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p$ of the dynamics (3.8), where the solutions are understood in the sense of Krasovskii (cf. Definition 2.6 in Ch. 2). For notational convenience, we use the set-valued map $F^b : \mathbb{R}^n \times \mathbb{R}^p \Rightarrow \mathbb{R}^n \times \mathbb{R}^p$ to refer to the saddle-point dynamics (3.8). Note that if $(x^*, \mu^*) \in \text{eq}(F^b)$, then $x^* \in X$ is a primal solution of the nonsmooth convex program (3.1) (cf. Proposition 2.5).
3.3. DISTRIBUTED COORDINATION ALGORITHM

3.3.2 Convergence Analysis

The following result characterizes asymptotic convergence properties of the saddle-point dynamics (3.8) synthesized for solving nonsmooth convex optimization problems using classical notions of stability analysis. In particular, Lyapunov theory and the LaSalle Invariance Principle for differential inclusions are used to establish that the primal-dual optimizer are globally asymptotically stable under the saddle-point dynamics (3.8) and that each solution of the dynamics converges to a point in the set of primal-dual solutions $X \times M$.

**Theorem 3.1** (Point-wise convergence). Let $\kappa \in \mathbb{R}_{\geq 0}$, $\rho \in \mathbb{R}_{> 0}$, $x^* \in X$, and $(\mu^*, \nu^*) \in M \times N$. Define the mapping $d : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}_{\geq 0}$ by

$$d(x, \mu) = \frac{1}{2} \|x - x^*\|^2 + \frac{1}{2} \|\mu - \mu^*\|^2. \quad (3.9)$$

If $\kappa > \|\nu^*\|_\infty$, then $(\mathcal{L}_F^\kappa d)(x, \mu) \subset (-\infty, 0]$ holds for all $(x, \mu) \in \mathbb{R}^n \times \mathbb{R}^p$ and any solution $(x, \mu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p$ of (3.8) converges asymptotically to a point in the set of primal-dual solutions $X \times M$.

**Proof.** Note that the set-valued map $F^\kappa$ is locally bounded, upper semi-continuous and takes nonempty, convex, and compact values (cf. Proposition 1.7 in Ch. 2). Moreover, by Proposition 2.5(i), the tuple $(x^*, \mu^*)$ identifies a saddle point of (3.4) when $\kappa \geq \|\nu^*\|_\infty$. Note that $d \in C^1(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R}_{\geq 0})$. In addition, the candidate Lyapunov function (3.9) is proper. Let $\psi \in (\mathcal{L}_F^\kappa d)(x, \mu)$. By definition of the Lie-derivative (cf. Definition 2.3 in Ch. 2), there exists

$$\xi = \left(-A^\top \left(\rho(Ax - b) + \mu\right) - \pi_f - \pi_g^+\right) \in F^\kappa,$$

where $\pi_f \in \partial f(x)$ and $\pi_g^+ \in \kappa \sum_{k \in K(x)} \partial [g_k(x)]^+$, such that

$$\psi = \langle \nabla d(x, \mu), \xi \rangle = \langle x - x^*, -A^\top \left(\rho(Ax - b) + \mu\right) - \pi_f - \pi_g^+ \rangle$$

$$+ \langle \mu - \mu^*, Ax - b \rangle. \quad (3.10)$$

Note that $\mathcal{L}^\kappa$ is convex in its primal variable $x$, and

$$A^\top \left(\rho(Ax - b) + \mu\right) + \pi_f + \pi_g^+ \in \partial_x \mathcal{L}^\kappa(x, \mu).$$

By the 1st-order convexity condition in $x$, it follows

$$\mathcal{L}^\kappa(x^*, \mu) \geq \mathcal{L}^\kappa(x, \mu) + \langle x - x^*, -A^\top \left(\rho(Ax - b) + \mu\right) - \pi_f - \pi_g^+ \rangle. \quad (3.11)$$

Similarly, since $\mathcal{L}^\kappa$ is concave (in fact, linear) in its dual variable $\mu$, and

$$Ax - b \in \partial_\mu \mathcal{L}^\kappa(x, \mu),$$

the 1st-order convexity condition in $\mu$ yields

$$\mathcal{L}^\kappa(x, \mu) = \mathcal{L}^\kappa(x, \mu^*) + \langle \mu - \mu^*, Ax - b \rangle. \quad (3.12)$$
Substituting (3.11) and (3.12) into equation (3.10) yields
\[
\psi \leq \mathcal{L}^\kappa(x^*, \mu) - \mathcal{L}^\kappa(x^*, \mu^*) + \mathcal{L}^\kappa(x^*, \mu^*) - \mathcal{L}^\kappa(x, \mu^*) \leq 0,
\] (3.13)
since \((x^*, \mu^*)\) is a saddle point of \(\mathcal{L}^\kappa\). Since \(\psi\) is chosen arbitrary, the inclusion \((\mathcal{L}_{F^p} d)(x, \mu) \subseteq (-\infty, 0)\) holds for all \((x, \mu) \in \mathbb{R}^n \times \mathbb{R}^p\). Hence, by Theorem 2.4 in Chapter 2, the point \((x^*, \mu^*) \in \text{eq}(F^p)\) is a strongly stable equilibrium point of (3.8).

In following proof strategy is based on verifying the hypothesis of LaSalle’s Invariance Principle (cf. Theorem 2.5 in Ch.2), and identifying the set of primal-dual solutions \(X \times M\) as the corresponding largest weakly invariant set. For any \(\gamma \in \mathbb{R}_{>0}\), the \(-\gamma\)-sublevel set \(d^{-1}(\leq \gamma)\) is strongly invariant with respect to (3.8). Since the mapping \(d : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}_{\geq 0}\) is proper, the \(-\gamma\)-sublevel set \(d^{-1}(\leq \gamma)\) is compact. By Theorem 2.5 in Chapter 2, any absolutely continuous trajectory \((x, \mu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p\) of (3.8) starting in \(d^{-1}(\leq \gamma)\) converges to the largest weakly invariant set
\[
\Omega \subset \text{cl}\{(x, \mu) \in \mathbb{R}^n \times \mathbb{R}^p \mid 0 \in (\mathcal{L}_{F^p} d)(x, \mu)\} \cap d^{-1}(\leq \gamma).
\]
Since the set-valued map \(F^p\) is upper semi-continuous and \(d \in C^1(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R}_{\geq 0})\), it follows that the set-valued map \((x, \mu) \mapsto (\mathcal{L}_{F^p} d)(x, \mu)\) is also upper semi-continuous. Hence, closedness of the set
\[
\text{cl}\{(x, \mu) \in \mathbb{R}^n \times \mathbb{R}^p \mid 0 \in (\mathcal{L}_{F^p} d)(x, \mu)\} \cap d^{-1}(\leq \gamma)
\]
is an immediate consequence (cf. [Cor08]).

To show that \(\Omega \subset X \times M\), take a point \((\bar{x}, \bar{\mu}) \in \Omega\). Then, from inequality (3.13), it follows that \(\mathcal{L}^\kappa(x^*, \mu^*) - \mathcal{L}^\kappa(\bar{x}, \mu^*) = 0\), which implies
\[
\tilde{\mathcal{L}}^\kappa(\bar{x}, \mu^*) - \frac{\rho}{2}\|A\bar{x} - b\|^2 = 0,
\] (3.14)
where the mapping \(\tilde{\mathcal{L}}^\kappa : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}\) is defined by
\[
\tilde{\mathcal{L}}^\kappa(\bar{x}, \mu^*) = f(x^*) - f(\bar{x}) - \langle \mu^*, A\bar{x} - b \rangle - \kappa(\mathbb{1}_m, [g(\bar{x})]^+).
\]
By Assumption 1.2, strong duality holds and we conclude that
\[
\tilde{\mathcal{L}}^\kappa(\bar{x}, \mu^*) = \phi(\mu^*, \nu^*) - f(\bar{x}) - \langle \mu^*, A\bar{x} - b \rangle - \kappa(\mathbb{1}_m, [g(\bar{x})]^+)
= \inf_{\tilde{x} \in \mathbb{R}^n} \left( f(\tilde{x}) + \langle \mu^*, A\tilde{x} - b \rangle + \langle \nu^*, g(\tilde{x}) \rangle \right)
- f(\bar{x}) - \langle \mu^*, A\bar{x} - b \rangle - \kappa(\mathbb{1}_m, [g(\bar{x})]^+)
\leq \langle \nu^*, g(\bar{x}) \rangle - \kappa(\mathbb{1}_m, [g(\bar{x})]^+).
\]
Note that if \(\kappa \geq \|\nu^*\|_{\infty}\), then \(\tilde{\mathcal{L}}^\kappa(\bar{x}, \mu^*) \leq 0\), which implies \(A\bar{x} - b = 0\). However, if \(\kappa > \|\nu^*\|_{\infty}\), then it follows
\[
\tilde{\mathcal{L}}^\kappa(\bar{x}, \mu^*) < \langle \nu^*, g(\bar{x}) \rangle - \langle \nu^*, [g(\bar{x})]^+ \rangle.
\]
For \(g(\bar{x}) \neq 0\), it follows \(\tilde{\mathcal{L}}^\kappa(\bar{x}, \mu^*) < 0\) which contradicts (3.14). Hence, it must be that \(g(\bar{x}) \leq 0\). From (3.14), we conclude \(\tilde{\mathcal{L}}^\kappa(\bar{x}, \mu^*) = f(x^*) - f(\bar{x}) = 0\). Thus, if \((\bar{x}, \bar{\mu}) \in \Omega\), then all primal feasibility conditions are satisfied and \(\bar{x}\) is a solution of (3.1).
3.3. DISTRIBUTED COORDINATION ALGORITHM

Since Ω is weakly invariant, there exists an absolutely continuous trajectory starting from (\(\bar{x}, \bar{\mu}\)) that remains in Ω only if \(0 \in \partial_{\bar{x}} \mathcal{L}^n(\bar{x}, \bar{\mu})\), and \(0 \in \partial_{\bar{\mu}} \mathcal{L}^n(\bar{x}, \bar{\mu})\), i.e.,

\[
\begin{align*}
-A^\top (\rho(A\bar{x}(t) - b) + \bar{\mu}(t)) & \in \partial f(\bar{x}(t)) \\
+ \kappa \sum_{k \in K(\bar{x})} \partial [g_k(\bar{x}(t))]^+ & = 0 = \rho(A\bar{x}(t) - b). 
\end{align*}
\]

(3.15a) (3.15b)

In what follows, we show that all dual feasibility conditions of (3.3) are satisfied. Since strong duality holds (cf. Assumption 1.2), we have

\[
\begin{align*}
f(x^*) = \inf_{\bar{x} \in \mathbb{R}^n} \left( f(\bar{x}) + \langle \bar{\mu}, A\bar{x} - b \rangle + \langle \bar{\nu}, g(\bar{x}) \rangle \right) \\
& \leq f(\bar{x}) + \langle \bar{\mu}, A\bar{x} - b \rangle + \langle \bar{\nu}, g(\bar{x}) \rangle \\
& \leq f(\bar{x}).
\end{align*}
\]

Note that the conditions \(A\bar{x} - b = 0\) and \(g(\bar{x}) \leq 0\) imply that \(\bar{\nu} \in \mathbb{R}^n_{\geq 0}\). Since \(f(x^*) = f(\bar{x})\), we conclude that \(\langle \bar{\nu}, g(\bar{x}) \rangle = 0\). Hence, for \(\bar{\nu} = 0\), we have \(g(\bar{x}) < 0\) and

\[
\kappa \sum_{k \in K(\bar{x})} \partial [g_k(\bar{x})]^+ = \{0\}.
\]

Since each \(g_k\) is regular at \(\bar{x} \in \mathbb{R}^n\) for \(k \in K(\bar{x})\), by Proposition 1.4(i)-(ii) in Chapter 2, inclusion (3.15a) implies

\[
A^\top \bar{\mu} \in -\partial f(\bar{x}).
\]

(3.16)

However, if \(\bar{\nu} \succ 0\), then \(g(\bar{x}) = 0\) and, by Proposition 1.4(iii) in Chapter 2, inclusion (3.15a) yields

\[
A^\top \bar{\mu} \in -\partial f(\bar{x}) - \kappa \sum_{k \in K(\bar{x})} \co (\{0\} \cup \partial g_k(\bar{x})).
\]

(3.17)

By definition of the normal cone \(N_G(\bar{x})\) (cf. Proposition 3.5 in Ch. 2), we have

\[
\kappa \sum_{k \in K(\bar{x})} \co (\{0\} \cup \partial g_k(\bar{x})) \subset \left\{ \sum_{k \in K(\bar{x})} \nu_k \pi_{g_k} \in \mathbb{R}^n \mid \pi_{g_k} \in \partial g_k(\bar{x}), \nu_k \geq 0, k \in K(\bar{x}) \right\}.
\]

Comparing (3.3) with (3.16)-(3.17), we conclude that all dual feasibility conditions are satisfied. Therefore, given \(\nu^* \in N\), the point \((\bar{x}, \bar{\mu}) \in \Omega\) satisfies the KKT-conditions (cf. Theorem 3.6 in Ch. 2) and hence, \(\Omega \subset X \times M\). Since the initial choice \(\gamma \in \mathbb{R}_{>0}\) is arbitrary, we deduce that convergence of solutions \((x, \mu) : [t_0, +\infty) \rightarrow \mathbb{R}^n \times \mathbb{R}^p\) of (3.8) to the set \(X \times M\) holds from any point in \(\mathbb{R}^n \times \mathbb{R}^p\).

Finally, we show asymptotic convergence of solutions \((x, \mu) : [t_0, +\infty) \rightarrow \mathbb{R}^n \times \mathbb{R}^p\) of (3.8) to a point in \(X \times M\). This is established by showing that the \(\omega\)-limit set

\[
\omega(x, \mu) = \{(y_x, y_\mu) \in d^{-1}(\leq \gamma) \mid \liminf_{t \to \infty} \| (x(t), \mu(t)) - (y_x, y_\mu) \| = 0 \} \subset X \times M
\]

of any trajectory \(t \mapsto (x(t), \mu(t))\) of (3.8) is a singleton set. Following the above
arguments, note that the $\omega$-limit set is nonempty and weakly invariant [Cor08]. By contradiction, assume that $(y_x, y_\mu), (z_x, z_\mu) \in \omega(x, \mu)$ with $(y_x, y_\mu) \neq (z_x, z_\mu)$, and define the mappings $d_1, d_2 : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}_{\geq 0}$ by

$$
d_1(x, \mu) = \frac{1}{2}\|x - y_x\|^2 + \frac{1}{2}\|\mu - y_\mu\|^2, \quad d_2(x, \mu) = \frac{1}{2}\|x - z_x\|^2 + \frac{1}{2}\|\mu - z_\mu\|^2.
$$

Since $(y_x, y_\mu), (z_x, z_\mu) \in X \times M$, the above discussion implies that the $\gamma$-sublevel sets $d_1^{-1}(\leq \gamma)$ and $d_2^{-1}(\leq \gamma)$ are strongly invariant under $F^b$, for any $\gamma > 0$. Pick $\gamma < \frac{1}{2}\|(y_x, y_\mu) - (z_x, z_\mu)\|$. Since $(y_x, y_\mu) \in \omega(x, \mu)$, the solution $(x, \mu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p$ enters $d_1^{-1}(\leq \gamma)$ at some finite time $t_1 \in [t_0, +\infty)$ and remains there afterwards because of the strong invariance of the $\gamma$-sublevel set. Similarly, for $(z_x, z_\mu) \in \omega(x, \mu)$, there exists a finite time $t_2 \in [t_0, +\infty)$ such that $(x(t), \mu(t)) \in d_2^{-1}(\leq \gamma)$ for all $t \geq t_2$. Therefore, for $t \geq \max\{t_1, t_2\}$, the solution belongs to $d_1^{-1}(\leq \gamma) \cap d_2^{-1}(\leq \gamma) = \emptyset$, which is a contradiction.

**Remark 3.2 (On the parameter $\kappa$).** The bound on $\kappa$ depends on the dual solution set $N$ as well as on the initial condition, since the result is only valid when the saddle-point dynamics start within the $\gamma$-sublevel set $d^{-1}(\leq \gamma)$. However, in our forthcoming discussion we introduce discontinuous saddle-point-like dynamics that do not rely on knowledge of the parameter $\kappa$ a priori.

### 3.4 Projected Saddle-Point-Like Dynamics

This section proposes discontinuous saddle-point-like dynamics that involve set-valued projection operations but do not incorporate knowledge of the dual solution $\nu^* \in N$ beforehand, as in Theorem 3.1. However, the dynamics proposed here enjoy the same convergence properties to a point in the set of primal-dual solutions $X \times M$ and are amenable to fully distributed implementation.

Recall that $G \subset \mathbb{R}^n$ denotes the viability set associated with the nonsmooth convex program (3.1). Let the set-valued flow $F : G \times \mathbb{R}^p \rightrightarrows \mathbb{R}^n$ be defined by

$$
F(x, \mu) = -A^\top(\rho(Ax - b) + \mu) - \partial f(x). \quad (3.18)
$$

The definition in (3.18) is motivated by the fact that, for $(x, \mu) \in \text{int}(G) \times \mathbb{R}^p$, it follows $\partial_x \mathcal{L}^\kappa = -F(x, \mu)$. Consider the saddle-point-like dynamics defined over $G \times \mathbb{R}^p$,

$$
\dot{x}(t) \in \begin{cases} 
F(x(t), \mu(t)), & \text{if } x \in \text{int}(G), \\
\Pi_{\mathcal{T}_G}(x(t), F(x(t), \mu(t))), & \text{if } x \in \text{bd}(G), 
\end{cases} \quad (3.19a)
$$

$$
\dot{\mu}(t) = \rho(Ax(t) - b), \quad (3.19b)
$$

for a.a. $t \in [t_0, +\infty)$ with initial condition $(x_0, \mu_0) \in G \times \mathbb{R}^p$. Note that the set $\Pi_{\mathcal{T}_G}(x, F(x, \mu))$ is not necessarily convex and that the set-valued map $(x, \mu) \mapsto $
\[ \Pi_{T_G}(x, F(x, \mu)) \] has no continuity properties. However, the existence of viable solutions \((x, \mu) : [t_0, +\infty) \to G \times \mathbb{R}^p\) (cf. Definition 2.7 in Ch. 2) of (3.19) is guaranteed (cf. Theorem 2.9 in Ch. 2). For notational convenience, we use the set-valued map \(F^\sharp : G \times \mathbb{R}^p \rightrightarrows G \times \mathbb{R}^p\) to refer to the saddle-point-like dynamics (3.19).

### 3.4.1 Set-valued Projection Operator

The objective here is to provide an alternative expression of the set-valued projection operator \(\Pi_{T_G}\) in (3.19a) that more clearly displays its amenability to distributed implementation over a network of agents.

**Definition 4.1 (Set of outward normals).** The set of outward normals to \(G\) at \(x \in \text{bd}(G)\) is defined by

\[
N^\sharp_G(x) = \left\{ \pi_g \in \sum_{k \in K(x)} \partial g_k(x) \Bigm\| \pi_g \Big\| = 1 \right\} \subset N_G(x),
\]

where \(N_G(x)\) denotes the normal cone of \(G\) at \(x \in G\) (cf. Proposition 3.5 in Ch. 2).

Following the idea of singleton-valued projections [NZ96], the set-valued projection operator (3.19a) for \((x, \mu) \in G \times \mathbb{R}^p\) can be expressed as

\[
\Pi_{T_G}(x, F(x, \mu)) = \bigcup_{\xi \in F(x, \mu)} \xi - \max \left\{0, \langle \xi, \pi_g^*(\xi) \rangle \right\} \pi_g^*(\xi), \tag{3.20}
\]

where \(\pi_g^*(\xi)\) is the unique maximizer determined by the (sub-)optimization problem

\[
\text{maximize}_{\pi_g \in \mathbb{R}^n} \quad \langle \xi, \pi_g \rangle \\
\text{subject to} \quad \pi_g \in N_G^\sharp(x). \tag{3.21}
\]

Note that whenever \((x, \mu) \in \text{int}(G) \times \mathbb{R}^p\), it follows \(\langle \xi, \pi_g^*(\xi) \rangle \leq 0\) for all \(\xi \in F(x, \mu)\) and therefore, \(\Pi_{T_G}(x, F(x, \mu)) = F(x, \mu)\). The following result states existence and uniqueness of the maximizer \(\pi_g^*(\xi)\) in (3.21).

**Lemma 4.2 (Existence and uniqueness).** Given \((x, \mu) \in \text{bd}(G) \times \mathbb{R}^p\). If there exists an \(\xi \in F(x, \mu)\) such that \(\sup_{\pi_g \in N_G^\sharp(x)} \langle \xi, \pi_g \rangle > 0\), then the maximizer \(\pi_g^*(\xi)\) of (3.21) exists and is unique.

