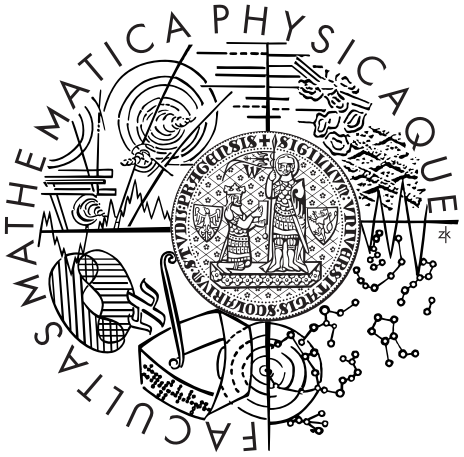
Charles University in Prague
Faculty of Mathematics and Physics

DOCTORAL THESIS



Lukáš Adam

Hierarchical Problems with Evolutionary Equilibrium Constraints

Department of Probability and Mathematical Statistics

Supervisor of the doctoral thesis: Jiří Outrata
Study programme: Mathematics
Specialization: Probability and Statistics, Econometrics and Financial Mathematics

Prague 2015

I would like to express my deepest thanks to

- my parents and Evča for everything they have done to me
- Jiří for his leadership and introducing me to many interesting people
- Vicky for her courage and ability to overcome all hardships in her life
- Roya for being almost my sister
- Dávid for always being in a good mood
- Mirek for his never–ending supply of chocolate
- Michal for constantly improving my patience and self–control
- and to all girls I liked; a list which would be too long to fit here

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague, 26 April 2015 Lukáš Adam

Název práce: Hierarchické úlohy s evolučními ekvilibriálními omezeními

Autor: Mgr. Lukáš Adam

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí disertační práce: doc. Ing. Jiří Outrata, DrSc., Ústav teorie informace a automatizace, AV ČR

Abstrakt: Předložená práce je věnována hierarchickým modelům s evolučními ekvilibriálními omezeními. Tyto modely přirozeně vznikají při optimálním řízení nebo identifikaci parametrů v časově závislém problému. Naším cílem je tyto problémy diskretizovat a vyřešit pomocí metody implicitního programování. Tato technika vyžaduje znalost zobecněné derivace řídícího zobrazení (solution mapping), které řídící proměnné či parametru přiřazuje stavovou proměnnou. Výpočet této zobecněné derivace je ekvivalentní výpočtu (limitního) normálového kužele ke grafu řídícího zobrazení.

V první části shrneme známé techniky pro výpočet normálového kužele k množině reprezentovatelné jako konečné sjednocení konvexních polyedrů. Poté navrhneme nový přístup založený na takzvané normálně přípustné stratifikaci a zjednodušíme získané formule pro případ časově závislých problémů. Teoretické výsledky jsou poté aplikovány pro získání kritéria citlivosti řídícího zobrazení a na dva prakticky motivované příklady. První se zabývá optimální řízením fronty u přepážky, zatímco druhý identifikací parametrů v modelu delaminace.

Klíčová slova: matematický program s ekvilibriálními omezeními, solution mapping, normálový kužel, koderivace, normally admissible stratification, analýza stability a sensitivity, delaminace

Title: Hierarchical Problems with Evolutionary Equilibrium Constraints

Author: Lukáš Adam

Department: Department of Probability and Mathematical Statistics

Supervisor: Prof. Jiří Outrata, Institute of Information Theory and Automation, Czech Academy of Sciences

Abstract: In the presented thesis, we are interested in hierarchical models with evolutionary equilibrium constraints. Such models arise naturally when a time–dependent problem is to be controlled or if parameters in such a model are to be identified. We intend to discretize the problem and solve it on the basis of the so–called implicit programming approach. This technique requires knowledge of a generalized derivative of the solution mapping which assigns the state variable to the control variable/parameter. The computation of this generalized derivative amounts equivalently to the computation of (limiting) normal cone to the graph of the solution mapping.

In the first part we summarize known techniques for computation of the normal cone to the set which can be represented as a finite union of convex polyhedra. Then we propose a new approach based on the so–called normally admissible stratification and simplify the obtained formulas for the case of time–dependent problems. The theoretical results are then applied first to deriving a criterion for the sensitivity analysis of the solution mapping and then to the solution of two practically motivated problems. The first one concerns optimal control of a queue at a service point while the other one deals with parameter identification in a delamination model.

Keywords: mathematical program with equilibrium constraints, solution mapping, normal cone, coderivative, normally admissible stratification, stability and sensitivity analysis, delamination

Contents

1. Introduction

In this dissertation thesis we study a class of hierarchical problems with evolutionary equilibrium constraints. In such problems we have two types of variables, the *parameter* or control variable $u \in U$ and the *state variable* $x \in X$, where U and X are Banach spaces. By equilibrium constraints we mean that the state variable x depends on the control variable u via a *solution mapping* S in such a way that $x \in S(u)$. For multifunction S usually only an implicit expression may exist. Typically, this multifunction is governed by a solution of an optimization problem, a complementarity system or another complicated system. Since we consider evolutionary equilibrium constraints, we assume that that at least the state variable x depends on time. Having some admissible control set $U_{ad} \subset U$, we try to find some $\bar{u} \in U_{ad}$ and a corresponding $\bar{x} \in S(\bar{u})$ such that $(\bar{u}, \bar{x})$ minimizes given functional $J(u, x)$ among all feasible pairs (u, x).

Writing this in a concise way, we want to solve the following problem

$$\begin{aligned} &\underset{u,x}{\text{minimize}}\ J(u,x) \\ &\text{subject to } x \in S(u), \\ &\qquad\qquad u \in U_{ad}. \end{aligned} \tag{1.1}$$

We will denote such problem *Mathematical Program with Equilibrium Constraints* (MPEC). Note that the usual definition of an MPEC is more specific and some authors would not consider problem (1.1) to be an MPEC. The constraint $x \in S(u)$, which models an equilibrium, is usually referred to as the lower level, while the rest of the problem is the upper level.

To be less abstract, we present an example from [5]. Since it will be later analyzed in Chapter 6, we describe it here only in its simplified form. Consider a body $\Omega \subset \mathbb{R}^d$ glued to a rigid obstacle. This body is dragged by a part of its boundary and its displacement $x_m : (0,T) \times \Omega \to \mathbb{R}^d$ is measured. Some physical parameters u of Ω are unknown and are to be identified from measurements x_m. The control variable (or to be more specific parameter) may be the elastic moduli of the adhesive, thus $u : \Gamma_c \to \mathbb{R}$, where $\Gamma_c \subset \partial\Omega$ is the contact boundary. The state variable $x : (0,T) \times \Omega \to \mathbb{R}^d$ may be the displacement. The solution mapping is described by a partial differential equation with complementarity conditions on the contact boundary. Hence, it is a rather difficult system, see (6.22) below.

Then the problem reads

$$\begin{aligned} &\underset{u,x}{\text{minimize}}\ \frac{1}{2}\int_\Omega \int_0^T (x(t,y) - x_m(t,y))^2 dt dy \\ &\text{subject to } x \in S(u), \\ &\qquad\qquad u \in U_{ad}. \end{aligned} \tag{1.2}$$

The objective function of (1.2) is the distance between the response x and the measured values x_m. As we have already said, the mapping $S : u \mapsto x$ is governed by a partial differential equation with prescribed initial and boundary conditions. Set U_{ad} gives constraints on the control variable, we may consider

$$U_{ad} = \{u |\ 0 \leq u(\cdot,\cdot)\}.$$

Note that we have used the $y \in \Omega$ for the space variable, the reasons for this are described in Section 2.3. It is not difficult to see that problem (1.2) is indeed of type (1.1).

There are plentiful applications of such hierarchical problems. Among others, we may mention deregulated electricity markets [103], financial market modeling [133], optimal pricing [24], optimal shape design [94], resource allocation [141] or waste management [10]. For further information about hierarchical equilibrium problems, see monographs or annotated papers [33, 34, 80, 94, 119, 136].

In the game theory, the control variable u is known as strategy of a player commonly referred to as *leader* and the state variable x as strategy of a *follower*. The reasoning behind this is that the leader chooses his strategy as the first one and the strategy of the follower is chosen only after that.

Problem (1.1) is written in the so–called *optimistic formulation*. The reason for this is that when $S(u)$ is multivalued, then the strategy of the follower is chosen in a way which is most beneficial to the leader. However, in some situations, the follower may want to choose his strategy $x \in S(u)$ in a way which harms the leader as much as possible. This leads to the so–called *pessimistic formulation* which reads

$$\begin{aligned} \underset{u}{\text{minimize}} \ \underset{x}{\text{maximize}} \ & J(u,x) \\ \text{subject to } & x \in S(u), \\ & u \in U_{ad}. \end{aligned}$$

In the majority of this thesis, the solution mapping will be single–valued. For such problems, the optimistic and pessimistic formulations naturally coincide. The only part in this thesis where S will be multivalued is a part of Chapter 6 but, after a regularizing procedure, we obtain that S is single–valued as well.

One of the most well–known instances of MPECs are the so–called *bilevel programs* where the lower level is an optimization problem. Such problems are well–known in the field of game theory under the name of *Stackelberg games* and date back to [133]. The problem reads

$$\begin{aligned} \underset{u,x}{\text{minimize}} \ & J(u,x) \\ \text{subject to } & x \in \underset{y}{\text{argmin}}\{f(u,y) |\ g(u,y) \leq 0,\ h(u,y) = 0\}, \\ & u \in U_{ad}. \end{aligned} \tag{1.3}$$

Provided some differentiability and a constraint qualification are present, we may reformulate problem (1.3) by writing the lower level as its Karush–Kuhn–Tucker conditions. Then the whole problem reads

$$\begin{aligned} \underset{u,x}{\text{minimize}} \ & J(u,x) \\ \text{subject to } & \nabla_x L(u,x,\lambda,\mu) = 0, \\ & \lambda \geq 0,\ g(u,x) \leq 0,\ \langle \lambda, g(u,x) \rangle_{Z_1^*,Z_1} = 0, \\ & h(u,x) = 0, \\ & u \in U_{ad}. \end{aligned} \tag{1.4}$$

where

$$L(u,x,\lambda,\mu) := f(u,x) + \langle \lambda, g(u,x) \rangle_{Z_1^*,Z_1} + \langle \mu, h(u,x) \rangle_{Z_2^*,Z_2}$$

is the Lagrangian function and multipliers λ and μ are associated with $g : U \times X \to Z_1$ and $h : U \times X \to Z_2$, respectively. Even though problem (1.4) can be understood as a standard nonlinear problem, the usual constraint qualifications (such as linear independence or Mangasarian–Fromovitz constraint qualifications) are violated at all feasible points of (1.4).