**Proof.** Let \((x, \mu) \in \text{bd}(G) \times \mathbb{R}^p\) and let \(\xi \in F(x, \mu)\). Note that the set \(N_G^\sharp(x)\) is closed and convex (cf. Definition 4.1). Existence of \(\pi_g^*(\xi)\) follows from compactness of \(N_G^\sharp(x)\). Let \(\pi_{g_1}^*(\xi)\) and \(\pi_{g_2}^*(\xi)\) be two distinct maximizer of (3.21) such that \(\langle \xi, \pi_{g_i}^*(\xi) \rangle > 0\) for any \(i \in \{1, 2\}\). Then, by convexity it follows

\[
\frac{\pi_{g_1}^*(\xi) + \pi_{g_2}^*(\xi)}{\| \pi_{g_1}^*(\xi) + \pi_{g_2}^*(\xi) \|} \in N_G^\sharp(x).
\]
Hence, it holds that
\[
\frac{\langle \xi, \pi_{g_1}^*(\xi) + \pi_{g_2}^*(\xi) \rangle}{\|\pi_{g_1}^*(\xi) + \pi_{g_2}^*(\xi)\|} = \frac{2\langle \xi, \pi_{g_1}^*(\xi) \rangle}{\|\pi_{g_1}^*(\xi) + \pi_{g_2}^*(\xi)\|} > \langle \xi, \pi_{g_1}^*(\xi) \rangle,
\]
which contradicts the fact that \(\pi_{g_1}^*(\xi)\) maximizes (3.21). This concludes the proof. \(\blacksquare\)

Remark 4.3 (Relationship to slow solutions). Recall that a slow solution \((x, \mu) : [t_0, +\infty) \to G \times \mathbb{R}^p\) of the differential inclusion (3.19) is a solution of
\[
\begin{align*}
\dot{x}(t) &\in \begin{cases} 
\text{Ln} \left(F(x(t), \mu(t))\right), & \text{if } x \in \text{int}(G), \\
\text{Ln} \left(\Pi_{TG}(x(t), F(x(t), \mu(t)))\right), & \text{if } x \in \text{bd}(G),
\end{cases} \\
\dot{\mu}(t) &\in \rho(Ax(t) - b),
\end{align*}
(3.22)
\]
for a.a. \(t \in [t_0, +\infty)\) with initial condition \((x_0, \mu_0) \in G \times \mathbb{R}^p\) (cf. Definition 2.8 in Ch. 2). Note that if the hypotheses in Theorem 2.9 in Chapter 2 are satisfied, the existence of slow solutions of (3.19) (respectively of (3.22)) is guaranteed. Therefore, whenever \((x, \mu) \in \text{bd}(G) \times \mathbb{R}^p\), it suffices to study the map \(\text{Ln} \left(\Pi_{TG}(x, F(x, \mu))\right)\) to guarantee viability of the set \(G \subset \mathbb{R}^n\).

Remark 4.4 (Projection operator for continuously differentiable functions). If \(f, g \in C^1\), then the projection operator (3.20) reduces to the singleton-valued map
\[
\Pi_{TG}(x, F(x, \mu)) = F(x, \mu) - \max \left\{0, \langle F(x, \mu), \pi_g^*(x, \mu) \rangle\right\} \pi_g^*(x, \mu),
\]
where \(F(x, \mu) = -A^\top (\rho(Ax - b) + \mu) - \nabla f(x)\), and
\[
\pi_g^*(x, \mu) = \sum_{k \in K(x)} \nu_k^* \nabla g_k(x) \in N_G^e(x),
\]
with \(\nu^* \in \mathbb{R}^{m \geq 0}\) as the unique maximizer of the (sub-)optimization problem
\[
\begin{align*}
\text{maximize } & \langle F(x, \mu), \sum_{k \in K(x)} \nu_k \nabla g_k(x) \rangle \\
\text{subject to } & \left\| \sum_{k \in K(x)} \nu_k \nabla g_k(x) \right\| = 1, \\
& \nu \succeq 0.
\end{align*}
(3.23)
\]
Since the flow \(F\) is singleton valued, it follows \(\Pi_{TG}(x, F(x, \mu)) = \text{Ln} \left(\Pi_{TG}(x, F(x, \mu))\right)\) and therefore, any solution \((x, \mu) : [t_0, +\infty) \to G \times \mathbb{R}^p\) of (3.19) is also a slow solution that satisfies the viability condition (cf. Section 2.2) with respect to the set \(G \subset \mathbb{R}^n.\) \(\blacksquare\)

### 3.4.2 Stability Analysis

In the following, we establish a relationship between the solutions of the saddle-point dynamics \(F^s\) and the saddle-point-like dynamics \(F^s\) that allows us to conclude convergence properties of the differential inclusion (3.19).
Lemma 4.5 (Relationship of trajectories). Given \((x, \mu) \in G \times \mathbb{R}^p\). Let \(\kappa \in \mathbb{R}_{\geq 0}\) satisfy the inequality
\[
\kappa \geq \sup_{\xi \in F(x, \mu)} \max \left\{ 0, \langle \xi, \pi_g^*(\xi) \rangle \right\},
\]
where \(\xi \in F(x, \mu)\) and \(\pi_g^*(\xi) = \arg\max_{\pi_g \in N_{G(x)}^g(\xi, \pi_g)} \langle \xi, \pi_g \rangle\). Then, the inclusion \(F^g(x, \mu) \subset F^\flat(x, \mu)\) holds for all \((x, \mu) \in G \times \mathbb{R}^p\).

Proof. Note that if \((x, \mu) \in \text{int}(G) \times \mathbb{R}^p\), then the condition \(F^\flat(x, \mu) = F^\flat(x, \mu)\) holds trivially. Let \((x, \mu) \in \partial(G) \times \mathbb{R}^p\), and let \(\pi_g^*(\xi) = \arg\max_{\pi_g \in N_{G(x)}^g(\xi, \pi_g)} \langle \xi, \pi_g \rangle\). Since \(\max \left\{ 0, \langle \xi, \pi_g^*(\xi) \rangle \right\}\) provides a minimum distance measure from \(\xi \in F(x, \mu)\) to the tangent cone \(T_G(x)\) at \((x, \mu) \in \partial(G) \times \mathbb{R}^p\), taking the supremum of \(\xi\) over the convex and compact set \(F(x, \mu)\) guarantees that the inclusions
\[
\Pi_{T_G}(x, \mu) \subset F(x, \mu) - \kappa \sum_{k \in K(x)} \co \left( \{0\} \cup \partial g_k(x) \right) \subset F(x, \mu) - N_G(x)
\]
hold for all \((x, \mu) \in \partial(G) \times \mathbb{R}^p\). This concludes the proof.

Note that the inclusion in Lemma 4.5 may be strict and, in general, the set of trajectories of (3.8) is richer than the set of trajectories of (3.19). Building on the previous result, our following contribution characterizes convergence of solutions \((x, \mu) : [t_0, +\infty) \to G \times \mathbb{R}^p\) of the differential inclusion (3.19) to a point \((x^*, \mu^*)\) in the set of primal-dual minimizer \(X \times M\) of the nonsmooth convex program (3.1).

Theorem 4.6 (Point-wise convergence). Any solution \((x, \mu) : [t_0, +\infty) \to G \times \mathbb{R}^p\) of (3.19) starting from \((x_0, \mu_0) \in G \times \mathbb{R}^p\) converges asymptotically to a point in \(X \times M\).

Proof. Let \(\gamma = d(x_0, \mu_0)\), where \(d : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}_{\geq 0}\) is defined as in (3.9). If \(\kappa > \|\nu^*\|_{\infty}\), then the \(\gamma\)-sublevel set \(d^{-1}(\leq \gamma)\) is strongly invariant with respect to (3.8) and any solution \((x, \mu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p\) of \(F^\flat\) converges asymptotically to a point in the set \(X \times M\) (cf. Theorem 3.1). By definition of (3.20), the set \(G \times \mathbb{R}^p\) is strongly invariant with respect to (3.19). Let \(\kappa \in \mathbb{R}_{\geq 0}\) satisfy the inequality
\[
\kappa > \max \left\{ \sup_{\xi \in F(x, \mu)} \max \left\{ 0, \langle \xi, \pi_g^*(\xi) \rangle \right\}, \|\nu^*\|_{\infty} \right\},
\]
where \(\pi_g^*(\xi) = \arg\max_{\pi_g \in N_{G(x)}^g(\xi, \pi_g)} \langle \xi, \pi_g \rangle\). Then, by Lemma 4.5, the inclusion \(F^\flat(x, \mu) \subset F^\flat(x, \mu)\) holds for all \((x, \mu) \in G \times \mathbb{R}^p\). Therefore, any solution \((x, \mu) : [t_0, +\infty) \to G \times \mathbb{R}^p\) of (3.19) starting in \(G \times \mathbb{R}^p\) is also a solution of (3.8). Point-wise convergence of the solutions of (3.19) to a point in \(X \times M\) follows from Theorem 3.1.

Remark 4.7 (Comparison with existing dynamics). The work [FP10] builds on saddle-point dynamics of a smooth Lagrangian function to deal with non-strict convex programs with inequality constraints, where a projection operator is introduced to constrain the nonnegative dual variables associated with the inequality constraints rather...
than exact penalty functions. However, the inequality constraints may only be satisfied asymptotically. The authors establish asymptotic convergence in the primal variables under the assumption that Slater’s constraints qualification certificate is satisfied. However, the dual variables converge to some unknown point that might not represent a solution of the dual problem. This is to be in contrast with the convergence properties of the dynamics proposed in (3.19). In addition, our algorithm respects the viability set induced by the inequality constraints at any time.

•

3.5 Distributed Implementation

This section emphasizes that the discontinuous set-valued dynamics (3.19) are well-suited for fully distributed implementation over a network of agents. This is in contrast to other consensus-based works proposed in the literature [NOP10, ZM12, WE11, JKJJ08] that rely on state estimations over the whole network.

Consider the network model in Section 3.1 to see under what conditions the dynamics can be implemented by an agent using local information. For each agent $i \in \{1,\ldots,n\}$, the set-valued flow (3.18) can be component-wise written as

$$F_i(x, \mu) = -\sum_{\{\ell : a_{\ell i} \neq 0\}} a_{\ell i} \left( \rho \left( \sum_{\{j : a_{\ell j} \neq 0\}} a_{\ell j} x_j - b_{\ell} \right) + \mu_{\ell} \right) - \partial f_i(x_i),$$

where $a_{\ell i} \in \mathbb{R}$ refer to elements of the matrix $A \in \mathbb{R}^{p \times n}$, and the dynamics (3.19b) related to the dual variables $\mu \in \mathbb{R}^p$ for each $\ell \in \{1,\ldots,m\}$ take the form

$$\dot{\mu}_\ell = \rho \sum_{\{i : a_{\ell i} \neq 0\}} a_{\ell i} x_i - b_{\ell}.$$

Hence, in order for agent $i \in \{1,\ldots,n\}$ to be able to implement its corresponding dynamics (3.19a), it also needs access to certain dual components $\mu_{\ell} \in \mathbb{R}$ for which $a_{\ell i} \neq 0$, and therefore needs to implement the corresponding dynamics. By Assumption 1.3, the dynamics (3.19) are fully distributed over the network topology described by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ when the following conditions are satisfied:

- (C1) $\mathcal{G}$ is compatible with the equality constraints $Ax = b$ and the inequality constraints $g(x) \preceq 0$ of the nonsmooth convex program (3.1),
- (C2) for each $i \in \{1,\ldots,n\}$, agent $i$ knows its own state $x_i \in \mathbb{R}$ and its local cost function $f_i : \mathbb{R} \rightarrow \mathbb{R}$,
- (C3) each agent $i \in \{1,\ldots,n\}$ has access to its immediate neighbors decision variables $x_j \in \mathbb{R}$ specified by the network topology $\mathcal{G}$, and
  - (i) their local cost functions $f_j : \mathbb{R} \rightarrow \mathbb{R}$,
  - (ii) every $b_{\ell} \in \mathbb{R}$ for which $a_{\ell i} \neq 0$,
  - (iii) the elements of every row of $A \in \mathbb{R}^{p \times n}$ for which the $i^{th}$ component, $a_{\ell i} \in \mathbb{R}$, is non-zero, and
(iv) has knowledge of the active constraint functions \( g_k(x_i \cup x_{\mathcal{N}(i)}) \) in which agent \( i \in \{1, \ldots, n\} \) is involved.

Note that the inequality constraints \( g_k(x) \leq 0 \) for \( k \in K(x) \) are incorporated by either the (sub-)optimization problem (3.21) or (3.23) that only require local information. In order to solve (3.21) or (3.23), full knowledge of the subgradient mappings \( x_i \mapsto \partial f_i(x_i) \) for some \( i \in \{1, \ldots, n\} \) and \( x \mapsto \partial g_k(x) \) for all \( k \in K(x) \) is required whenever the projection operator becomes active, i.e., for \( (x, \mu) \in \text{bd}(G) \times \mathbb{R}^p \). However, the (sub-)optimization problems (3.21) and (3.23) require only knowledge of neighboring agents that are involved in the active inequality constraints. Once an arbitrarily chosen agent \( i \in \{1, \ldots, n\} \) that is involved in the active constraints solves the (sub-)optimization problem, it propagates the solution to its immediate neighbors.

**Remark 5.1 (Scalability of the proposed saddle-point-like dynamics).** For the nonsmooth convex program of a sum of convex functions (3.1) one could use consensus-based algorithms developed in, for instance [NOP10, ZM12]. However, this approach would lead to agents storing and communicating with neighbors estimates of the entire solution vector in \( \mathbb{R}^n \), and hence, would not scale well with the number of agents in the network. In contrast to these works, to execute the discontinuous saddle-point-like dynamics (3.19), agents only need to store the component of the solution vector that they control and communicate it with its neighbors. Therefore, the proposed set-valued dynamics scale well with respect to the number of agents in the network.

### 3.6 Convergence and Performance Characterization

In this section, we slightly modify the distributed continuous-time coordination algorithm designed in Section 3.3 and establish asymptotic convergence of its trajectories to the set of primal-dual solutions \( X \times M \) using classical notions of stability theory. Under mild convexity and regularity conditions on the objective function of the nonsmooth convex program at hand, we further characterize exponential convergence properties of the proposed dynamics and provide a performance bound in the absence of inequality constraints.

#### 3.6.1 Asymptotic Convergence

The following result characterizes asymptotic convergence of modified saddle-point dynamics to a point in the set \( X \times M \) without relying on the LaSalle Invariance Principle. We then establish asymptotic convergence of the saddle-point dynamics (3.8) by relating its solutions with those of the modified saddle-point dynamics.

Consider the following modifications of the saddle points dynamics (3.7),

\[
\begin{align*}
\dot{x}(t) & \in -\partial_x \mathcal{L}^\kappa(x, \mu)(t) + \partial_x \mathcal{L}^\kappa(x^*, \mu^*), & x(t_0) = x_0 \in \mathbb{R}^n, \\
\dot{\mu}(t) & \in +\partial_\mu \mathcal{L}^\kappa(x, \mu)(t) - \partial_\mu \mathcal{L}^\kappa(x^*, \mu^*), & \mu(t_0) = \mu_0 \in \mathbb{R}^p,
\end{align*}
\]  

(3.25a)  

(3.25b)
for a.a. \( t \in [t_0, +\infty) \). Note that the KKT-conditions (cf. Theorem 3.6 in Chapter 2) imply that \( 0 \in \partial_x \mathcal{L}_\kappa^\pi(x^*, \mu^*) \) and \( 0 \in \partial_\mu \mathcal{L}_\kappa^\pi(x^*, \mu^*) \). In particular, the modified saddle-point dynamics (3.25) defined over \( \mathbb{R}^n \times \mathbb{R}^p \) take the form

\[
\dot{x}(t) + A^\top (A(x(t) - x^*) + A^\top (\mu(t) - \mu^*)) \in -\left( \partial f(x(t)) - \partial f(x^*) \right) \tag{3.26a}
\]

\[
+ \kappa \sum_{k \in K(x)} \partial [g_k(x(t))]^+ - \kappa \sum_{k \in K(x^*)} \partial [g_k(x^*)]^+ \right),
\]

\[
\dot{\mu}(t) = A(x(t) - x^*), \tag{3.26b}
\]

for a.a. \( t \in [t_0, +\infty) \) with initial condition \( (x_0, \mu_0) \in \mathbb{R}^n \times \mathbb{R}^p \). Proposition 1.7 in Chapter 2 and Lemma 2.2 in Chapter 2 guarantee the existence of solutions \( (x, \mu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p \) of the dynamics (3.26), where the solutions are understood in the sense of Krasovskii (cf. Definition 2.6 in Ch. 2). Here, we use the set-valued map \( F^{\text{bb}} : \mathbb{R}^n \times \mathbb{R}^p \Rightarrow \mathbb{R}^n \times \mathbb{R}^p \) to refer to the modified saddle-point dynamics (3.26).

**Theorem 6.1** (Point-wise convergence). Let \( x^* \in X \), \( (\mu^*, \nu^*) \in M \times N \), and let \( \kappa > \| \nu^* \|_\infty \). Define the mapping \( \tilde{d} : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^p \) by

\[
\tilde{d}(x, \mu) = \mathcal{L}_\kappa^\pi(x, \mu) - \mathcal{L}_\kappa^\pi(x^*, \mu^*) + \frac{1}{2} \| x - x^* \|^2 + \frac{1}{2} \| \mu - \mu^* \|^2. \tag{3.27}
\]

Then, the inclusion \( (\mathcal{L}_{F^{\text{bb}}} \tilde{d})(x, \mu) \subset (-\infty, 0) \) holds for all \( (x, \mu) \in \mathbb{R}^n \times \mathbb{R}^p \) and any solution \( (x, \mu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p \) of (3.26) converges asymptotically to a point in the set of primal-dual minimizer \( X \times M \).

**Proof.** Note that the set-valued map \( F^{\text{bb}} \) is locally bounded, upper semi-continuous and takes nonempty, convex, and compact values (cf. Proposition 1.7 in Ch. 2). Moreover, by Proposition 2.5(i), the tuple \( (x^*, \mu^*) \) identifies a saddle point of the augmented Lagrangian function \( \mathcal{L}_\kappa^\pi \) in (3.4) when \( \kappa \geq \| \nu^* \|_\infty \). Let \( \kappa > \| \nu^* \|_\infty \). Then, the solution sets \( X \times M \) of (3.1) and (3.6) coincide (cf. Lemma 3.7 in Ch. 2).

Note that the candidate Lyapunov function \( \tilde{d} : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}_{\geq 0} \) is Lipschitz continuous and regular. In particular, we have

\[
\tilde{d}(x, \mu) = f(x) + \frac{1}{2} \| A(x - x^*) \|^2 + \langle \mu, A(x - x^*) \rangle + \kappa \langle I_m, [g(x)]^+ \rangle
\]

\[
- f(x^*) - \kappa \langle I_m, [g(x^*)]^+ \rangle + \frac{1}{2} \| x - x^* \|^2 + \frac{1}{2} \| \mu - \mu^* \|^2.
\]

The KKT-conditions (cf. Theorem 3.6 in Ch. 2) imply \( 0 \in \partial_x \mathcal{L}_\kappa^\pi(x^*, \mu^*) \) and \( 0 \in \partial_\mu \mathcal{L}_\kappa^\pi(x^*, \mu^*) \), and therefore,

\[
\exists \pi_f^* \in \partial f(x^*), \exists \pi_g^+ \in \kappa \sum_{k \in K(x^*)} \partial [g(x^*)]^+ : -A^\top \mu^* - \pi_f^* - \pi_g^+ = 0. \tag{3.28}
\]

Pre-multiplication of (3.28) with \( \langle x - x^*, \cdot \rangle \) yields

\[
-\langle x - x^*, A^\top \mu^* \rangle - \langle x - x^*, \pi_f^* \rangle - \langle x - x^*, \pi_g^+ \rangle = 0.
\]
Hence, the candidate Lyapunov function takes the form
\[
\tilde{d}(x, \mu) = f(x) - f(x^*) - \langle x - x^*, \pi_f^\dagger \rangle + \kappa \langle 1_m, [g(x)]^\dagger \rangle - \kappa \langle 1_m, [g(x^*)]^\dagger \rangle - \langle x - x^*, \pi_g^\dagger \rangle + \frac{1}{2} \|A(x - x^*)\|^2 + \langle x - x^*, A^\top (\mu - \mu^*) \rangle + \frac{1}{2} \mu - \mu^*\|^2. \tag{3.29}
\]

By the 1st-order convexity condition in \(x\), it follows
\[
f(x) - f(x^*) - \langle x - x^*, \pi_f^\dagger \rangle \geq 0, \tag{3.30a}
\]
\[
\kappa \langle 1_m, [g(x)]^\dagger \rangle - \kappa \langle 1_m, [g(x^*)]^\dagger \rangle - \langle x - x^*, \pi_g^\dagger \rangle \geq 0. \tag{3.30b}
\]

Substituting (3.30a) and (3.30b) into equation (3.29) yields
\[
\tilde{d}(x, \mu) \geq \frac{1}{2} \|A(x - x^*)\|^2 + \frac{1}{2} \langle x - x^*, A^\top (\mu - \mu^*) \rangle + \frac{1}{2} \langle \mu - \mu^*, A(x - x^*) \rangle + \frac{1}{2} \|x - x^*\|^2 + \frac{1}{2} \|\mu - \mu^*\|^2
\]
\[
= \frac{1}{2} \left\langle \begin{pmatrix} x - x^* \\ \mu - \mu^* \end{pmatrix}, \begin{pmatrix} A^\top A + I_n & A^\top I_p \\ I_p & I_p \end{pmatrix} \begin{pmatrix} x - x^* \\ \mu - \mu^* \end{pmatrix} \right\rangle =: P \in \mathbb{R}^{n+p \times n+p}
\]

Note that the Schur complement (cf. [Ber09]) of \(P\) in \(P\) implies \(P > 0\) for all \((x, \mu) \in \mathbb{R}^n \times \mathbb{R}^p\), since \(I_p > 0\) and \(A^\top A + I_n - A^\top I_p^{-1} A > 0\). Hence, we have \(\tilde{d}(x, \mu) > 0\) for all \((x, \mu) \in (\mathbb{R}^n \times \mathbb{R}^p) \setminus (X \times M)\) and \(\tilde{d}(x, \mu) = 0\) if and only if \((x, \mu) \in X \times M\). Moreover, the candidate Lyapunov function \(\tilde{d}\) is radially unbounded.

Let \((\eta_x, \eta_\mu) \in \partial_x d(x, \mu) \times \partial_\mu d(x, \mu)\), and let \(\psi \in (\mathcal{L}^n_{F^{bb}} d)(x, \mu)\). By definition of the Lie-derivative (cf. Definition 2.3 in Ch. 2), there exists
\[
\begin{pmatrix} \xi_x \\ \xi_\mu \end{pmatrix} = \begin{pmatrix} -A^\top A(x - x^*) - A^\top (\mu - \mu^*) - \pi_f + \pi_f^\dagger + \pi_g^\dagger - \pi_g^+ - \pi_g^\dagger x \end{pmatrix} \in F^{bb},
\]
where \(\pi_f \in \partial f(x), \pi_f^\dagger \in \kappa \sum_{k \in K(x)} \partial [g_k(x)]^\dagger, \text{ and similarly, } \pi_f^\dagger \in \partial f(x^*), \pi_g^\dagger \in \kappa \sum_{k \in K(x^*)} \partial [g_k(x^*)]^\dagger\), such that
\[
\psi = \left\langle \begin{pmatrix} \eta_x \\ \eta_\mu \end{pmatrix}, \begin{pmatrix} \xi_x \\ \xi_\mu \end{pmatrix} \right\rangle = \langle A^\top A(x - x^*) + A^\top (\mu - \mu^*) + \pi_f - \pi_f^\dagger + \pi_g^+ - \pi_g^\dagger + \pi_g^\dagger x, \dot{x} \rangle + \langle A^\top (\mu - \mu^*), \dot{\mu} \rangle + \langle x - x^*, \dot{x} \rangle + \langle \mu - \mu^*, \dot{\mu} \rangle
\]
\[
= -\langle x - x^*, \pi_f - \pi_f^\dagger \rangle - \langle x - x^*, \pi_g - \pi_g^* \rangle - \|\dot{x}\|^2.
\]

By the 1st-order convexity condition in \(x\), it follows
\[
\langle x - x^*, \pi_f - \pi_f^\dagger \rangle \geq 0, \tag{3.32a}
\]
\[
\langle x - x^*, \pi_g^+ - \pi_g^\dagger \rangle \geq 0. \tag{3.32b}
\]
Substituting (3.32a) and (3.32b) into equation (3.31) yields
\[ \psi \leq -\| \dot{x} \|^2 = -\| A^T A (x - x^*) + A^T (\mu - \mu^*) + \pi_f - \pi_f^* + \pi_g^+ - \pi_g^+ \|^2 < 0, \]
for all \((x, \mu) \in (\mathbb{R}^n \times \mathbb{R}^p) \setminus (X \times M)\) and \(\psi = 0\) if and only if \((x, \mu) \in X \times M\). Since \(\psi\) is chosen arbitrary, the inclusion \((L_{F^{bb}}) (x, \mu) \subset (-\infty, 0)\) holds for all \((x, \mu) \in \mathbb{R}^n \times \mathbb{R}^p\). Hence, by Theorem 2.4 in Chapter 2, the point \((x^*, \mu^*) \in \text{eq}(F^{bb})\) is a strongly asymptotically stable equilibrium point of (3.26).

Finally, we show asymptotic convergence of solutions \((x, \mu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p\) of (3.26) to a point in the set \(X \times M\). Note that the \(\omega\)-limit set
\[ \omega(x, \mu) = \{ (y_x, y_\mu) \in \bar{d}^{-1}(\leq \gamma) \mid \lim \inf_{t \to \infty} \| (x(t), \mu(t)) - (y_x, y_\mu) \| = 0 \} \subset X \times M \]
is nonempty and weakly invariant. Our proof strategy is based on establishing that \(x, \mu\) in (3.26) to a point in the set \(X \times M\), converges asymptotically to a point in \(X \times M\). Since \((y_x, y_\mu), (z_x, z_\mu) \in \omega(x, \mu)\) with \((y_x, y_\mu) \neq (z_x, z_\mu)\), and define the functions \(\bar{d}_1, \bar{d}_2 : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}_{\geq 0}\) by
\[
\bar{d}_1(x, \mu) = L^\kappa(x, \mu) - L^\kappa(y_x, y_\mu) + \frac{1}{2} \| x - y_x \|^2 + \frac{1}{2} \| \mu - y_\mu \|^2,
\]
\[
\bar{d}_2(x, \mu) = L^\kappa(x, \mu) - L^\kappa(z_x, z_\mu) + \frac{1}{2} \| x - z_x \|^2 + \frac{1}{2} \| \mu - z_\mu \|^2.
\]
Since \((y_x, y_\mu), (z_x, z_\mu) \in X \times M\), the above discussion implies that the \(\gamma\)-sublevel sets \(\bar{d}_1^{-1}(\leq \gamma)\) and \(\bar{d}_2^{-1}(\leq \gamma)\) are strongly invariant under \(F^{bb}\), for any \(\gamma > 0\). Pick \(\gamma < \frac{1}{2} \| (y_x, y_\mu) - (z_x, z_\mu) \|\). Since \((y_x, y_\mu) \in \omega(x, \mu)\), the solution \((x, \mu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p\) enters \(\bar{d}_1^{-1}(\leq \gamma)\) at some finite time \(t_1 \in [t_0, +\infty)\) and remains there afterwards because of the strong invariance of the \(\gamma\)-sublevel set. Similarly, for \((z_x, z_\mu) \in \omega(x, \mu)\), there exists a finite time \(t_2 \in [t_0, +\infty)\) such that \((x(t), \mu(t)) \in \bar{d}_2^{-1}(\leq \gamma)\) for all \(t \geq t_2\). Therefore, for \(t \geq \max\{t_1, t_2\}\), the solution belongs to \(\bar{d}_1^{-1}(\leq \gamma) \cap \bar{d}_2^{-1}(\leq \gamma) = \emptyset\), which is a contradiction. This concludes the proof.

**Corollary 6.2 (Convergence of trajectories).** Any solution \((x, \mu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p\) of the saddle-point dynamics (3.8) starting from \((x_0, \mu_0) \in \mathbb{R}^n \times \mathbb{R}^p\) converges asymptotically to a point in \(X \times M\).

**Proof.** Let \(x^* \in X\), \((\mu^*, \nu^*) \in M \times N\), and let \(\kappa > \| \nu^* \|_{\infty}\). Note that the KKT-conditions (cf. Theorem 3.6 in Ch. 2) imply \(0 \in \partial_x L^\kappa(x^*, \mu^*)\) and \(0 \in \partial_\mu L^\kappa(x^*, \mu^*)\). By definition of the modified saddle-point dynamics (3.26), it follows
\[
-\partial_x L^\kappa(x, \mu) \subset -\partial_x L^\kappa(x^*, \mu^*) + \partial_x L^\kappa(x^*, \mu^*),
\]
\[
+\partial_\mu L^\kappa(x, \mu) = +\partial_\mu L^\kappa(x^*, \mu^*) - \partial_\mu L^\kappa(x^*, \mu^*).
\]
Hence, the inclusion \(F^b(x, \mu) \subset F^{bb}(x, \mu)\) holds for all \((x, \mu) \in \mathbb{R}^n \times \mathbb{R}^p\) and any solution \((x, \mu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p\) of \(F^b\) starting from \((x_0, \mu_0) \in \mathbb{R}^n \times \mathbb{R}^p\) is also a solution of \(F^{bb}\). Therefore, we deduce asymptotic convergence of trajectories of (3.8) to a point in \(X \times M\). This concludes the proof.

\[ \square \]
3.6.2 Performance Bound Characterization

The following result establishes exponential convergence of trajectories of the modified saddle-point dynamics and provides a performance bound under additional convexity and regularity conditions on the objective function of the nonsmooth convex optimization problem.

Recall that $G \subset \mathbb{R}^n$ denotes the viability set associated with the nonsmooth convex program (3.1). Suppose $f \in C^{1,1}(\mathbb{R}^n, \mathbb{R})$, and the trajectories of (3.26) satisfy $(x(t), \mu(t)) \in \text{int}(G) \times \mathbb{R}^p$ for all $t \in [t_0, +\infty)$. Under these assumptions, the augmented Lagrangian function (3.4) reduces to

$$
\mathcal{L}(x, \mu) = f(x) + \frac{1}{2}\|Ax - b\|^2 + \langle \mu, Ax - b \rangle,
$$

(3.34)

since $g(x) < 0$ for all $x \in \text{int}(G)$ implies $\kappa(\mathbf{1}_m, [g(x)]^+) = 0$ for all $x \in \text{int}(G)$.

Consider the following modifications of the saddle-point dynamics (3.8),

$$
\dot{x}(t) = -\nabla_x \mathcal{L}(x, \mu)(t) + \nabla_x \mathcal{L}(x^*, \mu^*), \quad x(t_0) = x_0 \in \text{int}(G),
$$

(3.35a)

$$
\dot{\mu}(t) = +\nabla_\mu \mathcal{L}(x, \mu)(t) - \nabla_\mu \mathcal{L}(x^*, \mu^*), \quad \mu(t_0) = \mu_0 \in \mathbb{R}^p,
$$

(3.35b)

for all $t \in [t_0, +\infty)$. Note that the KKT-conditions (cf. Theorem 3.6 in Ch. 2) imply $0 = \nabla_x \mathcal{L}(x^*, \mu^*)$ and $0 = \nabla_\mu \mathcal{L}(x^*, \mu^*)$. In particular, the modified saddle-point dynamics (3.26) take the form

$$
\dot{x}(t) = -A^\top A(x(t) - x^*) - A^\top (\mu(t) - \mu^*)
$$

(3.36a)

$$
- (\nabla f(x(t)) - \nabla f(x^*))
$$

(3.36b)

$$
\dot{\mu}(t) = A(x(t) - x^*),
$$

for all $t \in [t_0, +\infty)$ with initial condition $(x_0, \mu_0) \in \text{int}(G) \times \mathbb{R}^p$. Here, we use the mapping $F^\circ : \text{int}(G) \times \mathbb{R}^p \to \text{int}(G) \times \mathbb{R}^p$ to refer to the modified saddle-point dynamics (3.36). Note that the existence of classical solutions $(x, \mu) : [t_0, +\infty) \to \text{int}(G) \times \mathbb{R}^p$ of (3.36) is guaranteed since $F^\circ$ is Lipschitz continuous [Cor08].

Remark 6.3 (Relationship to saddle-point dynamics). Note that under the above assumptions, the modified saddle-point dynamics (3.36) coincide with the saddle-point dynamics (3.8) (in the absence of inequality constraints), since the KKT-conditions (cf. Theorem 3.6 in Ch. 2) imply $0 = \nabla_x \mathcal{L}(x^*, \mu^*)$ and $0 = \nabla_\mu \mathcal{L}(x^*, \mu^*)$.

The following result characterizes exponential convergence of trajectories of the modified saddle-point dynamics (3.36) to the primal-dual minimizer $X \times M$.

**Theorem 6.4** (Exponential convergence). Let $f \in C^{1,1}(\mathbb{R}^n, \mathbb{R})$, and let $(x^*, \mu^*) \in X \times M \subset \text{int}(G) \times \mathbb{R}^p$. Define the mapping $d^\circ : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}_{\geq 0}$ by

$$
d^\circ(x, \mu) = \mathcal{L}(x, \mu) - \mathcal{L}(x^*, \mu^*) + \frac{1}{2}\|x - x^*\|^2 + \frac{1}{2}\|\mu - \mu^*\|^2.
$$

(3.37)

If $\partial(\nabla f)(x) \succ 0$ for all $x \in \mathbb{R}^n$, and $\text{rank}(A) = p$, then any solution $(x, \mu) : [t_0, +\infty) \to \text{int}(G) \times \mathbb{R}^p$ of (3.36) starting from $(x_0, \mu_0) \in \text{int}(G) \times \mathbb{R}^p$ converges exponentially fast to the unique minimizer $(x^*, \mu^*) \in X \times M$. 

Moreover, the candidate Lyapunov function (3.37) is radially unbounded, since
\[ -I \]
Note that the Schur complement of \( A \) is invertible and the KKT-conditions imply \( \mu_* = -(A^T)^{-1}A\nabla f(x^*) \). Let \((x^*, \mu^*) \in X \times M\) denote the unique minimizer of (3.1).

Note that \( d^o \in C^{1,1}(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R}_{\geq 0}) \). In particular, we have for (3.37),
\[
d^o(x, \mu) = f(x) - f(x^*) + \frac{1}{2} \| A(x - x^*) \|^2 + \langle \mu, A(x - x^*) \rangle + \frac{1}{2} \| x - x^* \|^2 + \frac{1}{2} \| \mu - \mu^* \|^2.
\]
By the KKT-conditions, we have \(-\langle x - x^*, A^T \mu^* \rangle - \langle x - x^*, \nabla f(x^*) \rangle = 0 \). Hence, the candidate Lyapunov function \( d^o \) takes the form
\[
d^o(x, \mu) = f(x) - f(x^*) - \langle x - x^*, \nabla f(x^*) \rangle + \frac{1}{2} \| A(x - x^*) \|^2
\]
\[
+ \langle x - x^*, A^T(\mu - \mu^*) \rangle + \frac{1}{2} \| x - x^* \|^2 + \frac{1}{2} \| \mu - \mu^* \|^2.
\]
Note that the 1st-order convexity condition implies \( f(x) - f(x^*) - \langle x - x^*, \nabla f(x^*) \rangle \geq 0 \). Therefore, it follows from (3.38),
\[
d^o(x, \mu) \geq \frac{1}{2} \| A(x - x^*) \|^2 + \frac{1}{2} \langle x - x^*, A^T(\mu - \mu^*) \rangle + \frac{1}{2} \langle \mu - \mu^*, A(x - x^*) \rangle
\]
\[
+ \frac{1}{2} \| x - x^* \|^2 + \frac{1}{2} \| \mu - \mu^* \|^2
\]
\[
= \frac{1}{2} \left\langle \begin{pmatrix} x - x^* \\ \mu - \mu^* \end{pmatrix}, \begin{pmatrix} A^T A + I_n \\ A^T \end{pmatrix} \begin{pmatrix} x - x^* \\ \mu - \mu^* \end{pmatrix} \right\rangle =: P \in \mathbb{R}^{n+p \times n+p}.
\]
Note that the Schur complement of \( I_p \) in \( P \) implies \( P > 0 \) for all \((x, \mu) \in \mathbb{R}^n \times \mathbb{R}^p\) since \( I_p > 0 \) and \( A^T A + I_n - A^T I_p^{-1} A > 0 \). Hence, we have \( d^o(x, \mu) > 0 \) for all \((x, \mu) \in (\mathbb{R}^n \times \mathbb{R}^p) \setminus \{(x^*, \mu^*)\}\) and \( d^o(x, \mu) = 0 \) if and only if \((x, \mu) = (x^*, \mu^*)\). Moreover, the candidate Lyapunov function (3.37) is radially unbounded, since
\[
d^o(x, \mu) \geq \frac{1}{2} \lambda_{\min}(P) \left\| \begin{pmatrix} x - x^* \\ \mu - \mu^* \end{pmatrix} \right\|^2.
\]
(3.39)

The Lie-derivative \( (L_{F^o} d^o)(x, \mu) : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R} \) of \( d^o \) with respect to \( F^o \) at \((x, \mu)\) is
\[
(L_{F^o} d^o)(x, \mu) = \langle \nabla f(x), \dot{x} \rangle + \langle A^T A(x - x^*), \dot{x} \rangle + A^T \mu, \dot{x} \rangle
\]
\[
+ \langle A(x - x^*), \dot{\mu} \rangle + \langle x - x^*, \dot{x} \rangle + \langle \mu - \mu^*, \dot{\mu} \rangle.
\]
Using the fact that \(-\langle A^T \mu, \dot{x} \rangle - \langle \nabla f(x^*), \dot{x} \rangle = 0 \) (cf. Theorem 3.6 in Ch. 2),
\[
(L_{F^o} d^o)(x, \mu) = \langle A^T A(x - x^*) + A^T(\mu - \mu^*) + (\nabla f(x) - \nabla f(x^*)), \dot{x} \rangle
\]
\[
+ \langle A(x - x^*), \dot{\mu} \rangle + \langle x - x^*, \dot{x} \rangle + \langle \mu - \mu^*, \dot{\mu} \rangle
\]
\[
= -\langle x - x^*, \nabla f(x) - \nabla f(x^*) \rangle - \| \nabla_x \mathcal{L}(x, \mu) - \nabla_x \mathcal{L}(x^*, \mu^*) \|^2.
\]
(3.40)
3.6. CONVERGENCE AND PERFORMANCE CHARACTERIZATION

Substituting (3.36a) into equation (3.40) yields
\[
(L_{F\circ d^0})(x, \mu) = -\langle x - x^*, \nabla f(x) - \nabla f(x^*) \rangle - \langle A^\top A(x - x^*), A^\top A(x - x^*) \rangle
- \langle A^\top A(x - x^*), A^\top (\mu - \mu^*) \rangle - \langle A^\top (\mu - \mu^*), A^\top A(x - x^*) \rangle
- \langle A^\top (\mu - \mu^*), \nabla f(x) - \nabla f(x^*) \rangle - \langle \nabla f(x) - \nabla f(x^*), A^\top (\mu - \mu^*) \rangle
- \langle \nabla f(x) - \nabla f(x^*), A^\top A(x - x^*) \rangle
- \langle \nabla f(x) - \nabla f(x^*), \nabla f(x) - \nabla f(x^*) \rangle.
\]

Since \( f \in C^{1,1}(\mathbb{R}^n, \mathbb{R}) \), the gradient (mapping) \( \nabla f : \mathbb{R}^n \to \mathbb{R}^n \) is Lipschitz continuous. By the extended Mean-Value Theorem (cf. Proposition 1.9 in Ch. 2), it follows
\[
\nabla f(x) - \nabla f(x^*) \in \text{co} \{ \partial(\nabla f([x, x^*])) \} (x - x^*),
\]
where \([x, x^*] = x + \theta(x^* - x)\) for \( \theta \in [0, 1] \), and \( \partial(\nabla f)(x) \) denotes the generalized Hessian of \( f \) at \( x \in \mathbb{R}^n \) (cf. Definition 1.8 in Ch. 2). Therefore, we have
\[
(L_{F\circ d^0})(x, \mu) \leq \sup_{M \in \text{co} \{ \partial(\nabla f([x, x^*])) \}} - \left\langle \left( \begin{array}{cc} x - x^* \\ \mu - \mu^* \end{array} \right), \left( \begin{array}{cc} B & C \\ * & D \end{array} \right) \left( \begin{array}{c} x - x^* \\ \mu - \mu^* \end{array} \right) \right\rangle,
\]
where the partitioned matrix \( Q = Q^\top \) is composed of the submatrices
\[
\begin{align*}
\mathbb{R}^{n \times n} &\ni B = (M + A^\top A)^\top (M + A^\top A) + M, \\
\mathbb{R}^{n \times p} &\ni C = (M + A^\top A)^\top A^\top, \\
\mathbb{R}^{p \times p} &\ni D = AA^\top.
\end{align*}
\]