Bilevel programs have foundation in the Cournot duopoly model which appeared already in 1838 in the famous monograph [30]. This model consider two companies choosing their strategies at the same time. A bilevel program is formed whenever one of these companies takes the role of the leader. This was later generalized into Cournot oligopoly model where instead of one follower, multiple followers may appear. This means that on the lower level there is no longer an optimization problem but an equilibrium has to be found among the followers. Another possibility is to consider an inclusion on the lower level, a situation which will be considered in the majority of this thesis.

Note that when we employ multiplier–free optimality conditions for the lower level of bilevel problem (1.3), we obtain problem

$$\begin{aligned} &\underset{u,x}{\text{minimize}}\ J(u,x) \\ &\text{subject to } 0 \in \nabla_x f(u,x) + \mathrm{N}_{C(u)}(x), \\ &\qquad u \in U_{ad}, \end{aligned} \tag{1.5}$$

where $\mathrm{N}_{C(u)}(x)$ denotes the normal cone to $C(u)$ at x, see Definition 2.1.5 below for finite–dimensional or [87, Definition 1.1] for infinite–dimensional definition, and

$$C(u) := \{x|\ g(u,x) \leq 0,\ h(u,x) = 0\}.$$

Thus, we may say that MPECs with S governed by an inclusion encompass bilevel problems with the lower level being written via necessary optimality conditions. Since normal cone is the respective subdifferential to an indicator function due to [87, Proposition 1.79], to analyze problem (1.5), we have to work with objects of second–order variational analysis.

Many approaches were proposed to solve (1.4). Among others, we mention global methods [14], interior point methods [76], methods using merit function [80], nondifferentiable methods [94], penalty methods [80, 122], quasi–Newton methods [62], smoothening of the complementarity condition [44, 59, 120] or SQP methods [47]. Application of standard methods of nonlinear optimization have been studied in [46]. For a nice benchmark for MPECs we mention the collection MacMPEC [75].

Note that the previous approaches did not exploit the structural difference between u and x. However, if S is single–valued, another approach may be used. The basic idea is to replace the original problem (1.1) by the single–level problem

$$\begin{aligned} &\underset{u}{\text{minimize}}\ J(u, S(u)) \\ &\text{subject to } u \in U_{ad} \end{aligned}$$

and then to employ a technique which uses (generalized) derivative of the new composite objective function. Such approach is known as *implicit programming approach*. We will investigate this approach thoroughly is Section 2.2 after we define some objects of nonsmooth calculus in Section 2.1. A comparison of these

techniques can be found in [66]. Generally, we can say that implicit programming approach performs in a very good way whenever the solution mapping is single–valued, Lipschitzian, there is an efficient solver for the lower level problem and the dimension of control variable is smaller than the dimension of the state variable. This follows directly from the fact that instead of minimizing with respect to (u, x) we minimize only with respect to u in the problem above.

Apart from solving problem (1.1), we will be interested in performing stability and sensitivity analysis of the solution mapping S. These two topics are closely connected. Besides being interesting for its own merits, sensitivity analysis provides a useful insight into the evaluation of $\partial\tilde{J}(\bar{u})$, where $\tilde{J}(u) := J(u, S(u))$ is the composite/reduced objective function of problem (1.1). This is useful both for deriving optimality conditions, as well as for solving the problem via a technique of nonsmooth optimization.

This thesis consists of this introductory chapter, preliminaries, four main chapters, conclusion and some supplementary material. The preliminary Chapter 2 is divided into three parts. In the first one, we define selected objects of nonsmooth calculus, starting with normal cones to convex sets and subdifferentials to convex functions and then show possible generalizations of these objects to a nonconvex world. In the second part, we present the implicit programming approach, first for a continuously differentiable solution map and then for only a Lipschitzian one. In the final part, we introduce the basis notation.

The main Chapters 3, 4, 5 and 6 are based on the author's papers [6], [3], [4] and [5], respectively. Even though there is a strong link between them, all these chapters can be read as individual parts. This means that every of these chapters contains its separate introduction which is specifically tailored to the investigated phenomenon.

As we have already mentioned, when applying the implicit programming approach, we need to compute objects of second–order variational analysis. In Chapter 3 we are interested in the computation of such objects. We consider a set Γ which is a union of finite number of polyhedral sets, define a special partition which we name *normally admissible stratification* and on its basis compute the regular (Fréchet) and limiting (Mordukhovich) normal cones to Γ. Further, we compare our approach with known results [25, 37, 54] and apply it to the case of time–dependent problems, where a significant simplification is possible. By doing so, we have managed to enrich the calculus for limiting normal cones because we were able to obtain equalities in cases where the standard calculus rules would result only in inclusions. This chapter forms the backbone of the whole thesis and single results from this chapter are used in all remaining chapters.