Note that the set-valued map \( x \mapsto \partial(\nabla f)(x) \) takes nonempty, convex and compact values (cf. [Cla83]). Since \( \partial(\nabla f)(x) \ni 0 \) for all \( x \in \mathbb{R}^n \), it follows \( \text{co} \{ \partial(\nabla f([x, x^*])) \} \ni 0 \) for all \( x \in \mathbb{R}^n \) and \( \theta \in [0, 1] \). Hence, \( M \ni 0 \) holds for all \( x \in \mathbb{R}^n \). Moreover, any matrix \( M \in \text{co} \{ \partial(\nabla f([x, x^*])) \} \subset \mathbb{R}^{n \times n} \) is symmetric [HUSN84]. Let
\[
\overline{M} = \argmax_{M \in \text{co} \{ \partial(\nabla f([x, x^*])) \}} - \left\langle \left( \begin{array}{cc} x - x^* \\ \mu - \mu^* \end{array} \right), \left( \begin{array}{cc} B & C \\ * & D \end{array} \right) \left( \begin{array}{c} x - x^* \\ \mu - \mu^* \end{array} \right) \right\rangle,
\]
where the maximum \( \overline{M} \) is achieved by compactness of the set \( \text{co} \{ \partial(\nabla f([x, x^*])) \} \). Since \( \text{rank}(A) = p \), it follows \( D \ni 0 \). The Schur complement of \( D \) in \( Q \) implies
\[
\mathbb{R}^{n \times n} \ni R = (\overline{M} + A^\top A)^\top \left( I_n - A^\top (AA^\top)^{-1} A \right) (\overline{M} + A^\top A) + \overline{M}.
\]
Since \( \overline{M} + A^\top A \ni 0 \) for all \( x \in \mathbb{R}^n \), and
\[
\left( \begin{array}{cc} I_n & A^\top \\ * & AA^\top \end{array} \right) \ni 0,
\]
we conclude that $Q > 0$ holds for all $x \in \mathbb{R}^n$. Therefore, it follows
\begin{equation}
(L_{F^o}d^o)(x, \mu) \leq -\inf_{(x, \mu) \in d^{o-1}(\leq d^o(x_0, \mu_0))} \left( \lambda_{\min}(Q(x)) \left\| \left( \begin{array}{c} x - x^* \\ \mu - \mu^* \end{array} \right) \right\|^2, \right.
\end{equation}
where the minimum in (3.41) is achieved since the sublevel set $d^{o-1}(\leq d^o(x_0, \mu_0))$ is compact and invariant under $F^o$. This is an immediate consequence of (3.40) and the 1st-order convexity condition of $f$ in $x$, i.e., $(x - x^*, \nabla f(x) - \nabla f(x^*)) \geq 0$, implying $(L_{F^o}d^o)(x, \mu) \leq -\|\nabla_x L(x, \mu) - \nabla_x L(x^*, \mu^*)\|^2 < 0, \ \forall (x, \mu) \in (\mathbb{R}^n \times \mathbb{R}^p) \setminus \{(x^*, \mu^*)\}.$

Consider again the Lyapunov function (3.38),
\begin{align*}
d^o(x, \mu) &= f(x) - f(x^*) - \langle x - x^*, \nabla f(x^*) \rangle \\
&\quad + \frac{1}{2} \left\langle \left( \begin{array}{c} x - x^* \\ \mu - \mu^* \end{array} \right), \left( \begin{array}{cc} A^T A + I_n & A^T \\
* & I_p \end{array} \right) \left( \begin{array}{c} x - x^* \\ \mu - \mu^* \end{array} \right) \right\rangle.
\end{align*}
Since $f$ is strictly convex in $x$, it follows $f(x) - f(x^*) < \langle x - x^*, \nabla f(x^*) \rangle$, and therefore,
\begin{align*}
d^o(x, \mu) &\leq \langle x - x^*, \nabla f(x) - \nabla f(x^*) \rangle \\
&\quad + \frac{1}{2} \left\langle \left( \begin{array}{c} x - x^* \\ \mu - \mu^* \end{array} \right), \left( \begin{array}{cc} A^T A + I_n & A^T \\
* & I_p \end{array} \right) \left( \begin{array}{c} x - x^* \\ \mu - \mu^* \end{array} \right) \right\rangle.
\end{align*}
By the Mean-Value Theorem (cf. Proposition 1.9 in Ch. 2), it follows
\begin{equation}
d^o(x, \mu) < \inf_{M \in co(\partial(\nabla f)([x, x^*])} \frac{1}{2} \left\langle \left( \begin{array}{c} x - x^* \\ \mu - \mu^* \end{array} \right), \left( \begin{array}{cc} E & A^T \\
* & I_p \end{array} \right) \left( \begin{array}{c} x - x^* \\ \mu - \mu^* \end{array} \right) \right\rangle,
\end{equation}
where $\mathbb{R}^{n \times n} \ni E = 2M + A^T A + I_n \succ 0$ for all $x \in \mathbb{R}^n$. Let
\begin{equation}
M = \arg\min_{M \in co(\partial(\nabla f)([x, x^*])} \frac{1}{2} \left\langle \left( \begin{array}{c} x - x^* \\ \mu - \mu^* \end{array} \right), \left( \begin{array}{cc} E & A^T \\
* & I_p \end{array} \right) \left( \begin{array}{c} x - x^* \\ \mu - \mu^* \end{array} \right) \right\rangle.
\end{equation}
Then, the Schur complement of $I_p$ in $S$ implies
\begin{equation}
\mathbb{R}^{n \times n} \ni T = 2M + A^T A + I_n - A^T A \succ 0, \ \forall x \in \mathbb{R}^n,
\end{equation}
and we deduce that $S \succ 0$ holds for all $x \in \mathbb{R}^n$. Therefore, it follows
\begin{equation}
\sup_{(x, \mu) \in d^{o-1}(\leq d^o(x_0, \mu_0))} \frac{1}{2} (\lambda_{\max}(S(x))) \left\| \left( \begin{array}{c} x - x^* \\ \mu - \mu^* \end{array} \right) \right\|^2,
\end{equation}
where the maximum in (3.42) is achieved since the sublevel set $d^{o-1}(\leq d^o(x_0, \mu_0))$ is compact and invariant under $F^o$. By the exponential convergence theorem (cf. Theorem 4.10 in [Kha02]), it follows
\begin{equation}
\left\| \left( \begin{array}{c} x(t) - x^* \\ \mu(t) - \mu^* \end{array} \right) \right\| < \sqrt{\frac{k_2}{k_1}} \exp(-\delta t) \left\| \left( \begin{array}{c} x(t_0) - x^* \\ \mu(t_0) - \mu^* \end{array} \right) \right\|,
\end{equation}
for all $t \in [t_0, +\infty)$ with initial condition $(x_0, \mu_0) \in int(G) \times \mathbb{R}^p$, where $\delta = k_3/2k_2$ is...
This section summarizes our previous studies and illustrates the convergence properties of the discontinuous saddle-point-like dynamics (3.19).

Consider a linear model predictive control (MPC) setup for a network of agents with coupling constraints, whose aim is to compute a control input sequence that simultaneously minimizes the actuation effort and the network state in a distributed
way. Formally, consider the discrete-time optimal control problem

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \sum_{k=t}^{t+N-1} \|x(k+1|t)\|_Q^2 + \|u(k|t)\|_R^2 \\
\text{subject to} & \quad x(k+1|t) = Sx(k|t) + Pu(k|t), \quad x(t) = x(t), \\
& \quad Cx(k|t) + Du(k|t) \leq 1_m, \quad \forall k \in [t, t+N-1],
\end{align*}
\]

where \( x(k|t) \in \mathbb{R}^n \) and \( u(k|t) \in \mathbb{R}^n \) are the network state and control input, predicted at time \( t \in \mathbb{R} \) over the horizon \( N \in \mathbb{Z}_{>0} \). Let the weighing matrices \( Q, R \in \mathbb{R}^{n \times n} \) be such that \( Q = Q^T \succeq 0 \) and \( R = R^T > 0 \). The initial condition \( x_i(t|t) \) is assumed to be known to agent \( i \in \{1, \ldots, n\} \) and its neighbors. Note that the network topology is encoded in the sparsity structure of the matrices \( S, P \in \mathbb{R}^{n \times n} \) and \( C, D \in \mathbb{R}^{m \times n} \). Each agent can be interpreted as a subsystem whose dynamics are influenced by the predicted states of neighboring agents. Here, every agent knows the dynamics of its own subsystem and its neighbors subsystem, but not the entire network dynamics. Moreover, only discrete-time system dynamics fully characterized by the tuple \( (S, P) \) that are controllable are considered. The prediction horizon \( N \in \mathbb{Z}_{>0} \) is chosen such that stability of the overall model predictive control setup is guaranteed. For more details regarding to stability in MPC, see e.g., the works [MRRS00, RM09].

At every iteration step, the solution of the MPC problem is a sequence of predicted open-loop optimal controls

\[
u^*(\cdot | t) = \{u^*(k|t), \ldots, u^*(t+N-1|t)\},
\]

where only \( u^*(t) = u^*(t|t) \) is implemented to the system. To express the model predictive control problem as a (static) convex program of the form (3.1), we introduce the vector of variables

\[
x(\cdot | t) = (x(k+1|t), \ldots, x(t+N|t), u(k|t), \ldots, u(t+N-1|t)) \in \mathbb{R}^{2Nn}.
\]

The equality constraints of (3.1) are characterized by the matrix \( A \in \mathbb{R}^{Nn \times 2Nn} \),

\[
A = \begin{pmatrix}
I_n & 0_n & \cdots & 0_n \\
-S & I_n & \cdots & 0_n \\
\vdots & \ddots & \ddots & \vdots \\
0_n & 0_n & \cdots & I_n
\end{pmatrix}
\begin{pmatrix}
-P & 0_n & \cdots & 0_n \\
0_n & -P & \cdots & 0_n \\
\vdots & \ddots & \ddots & \vdots \\
0_n & 0_n & \cdots & -P
\end{pmatrix},
\]

and the vector \( b = (Sx(t|t), 0, \ldots, 0) \in \mathbb{R}^{Nn} \), where \( I_n, 0_n \in \mathbb{R}^{n \times n} \) denote the identity and zero matrices, respectively. Similarly, the inequality constraint function \( g : \mathbb{R}^{2Nn} \rightarrow \mathbb{R}^{Nm} \) is characterized by

\[
g(x) = \begin{pmatrix}
C & 0_m & \cdots & 0_m \\
0_m & C & \cdots & 0_m \\
\vdots & \ddots & \ddots & \vdots \\
0_m & 0_m & \cdots & C
\end{pmatrix}
\begin{pmatrix}
D & 0_m & \cdots & 0_m \\
0_m & D & \cdots & 0_m \\
\vdots & \ddots & \ddots & \vdots \\
0_m & 0_m & \cdots & D
\end{pmatrix}
\begin{pmatrix}
1_m \\
1_m \\
\vdots \\
1_m
\end{pmatrix}
succeq 0_{Nm},
\]

where \( 0_m \in \mathbb{R}^{m \times n} \). Note that the objective function in the MPC setup is separable.
only for diagonal weighting matrices $Q$ and $R$. In the following, suppose $Q = R = I_n$. Hence, the iterative optimal control problem can be equivalently stated as

$$\begin{aligned}
\text{minimize} & & \frac{1}{2} \sum_{k=t}^{t+N-1} \sum_{i=1}^{n} x_i(k + 1|t)^2 + u_i(k|t)^2 \\
\text{subject to} & & Ax - b = 0, \quad g(x) \preceq 0.
\end{aligned}$$

In what follows, consider the network dynamics and constraints as in Figure 3.1(a), with a prediction horizon of $N = 4$. The total number of primal-dual variables of algorithm (4.1) is $2Nn + N = 68$. In this example, each agent $i \in \{1, \ldots, 8\}$ is responsible for its own $2N = 8$ variables in $x$ and $\mu$, and also for its neighboring variables with respect to the network topology, illustrated in Figure 3.1(b). Thus, the communication is local, independent of the network size. Note that agent 5 and its neighbors incorporate additional $N = 4$ variables which are related to the control input sequence over the prediction horizon $N$. When implementing the dynamics (3.19), a first-order Euler approximation with stepsize $0.008$ is used. The regularization parameter is $\rho = 3$. For every iteration step, the stopping criteria used is based on the violation of the equality constraints, i.e., $\|Ax - b\| \leq \varepsilon$ with $\varepsilon = 0.001$. Also, a warm-start of the algorithm with the steady-state primal variables obtained from the previous iteration step is performed. At every iteration step, the implementation of (3.19) requires agent 5 to compute its optimal control sequence over the prediction horizon, i.e., $\{u_5(k|t), \ldots, u_5(t+N-1|t)\}$. Furthermore, all agents $i \in \{1, \ldots, 8\}$ compute their predicted open-loop state evolutions over the horizon, i.e., $\{x_i(k+1|t), \ldots, x_i(t+N|t)\}$.

\[ S = \begin{bmatrix}
\frac{1}{10} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{10} & 0 & 0 & 0 & \frac{1}{5} & 0 & 0 \\
0 & 0 & \frac{1}{10} & 0 & 0 & 0 & \frac{1}{5} & 0 \\
0 & 0 & 0 & \frac{6}{10} & \frac{1}{10} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2 & \frac{1}{10} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{10} & 0 & \frac{1}{5} \\
0 & 0 & 0 & 0 & \frac{1}{10} & 0 & \frac{1}{5} & 0 \\
0 & 0 & 0 & \frac{1}{5} & 0 & \frac{1}{5} & 0 & \frac{1}{10}
\end{bmatrix}, \quad P = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}, \quad C = 0_{2 \times 8}, \quad D = \begin{bmatrix}
1 \\
-1
\end{bmatrix}. \]

(a) Network and constraint matrices. (b) Network topology.

Figure 3.1: Network and constraint matrices (a) and associated graph topology (b) of the multi-agent system over which the algorithm (3.19) is fully distributed. Note that the tuple $(S, P)$ is controllable. The network of agents is controlled through the dynamics of agent 5 ((b) black node). Also, note that the choice of $D$ implies that the box constraints $\|(u(k|t), \ldots, (t + N - 1|t))\|_\infty \leq 1$ are satisfied.

Figure 3.2 shows the results of the iterative implementation of algorithm (3.19). Once the optimal control and network states over the prediction horizon are computed
Figure 3.2: The predicted open-loop and closed-loop network evolution of the linear model predictive control setup is depicted in plot (a) over 8 iteration steps. Plot (b) shows the associated open-loop and closed-loop optimal control sequence of agent 5, with a prediction horizon of $N = 4$. At every iteration step, the network state evolution (d) and optimal control input sequence (e) are computed by algorithm (3.19). The steady-state values achieved by these trajectories at every iteration step correspond to the optimal network state evolution and control input over the entire prediction horizon. Plot (c) shows the equality constraint violation and (f) depicts the evolution of the LaSalle function (3.9) at every iteration.
(corresponding to the steady-state values in the Figures 3.2(d)-(e) for every iteration), they are implemented by their respective agents. This results in the predicted open-loop and closed-loop network evolution depicted in Figure 3.2(a) and the associated open-loop and closed-loop optimal control input as shown in Figure 3.2(b). Note that the projection operator used in the algorithm prevents the trajectories of (3.19) from violating the viability condition with respect to the set \( G = \{ x \in \mathbb{R}^n \mid g(x) \preceq 0 \} \subset \mathbb{R}^n \) at any time. The equality constraint violation is plotted in Figure 3.2(c) and the LaSalle function (3.9) is illustrated in Figure 3.2(f). Clearly, the distributed algorithm (3.19) converges asymptotically at every iteration step and thus, agent 5 steers the network state to zero under minimal actuation effort.

### 3.8 Summary

In this chapter, we have developed continuous-time coordination algorithms for networks of agents that seek to collectively solve nonsmooth convex optimization problems with an inherent distributed structure. We have shown that the proposed saddle-point(-like) dynamics based on an augmented Lagrangian function asymptotically converge to a point in the set of solutions of its associated nonsmooth convex program. The dynamics are fully amenable to distributed implementation in the sense that each individual agent asymptotically finds its component of the optimal solution vector using only local information provided by its neighbors. We have also shown that under mild convexity and regularity conditions on the objective function of the nonsmooth convex program, convergence of trajectories of the modified dynamics to the unique optimal solution of the underlying optimization problem is exponential. Specifically, we have characterized a performance bound on the trajectories that strictly evolve within the interior of the viability set.
Chapter 4

Robust Distributed Optimization

In this chapter, we consider robust, i.e., semi-infinite optimization scenarios over networks defined by a – possibly nonsmooth – additively separable objective function with linear uncertain inequality constraints. Our objective in this chapter is to develop provably correct distributed continuous-time coordination algorithms that allow individual agents in a group to compute their own component of the optimal solution vector through local interactions with their neighbors.

We first turn our attention to the question on how to find a representation of the polyhedral uncertainty sets of the robust optimization problem that respects the distributed nature of the underlying network topology. In particular, we introduce the concept of so-called polyhedral lifting that allows us to formulate the problem as a nonsmooth convex program with an inherent distributed structure comprising information of the uncertainty sets.

We then focus on both the design and analysis of distributed continuous-time coordination algorithms that build on the characterization of the solutions of the robust optimization problem as saddle points of an augmented Lagrangian function. Specifically, we establish convergence properties of the resulting saddle-point algorithm using classical notions of stability theory. As in Section 3.3, the proposed dynamics are not fully distributed over the associated network topology because of a global parameter. Nevertheless, due to linearity of the inequality constraints of the robust optimization problem, we further develop explicit projected dynamics that do not incorporate (sub-)optimization problems and are amenable for distributed implementation.

Simulations in a (trivial) network of two agents that collectively solve a robust optimization problem with affine equality and linear inequality couplings constraints illustrate our findings and conclude this chapter.

4.1 Network Model and Problem Statement

This section describes the network model of agents associated to a nonsmooth robust optimization problem we set out to solve in a distributed way using continuous-time coordination algorithms.
Consider a group of \( n \in \mathbb{Z}_{>0} \) agents with identities \( i = \{1, \ldots, n\} \) that communicate over a weighted undirected graph \( G = (\mathcal{V}, \mathcal{E}, \mathcal{W}) \). Let the state of agent \( i \) be denoted by \( x_i \in \mathbb{R} \) and let its set of neighbors be \( \mathcal{N}(i) = \{j \in \mathcal{V} | (i,j) \in \mathcal{E}\} \). In this chapter, the objective of the agents is to cooperatively solve the semi-infinite optimization problem

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{n} f_i(x_i) \\
\text{subject to} & \quad \langle a_k, x \rangle \leq b_k, \quad \forall a_k \in \mathcal{A}_k,
\end{align*}
\]

(4.1)

for \( k \in \{1, \ldots, m\} \), where \( f_i : \mathbb{R} \to \mathbb{R} \) is a convex and Lipschitz continuous objective function associated with agent \( i \in \{1, \ldots, n\} \), \( a_k \in \mathbb{R}^n \) denotes the \( k \)th row of \( A \in \mathbb{R}^{m \times n} \), and \( b_k \in \mathbb{R} \). Let \( \mathcal{A}_k \subset \mathbb{R}^n \) denote the bounded uncertainty set. For convenience, we denote the network state by \( x = (x_1, \ldots, x_n) \in \mathbb{R}^n \) and the aggregate objective function by \( f(x) = \sum_{i=1}^{n} f_i(x_i) \). The robust feasibility set of (4.1) is

\[
C(\mathcal{A}) = \{ x \in \mathbb{R}^n \mid \langle a_k, x \rangle \leq b_k, \forall a_k \in \mathcal{A}_k, k \in \{1, \ldots, m\}\},
\]

where \( \mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_m \). The robust solution set of (4.1) is given by

\[
X(\mathcal{A}) = \{ x^* \in \mathbb{R}^n \mid f(x^*) \leq f(x), \forall x \in C(\mathcal{A})\}.
\]

The robust counterpart (cf. Definition 3.9 in Ch. 2) associated with (4.1) is the bi-level structured optimization problem

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{n} f_i(x_i) \\
\text{subject to} & \quad \max_{a_k \in \mathcal{A}_k} \langle a_k, x \rangle \leq b_k, \quad k \in \{1, \ldots, m\}.
\end{align*}
\]

(4.2)

Suppose that the uncertainty set \( \mathcal{A}_k \subset \mathbb{R}^n \) possesses a polyhedral representation, i.e.,

\[
\mathcal{A}_k = \bigcap_{\ell \in \{1, \ldots, p_k\}} \{ a_k \in \mathbb{R}^n \mid \langle d_{k,\ell}, a_k \rangle \leq e_{k,\ell} \} = \{ a_k \in \mathbb{R}^n \mid D_k a_k \preceq e_k \},
\]

where \( d_{k,\ell} \in \mathbb{R}^n \) denotes the \( \ell \)th row of \( D_k \in \mathbb{R}^{p_k \times n} \), \( e_{k,\ell} \) denotes the \( \ell \)th element of \( e_k \in \mathbb{R}^{p_k} \), and \( p_k \in \mathbb{Z}_{>0} \) is the number of halfspaces that enclose the uncertainty vector \( a_k \in \mathbb{R}^n \) for each \( k \in \{1, \ldots, m\} \). Note that Proposition 3.10 in Chapter 2 provides an explicit and convex reformulation of the robust counterpart (4.2) as

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{n} f_i(x_i) \\
\text{subject to} & \quad \langle e_k, \lambda_k \rangle \leq b_k, \quad k \in \{1, \ldots, m\}, \\
& \quad D_k^\top \lambda_k = x, \quad k \in \{1, \ldots, m\}, \\
& \quad \lambda_k \succeq 0, \quad k \in \{1, \ldots, m\}.
\end{align*}
\]

(4.3)

The following assumptions are made throughout the chapter:

**Assumption 1.1 (Robust feasibility).** The convex program (4.3) is robust feasible, i.e.,
C(A) \neq \emptyset$, and possesses a finite robust optimal value $f(x^*)$ for some robust minimizer $x^* \in X(A)$, i.e., $|f(x^*)| < +\infty$.

**Assumption 1.2** (Constraint qualification). The convex program (4.3) satisfies the refined Slater constraint qualification certificate [BV09].