In Chapter 4, we are interested in stability analysis of the solution mapping $S : u \mapsto x$ which is governed by a differential inclusion

$$g(t, u, x(t), \dot{x}(t)) \in \Lambda(t), \quad t \in [0, T] \text{ a.e.} \tag{1.6}$$

with known initial value $x(0) = a$. This differential inclusion is parameterized by a stationary parameter u. The values $\Lambda(t)$ can be sets of a possible complicated structure, for example we consider $\Lambda(t) = \operatorname{gph} \mathrm{N}_{C(t)}$ for polyhedral sets $C(t)$. We first discretize the differential inclusion (1.6), derive stability criteria for the Aubin property [11] of the discretized solution mapping S^K and under the assumption

of single–valuedness of both the discretized and original solution mappings, we derive a stability criterion for the original infinite–dimensional solution mapping as well. This theoretical result is then applied to a simple case of differential equation and later to a more sophisticated case of a modified sweeping process, which arises in the analysis of electrical circuits [7].

Stability analysis amounts to the computation of the normal cone to the solution mapping. Since $\operatorname{gph} \mathrm{N}_{C(t)}$ is a finite union of polyhedral sets whenever $C(t)$ is a polyhedral set, we again arrive at computation of normal cone to a union of polyhedral sets, which has been already thoroughly investigated in Chapter 3. In this chapter, we also derive estimates for the solution of the adjoint system; a result which is later employed in both Chapters 5 and 6.

Chapter 5 deal with investigation of an optimal control problem

$$\begin{aligned} \underset{u,y,z}{\text{minimize}} \ & \int_0^T [L_1(t, y(t), z(t)) + L_2(u(t))]dt + L_3(y(T), z(T)) \\ \text{subject to} \ & -\dot{z}(t) + R\dot{y}(t) \in \mathrm{N}_{Z(t)}(z(t)), \quad t \in [0, T] \text{ a.e.} \\ & \dot{y}(t) = f(t, y(t), z(t)) + Bu(t), \quad t \in [0, T] \text{ a.e.} \\ & u(t) \in \Omega \\ & y(0) = a, \ z(0) = b, \end{aligned} \tag{1.7}$$

which was recently proposed in [22] with possible applications in processes with rate–independent memory in mechanics of elastoplastic and thermoelastoplastic materials as well as in ferromagnetism, piezoelectricity or phase transitions. In the mentioned paper, the problem was regularized and necessary optimality conditions were derived for both regularized and original problems. We take another route, depart from the infinite–dimensional setting, discretize the problem and prove that the solutions of the discretized problems converge to solution of the original problem. Based on the implicit programming approach from Section 2.2 we propose a solution method for the discretized problem and apply it to an academic example arising in the queuing theory where the control parameter is the service point operating rate.

Note that the differential inclusion, which makes the major difficulty in (1.7), is a special case of the general differential inclusion (1.6), and thus we were able to use several results from Chapter 4 in this chapter. However, the goals were different in both chapters. In Chapter 4, the main goal was to compute $\mathrm{N}_{\operatorname{gph} S^K}$, on its basis to derive the adjoint system and from it to compute the Lipschitzian modulus of the discretized solution mapping S^K. This was later used to derive the Lipschitzian modulus of the original solution mapping S. The structure of Chapter 5 consists of several similar steps: we first computed $\mathrm{N}_{\operatorname{gph} S^K}$ and then derived the adjoint system. However, at this point the similarities ended and a different route was taken. Instead of computing the Lispchitzian modulus of S^K, we were interested in finding a solution to the adjoint system which helped us to compute a (upper estimate of) generalized derivative of S^K. With the help of this derivative information, the discretized version of problem (1.7) was solved via the implicit programming approach described in Section 2.2. Since we know that the discretized solutions converge to solution of problem (1.7), by doing so, we have managed to solve the original problem (1.7).

Chapter 6 is probably the most involved one because it contains a delamination system [114] which is infinite–dimensional both in time and space. The

problem is stated as an inverse problem: knowing some measurements x_m, the goal is to estimate some physical parameters u of a body such that the computed response $x = S(u)$ is as close to x_m as possible (compare to problem (1.2)). We basically repeat the approach from the previous chapter, discretize the problem but instead of showing proper convergence proofs, we only suggest how the proofs may be performed.

The principal difficulty is hidden in the lower level problem which does not satisfy even the Mangasarian–Fromovitz constraint qualification. Because of this, we were not able to use the standard methods to compute $\mathrm{N}_{\mathrm{gph}\, S^K}$ but we had to use the results from Chapter 3, specifically the application of general results from that chapter to time–dependent problems. Having further used results from Chapter 4, we have managed to show the Lipschitzian property of S^K and based on a similar procedure as in Chapter 5, we were able to find a solution to the adjoint system. Having done so, we again applied the implicit programming approach to identify the unknown parameters.

Another main difference between Chapters 5 and 6 is that the former chapter contains a differential inclusion whose values are finite–dimensional while the latter one contains a differential inclusion with infinite–dimensional values. This means that the numerical implementation was rather demanding in Chapter 6 and required the use of finite element method for elasticity.

We again emphasize that more comprehensive introductions are given at the beginnings of each chapter.

2. Preliminaries

In this preliminary chapter, we will present selected known results of variational analysis which will be crucial later in the text. This chapter is divided into three parts. In the first one, we will show selected notions of modern variational analysis such as various generalizations of the classical convex subdifferential. In the second one, we will present the so–called implicit programming approach which is a method tailored to solving (1.1) provided that the solution mapping S possess some special properties. In the last part, we will specify the basic notation.

2.1 Selected notions of variational analysis

In this section, we will define regular, limiting and Clarke normal cones and the respective subdifferentials. Note that all normal cones coincide with the standard normal cone in the sense of convex analysis whenever a set is convex and, similarly, we obtain coincidence for all subdifferentials in the case of convex function. We will start first in the convex case and then slowly depart to a nonconvex setting. After doing so, we concentrate on multifunctions and define some concepts of derivative and of Lipschitzian continuity.

Note that even though we will later work with infinite–dimensional MPECs, the definition and calculus rules for these objects are more developed in finite dimension. It is known that limiting subdifferential can be "trusted" only in Asplund spaces [39, Definition 4.4 and discussion below]. This means that calculus rules will be weaker in general Banach spaces, which rules out commonly used spaces such as Lebesgue spaces $L^1(\mathbb{R}^n)$, $L^\infty(\mathbb{R}^n)$, similar Bochner spaces or the space of continuous functions $\mathcal{C}(\mathbb{R}^n)$. Another difficulty is that it may be sometimes difficult to verify the so–called sequential normal compactness condition [87, Definition 1.20].

For these reasons, we will always discretize the problem first and apply the variational analysis theory purely to these discretized problems. Thus, all objects in this section are considered to be finite–dimensional.

We define now several well–known objects. We say that a set $C \subset \mathbb{R}^n$ is a cone provided that $0 \in C$ and $tx \in C$ whenever $x \in C$ and $t \geq 0$. For a general set C, by a conic hull $\operatorname{cone} C$, we understand the smaller superset of C which is a cone. Similarly, by $\operatorname{cl} C$ and $\operatorname{co} C$ we understand the closure of a set or its convex hull, respectively.

Definition 2.1.1. Assume that $C \subset \mathbb{R}^n$ is a convex set and fix any $\bar{x} \in C$. Then we define the tangent and normal cones to C at $\bar{x}$ as follows

$$\mathrm{T}_C(\bar{x}) = \operatorname{cl}\operatorname{cone}\{x - \bar{x} |\ x \in C\}, \tag{2.1a}$$

$$\mathrm{N}_C(\bar{x}) = \{x^* |\ \langle x^*, x - \bar{x}\rangle \leq 0 \text{ for all } x \in C\}. \tag{2.1b}$$

We provide an example of tangent and normal cones to a simple convex set.

Example 2.1.2. Consider a convex set

$$C := \{(x, y) |\ 0 \leq y \leq \sqrt{x}\}$$

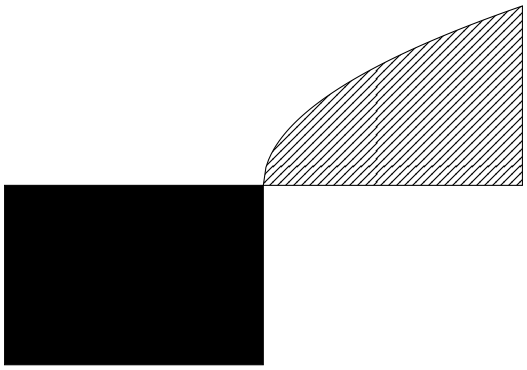


Figure 2.1: Convex set and its normal cone

which is depicted in Figure 2.1. We are interested in point $(\bar{x}, \bar{y}) = (0, 0)$. Roughly speaking, the tangent cone consists of all directions pointing to the interior of the set, and thus $\mathrm{T}_C(\bar{x}, \bar{y}) = \mathbb{R}_+ \times \mathbb{R}_+$. Similarly, the normal cone is the set of all vectors which have an obtuse angle with all vectors from the tangent cone. Thus, $\mathrm{N}_C(\bar{x}, \bar{y}) = \mathbb{R}_- \times \mathbb{R}_-$ as depicted by the black set in Figure 2.1. $\triangle$

Next, we define the polar set to C as

$$C^* := \{x^* | \, \langle x^*, x \rangle \leq 1 \text{ for all } x \in C\}.$$

Provided that C is a cone, this relation may be simplified into

$$C^* := \{x^* | \, \langle x^*, x \rangle \leq 0 \text{ for all } x \in C\}. \tag{2.2}$$

From (2.1) and (2.2) it can be seen that the tangent and normal cones to a convex set enjoy the full polarity, thus we have $\mathrm{N}_C(\bar{x}) = (\mathrm{T}_C(\bar{x}))^*$ and $\mathrm{T}_C(\bar{x}) = (\mathrm{N}_C(\bar{x}))^*$.

Now, we provide the definition of a subdifferential to a convex function.

Definition 2.1.3. For a convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ and a point $\bar{x} \in \operatorname{dom} f$ we define subdifferential $\partial f(\bar{x})$ of f at $\bar{x}$ as

$$\partial f(\bar{x}) = \{x^* | \, f(x) - f(\bar{x}) \geq \langle x^*, x - \bar{x} \rangle \text{ for all } x\}.$$

Thus, the convex subdifferential is created by normal vectors to all supporting hyperplanes to epigraph of f at $(\bar{x}, f(\bar{x}))$, which is defined as

$$\operatorname{epi} f := \{(x, \alpha) | \; f(x) \leq \alpha\}.$$

There is a close connecting between the subdifferential and the normal cone, namely

$$\partial f(\bar{x}) = \{x^* | \, (x^*, -1) \in \mathrm{N}_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))\}, \tag{2.3}$$

This relation is depicted in Figure 2.2.