**Assumption 1.3** (Compatibility with network topology). The uncertainties $a_k \in A_k \subset \mathbb{R}^n$ in (4.1) are compatible with $G$ if $a_{k_i}^i = a_{k_j}^j \neq 0 \Rightarrow (i, j) \in E$ for $k \in \{1, \ldots, m\}$ and $i, j \in V$, where $a_k^i \in \mathbb{R}$ denotes the $i^{th}$ element of $a_k \in \mathbb{R}^n$. Only the non-zero elements of $a_k \in \mathbb{R}^n$ are subject to uncertainty, i.e., the zero entries remain invariant.

A natural question which arises is how to choose the polyhedral representation of the uncertainty set $A_k$ that respects the distributed nature of the underlying graph topology (cf. Assumption 1.3) and therefore, preserves the 0-pattern of $a_k \in A_k \subset \mathbb{R}^n$ once row-wise uncertainty is present. This motivates the following definition leading to the procedure on polyhedral lifting.

**Definition 1.4** (Reduced polyhedral uncertainty representation). Let the reduced polyhedral uncertainty set $\tilde{A}_k$ be defined by

$$\tilde{A}_k = \{ \tilde{a}_k \in \mathbb{R}^{n_k} \mid \tilde{D}_k \tilde{a}_k \preceq \tilde{e}_k \},$$

where $n_k \in \mathbb{Z}_{>0}$ denotes the number of non-zero elements in $a_k \in \mathbb{R}^n$, $\tilde{D}_k \in \mathbb{R}^{r_k \times n_k}$, and $\tilde{e}_k \in \mathbb{R}^{r_k}$, with $r_k \in \mathbb{Z}_{>0}$ as the number of halfspaces that enclose the reduced uncertainty vector $\tilde{a}_k \in \mathbb{R}^{n_k}$.

W.l.o.g., the reduced polyhedral uncertainty set $\tilde{A}_k \subset \mathbb{R}^{n_k}$ can be lifted into the space $\tilde{A}_k \subset \mathbb{R}^n$ by introducing

$$D_k = \begin{pmatrix} \tilde{D}_k & 0_{r_k \times (n-n_k)} \\ 0_{2(n-n_k) \times n_k} & I_{(n-n_k)} \otimes \begin{pmatrix} 1 \\ -1 \end{pmatrix} \end{pmatrix} \in \mathbb{R}^{p_k \times n}, \quad k \in \{1, \ldots, m\},$$

and $e_k = (\tilde{e}_k, 0_{2(n-n_k)}) \in \mathbb{R}^{p_k}$, where $p_k = r_k + 2(n - n_k)$. Note that the above partitioning is valid by relabelling the network state $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ according to the given 0-pattern of the uncertainty vector $a_k \in A_k \subset \mathbb{R}^n$. The following example illustrates the procedure on polyhedral lifting.

**Example 1** (Polyhedral lifting). Consider an uncertain optimization problem of the form (4.1). Suppose $A \in A = (A_1 \times \cdots \times A_m) \subset \mathbb{R}^{m \times n}$ is given by the matrix

$$A = \begin{pmatrix} - & a_{11} & - \\ - & a_{12} & - \\ & a_{21} & a_{22} \end{pmatrix},$$

where $a_{12} = 0$, $n = 2$, $m = 2$, $n_1 = 1$, and $n_2 = 2$. The undirected (complete) graph compatible with $A \in A \subset \mathbb{R}^{m \times n}$ is denoted by $K_2$. Let the reduced polyhedral uncertainty sets $\tilde{A}_1, \tilde{A}_2 \subset \mathbb{R}^{n_k}$ be defined by

$$\tilde{A}_1 = \left\{ (a_{11}) \in \mathbb{R}^{n_1} \mid \begin{pmatrix} d_{11}^{[1]} \\ d_{21}^{[1]} \end{pmatrix} (a_{11}) \preceq \begin{pmatrix} e_{1}^{[1]} \\ e_{2}^{[1]} \end{pmatrix} \right\},$$

and
and, respectively,
\[ \tilde{A}_2 = \left\{ \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} \in \mathbb{R}^{n_2} \mid \begin{pmatrix} d_{11}^{[2]} & d_{12}^{[2]} \\ d_{21}^{[2]} & d_{22}^{[2]} \end{pmatrix} \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} \preceq \begin{pmatrix} e_{1}^{[2]} \\ e_{2}^{[2]} \end{pmatrix} \right\}, \]
where the number of halfspaces is \( r_1 = r_2 = 2 \). Then, by lifting \( \tilde{A}_k \subset \mathbb{R}^{n_k} \) into the superspace \( A_k \subset \mathbb{R}^n \) for \( k \in \{1, 2\} \), it follows
\[ A_1 = \left\{ a_1 \in \mathbb{R}^n \mid \begin{pmatrix} d_{11}^{[1]} & 0 \\ d_{21}^{[1]} & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{12} \end{pmatrix} \preceq \begin{pmatrix} e_{1}^{[1]} \\ 0 \\ 0 \end{pmatrix} \right\}, \]
and \( A_2 = \tilde{A}_2 \), respectively.

### 4.2 Robust Distributed Coordination Algorithm

This section proposes a framework that establishes optimal solutions of the robust counterpart (4.2), respectively (4.3), in terms of saddle points of an augmented Lagrangian function. The continuous-time coordination algorithm presented in this section builds upon saddle-point dynamics that are guaranteed to converge asymptotically to a point in the set of robust primal-dual solutions.

Let the stacked version of the convex reformulation (4.3) of the robust counterpart (4.2) take the form

\[
\begin{align*}
\text{minimize} &\quad f(x) \\
\text{subject to} &\quad E^\top \lambda + \vartheta = b, \\
&\quad D^\top \lambda = (\mathbb{1}_m \otimes x), \\
&\quad \lambda \succeq 0, \quad \vartheta \succeq 0,
\end{align*}
\]

with the block-structured matrices
\[
E^\top = \bigoplus_{k \in \{1, \ldots, m\}} e_k^\top, \quad \text{and} \quad D^\top = \bigoplus_{k \in \{1, \ldots, m\}} D_k^\top,
\]
where \( E^\top \in \mathbb{R}^{m \times p}, \ D^\top \in \mathbb{R}^{nm \times p} \), and \( b \in \mathbb{R}^m \). Note that the slack variable \( \vartheta \in \mathbb{R}^m \) is further introduced in (4.4). In addition, let \( \lambda = (\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^p \), where \( \mathbb{R}^p = \mathbb{R}^{p1} \times \cdots \times \mathbb{R}^{pm} \). The robust feasibility set of (4.4) is denoted by
\[
C = \{(x, \lambda, \vartheta) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \mid E^\top \lambda + \vartheta = b, \ D^\top \lambda = (\mathbb{1}_m \otimes x), \ \lambda \succeq 0, \ \vartheta \succeq 0\},
\]
and the robust solution set of (4.4) is denoted by
\[
X \times \Lambda \times \Theta = \{(x^*, \lambda^*, \vartheta^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \mid f(x^*) \leq f(x), \ \forall (x, \lambda, \vartheta) \in C\}.
\]

**Definition 2.1** (Lagrangian function). The *Lagrangian function* associated with the robust optimization problem (4.4) is the function \( \mathcal{L} : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m \times \)
4.2. ROBUST DISTRIBUTED COORDINATION ALGORITHM

\[ \mathbb{R}_{\geq 0}^p \times \mathbb{R}_{\geq 0}^m \to \mathbb{R} \] defined by

\[
\mathcal{L}(x, \lambda, \vartheta, \mu, \nu, \psi, \xi) = f(x) + \langle \mu, E^\top \lambda + \vartheta - b \rangle - \langle \psi, \lambda \rangle \\
+ \langle \nu, D^\top \lambda - (1_m \otimes x) \rangle - \langle \xi, \vartheta \rangle,
\]

where \( \mu = (\mu_1, \ldots, \mu_m) \in \mathbb{R}^m \), \( \nu = (\nu_1, \ldots, \nu_m) \in (\mathbb{R}^n)^m \), \( \psi = (\psi_1, \ldots, \psi_p) \in \mathbb{R}^p \geq 0 \), and \( \xi = (\xi_1, \ldots, \xi_m) \in \mathbb{R}^m_{\geq 0} \) are Lagrange multipliers.

**Definition 2.2** (Lagrangian dual function). The Lagrangian dual function associated with (4.4) is the function \( \phi : \mathbb{R}^m \times (\mathbb{R}^n)^m \to \mathbb{R} \cup \{-\infty, +\infty\} \) defined by

\[
\phi(\mu, \nu) = \inf_{(x, \lambda, \vartheta) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m} \mathcal{L}(x, \lambda, \vartheta, \mu, \nu, \psi, \xi)
\]

\[
= \begin{cases} 
- \langle \mu, b \rangle & \text{if } (1_m^\top \otimes I_n) \nu \in \partial f(x), \\
-\infty & \text{otherwise},
\end{cases}
\]

where \( \partial f(x) \) is the generalized gradient of \( f \) at \( x \in \mathbb{R}^n \) (cf. Definition 1.3 in Ch. 2). •

The Lagrangian dual problem associated with the primal problem (4.4) is

\[
\begin{align*}
\text{minimize} & \quad \langle \mu, b \rangle \\
\text{subject to} & \quad (1_m^\top \otimes I_n) \nu \in \partial f(x), \\
& \quad E\mu + D\nu \succeq 0, \quad \mu \succeq 0, \quad \nu \succeq 0.
\end{align*}
\]

Let the pair \( (\mu^*, \nu^*) \) denote a robust solution of the dual problem (4.5) and let the set of robust solutions be denoted by \( M \times N \subset \mathbb{R}^m \times (\mathbb{R}^n)^m \). The following result provides first order necessary and sufficient optimality conditions (cf. Theorem 3.6 in Ch. 2) for the robust optimization problem (4.4).

**Theorem 2.3** (KKT optimality conditions). Let the refined Slater constraint qualification certificate be satisfied by (4.4). The triplet \( (x^*, \lambda^*, \vartheta^*) \in X \times \Lambda \times \Theta \) is a KKT point of (4.4) iff there exist Lagrange multiplier \( \mu^* \in \mathbb{R}^m \) and \( \nu^* \in (\mathbb{R}^n)^m \), s.t.

\[
\begin{align*}
(1_m^\top \otimes I_n) \nu^* \in \partial f(x^*), \\
E\mu^* + D\nu^* & \succeq 0, \\
\mu^* & \succeq 0,
\end{align*}
\]

(stationarity conditions)

and

\[
\begin{align*}
E^\top \lambda^* + \vartheta^* & = b, \\
D^\top \lambda^* - (1_m^\top \otimes I_n) & = 0, \\
\lambda^* & \succeq 0, \quad \vartheta^* \succeq 0,
\end{align*}
\]

(primal feasibility)

and

\[
\begin{align*}
\langle \mu^*, \vartheta^* \rangle & = 0, \\
\langle E\mu^* + D\nu^*, \lambda^* \rangle & = 0,
\end{align*}
\]

(complementary slackness)
Remark 2.4 (Strong duality). Note that under Assumption 1.1 and 1.2, the robust optimal primal-dual values of (4.4) and respectively (4.5) are finite and coincide, i.e., strong duality holds.

The following result characterizes the robust primal-dual solutions of (4.4) and (4.5) in terms of saddle points (cf. Definition 2.3 in Ch. 3) of an augmented Lagrangian function associated to the robust optimization problem (4.4).

Proposition 2.5 (Robust primal-dual solutions via saddle-point characterization). Given \( \kappa_1, \kappa_2 \in \mathbb{R}_{\geq 0} \), let the augmented Lagrangian function \( \mathcal{L}^\kappa : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^m \times (\mathbb{R}^n)^m \to \mathbb{R} \) be defined by

\[
\mathcal{L}^\kappa(x, \lambda, \vartheta, \mu, \nu) = f(x) + \frac{1}{2} \| E^\top \lambda + \vartheta - b \|^2 + \frac{1}{2} \| D^\top \lambda - (1_m \otimes x) \|^2 \\
+ \langle \mu, E^\top \lambda + \vartheta - b \rangle + \langle \nu, D^\top \lambda - (1_m \otimes x) \rangle \\
+ \kappa_1 \langle 1_p, [-\lambda]^+ \rangle + \kappa_2 \langle 1_m, [-\vartheta]^+ \rangle.
\]

Then, the function \((x, \lambda, \vartheta, \mu, \nu) \mapsto \mathcal{L}^\kappa(x, \lambda, \vartheta, \mu, \nu)\) is convex-concave, and

(i) if \((x^*, \lambda^*, \vartheta^*) \in X \times \Lambda \times \Theta\) and \((\mu^*, \nu^*) \in M \times N\) are robust primal-dual solutions of (4.4) and (4.5), then \((x^*, \lambda^*, \vartheta^*, \mu^*, \nu^*) \in X \times \Lambda \times \Theta \times M \times N\) is a saddle point of \( \mathcal{L}^\kappa \) for any \( \kappa_1 \geq \| E\mu^* + D\nu^* \|_\infty \) and \( \kappa_2 \geq \| \mu^* \|_\infty \),

(ii) if \((\tilde{x}, \tilde{\lambda}, \tilde{\vartheta}, \tilde{\mu}, \tilde{\nu}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^m \times (\mathbb{R}^n)^m\) is a saddle point of \( \mathcal{L}^\kappa \) with \( \kappa_1 > \| E\mu^* + D\nu^* \|_\infty \) and \( \kappa_2 > \| \mu^* \|_\infty \) for some \((\mu^*, \nu^*) \in M \times N\), then \((\tilde{x}, \tilde{\lambda}, \tilde{\vartheta}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m\) is a robust primal solution of (4.4).

Proof. Note that the augmented Lagrangian function \( \mathcal{L}^\kappa \) in (4.6) is strictly convex in its primal variables \((x, \lambda, \vartheta)\) (due to the regularization terms) and concave (in fact, linear) in its co-variables \((\mu, \nu)\).

(i) Let \((x^*, \lambda^*, \vartheta^*) \in X \times \Lambda \times \Theta\) and \((\mu^*, \nu^*) \in M \times N\) be robust primal-dual solutions of (4.4) and (4.5), respectively. By strong duality and the bounds \( \kappa_1 \geq \| E\mu^* + D\nu^* \|_\infty \) and \( \kappa_2 \geq \| \mu^* \|_\infty \), for any \((x, \lambda, \vartheta) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m\), it follows

\[
\mathcal{L}^\kappa(x, \lambda, \vartheta, \mu^*, \nu^*) = f(x) + \frac{1}{2} \| E^\top \lambda + \vartheta - b \|^2 + \frac{1}{2} \| D^\top \lambda - (1_m \otimes x) \|^2 \\
+ \langle \mu^*, E^\top \lambda + \vartheta - b \rangle + \langle \nu^*, D^\top \lambda - (1_m \otimes x) \rangle \\
+ \kappa_1 \langle 1_p, [-\lambda]^+ \rangle + \kappa_2 \langle 1_m, [-\vartheta]^+ \rangle \\
\geq f(x) + \langle \mu^*, E^\top \lambda + \vartheta - b \rangle + \langle \nu^*, D^\top \lambda - (1_m \otimes x) \rangle \\
+ \underbrace{\langle E\mu^* + D\nu^*, [-\lambda]^+ \rangle}_{=: \psi^*} + \underbrace{\langle \mu^*, [-\vartheta]^+ \rangle}_{=: \xi^*} \\
\geq \inf_{(x, \lambda, \vartheta) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m} (f(x) + \langle \mu^*, E^\top \lambda + \vartheta - b \rangle - \langle \psi^*, \lambda \rangle) \\
+ \langle \nu^*, D^\top \lambda - (1_m \otimes x) \rangle - \langle \xi^*, \vartheta \rangle \\
= \phi(\mu^*, \nu^*) = \mathcal{L}(x^*, \lambda^*, \vartheta^*, \mu^*, \nu^*, \psi^*, \xi^*) = f(x^*) \\
= \mathcal{L}^\kappa(x^*, \lambda^*, \vartheta^*, \mu^*, \nu^*).
4.2. ROBUST DISTRIBUTED COORDINATION ALGORITHM

Since $\mathcal{L}^\kappa$ is linear in the dual variables $(\mu, \nu)$, the fact that $\mathcal{L}^\kappa(x^*, \lambda^*, \vartheta^*, \mu, \nu) = \mathcal{L}^\kappa(x^*, \lambda^*, \vartheta^*, \mu^*, \nu^*)$ for all $(\mu, \nu) \in \mathbb{R}^m \times (\mathbb{R}^n)^m$ is an immediate consequence.

(ii) By contradiction, assume that $(\bar{x}, \lambda, \bar{\vartheta}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ is a saddle point of the augmented Lagrangian function (4.6) with the bounds $\kappa_1 > \| E\mu^* + D\nu^* \| \infty$ and $\kappa_2 > \| \mu^* \| \infty$ for some $(\mu^*, \nu^*) \in M \times N$. Furthermore, suppose the triplet $(\bar{x}, \lambda, \bar{\vartheta})$ is not a robust primal solution of (4.4). Let $(x^*, \lambda^*, \vartheta^*) \in X \times \Lambda \times \Theta$. For the fixed point $(x, \lambda, \vartheta)$, the map $(\mu, \nu) \mapsto \mathcal{L}^\kappa(x, \lambda, \vartheta, \mu, \nu)$ is concave and differentiable. Necessary conditions for $(\bar{x}, \lambda, \bar{\vartheta}, \bar{\mu}, \bar{\nu})$ to be a saddle point of $\mathcal{L}^\kappa$ are that $E^\top \lambda + \bar{\vartheta} - b = 0$ and $D^\top \lambda - (\mathbb{1}_m \otimes \bar{x}) = 0$. Hence, $\mathcal{L}^\kappa(x^*, \lambda^*, \vartheta^*, \bar{\mu}, \bar{\nu}) \geq \mathcal{L}^\kappa(\bar{x}, \lambda, \bar{\vartheta}, \bar{\mu}, \bar{\nu})$, and

$$f(x^*) \geq f(\bar{x}) + \kappa_1 \langle 1_p, [-\bar{x}]^+ \rangle + \kappa_2 \langle 1_m, [-\bar{\vartheta}]^+ \rangle. \quad (4.7)$$

If $\bar{\lambda} \succeq 0$ and $\bar{\vartheta} \succeq 0$, then $f(x^*) \geq f(\bar{x})$, and $(\bar{x}, \lambda, \bar{\vartheta}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ would be a robust primal solution. However, if $\lambda \nless 0$ and $\vartheta \nless 0$, it follows

$$f(\bar{x}) = f(\bar{x}) + \langle \mu^*, E^\top \lambda + \bar{\vartheta} - b \rangle - \langle \mu^*, E^\top \hat{\lambda} + \bar{\vartheta} - b \rangle$$

$$+ \langle \nu^*, D^\top \hat{\lambda} - (\mathbb{1}_m \otimes \bar{x}) \rangle - \langle \nu^*, D^\top \hat{\lambda} - (\mathbb{1}_m \otimes \bar{x}) \rangle$$

$$+ \langle \psi^*, \lambda \rangle - \langle \psi^*, \hat{\lambda} \rangle + \langle \xi^*, \bar{\vartheta} \rangle - \langle \xi^*, \bar{\vartheta} \rangle$$

$$\geq \inf_{(\bar{x}, \lambda, \bar{\vartheta}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m} (f(\bar{x}) + \langle \mu^*, E^\top \lambda + \bar{\vartheta} - b \rangle - \langle \psi^*, \lambda \rangle)$$

$$+ \langle \nu^*, D^\top \hat{\lambda} - (\mathbb{1}_m \otimes \bar{x}) \rangle - \langle \xi^*, \bar{\vartheta} \rangle) - \langle \mu^*, E^\top \hat{\lambda} + \bar{\vartheta} - b \rangle$$

$$- \langle \psi^*, D^\top \hat{\lambda} - (\mathbb{1}_m \otimes \bar{x}) \rangle + \langle \psi^*, \hat{\lambda} \rangle + \langle \xi^*, \bar{\vartheta} \rangle$$

$$= \phi(\mu^*, \nu^*) + \langle \psi^*, \lambda \rangle + \langle \xi^*, \bar{\vartheta} \rangle$$

$$> f(x^*) - \kappa_1 \langle 1_p, [-\bar{\lambda}]^+ \rangle - \kappa_2 \langle 1_m, [-\bar{\vartheta}]^+ \rangle,$$

which contradicts (4.7). Similar arguments hold for both cases $\bar{\lambda} \succeq 0$ and $\bar{\vartheta} \nless 0$, or $\lambda \nless 0$ and $\vartheta \succeq 0$. This concludes the proof. □

4.2.1 Robust Distributed Saddle-Point Dynamics

Here, we build on the previous result in Proposition 2.5 that allows us to search for saddle points of $\mathcal{L}^\kappa$ via its associated saddle-point dynamics. The nonsmooth character of the mapping $\mathcal{L}^\kappa : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m \rightarrow \mathbb{R}$ suggests that the dynamics take a set-valued form, i.e.,

$$\dot{x}(t) \in -\partial_x \mathcal{L}^\kappa(x, \lambda, \vartheta, \mu, \nu)(t), \quad x(t_0) = x_0 \in \mathbb{R}^n, \quad (4.8a)$$

$$\dot{\lambda}(t) \in -\partial_\lambda \mathcal{L}^\kappa(x, \lambda, \vartheta, \mu, \nu)(t), \quad \lambda(t_0) = \lambda_0 \in \mathbb{R}^p, \quad (4.8b)$$

$$\dot{\vartheta}(t) \in -\partial_\vartheta \mathcal{L}^\kappa(x, \lambda, \vartheta, \mu, \nu)(t), \quad \vartheta(t_0) = \vartheta_0 \in \mathbb{R}^m, \quad (4.8c)$$

$$\dot{\mu}(t) \in +\partial_\mu \mathcal{L}^\kappa(x, \lambda, \vartheta, \mu, \nu)(t), \quad \mu(t_0) = \mu_0 \in \mathbb{R}^m, \quad (4.8d)$$

$$\dot{\nu}(t) \in +\partial_\nu \mathcal{L}^\kappa(x, \lambda, \vartheta, \mu, \nu)(t), \quad \nu(t_0) = \nu_0 \in (\mathbb{R}^n)^m, \quad (4.8e)$$
for a.a. \( t \in [t_0, +\infty) \). Specifically, the set-valued dynamics (4.8) defined over \( \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m \to \mathbb{R} \) take the form

\[
\begin{align*}
\dot{x}(t) + (\mathbb{1}_m, \mathbb{1}_m) x(t) - (\mathbb{1}_m \otimes I_n) (D^T \lambda(t) + \nu(t)) &\in -\partial f(x(t)), \\
\dot{\lambda}(t) + D(D^T \lambda(t) - (\mathbb{1}_m \otimes x(t)) + \nu(t)) + E\mu(t) &\in -\kappa_1 \partial [-\lambda(t)]^+, \\
\dot{\nu}(t) + (E^T \lambda(t) + \dot{\vartheta}(t) - b) &\in -\kappa_2 \partial [-\dot{\vartheta}(t)]^+,
\end{align*}
\] (4.9a, 4.9b, 4.9c)

for a.a. \( t \in [t_0, +\infty) \) with initial condition \( (x_0, \lambda_0, \dot{\vartheta}_0, \mu_0, \nu_0) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m \). Propositions 1.7 in Chapter 2 and Lemma 2.2 in Chapter 2 guarantee the existence of robust solutions \( (x, \lambda, \dot{\vartheta}, \mu, \nu) \) of (4.9), where the solutions are understood in the sense of Krasovskii (cf. Definition 2.6 in Ch. 2). We use the mapping \( F^\circ : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m \to \mathbb{R}^n \times \mathbb{R}^p \times (\mathbb{R}^n)^m \) to refer to (4.9). Note that if \( (x^*, \lambda^*, \dot{\vartheta}^*, \mu^*, \nu^*) \in \text{eq}(F^\circ) \), then \( (x^*, \lambda^*, \dot{\vartheta}^*) \in \mathcal{X} \times \Lambda \times \Theta \) is a robust primal solution of (4.4) (cf. Proposition 2.5).