When C is not convex, the relations are no longer so simple. We will consecutively define three normal cones, the regular, limiting and Clarke one. Based on these three normal cones, we will then define the corresponding subdifferentials for single–valued functions and coderivatives for multifunctions.

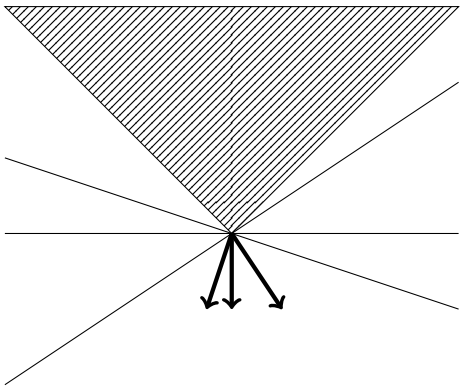


Figure 2.2: Subdifferential to a convex nonsmooth function

Definition 2.1.4. To a closed set $C \subset \mathbb{R}^n$ and $\bar{x} \in C$ we define the *Bouligand tangent (contingent) cone* to C and $\bar{x}$ as

$$\mathrm{T}_C(\bar{x}) = \{h \mid \exists h_k \to h,\ \lambda_k \searrow 0,\ \bar{x} + \lambda_k h_k \in C\}.$$

Definition 2.1.5. For a closed set $C \subset \mathbb{R}^n$ and $\bar{x} \in C$ we define the *regular (Fréchet), limiting (Mordukhovich)* and *Clarke normal cone* as

$$\begin{aligned}
&\hat{\mathrm{N}}_C(\bar{x}) = (\mathrm{T}_C(\bar{x}))^*,\\
&\mathrm{N}_C(\bar{x}) = \{x^* \mid \exists x_k \in C,\ x_k \to \bar{x},\ \exists x_k^* \in \hat{\mathrm{N}}_C(x_k),\ x_k^* \to x^*\},\\
&\overline{\mathrm{N}}_C(\bar{x}) = \mathrm{cl}\,\mathrm{co}\,\mathrm{N}_C(\bar{x}).
\end{aligned}$$

We present the basic differences between the above–defined normal cones in the following example.

Example 2.1.6. Consider the following nonconvex set

$$C := \left\{(x, y) \;\middle|\; y \geq -\frac{1}{2}|x|\right\},$$

which is depicted in Figure 2.3. Again, we are interested in the point $(\bar{x}, \bar{y}) = (0, 0)$. Since C is a cone, we have $\mathrm{T}_C(\bar{x}, \bar{y}) = C$, and thus for the regular normal cone we obtain $\hat{\mathrm{N}}_C(\bar{x}, \bar{y}) = \{0\}$. Since the limiting normal cone is obtained as Painlevé–Kuratowski upper (outer) limit of regular normal cones, it consists of two halflines which are depicted in Figure 2.3. Finally, the Clarke normal cone is the convex hull of these two halflines. △

From the previous example we see that the regular normal cone can consist only of zero. This cannot happen for the limiting normal cone and thus neither for the Clarke normal cone provided that the point in question lies on the boundary of the set. Both regular and Clarke normal cones are always closed and convex. As we have seen in the previous example, this is generally not true for limiting normal cone. It also implies that the limiting normal cone cannot be obtained by polarity relations which were presented for the regular normal cone and which also exist for the Clarke one.

The following formula can be derived

$$\hat{\mathrm{N}}_C(\bar{x}) = \left\{x^* \;\middle|\; \limsup_{x \xrightarrow{C} \bar{x}} \frac{\langle x^*, x - \bar{x}\rangle}{\|x - \bar{x}\|} \leq 0\right\},$$

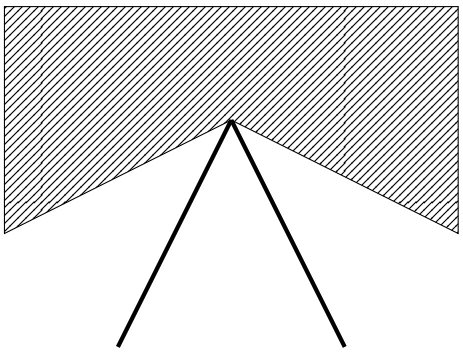


Figure 2.3: Normal cones to a nonconvex set

where $x \overset{C}{\to} \bar{x}$ stands for $x \to \bar{x}$ with $x \in C$. This sheds some light on the relation between the normal cone of convex analysis and the newly defined normal cones, see (2.1b). As we have already mentioned, for a convex set C all normal cones $\hat{\mathrm{N}}_C$, N_C and $\overline{\mathrm{N}}_C$ amount to the normal cone of convex analysis.

Similarly as in formula (2.3), we can define subdifferentials to a nonconvex function.

Definition 2.1.7. For a lower semicontinuous function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ and some $\bar{x} \in \operatorname{dom} f$ we define the *regular (Fréchet), limiting (Mordukhovich)* and *Clarke subdifferential* at $\bar{x}$ respectively as

$$\begin{aligned}
\hat{\partial} f(\bar{x}) &= \{x^* |\ (x^*, -1) \in \hat{\mathrm{N}}_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))\} \\
\partial f(\bar{x}) &= \{x^* |\ (x^*, -1) \in \mathrm{N}_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))\} \\
\overline{\partial} f(\bar{x}) &= \{x^* |\ (x^*, -1) \in \overline{\mathrm{N}}_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))\}.
\end{aligned}$$

We further define the *singular subdifferential* by

$$\partial^\infty f(\bar{x}) = \{x^* |\ (x^*, 0) \in \mathrm{N}_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))\}.$$

Single elements of a subdifferential are called subgradients.

Similarly to the convex setting for a set, we obtain that if f is convex, then all defined subdifferentials coincide with the subdifferential of the convex analysis. Furthermore, if f is continuously differentiable at $\bar{x}$, then $\hat{\partial} f(\bar{x}) = \partial f(\bar{x}) = \overline{\partial} f(\bar{x}) = \{\nabla f(\bar{x})\}$.

Example 2.1.8. We return back to Example 2.1.6 and set $f(x) = -\frac{1}{2}|x|$. Then we can see that

$$\begin{aligned}
\hat{\partial} f(\bar{x}) &= \emptyset, \\
\partial f(\bar{x}) &= \left\{-\frac{1}{2}, \frac{1}{2}\right\}, \\
\overline{\partial} f(\bar{x}) &= \left[-\frac{1}{2}, \frac{1}{2}\right].
\end{aligned}$$

△

By a multifunction $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ we understand a mapping from $\mathbb{R}^n$ whose values are (possibly empty) subsets of $\mathbb{R}^m$. For a multifunction, we define its domain by $\operatorname{dom} M = \{x |\ M(x) \neq \emptyset\} \subset \mathbb{R}^n$ and graph by $\operatorname{gph} M = \{(x, y) |\ y \in M(x)\} \subset \mathbb{R}^{n+m}$.

Definition 2.1.9. For a multifunction $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and for any $(\bar{x}, \bar{y}) \in \operatorname{gph} M$ we define the *regular coderivative* $\hat{D}^*M(\bar{x}, \bar{y}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ and *limiting coderivative* $D^*M(\bar{x}, \bar{y}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ at $(\bar{x}, \bar{y})$ as follows

$$\hat{D}^*M(\bar{x}, \bar{y})(y^*) = \left\{x^* \mid (x^*, -y^*) \in \hat{\mathrm{N}}_{\operatorname{gph} M}(\bar{x}, \bar{y})\right\},$$
$$D^*M(\bar{x}, \bar{y})(y^*) = \left\{x^* \mid (x^*, -y^*) \in \mathrm{N}_{\operatorname{gph} M}(\bar{x}, \bar{y})\right\}.$$

If M is single-valued, we write only $D^*M(\bar{x})(y^*)$ instead of $D^*M(\bar{x}, M(\bar{x}))(y^*)$. In this is the case and if $m = 1$, then we have $\partial M(\bar{x}) = D^*M(\bar{x})(1)$. If M is single-valued and smooth, then its coderivative amounts to the adjoint Jacobian. The following stability properties are generalization of Lipschitzian property from functions to multifunctions.

Definition 2.1.10. A multifunction $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ has the *Aubin property* around $(\bar{x}, \bar{y}) \in \operatorname{gph} M$ if there exist a nonnegative modulus L and neighborhoods $\mathcal{U}$ of $\bar{x}$ and $\mathcal{V}$ of $\bar{y}$ such that for all $x, x' \in \mathcal{U}$ the following inclusion holds true

$$M(x) \cap \mathcal{V} \subset M(x') + L\|x - x'\|B(0, 1), \tag{2.4}$$

where $B(0, 1)$ is the closed unit ball in the corresponding space.

We say that M is *calm* at $(\bar{x}, \bar{y}) \in \operatorname{gph} M$ if there exist a nonnegative calmness modulus L and neighborhoods $\mathcal{U}$ of $\bar{x}$ and $\mathcal{V}$ of $\bar{y}$ such that for all $x \in \mathcal{U}$ the following inclusion holds true

$$M(x) \cap \mathcal{V} \subset M(\bar{x}) + L\|x - \bar{x}\|B(0, 1). \tag{2.5}$$

The difference between the Aubin property and calmness is that in the second definition we consider only fixed point $x' = \bar{x}$. Thus, calmness is significantly weaker than the Aubin property. For example, any polyhedral multifunction satisfies the calmness property at any point of its graph while the Aubin property may not be satisfied. If M is single-valued, then the Aubin property reduces exactly to local Lipschitzian property. The infimum of all L satisfying the definition of Aubin property for some $\mathcal{U}$ and $\mathcal{V}$ will be called the modulus of Aubin property.

The use of Aubin property and calmness is twofold. Firstly, they play role of useful stability criteria and as we have already said, the former one reduces to local Lipschitzian continuity whenever M is single-valued. Secondly, they are often used in constraint qualifications for calculus rules. Note that M has the Aubin property around $(\bar{x}, \bar{y})$ if only if M^{-1} is *metrically regular* around $(\bar{y}, \bar{x})$. Similarly, M is calm at $(\bar{x}, \bar{y})$ if only if M^{-1} is *metrically subregular* at $(\bar{y}, \bar{x})$, see [38].