### 4.2.2 Convergence Analysis

Here, we establish asymptotic convergence properties of the saddle-point dynamics (4.9) using classical notions of stability analysis. Specifically, we use LaSalle’s Invariance Principle for differential inclusions to prove that the robust primal-dual minimizer are globally asymptotically stable under \( F^\circ \) and each solution of the dynamics converges to the set of robust primal-dual solutions \( \mathcal{X} \times \Lambda \times \Theta \times M \times N \).

**Theorem 2.6** (Asymptotic convergence). Let \( \kappa_1, \kappa_2 \in \mathbb{R}_{\geq 0} \), \( (x^*, \lambda^*, \dot{\vartheta}^*) \in \mathcal{X} \times \Lambda \times \Theta \), and \( (\mu^*, \nu^*) \in M \times N \). Define the map \( d^\circ : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m \to \mathbb{R}_{\geq 0} \) by

\[
d^\circ(x, \lambda, \vartheta, \mu, \nu) = \frac{1}{2} \|x - x^*\|^2 + \frac{1}{2} \|\lambda - \lambda^*\|^2 + \frac{1}{2} \|\vartheta - \dot{\vartheta}^*\|^2 + \frac{1}{2} \|\mu - \mu^*\|^2 + \frac{1}{2} \|\nu - \nu^*\|^2.
\] (4.10)

If \( \kappa_1 > \|E\mu^* + D\nu^*\|_{\infty} \) and \( \kappa_2 > \|\mu^*\|_{\infty} \), then \( (\mathcal{L}_{F^\circ} d^\circ)(x, \lambda, \vartheta, \mu, \nu) \subset (-\infty, 0] \) holds for all \( (x, \lambda, \vartheta, \mu, \nu) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m \) and any trajectory \( (x(t), \lambda(t), \vartheta(t), \mu(t), \nu(t)) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m \) of (4.9) converges asymptotically to the set of robust primal-dual minimizer \( \mathcal{X} \times \Lambda \times \Theta \times M \times N \).

**Proof.** The following proof is similar to the first part of the proof of Theorem 3.1 in Chapter 3. However, we provide the proof for reasons of completeness. By definition of (4.9), the set-valued map \( F^\circ \) is locally bounded, upper semi-continuous and takes nonempty, convex, and compact values (cf. Proposition 1.7 in Ch. 2). Moreover, by Proposition 2.5(i), the quintuple \( (x^*, \lambda^*, \dot{\vartheta}^*, \mu^*, \nu^*) \) identifies a saddle point of (4.6) if \( \kappa_1 \geq \|E\mu^* + D\nu^*\|_{\infty} \) and \( \kappa_2 \geq \|\mu^*\|_{\infty} \). Note that \( d^\circ \in C^1(\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m \times (\mathbb{R}^n)^m) \).
Similarly, since \((\mathbb{R}^n)^m, \mathbb{R}_{\geq 0}\). Furthermore, the candidate Lyapunov function (4.10) is proper. Let \(\psi \in (\mathcal{L}_{F^\diamond} d^\Diamond)(x, \lambda, \vartheta, \mu, \nu)\). By definition of the Lie-derivative (cf. Definition 2.3 in Ch. 2), there exists

\[
\xi = \begin{pmatrix}
\langle \mathbb{1}_m^\top \otimes I_n \rangle (D^\top \lambda + \nu) - \langle \mathbb{1}_m, \mathbb{1}_m \rangle x - \pi_f \\
-D(D^\top \lambda - (\mathbb{1}_m \otimes x) + \nu) - E\mu - E(E^\top \lambda + \vartheta - b) - \pi_\lambda^+ \\
-E^\top \lambda + \vartheta - b \\
D^\top \lambda - (\mathbb{1}_m \otimes x)
\end{pmatrix} \in F^\Diamond,
\]

where \(\pi_f \in \partial f(x)\), \(\pi_\lambda^+ \in \kappa_1 \partial [\lambda]^+\), and \(\pi_\varphi^+ \in \kappa_2 \partial [-\vartheta]^+\), such that

\[
\psi = \langle \nabla d^\Diamond (x, \lambda, \vartheta, \mu, \nu), \xi \rangle
= \langle x - x^*, (\mathbb{1}_m^\top \otimes I_n) (D^\top \lambda + \nu) - \langle \mathbb{1}_m, \mathbb{1}_m \rangle x - \pi_f \rangle
+ \langle \lambda - \lambda^*, -D(D^\top \lambda - (\mathbb{1}_m \otimes x) + \nu) - E\mu - E(E^\top \lambda + \vartheta - b) - \pi_\lambda^+ \rangle
+ \langle \vartheta - \vartheta^*, -(E^\top \lambda + \vartheta - b) - \mu - \pi_\varphi^+ \rangle
+ \langle \mu - \mu^*, E^\top \lambda + \vartheta - b \rangle
+ \langle \nu - \nu^*, D^\top \lambda - (\mathbb{1}_m \otimes x) \rangle.
\]

Note that \(\mathcal{L}^\kappa\) is convex in its primal variables \((x, \lambda, \vartheta)\), and

\[
-\langle \mathbb{1}_m^\top \otimes I_n \rangle (D^\top \lambda + \nu) + \langle \mathbb{1}_m, \mathbb{1}_m \rangle x + \pi_f \in \partial x \mathcal{L}^\kappa,
D(D^\top \lambda - (\mathbb{1}_m \otimes x) + \nu) + E\mu + E(E^\top \lambda + \vartheta - b) + \pi_\lambda^+ \in \partial \lambda \mathcal{L}^\kappa,
(E^\top \lambda + \vartheta - b) + \mu + \pi_\varphi^+ \in \partial \vartheta \mathcal{L}^\kappa.
\]

By the 1\(^{st}\)-order convexity condition in \((x, \lambda, \vartheta)\), it follows

\[
\mathcal{L}^\kappa (x^*, \lambda^*, \vartheta^*, \mu, \nu) \geq \mathcal{L}^\kappa (x, \lambda, \vartheta, \mu, \nu) + \langle x - x^*, (\mathbb{1}_m^\top \otimes I_n) (D^\top \lambda + \nu) - \langle \mathbb{1}_m, \mathbb{1}_m \rangle x - \pi_f \rangle
+ \langle \lambda - \lambda^*, -D(D^\top \lambda - (\mathbb{1}_m \otimes x) + \nu) - E\mu \rangle
- E(E^\top \lambda + \vartheta - b) - \pi_\lambda^+ + \langle \vartheta - \vartheta^*, -(E^\top \lambda + \vartheta - b) - \mu - \pi_\varphi^+ \rangle.
\]

Similarly, since \(\mathcal{L}^\kappa\) is concave (in fact, linear) in its dual variables \((\mu, \nu)\), and

\[
E^\top \lambda + \vartheta - b \in \partial \mu \mathcal{L}^\kappa,
D^\top \lambda - (\mathbb{1}_m \otimes x) \in \partial \nu \mathcal{L}^\kappa,
\]

the 1\(^{st}\)-order convexity condition in \((\mu, \nu)\) yields

\[
\mathcal{L}^\kappa (x, \lambda, \vartheta, \mu, \nu) = \mathcal{L}^\kappa (x, \lambda, \vartheta, \mu^*, \nu^*) + \langle \mu - \mu^*, E^\top \lambda + \vartheta - b \rangle
+ \langle \nu - \nu^*, D^\top \lambda - (\mathbb{1}_m \otimes x) \rangle.
\]
Substituting (4.12) and (4.13) into equation (4.11) yields
\[
\psi \leq \mathcal{L}^\kappa(x^*, \lambda^*, \vartheta^*, \mu, \nu) - \mathcal{L}^\kappa(x^*, \lambda^*, \vartheta^*, \mu^*, \nu^*) \\
+ \mathcal{L}^\kappa(x^*, \lambda^*, \vartheta^*, \mu^*, \nu^*) - \mathcal{L}^\kappa(x, \lambda, \vartheta, \mu^*, \nu^*) \leq 0,
\]
(4.14)
since \((x^*, \lambda^*, \vartheta^*, \mu^*, \nu^*)\) is a saddle point of \(\mathcal{L}^\kappa\). Since \(\psi\) is chosen arbitrary, the inclusion \((\mathcal{L}_{F \circ d^\kappa}(x, \lambda, \vartheta, \mu, \nu) \subset (-\infty, 0]\) holds for all \((x, \lambda, \vartheta, \mu, \nu) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m\). Hence, by Theorem 2.4 in Chapter 2, the point \((x^*, \lambda^*, \vartheta^*, \mu^*, \nu^*) \in \text{eq}(F^\kappa)\) is a strongly stable equilibrium point of (4.9).

For any \(\gamma \in \mathbb{R}_{>0}\), the \(\gamma\)-sublevel set \(d^\kappa - 1(\leq \gamma)\) is strongly invariant with respect to (4.9). Since \(d^\kappa\) is proper, the \(\gamma\)-sublevel set \(d^\kappa - 1(\leq \gamma)\) is compact. By Theorem 2.5 in Ch. 2, any absolutely continuous trajectory \((x, \lambda, \vartheta, \mu, \nu) : [t_0, +\infty) \rightarrow \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m\) of (4.9) starting in \(d^\kappa - 1(\leq \gamma)\) converges to the largest weakly invariant set
\[
\Omega \subset \text{cl}\{(x, \lambda, \vartheta, \mu, \nu) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m \mid 0 \in (\mathcal{L}_{F \circ d^\kappa}(x, \lambda, \vartheta, \mu, \nu)\}
\cap d^\kappa - 1(\leq \gamma).
\]
Since the set-valued map \(F^\kappa\) is upper semi-continuous and \(d^\kappa \in C^1(\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m, \mathbb{R}_{\geq 0})\), it follows that \((x, \lambda, \vartheta, \mu, \nu) \mapsto (\mathcal{L}_{F \circ d^\kappa}(x, \lambda, \vartheta, \mu, \nu)\) is also upper semi-continuous. Hence, closedness of the set
\[
\text{cl}\{(x, \lambda, \vartheta, \mu, \nu) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m \mid 0 \in (\mathcal{L}_{F \circ d^\kappa}(x, \lambda, \vartheta, \mu, \nu)\} \cap d^\kappa - 1(\leq \gamma)
\]
is an immediate consequence (cf. [Cor08]).

To show that \(\Omega \subset X \times \Lambda \times \Theta \times M \times N\), take a point \((\bar{x}, \bar{\lambda}, \bar{\vartheta}, \bar{\mu}, \bar{\nu}) \in \Omega\). Then, from inequality (4.14), it follows \(\mathcal{L}^\kappa(x^*, \lambda^*, \vartheta^*, \mu^*, \nu^*) - \mathcal{L}^\kappa(\bar{x}, \bar{\lambda}, \bar{\vartheta}, \bar{\mu}, \nu^*) = 0\). Hence,
\[
\tilde{\mathcal{L}}^\kappa(\bar{x}, \bar{\lambda}, \bar{\vartheta}, \bar{\mu}, \nu^*) - \frac{1}{2} \|D^\top \bar{\lambda} - (\mathbb{1}_m \otimes \bar{x})\|^2 - \frac{1}{2} \|E^\top \bar{\lambda} + \bar{\vartheta} - b\|^2 = 0,
\]
(4.15)
where the mapping \(\tilde{\mathcal{L}}^\kappa : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m \rightarrow \mathbb{R}\) is defined by
\[
\tilde{\mathcal{L}}^\kappa(\bar{x}, \bar{\lambda}, \bar{\vartheta}, \bar{\mu}, \nu^*) = f(x^*) - f(\bar{x}) - \langle \mu^*, E^\top \bar{\lambda} + \bar{\vartheta} - b \rangle - \langle \nu^*, D^\top \bar{\lambda} - (\mathbb{1}_m \otimes \bar{x}) \rangle \\
- \kappa_1 \langle \mathbb{1}_p, [-\bar{\lambda}]^+ \rangle - \kappa_2 \langle \mathbb{1}_m, [-\bar{\vartheta}]^+ \rangle.
\]
By Assumption 1.2, strong duality holds and we conclude
\[
\tilde{\mathcal{L}}^\kappa(\bar{x}, \bar{\lambda}, \bar{\vartheta}, \bar{\mu}, \nu^*) = \phi(\mu^*, \nu^*) - f(\bar{x}) - \langle \mu^*, E^\top \bar{\lambda} + \bar{\vartheta} - b \rangle - \langle \nu^*, D^\top \bar{\lambda} - (\mathbb{1}_m \otimes \bar{x}) \rangle \\
- \kappa_1 \langle \mathbb{1}_p, [-\bar{\lambda}]^+ \rangle - \kappa_2 \langle \mathbb{1}_m, [-\bar{\vartheta}]^+ \rangle
\]
\[
= \inf_{(\bar{x}, \bar{\lambda}, \bar{\vartheta}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m} \left( f(\bar{x}) + \langle \mu^*, E^\top \bar{\lambda} + \bar{\vartheta} - b \rangle + \langle \nu^*, D^\top \bar{\lambda} - (\mathbb{1}_m \otimes \bar{x}) \rangle \\
+ \langle E \mu^* + D \nu^* - \bar{\lambda} \rangle + \langle \mu^* - \bar{\vartheta} \rangle - f(\bar{x}) - \langle \mu^*, E^\top \bar{\lambda} + \bar{\vartheta} - b \rangle \\
- \langle \nu^*, D^\top \bar{\lambda} - (\mathbb{1}_m \otimes \bar{x}) \rangle - \kappa_1 \langle \mathbb{1}_p, [-\bar{\lambda}]^+ \rangle - \kappa_2 \langle \mathbb{1}_m, [-\bar{\vartheta}]^+ \rangle \
\leq \langle E \mu^* + D \nu^* - \bar{\lambda} \rangle - \kappa_1 \langle \mathbb{1}_p, [-\bar{\lambda}]^+ \rangle + \langle \mu^*, -\bar{\vartheta} \rangle - \kappa_2 \langle \mathbb{1}_m, [-\bar{\vartheta}]^+ \rangle.
\]
4.2. ROBUST DISTRIBUTED COORDINATION ALGORITHM

If \( \kappa_1 \geq \|E\mu^* + Du^*\|_\infty \) and \( \kappa_2 \geq \|\mu^*\|_\infty \), then \( \tilde{L}^\kappa(\tilde{x}, \bar{\lambda}, \bar{\vartheta}, \mu^*, \nu^*) \leq 0 \), which implies \( D^\top \tilde{\lambda} - (\mathbb{1}_m \otimes \tilde{x}) = 0 \) and \( E^\top \tilde{\lambda} + \bar{\vartheta} - b = 0 \). However, if \( \kappa_1 > \|E\mu^* + Du^*\|_\infty \) and \( \kappa_2 > \|\mu^*\|_\infty \), then it follows

\[
\begin{align*}
\tilde{L}^\kappa(\tilde{x}, \bar{\lambda}, \bar{\vartheta}, \mu^*, \nu^*) &< \langle E\mu^* + Du^*, -\tilde{\lambda} \rangle - \langle E\mu^* + Du^*, [-\tilde{\lambda}]^+ \rangle \\
&+ \langle \mu^*, -\bar{\vartheta} \rangle - \langle \mu^*, [-\bar{\vartheta}]^+ \rangle.
\end{align*}
\]

For \( \tilde{\lambda} \not\geq 0 \) and \( \bar{\vartheta} \not\geq 0 \), we have \( \tilde{L}^\kappa(\tilde{x}, \bar{\lambda}, \bar{\vartheta}, \mu^*, \nu^*) < 0 \) which contradicts (4.15). Hence, it must be that \( \tilde{\lambda} \geq 0 \) and \( \bar{\vartheta} \geq 0 \). Similar arguments hold for both cases \( \tilde{\lambda} \geq 0 \) and \( \bar{\vartheta} \not\geq 0 \), or \( \tilde{\lambda} \not\geq 0 \) and \( \bar{\vartheta} \geq 0 \). From (4.15), we conclude \( \tilde{L}^\kappa(\tilde{x}, \bar{\lambda}, \bar{\vartheta}, \mu^*, \nu^*) = f(x^*) - f(\tilde{x}) = 0 \). Thus, if \( (\tilde{x}, \bar{\lambda}, \bar{\vartheta}, \mu^*, \nu^*) \in \Omega \), then all primal feasibility conditions are satisfied and \( (\tilde{x}, \bar{\lambda}, \bar{\vartheta}) \) is a robust primal solution of (4.4).

Since \( \Omega \) is weakly invariant, there exists an absolutely continuous trajectory starting from \( (\tilde{x}, \bar{\lambda}, \bar{\vartheta}, \mu^*, \nu^*) \in \Omega \) that remains in \( \Omega \) only if \( 0 \in \partial_x L^\kappa(\tilde{x}, \bar{\lambda}, \bar{\vartheta}, \mu^*, \nu^*) \), \( 0 \in \partial_{\bar{\lambda}} L^\kappa(\tilde{x}, \bar{\lambda}, \bar{\vartheta}, \mu^*, \nu^*) \), \( 0 \in \partial_{\bar{\vartheta}} L^\kappa(\tilde{x}, \bar{\lambda}, \bar{\vartheta}, \mu^*, \nu^*) \), and furthermore \( 0 \in \partial_{\nu^*} L^\kappa(\tilde{x}, \bar{\lambda}, \bar{\vartheta}, \mu^*, \nu^*) \), i.e.,

\[
\begin{align*}
(\mathbb{1}_m \otimes I_n)(D^\top \bar{\lambda} + \nu) - (\mathbb{1}_m, \mathbb{1}_m)\tilde{x} &\in \partial f(\tilde{x}), \quad (4.16a) \\
-D(D^\top \bar{\lambda} - (\mathbb{1}_m \otimes \tilde{x}) + \nu) - E\bar{\mu} - E(E^\top \bar{\lambda} + \bar{\vartheta} - b) &\in \kappa_1 \partial[-\tilde{\lambda}]^+, \quad (4.16b) \\
-(E^\top \bar{\lambda} + \bar{\vartheta} - b) - \bar{\mu} &\in \kappa_2 \partial[-\bar{\vartheta}]^+, \quad (4.16c) \\
E^\top \bar{\lambda} + \bar{\vartheta} - b = 0, \quad (4.16d) \\
D^\top \bar{\lambda} - (\mathbb{1}_m \otimes \tilde{x}) = 0. \quad (4.16e)
\end{align*}
\]

In what follows, we show that all dual feasibility conditions of (4.5) are satisfied. Since strong duality holds (cf. Assumption 1.2), we have

\[
f(x^*) = \inf_{(\tilde{x}, \bar{\lambda}, \bar{\vartheta}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m} \langle f(\tilde{x}) + \langle \bar{\mu}, E^\top \bar{\lambda} + \bar{\vartheta} - b \rangle + \langle E\bar{\mu} + D\nu, -\tilde{\lambda} \rangle \\
+ \langle \nu, D^\top \tilde{\lambda} - (\mathbb{1}_m \otimes \tilde{x}) \rangle + \langle \bar{\mu}, -\bar{\vartheta} \rangle \rangle \\
\leq f(\tilde{x}) + \langle \bar{\mu}, E^\top \bar{\lambda} + \bar{\vartheta} - b \rangle + \langle E\bar{\mu} + D\nu, -\tilde{\lambda} \rangle \\
+ \langle \nu, D^\top \tilde{\lambda} - (\mathbb{1}_m \otimes \tilde{x}) \rangle + \langle \bar{\mu}, -\bar{\vartheta} \rangle \\
\leq f(\tilde{x}).
\]

Note that the conditions \( E^\top \bar{\lambda} + \bar{\vartheta} - b = 0, D^\top \bar{\lambda} - (\mathbb{1}_m \otimes \tilde{x}) = 0, \bar{\lambda} \geq 0 \), and \( \bar{\vartheta} \geq 0 \) imply that \( E\bar{\mu} + D\nu \geq 0 \) and \( \bar{\mu} \geq 0 \). Since \( f(x^*) = f(\tilde{x}) \), we conclude that \( \langle E\bar{\mu} + D\nu, -\tilde{\lambda} \rangle = 0 \) and \( \langle \bar{\mu}, -\bar{\vartheta} \rangle = 0 \). Hence, for (4.16a)-(4.16c), we have

\[
\begin{align*}
(\mathbb{1}_m \otimes I_n)\tilde{\nu} &\in \partial f(\tilde{x}), \quad (4.17a) \\
-(E\bar{\mu} + D\nu) &\in \kappa_1 \partial[-\tilde{\lambda}]^+, \quad (4.17b) \\
-\bar{\mu} &\in \kappa_2 \partial[-\bar{\vartheta}]^+. \quad (4.17c)
\end{align*}
\]

Consider now the inclusion (4.17b). If \( E\bar{\mu} + D\nu = 0 \), then \( \tilde{\lambda} > 0 \) and therefore,
\(\kappa_1 \partial[-\tilde{\lambda}]^+ = \{0\}\). However, if \(E\tilde{\mu} + D\tilde{\nu} \succ 0\), then \(\tilde{\lambda} = 0\), and \(- (E\tilde{\mu} + D\tilde{\nu}) \in [-\kappa_1, 0] \subset (-\infty, 0]\). We conclude that \(E\tilde{\mu} + D\tilde{\nu} \succeq 0\). Similar arguments hold for inclusion (4.17c). Therefore, all dual feasibility conditions are satisfied. Thus, given \((\mu^*, \nu^*) \in M \times N\), the point \((\tilde{x}, \tilde{\lambda}, \tilde{\vartheta}, \tilde{\mu}, \tilde{\nu}) \in \Omega\) satisfies the KKT-conditions (cf. Theorem 3.6 in Ch. 2) and hence, \(\Omega \subset X \times \Lambda \times \Theta \times M \times N\). Since the initial choice \(\gamma \in \mathbb{R}_{>0}\) is arbitrary, we deduce that convergence of solutions \((x, \lambda, \vartheta, \mu, \nu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^m \times (\mathbb{R}^n)^m\) of (4.9) to the set \(X \times \Lambda \times \Theta \times M \times N\) holds from any point in \(\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^m \times (\mathbb{R}^n)^m\), concluding the proof.

The following result states that the saddle-point dynamics (4.9) that possibly possess non-isolated equilibria [BB03], asymptotically converge to a point in the set of primal-dual minimizer \(X \times \Lambda \times \Theta \times M \times N\).

**Corollary 2.7 (Point-wise convergence).** Let \(\kappa_1, \kappa_2 \in \mathbb{R}_{\geq 0}\), and let \(\gamma \in \mathbb{R}_{>0}\). If

\[
\kappa_1 \geq \sup_{(x, \lambda, \vartheta, \mu, \nu) \in (X \times \Lambda \times \Theta \times M \times N)} \|E\mu + D\nu\|_{\infty}, \\
\kappa_2 \geq \sup_{(x, \lambda, \vartheta, \mu, \nu) \in (X \times \Lambda \times \Theta \times M \times N)} \|\mu^*\|_{\infty},
\]

then any trajectory \((x, \lambda, \vartheta, \mu, \nu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^m \times (\mathbb{R}^n)^m\) starting in \(d^\infty(\leq \gamma)\) converges asymptotically to a point in \(X \times \Lambda \times \Theta \times M \times N\).

**Proof.** Note that if \(\kappa_1\) and \(\kappa_2\) satisfy the bounds (4.18a) and (4.18b), respectively, then in particular \(\kappa_1 \geq \|E\mu^* + D\nu^*\|_{\infty}\) and \(\kappa_2 \geq \|\mu^*\|_{\infty}\). Thus, the \(\gamma\)-sublevel set \(d^\infty(\leq \gamma)\) is strongly invariant under \(F^\infty\) since \((L_{F^\infty}d^\infty)(x, \lambda, \vartheta, \mu, \nu) \subset (-\infty, 0]\) for all \((x, \lambda, \vartheta, \mu, \nu) \in d^\infty(\leq \gamma)\) (cf. Theorem 2.6). Note that \(d^\infty \in C^1(\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^m \times (\mathbb{R}^n)^m, \mathbb{R}_{\geq 0})\) is proper. Therefore, the Bolzano-Weierstraß Theorem [Rud53] implies that the \(\omega\)-limit set

\[
\omega(x, \lambda, \vartheta, \mu, \nu) = \{(y_x, y_\lambda, y_\vartheta, y_\mu, y_\nu) \in d^\infty(\leq \gamma) | \\
\quad \liminf_{t \to \infty} \|(x(t), \lambda(t), \vartheta(t), \mu(t), \nu(t)) - (y_x, y_\lambda, y_\vartheta, y_\mu, y_\nu)\| = 0\}
\]

is nonempty and weakly invariant (cf. [Fil88]). Our proof strategy relies on establishing that the \(\omega\)-limit set \(\omega(x, \lambda, \vartheta, \mu, \nu) \subset X \times \Lambda \times \Theta \times M \times N\) of any solution \((x, \lambda, \vartheta, \mu, \nu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^m \times (\mathbb{R}^n)^m\) of \(F^\infty\) is a singleton set. By contradiction, assume that \((y_x, y_\lambda, y_\vartheta, y_\mu, y_\nu), (z_x, z_\lambda, z_\vartheta, z_\mu, z_\nu) \in \omega(x, \lambda, \vartheta, \mu, \nu)\) with \((y_x, y_\lambda, y_\vartheta, y_\mu, y_\nu) \neq (z_x, z_\lambda, z_\vartheta, z_\mu, z_\nu)\), and define the mappings \(d^\infty_1, d^\infty_2 : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^m \times (\mathbb{R}^n)^m \to \mathbb{R}_{\geq 0}\) by

\[
d^\infty_1(x, \lambda, \vartheta, \mu, \nu) = \frac{1}{2}\|x - y_x\|^2 + \frac{1}{2}\|\lambda - y_\lambda\|^2 + \frac{1}{2}\|\vartheta - y_\vartheta\|^2
\]

\[
+ \frac{1}{2}\|\mu - y_\mu\|^2 + \frac{1}{2}\|\nu - y_\nu\|^2,
\]

\[
d^\infty_2(x, \lambda, \vartheta, \mu, \nu) = \frac{1}{2}\|x - z_x\|^2 + \frac{1}{2}\|\lambda - z_\lambda\|^2 + \frac{1}{2}\|\vartheta - z_\vartheta\|^2
\]

\[
+ \frac{1}{2}\|\mu - z_\mu\|^2 + \frac{1}{2}\|\nu - z_\nu\|^2.
\]
Since \((y_x, y_\lambda, y_\theta, y_\mu, y_\nu), (z_x, z_\lambda, z_\theta, z_\mu, z_\nu) \in X \times \Lambda \times \Theta \times M \times N\), the above discussion implies that the \(\gamma\)-sublevel sets \(d_1^{\gamma-1}(\leq \gamma)\) and \(d_2^{\gamma-1}(\leq \gamma)\) are strongly invariant under \(F^0\), for any \(\gamma > 0\). Pick \(\gamma < \frac{1}{2} \|(y_x, y_\lambda, y_\theta, y_\mu, y_\nu) - (z_x, z_\lambda, z_\theta, z_\mu, z_\nu)\|\).

Since \((y_x, y_\lambda, y_\theta, y_\mu, y_\nu) \in \omega(x, \lambda, \vartheta, \mu, \nu)\), the solution \((x(t), \lambda, \vartheta(t), \mu(t), \nu(t)) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^m \times (\mathbb{R}^n)^m\) enters \(d_1^{\gamma-1}(\leq \gamma)\) at some finite time \(t_1 \in [t_0, +\infty)\) and remains there afterwards because of the strong invariance of the \(\gamma\)-sublevel set. Similarly, for \((z_x, z_\lambda, z_\theta, z_\mu, z_\nu) \in \omega(x, \lambda, \vartheta, \mu, \nu)\), there exists a finite time \(t_2 \in [t_0, +\infty)\) such that \((x(t), \lambda(t), \vartheta(t), \mu(t), \nu(t)) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^m \times (\mathbb{R}^n)^m\) for all \(t \geq t_2\). Therefore, for \(t \geq \max\{t_1, t_2\}\), the solution belongs to \(d_1^{\gamma-1}(\leq \gamma) \cap d_2^{\gamma-1}(\leq \gamma) = \emptyset\), which is a contradiction. \(\square\)

Remark 2.8 (On the bounds of \(\kappa_1\) and \(\kappa_2\)). Note that (4.18a) and (4.18b) depend on the primal-dual set \(X \times \Lambda \times \Theta \times M \times N\) and on the initial condition due to \(d_1^{\gamma-1}(\leq \gamma)\). However, if the primal-dual solution set is compact, then \(\kappa_1\) and \(\kappa_2\) can be chosen independently of the initial condition since both maxima in (4.18) are achieved. •

### 4.3 Robust Projected Dynamics

In this section, we develop discontinuous saddle-point-like dynamics that enjoy the same convergence properties as the proposed set-valued dynamics (4.9) but do not require knowledge on the global parameters a priori. Moreover, the dynamics provided involve explicit projection operators that do not require solutions of (sub-)optimization problems (as in Section 3.4) and are amenable for distributed implementation.

Consider the flow functions \(F^\gamma : \mathbb{R}^n \times \mathbb{R}_{\geq 0}^p \times \mathbb{R}_{\geq 0}^m \times \mathbb{R}^m \times (\mathbb{R}^n)^m \to \mathbb{R}^p\) and \(F^\delta : \mathbb{R}_{\geq 0}^p \times \mathbb{R}_{\geq 0}^m \times \mathbb{R}^m \to \mathbb{R}^m\) defined by

\[
F^\gamma(x, \lambda, \vartheta, \mu, \nu) = -D(D^\top \lambda - (1_m \otimes x) + \nu) - E \mu - E(E^\top \lambda + \vartheta - b), \tag{4.19a}
\]

\[
F^\delta(\lambda, \vartheta, \mu) = -(E^\top \lambda + \vartheta - b) - \mu. \tag{4.19b}
\]

The definitions in (4.19) are motivated by the fact that, for \((x, \lambda, \vartheta, \mu, \nu) \in \mathbb{R}_+^n \times \mathbb{R}_{\geq 0}^p \times \mathbb{R}_{\geq 0}^m \times \mathbb{R}^m \times (\mathbb{R}^n)^m\), it follows \(\partial_\lambda L^\gamma = -F^\gamma(x, \lambda, \vartheta, \mu, \nu)\) and \(\partial_\vartheta L^\gamma = -F^\delta(\lambda, \vartheta, \mu)\). Consider the saddle-point-like dynamics defined over \(\mathbb{R}^n \times \mathbb{R}_{\geq 0}^p \times \mathbb{R}_{\geq 0}^m \times \mathbb{R}^m \times (\mathbb{R}^n)^m\):

\[
\dot{x}(t) + (1_m, 1_m)x(t) - (1_m^\top I_n)(D^\top \lambda(t) + \nu(t)) \in -\partial f(x(t)), \tag{4.20a}
\]

\[
\dot{\lambda}(t) = \begin{cases} 
F^\gamma(x(t), \lambda(t), \vartheta(t), \mu(t), \nu(t)), & \text{if } \lambda > 0, \\
\{F^\gamma(x(t), \lambda(t), \vartheta(t), \mu(t), \nu(t))\}^+, & \text{if } \lambda = 0,
\end{cases} \tag{4.20b}
\]

\[
\dot{\vartheta}(t) = \begin{cases} 
F^\delta(\lambda(t), \vartheta(t), \mu(t)), & \text{if } \vartheta > 0, \\
\{F^\delta(\lambda(t), \vartheta(t), \mu(t))\}^+, & \text{if } \vartheta = 0,
\end{cases} \tag{4.20c}
\]

\[
\dot{\mu}(t) = E^\top \lambda(t) + \vartheta(t) - b, \tag{4.20d}
\]

\[
\dot{\nu}(t) = D^\top \lambda(t) - (1_m \otimes x(t)), \tag{4.20e}
\]

for a.a. \(t \in [t_0, +\infty)\) with initial condition \((x_0, \lambda_0, \vartheta_0, \mu_0, \nu_0) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0}^p \times \mathbb{R}_{\geq 0}^m \times \mathbb{R}^m \times (\mathbb{R}^n)^m\). Proposition 1.7 in Chapter 2 and Lemma 2.2 in Chapter 2 guarantee
the existence of solutions \((x, \lambda, \vartheta, \mu, \nu) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^p_{\geq 0} \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^m \times (\mathbb{R}^n)^m\) of the dynamics (4.20), where the solutions are understood in the sense of Krasovskii (cf. Definition 2.6 in Ch. 2). Here, we use the set-valued map \(F^\sharp : \mathbb{R}^n \times \mathbb{R}^p_{\geq 0} \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^m \times (\mathbb{R}^n)^m \Rightarrow \mathbb{R}^n \times \mathbb{R}^p_{\geq 0} \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^m \times (\mathbb{R}^n)^m\) to refer to (4.20).

The following result establishes a relationship between the solutions of the robust saddle-point dynamics \(F^\diamond\) and the Krasovskii set-valued map of the saddle-point-like dynamics \(F^\sharp\) that allows us to conclude convergence properties of (4.20).

**Lemma 3.1 (Relationship of trajectories).** Let \(k_1, k_2 \in \mathbb{R}_{\geq 0}, \gamma \in \mathbb{R}_{> 0}\) and let \((x^*, \lambda^*, \vartheta^*, \mu^*, \nu^*) \in X \times \Lambda \times \Theta \times M \times N\). Consider the function \(d^\diamond \in C^1(\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times (\mathbb{R}^n)^m, \mathbb{R}_{\geq 0})\) defined as in Theorem 2.6. If

\[
\kappa_1 \geq \kappa_1 := \sup_{(x, \lambda, \vartheta, \mu, \nu) \in d^\diamond(\leq \gamma)} \|F^\diamond (x, \lambda, \vartheta, \mu, \nu)\|_\infty, \quad (4.21a)
\]

\[
\kappa_2 \geq \kappa_2 := \sup_{(x, \lambda, \vartheta, \mu, \nu) \in d^\diamond(\leq \gamma)} \|F^\diamond (\lambda, \vartheta, \mu)\|_\infty, \quad (4.21b)
\]

then \(\mathcal{K}[F^\sharp](x, \lambda, \vartheta, \mu, \nu) \subset F^\diamond(x, \lambda, \vartheta, \mu, \nu)\) holds for all \((x, \lambda, \vartheta, \mu, \nu) \in d^\diamond(\leq \gamma)\).

**Proof.** Consider the Krasovskii set-valued map of the saddle-point-like dynamics \(F^\sharp\), i.e., the set-valued map \(\mathcal{K}[F^\sharp] : \mathbb{R}^n \times \mathbb{R}^p_{\geq 0} \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^m \times (\mathbb{R}^n)^m \Rightarrow \mathbb{R}^n \times \mathbb{R}^p_{\geq 0} \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^m \times (\mathbb{R}^n)^m\) defined by

\[
\mathcal{K}[F^\sharp](\cdot) = \left\{ \begin{array}{ll}
(1_m \otimes I_n)(D^\top \lambda + \nu) - (1_m, 1_m)x - \vartheta f(x), \\
\{F^\vartheta(x, \lambda, \vartheta, \mu, \nu)\}, \\
[F^\vartheta(x, \lambda, \vartheta, \mu, \nu)]^+ \\
[F^\vartheta(x, \lambda, \vartheta, \mu, \nu)]^+, \\
\{F^\vartheta(\lambda, \vartheta, \mu)\}, \\
[F^\vartheta(\lambda, \vartheta, \mu)]^+, \\
E^\top \lambda + \vartheta - b, \\
D^\top \lambda - (1_m \otimes x) \end{array} \right. 
\]

Note that if \((x, \lambda, \vartheta, \mu, \nu) \in \mathbb{R}^n \times \mathbb{R}^p_{\geq 0} \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^m \times (\mathbb{R}^n)^m\), then the condition \(\mathcal{K}[F^\sharp](x, \lambda, \vartheta, \mu, \nu) = F^\diamond(x, \lambda, \vartheta, \mu, \nu)\) holds trivially. Let \(\lambda_\ell = 0\) for some index \(\ell \in \{1, \ldots, p\}\). Then, it follows

\[
\text{proj}_\ell(\mathcal{K}[F^\sharp](x, \lambda, \vartheta, \mu, \nu)) = [F^\vartheta_\ell(\cdot), [F^\vartheta_\ell(\cdot)]^+] \subset [F^\vartheta_\ell(\cdot), F^\vartheta_\ell(\cdot) + |F^\vartheta_\ell(\cdot)|],
\]

\[
\text{proj}_\ell(F^\diamond(x, \lambda, \vartheta, \mu, \nu)) = [F^\vartheta_\ell(\cdot), F^\vartheta_\ell(\cdot) + \kappa_1],
\]

where \(\text{proj}_i(F)\) denotes the projection onto the \(i\)th-component of the set-valued map \(F\). Similarly, if \(\vartheta_k = 0\) for some index \(k \in \{1, \ldots, m\}\), then

\[
\text{proj}_k(\mathcal{K}[F^\sharp](x, \lambda, \vartheta, \mu, \nu)) = [F^\vartheta_k(\cdot), [F^\vartheta_k(\cdot)]^+] \subset [F^\vartheta_k(\cdot), F^\vartheta_k(\cdot) + |F^\vartheta_k(\cdot)|],
\]

\[
\text{proj}_k(F^\diamond(x, \lambda, \vartheta, \mu, \nu)) = [F^\vartheta_k(\cdot), F^\vartheta_k(\cdot) + \kappa_2].
\]

Hence, if \(\kappa_1 \geq |F^\vartheta_\ell(x, \lambda, \vartheta, \mu, \nu)|\) for all \(\ell \in \{1, \ldots, p\}\), and \(\kappa_2 \geq |F^\vartheta_k(\lambda, \vartheta, \mu)|\) for all \(k \in \{1, \ldots, m\}\), then the inclusion \(\mathcal{K}[F^\sharp](x, \lambda, \vartheta, \mu, \nu) \subset F^\diamond(x, \lambda, \vartheta, \mu, \nu)\) holds for all
(x, λ, θ, μ, ν) ∈ \mathbb{R}^n \times \mathbb{R}^p_0 × \mathbb{R}^m_0 × \mathbb{R}^m × (\mathbb{R}^n)^m. Since d_0^{-1}(≤ γ) is compact and the maps F^γ, F^δ are continuous, the bounds in (4.21a) and (4.21b) guarantee that the inclusion \mathcal{K}[F^\gamma](x, λ, θ, μ, ν) ⊂ F^\gamma(x, λ, θ, μ, ν) holds for all (x, λ, θ, μ, ν) ∈ d_0^{-1}(≤ γ). Note that the \gamma-sublevel set d_0^{-1}(≤ γ) is strongly invariant under F^\gamma (cf. Theorem 2.6), and by the inclusions F^\gamma(x, λ, θ, μ, ν) ⊂ \mathcal{K}[F^\gamma](x, λ, θ, μ, ν) ⊂ F^\gamma(x, λ, θ, μ, ν), it is also strongly invariant under (4.20). Therefore, any trajectory (x, λ, θ, μ, ν) : [t_0, +∞) → \mathbb{R}^n × R^p_0 × R^m_0 × R^m × (R^n)^m of F^\gamma starting in d_0^{-1}(≤ γ) is also a trajectory of F^\gamma, concluding the proof.

Building on the previous result, our next contribution characterizes asymptotic convergence of trajectories of F^\gamma to a point in the set of primal-dual minimizer X × Λ × Θ × M × N.

**Theorem 3.2 (Point-wise convergence).** Any solution (x, λ, θ, μ, ν) : [t_0, +∞) → \mathbb{R}^n \mathbb{R}^p_0 × \mathbb{R}^m_0 × \mathbb{R}^m × (\mathbb{R}^n)^m starting from (x_0, λ_0, θ_0, μ_0, ν_0) ∈ \mathbb{R}^n \mathbb{R}^p_0 × \mathbb{R}^m_0 × \mathbb{R}^m × (\mathbb{R}^n)^m converges asymptotically to a point in X × Λ × Θ × M × N.

**Proof.** Let d_0 be defined as in Theorem 2.6. Given (x_0, λ_0, θ_0, μ_0, ν_0) ∈ \mathbb{R}^n × R^p_0 × R^m_0 × R^m × (R^n)^m, let (x, λ, θ, μ, ν) : [t_0, +∞) → \mathbb{R}^n × R^p_0 × R^m_0 × R^m × (R^n)^m be a solution of (4.20) starting from (x_0, λ_0, θ_0, μ_0, ν_0), and let γ = d_0((x_0, λ_0, θ_0, μ_0, ν_0)). Let κ_1, κ_2 ∈ \mathbb{R} ≥ 0 satisfy the inequalities

\[
κ_1 ≥ \max \left\{ K_1, \sup_{(x^*, λ^*, θ^*, μ^*, ν^*) ∈ (X × Λ × Θ × M × N) ∩ d_0^{-1}(≤ γ)} \| Eμ^* + Dν^* \|_\infty \right\},
\]

\[
κ_2 ≥ \max \left\{ K_2, \sup_{(x^*, λ^*, θ^*, μ^*, ν^*) ∈ (X × Λ × Θ × M × N) ∩ d_0^{-1}(≤ γ)} \| μ^* \|_\infty \right\}.
\]

Then, by Lemma (3.1), the inclusion F^\gamma(x, λ, θ, μ, ν) ⊂ F^\gamma(x, λ, θ, μ, ν) holds for all (x, λ, θ, μ, ν) ∈ \mathbb{R}^n × R^p_0 × R^m_0 × R^m × (R^n)^m. Therefore, any solution (x, λ, θ, μ, ν) : [t_0, +∞) → \mathbb{R}^n × R^p_0 × R^m_0 × R^m × (R^n)^m of F^\gamma starting in \mathbb{R}^n × R^p_0 × R^m_0 × R^m × (R^n)^m is also a solution of F^\gamma. Point-wise convergence of the solutions of (4.20) to a point in the set X × Λ × Θ × M × N follows from Corollary 2.7.

**Remark 3.3 (Exponential convergence).** Note that if f ∈ C^{1,1}(\mathbb{R}^n, \mathbb{R}) and furthermore (x(t), λ(t), θ(t), μ(t), ν(t)) ∈ \mathbb{R}^n × R^p_0 × R^m_0 × R^m × (R^n)^m for all t ∈ [t_0, +∞), then exponential convergence of trajectories of F^\gamma can be established by using the candidate Lyapunov function d_0 : \mathbb{R}^n × R^p × R^m × R^m × (R^n)^m → R_0 similarly defined as in Theorem 6.4 in Chapter 3.

### 4.4 Distributed Implementation

In this section, we show that the saddle-point-like dynamics (4.20) are amenable for fully distributed implementation over a network of agents associated with the robust optimization problem (4.4).
Consider the network model in Section 4.1. For each agent \( i \in \{1, \ldots, n\} \) in the network, the set-valued dynamics (4.20a) can be component-wise written as
\[
\dot{x}_i \in \sum_{\{j: w_{ij} \neq 0\}} w_{ij} \left( \sum_{\{\ell: d_{j\ell} \neq 0\}} d_{j\ell} \lambda_\ell + \nu_j \right) - \left\langle 1_m, 1_m \right\rangle x_i - \partial f_i(x_i),
\]
where \( w_{ij} \) denotes the \( ij \)-th entry of \((1_m^\top \otimes I_n) \in \mathbb{R}^{n \times nm}\). For each \( \ell \in \{1, \ldots, p\} \), the discontinuous dynamics (4.20b) can be component-wise written as
\[
\dot{\lambda}_\ell = - \sum_{\{j: d_{j\ell} \neq 0\}} d_{j\ell} \left( \sum_{\{h: d_{jh} \neq 0\}} d_{jh} \lambda_h - y_j + \nu_j \right) - \sum_{\{k: e_{k\ell} \neq 0\}} e_{k\ell} \mu_k,
\]
where \( y_j \) denotes the \( j \)-th element of \((1_m \otimes x) \in (\mathbb{R}^n)^m\) and \( h \in \{1, \ldots, p\} \). From (4.20c), we have for each \( k \in \{1, \ldots, m\} \),
\[
\dot{\vartheta}_k = - \sum_{\{\ell: e_{k\ell} \neq 0\}} e_{k\ell} \lambda_\ell + \vartheta_k - b_k - \mu_k.
\]
Similarly, from (4.20d), it follows for every \( k \in \{1, \ldots, m\} \),
\[
\dot{\mu}_k = \sum_{\{\ell: e_{k\ell} \neq 0\}} e_{k\ell} \lambda_\ell + \vartheta_k - b_k.
\]
Finally, for every \( j \in \{1, \ldots, nm\} \) in (4.20e), we have
\[
\dot{\nu}_j = \sum_{\{\ell: d_{j\ell} \neq 0\}} d_{j\ell} \lambda_\ell - y_j.
\]
In order for agent \( i \in \{1, \ldots, n\} \) to be able to implement its dynamics \( \dot{x}_i \), it also needs access to some components of \( \lambda \in \mathbb{R}^p, \vartheta \in \mathbb{R}^m, \mu \in \mathbb{R}^m \), and \( \nu \in (\mathbb{R}^n)^m \). Thus, agent \( i \) also needs to implement their corresponding dynamics.

**Remark 4.1 (Comparison with existing dynamics).** The work [RC15] builds on saddle-point dynamics of an augmented Lagrangian function for linear rather than general convex programs. However, a direct transcription would require the robust counterpart of an uncertain linear program to be in standard form, that is
\[
\begin{aligned}
\min_{x^+, x^- \in \mathbb{R}^n, \lambda \in \mathbb{R}^p, \vartheta \in \mathbb{R}^m} & \quad \langle c, x^+ - x^- \rangle \\
\text{subject to} & \quad \begin{pmatrix}
0_{m \times n} & 0_{m \times n} & E^\top & I_m \\
-1_m \otimes I_n & 1_m \otimes I_n & D^\top & 0_{nm \times m}
\end{pmatrix}
\begin{pmatrix}
x^+ \\
x^- \\
\lambda \\
\vartheta
\end{pmatrix}
= \begin{pmatrix}
b \\
0_{nm}
\end{pmatrix}
\end{aligned}
\]
\( x^+ \succeq 0, \quad x^- \succeq 0, \quad \lambda \succeq 0, \quad \vartheta \succeq 0. \)
For convenience, let $x = (x^+, x^-, \lambda, \vartheta) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$. The set-valued saddle-point dynamics corresponding to the above linear program in standard form are

$$\dot{x}(t) + G^\top (Gx(t) - h + \mu(t)) + \tilde{c} \in -\kappa \partial [x(t)]^+, \quad (4.22a)$$
$$\dot{\mu}(t) = Gx(t) - h, \quad (4.22b)$$

for a.a. $t \in [t_0, +\infty)$ with initial condition $(x_0, \mu_0) \in \mathbb{R}^{2n+p+m} \times \mathbb{R}^{nm+m}$ and $\tilde{c} = (c, -c, 0, 0) \in \mathbb{R}^{2n+p+m}$. From a global point of view, this approach requires the network $\mathcal{G}$ to run a total number of $(2n + p + 2m + nm)$ dynamics. In contrast, our proposed dynamics (4.9) require agents in the network to solve only $(n + p + 2m + nm)$ equations. Therefore, our approach to solve robust optimization problems scales better with the number of agents in the network.

### 4.5 Illustrative Example

This section summarizes our studies on robust distributed optimization and illustrates the convergence properties of the proposed saddle-point-like dynamics (4.20).

Consider a network of agents whose objective is to cooperatively solve the robust optimization problem with linear inequality constraints

$$\text{minimize} \quad \langle c, x \rangle$$
$$\text{subject to} \quad \langle a_k, x \rangle \leq b_k, \quad \forall a_k \in A_k, \quad (4.23)$$

for $k \in \{1, \ldots, m\}$, where $x \in \mathbb{R}^n$ is the network state, $c \in \mathbb{R}^n$, $b_k \in \mathbb{R}$, and $a_k \in \mathbb{R}^n$ denotes the $k^{th}$ row of the uncertain matrix $A \in \mathbb{R}^{m \times n}$. Let the uncertainty set $A_k \subset \mathbb{R}^n$ possess a polyhedral representation, i.e., $A_k = \{a_k \in \mathbb{R}^n \mid D_k a_k \leq e_k\}$, where $D_k \in \mathbb{R}^{pk \times n}$, $e_k \in \mathbb{R}^{pk}$, and $pk \in \mathbb{Z}_{>0}$ denotes the number of halfspaces that enclose the uncertainty vector $a_k \in \mathbb{R}^n$. Consider the following setup:

- **Number of agents:** $n = 2$
- **Number of constraints:** $m = 6$

<table>
<thead>
<tr>
<th>$k$</th>
<th>Nominal values</th>
<th>Uncertainty intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{11}$ = $(a_{11}, a_{12})$</td>
<td>$a_{11} = -2$, $a_{12} = -2$</td>
<td>$a_{11} \in [-\frac{11}{5}, -\frac{9}{5}]$, $a_{12} \in [-\frac{11}{5}, -\frac{9}{5}]$</td>
</tr>
<tr>
<td>$a_{21}$ = $(a_{21}, a_{22})$</td>
<td>$a_{21} = +6$, $a_{22} = -1$</td>
<td>$a_{21} \in [\frac{29}{5}, \frac{31}{5}]$, $a_{22} \in [-\frac{6}{5}, -\frac{4}{5}]$</td>
</tr>
<tr>
<td>$a_{31}$ = $(a_{31}, a_{32})$</td>
<td>$a_{31} = -1$, $a_{32} = +1$</td>
<td>$a_{31} \in [-\frac{6}{5}, -\frac{4}{5}]$, $a_{32} \in [+\frac{4}{5}, +\frac{6}{5}]$</td>
</tr>
<tr>
<td>$a_{41}$ = $(a_{41}, a_{42})$</td>
<td>$a_{41} = +2$, $a_{42} = +4$</td>
<td>$a_{41} \in [+\frac{9}{5}, +\frac{11}{5}]$, $a_{42} \in [+\frac{19}{5}, +\frac{21}{5}]$</td>
</tr>
<tr>
<td>$a_{51}$ = $(a_{51}, a_{52})$</td>
<td>$a_{51} = -1$, $a_{52} = +0$</td>
<td>$a_{51} \in [-\frac{6}{5}, -\frac{4}{5}]$</td>
</tr>
<tr>
<td>$a_{61}$ = $(a_{61}, a_{62})$</td>
<td>$a_{61} = +0$, $a_{62} = -1$</td>
<td>$a_{62} \in [-\frac{6}{5}, -\frac{4}{5}]$</td>
</tr>
</tbody>
</table>

Table 4.1: Nominal values and uncertainty intervals of elements $a_{ki} \in \mathbb{R}$ of the uncertainty matrix $A \in \mathbb{R}^{m \times n}$.

Moreover, let $c = (1, 2)$ and $b = (-6, 15, 4, 26, -1, -1)$. In order to preserve the 0-pattern in $a_k \in A_k$, we introduce the reduced polyhedral uncertainty sets defined
by $\tilde{A}_k = \{\tilde{a}_k \in \mathbb{R}^{n_k} | \tilde{D}_k \tilde{a}_k \preceq \tilde{e}_k\}$, where $n_k \in \mathbb{Z}_{>0}$ denotes the number of non-zero elements in $a_k \in \mathbb{R}^n$, $\tilde{D}_k \in \mathbb{R}^{r_k \times n_k}$, and $\tilde{e}_k \in \mathbb{R}^{r_k}$, with $r_k \in \mathbb{Z}_{>0}$ as the number of halfspaces that enclose the reduced uncertainty vector $\tilde{a}_k \in \mathbb{R}^{n_k}$ (cf. Definition 1.4).

In particular, the uncertainty intervals in Table 4.1 yield

$$D_1 = D_2 = D_3 = D_4 = \begin{pmatrix} -5 & 0 \\ 0 & -5 \\ 5 & 0 \\ 0 & 5 \end{pmatrix} \in \mathbb{R}^{4 \times 2}, \quad \tilde{D}_5 = \tilde{D}_6 = \begin{pmatrix} -5 \\ 0 \end{pmatrix} \in \mathbb{R}^{2 \times 1},$$

$$e_1 = (11, 11, -9, -9), \quad e_2 = (-29, 6, 31, -4), \quad e_3 = (6, -4, -4, 6),$$

$$e_4 = (-9, -19, 11, 21), \quad \tilde{e}_5 = \tilde{e}_6 = (6, -4).$$

The procedure of polyhedral lifting (cf. Example 1) on $\tilde{D}_5$, $\tilde{D}_6$ and $\tilde{e}_5$, $\tilde{e}_6$ implies

$$D_5 = \begin{pmatrix} -5 & 0 \\ 5 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix} \in \mathbb{R}^{4 \times 2}, \quad D_6 = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & -5 \\ 0 & 5 \end{pmatrix} \in \mathbb{R}^{4 \times 2},$$

$$e_5 = (6, -4, 0, 0), \quad e_6 = (0, 0, 6, -4).$$

The robust counterpart (cf. Definition 3.9 in Ch. 2) associated with (4.23) is the bi-level structured optimization problem

$$\begin{align*}
\text{minimize} & \quad \langle c, x \rangle \\
\text{subject to} & \quad \max_{D_k a_k \preceq e_k} \langle a_k, x \rangle \leq b_k, \quad k \in \{1, \ldots, m\}. \quad (4.24)
\end{align*}$$

By Proposition 3.10 in Chapter 2, the robust counterpart (4.24) can be explicitly reformulated as a linear program of the form

$$\begin{align*}
\text{minimize} & \quad \langle c, x \rangle \\
\text{subject to} & \quad E^\top \lambda + \vartheta = b, \\
& \quad D^\top \lambda = (\mathbbm{1}_m \otimes x), \\
& \lambda \succeq 0, \quad \vartheta \succeq 0,
\end{align*} \quad (4.25)$$

with the block-structured matrices

$$E^\top = \bigoplus_{k \in \{1, \ldots, m\}} e_k^\top, \quad \text{and} \quad D^\top = \bigoplus_{k \in \{1, \ldots, m\}} D_k^\top,$$

where $E^\top \in \mathbb{R}^{m \times p}$, $D^\top \in \mathbb{R}^{nm \times p}$, and $b \in \mathbb{R}^m$. Note that the slack variable $\vartheta \in \mathbb{R}^m$ is further introduced in (4.25). In addition, let $\lambda = (\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^p$, where $\mathbb{R}^p = \mathbb{R}^{p_1} \times \cdots \times \mathbb{R}^{p_m}$.

In what follows, we solve the linear program (4.25) using the saddle-point-like dynamics (4.20). Note that the network topology compatible with (4.25) is simply the complete undirected graph denoted by $K_2$. In this example, the total number of
primal-dual variables is $n + p + 2m + nm = 50$. Each agent $i \in \{1, 2\}$ is responsible to run its own dynamics $\dot{x}_i$ and parts of the primal-dual dynamics $\dot{\lambda}, \dot{\vartheta}, \dot{\mu}$ and $\dot{\nu}$ distributed according to Section 4.4. When implementing the dynamics (4.20), a first-order Euler approximation with stepsize 0.0007 is used.

Figure 4.1: The network state evolution of the linear program is depicted in plot (a). The plots (b) – (f) show the evolutions of the primal-dual variables according to the saddle-point-like dynamics (4.20). The steady-state values are achieved at time $t = 350s$.

Figure 4.1 shows the results of the distributed implementation of the discontinuous saddle-point-like dynamics (4.20). In Figure 4.1(a), the network state convergence
to the (worst-case) optimal solution – shown as ♦ – of the linear program (4.25) is depicted. The nominal solution of the linear program is illustrated by ⚫. Moreover, the nominal and uncertain polyhedral constraint sets of (4.25) are shown. The trajectories of the continuous-time coordination algorithm (4.20) are shown in Figures 4.1(b)–(f). Clearly, the distributed algorithm (4.20) converges to the primal-dual solution set of (4.25). Note that the explicit projection operator used in the algorithm prevents the trajectories $t \mapsto \lambda(t), \vartheta(t)$ from violating the non-negativity constraints in (4.25) at any time.

4.6 Summary

In this chapter, we built on our results from Chapter 3 and developed distributed continuous-time coordination algorithms that solve robust (i.e., semi-infinite) optimization problems with a convex and Lipschitz continuous objective function and uncertainty-affine inequality constraints. We have provided a framework that allows to reformulate the mathematical program under infinitely many constraints as a convex program with an inherent distributed structure by considering the worst-case solution among all feasible realizations of uncertainty. Our algorithmic design is based on set-valued saddle-point dynamics derived from an augmented Lagrangian function associated with the robust convex program. We show asymptotic convergence of the proposed algorithms to a point in the set of robust primal-dual minimizer by relying on classical notions of Lyapunov stability theory. In contrast to the discontinuous dynamics developed in Section 3.4, the algorithm proposed in Section 4.3 does not require to solve (sub-)optimization problems since the projection operator used possesses an explicit structure for the linear inequality constraints.
Chapter 5

Conclusions

In this chapter, we summarize the main results of the thesis and indicate possible directions for future research.

5.1 Summary and Discussion

In this thesis, we developed continuous-time coordination algorithms for networked systems that give rise to general nonsmooth convex and robust optimization problems with an inherent distributed structure. In particular, we investigated convex programs with an additively separable objective function and coupling equality and inequality constraints. Both the objective function and the inequality constraints were Lipschitz continuous. In the context of semi-infinite mathematical programming, we studied robust programs with a convex, separable and Lipschitz continuous objective function and uncertainty-affine inequality constraints that take values in bounded polyhedral uncertainty sets.

In Chapter 3, we considered continuous-time distributed dynamics that solve general nonsmooth convex optimization problems. In this context, the agents’ common objective was to asymptotically converge to a solution of the convex program by interchanging local information among each other. We proposed a framework based on Lagrangian duality theory that allows to find optimal solutions via saddle points of an augmented nonsmooth Lagrangian function. This approach naturally let us to study the associated set-valued saddle-point dynamics for which we established point-wise asymptotic convergence to the set of primal-dual solutions of a nonsmooth convex program using classical notions of stability theory. In a next step, we introduced an alternative algorithm that enjoys the same convergence properties but is amenable for fully distributed implementation over a network of agents. In this chapter, we also established properties of the algorithms beyond asymptotic convergence and characterized a performance bound under mild convexity and regularity conditions on the objective function of a nonsmooth convex program.

In Chapter 4, we built on our previous results to synthesize provably correct distributed algorithms that allow agents in a network to cooperatively solve robust opti-
mization problems. We proposed a framework that allows to reformulate semi-infinite optimization problems as conventional convex programs by considering worst-case solutions among all feasible realizations of uncertainty. In particular, we introduced the concept of polyhedral lifting to guarantee that the distributed structure of the uncertainty-affected inequality constraints remains invariant. Based on this approach, we then developed saddle-point(-like) algorithms specifically tailored to solve robust optimization problems. Similar to the analysis provided in Chapter 3, we established asymptotic convergence of the proposed continuous-time dynamics to a point in the set of robust primal-dual solution.

5.2 Outlook

The results in this thesis provide a constructive framework for future developments regarding to distributed convex and robust optimization and therefore, leave open the possibility of numerous extensions and additional analysis. In what follows, we outline some potential future research directions.

With regards to distributed convex optimization, a performance bound of the proposed saddle-point algorithms in the presence of Lipschitz continuous inequality constraint functions has yet to be established. Specifically, we observed in numerical studies that exponential convergence still holds even when the projection operator becomes active. This would particularly open up the study to characterize robustness properties of the dynamics – such as ISS or iISS (see e.g., [Son89, ASW00, Son98] for further details) – against (state) disturbance and noise. Another direction of interest would be to extend our results to possibly time-varying nonsmooth convex optimization scenarios. Also, maintaining user privacy in networks of agents has become more crucial. A distributed approach to privately solving convex programs using the algorithms in this thesis could provide new insights in that direction.

In terms of distributed semi-infinite mathematical programming, we would like to extend our results to other, yet computationally tractable representations of uncertainty sets, such as ellipsoidal and norm uncertainty [BBC11]. Also, we plan to design and analyze distributed continuous-time coordination algorithms for robust multi-stage optimization problems [BTGN09]. Finally, implementing our results of Chapter 3 and 4 would potentially provide insight into implementation issues that may exist.
Appendix A

Erklärung

Erklärung des Autors der Masterarbeit mit dem Titel:

Distributed Continuous-Time Coordination for
Nonsmooth Convex & Robust Optimization


 Ort, Datum

Unterschrift
Bibliography