We will briefly mention what calculus rules can be expected from regular, limiting and Clarke normal cones and subdifferentials. Consider first a continuously differentiable mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ and closed sets $C \subset \mathbb{R}^m$ and $D \subset \mathbb{R}^n$ such that

$$D = f^{-1}(C) := \{x| \ f(x) \in C\}.$$

Then we have

$$\begin{array}{ccc}
\hat{\mathrm{N}}_D(\bar{x}) & \supset & \nabla f(\bar{x})^\top \hat{\mathrm{N}}_C(f(\bar{x})) \\
\cap & & \cap \\
\mathrm{N}_D(\bar{x}) & \subset & \nabla f(\bar{x})^\top \mathrm{N}_C(f(\bar{x})) \\
\cap & & \cap \\
\overline{\mathrm{N}}_D(\bar{x}) & \subset & \nabla f(\bar{x})^\top \overline{\mathrm{N}}_C(f(\bar{x})),
\end{array} \tag{2.6}$$

where for both $\subset$ horizontal inclusions some mild constraint qualification is needed. Inclusion $\supset$ and the upper $\subset$ inclusion can be found in [110, Theorem 6.14] while the lower $\subset$ inclusion follows from passing to a convex hull.

We obtain similar results for subdifferential chain rule. Assume that $f = g \circ h$ where $g : \mathbb{R}^m \to \mathbb{R}$ is continuously differentiable and $h : \mathbb{R}^n \to \mathbb{R}^m$ is locally Lipschitz. Then we have the following scheme

$$\begin{array}{ccc} \hat{\partial} f(\bar{x}) & \supset & \hat{D}^* h(\bar{x})(\nabla g(h(\bar{x}))) \\ \cap & & \cap \\ \partial f(\bar{x}) & = & D^* h(\bar{x})(\nabla g(h(\bar{x}))) \\ \cap & & \cap \\ \overline{\partial} f(\bar{x}) & = & \operatorname{co} D^* h(\bar{x})(\nabla g(h(\bar{x}))). \end{array} \tag{2.7}$$

Inclusion $\supset$ can be found in [110, Theorem 10.49] and the upper equality in [87, Theorem 1.110]. The lower equality either follows from passing to a convex hull or can be found in literature as a combination of [27, Theorem 2.6.6] and [86, Proposition 2.11].

Schemes (2.6) and (2.7) lead to a concept of regularity, which together with a constraint qualification ensures equalities in both schemes.

Definition 2.1.11. We say that a closed set C is regular at $\bar{x} \in C$ if $\hat{\mathrm{N}}_C(\bar{x}) = \mathrm{N}_C(\bar{x})$. Similarly, a lower semicontinuous function f is regular at $\bar{x}$ if its epigraph is regular at $(\bar{x}, f(\bar{x}))$. In the opposite case, we say that $\bar{x}$ is a nonregular point of C.

We will finish this subsection with a short example showing that the estimates in (2.6) may not be very sharp.

Example 2.1.12. For notational simplicity define first

$$\begin{aligned} \Gamma_1 &= \mathbb{R}_- \times \mathbb{R}_+, \\ \Gamma_2 &= \mathbb{R}_+ \times \{0\}, \\ \Gamma_3 &= \{0\} \times \mathbb{R}_-. \end{aligned}$$

Now consider set $C = \operatorname{gph} \mathrm{N}_{[0,\infty)} = \Gamma_2 \cup \Gamma_3$, point $(\bar{u}, \bar{x}) = (0, 0)$ and function

$$f(u, x) = \begin{pmatrix} 0 \\ u + x \end{pmatrix}.$$

Then we put $D = \{(u, x) |\ u + x \leq 0\}$, which is a convex set and thus

$$\hat{\mathrm{N}}_D(\bar{u}, \bar{x}) = \mathrm{N}_D(\bar{u}, \bar{x}) = \overline{\mathrm{N}}_D(\bar{u}, \bar{x}) = \{(b, b) |\ b \geq 0\}. \tag{2.8}$$

Since the calmness constraint qualification from [60, Proposition 3.4] is satisfied, we can use scheme (2.6) to obtain

$$\hat{\mathrm{N}}_D(\bar{u}, \bar{x}) \supset \left\{ (b, b) \,\middle|\, \begin{pmatrix} a \\ b \end{pmatrix} \in \Gamma_1 \text{ for some } a \right\} = \{(b, b) \mid b \geq 0\}, \tag{2.9a}$$

$$\mathrm{N}_D(\bar{u}, \bar{x}) \subset \left\{ (b, b) \,\middle|\, \begin{pmatrix} a \\ b \end{pmatrix} \in \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \text{ for some } a \right\} = \{(b, b) \mid b \in \mathbb{R}\}, \tag{2.9b}$$

$$\overline{\mathrm{N}}_D(\bar{u}, \bar{x}) \subset \left\{ (b, b) \,\middle|\, \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2 \text{ for some } a \right\} = \{(b, b) \mid b \in \mathbb{R}\}. \tag{2.9c}$$

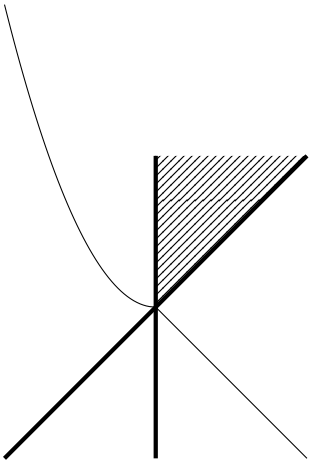


Figure 2.4: Limiting normal cone to a graph of a Lipschitz function

By comparing (2.8) and (2.9), we see that we have obtained the exact estimate only for the regular normal cone (2.9a). For the limiting and Clarke normal cones, we have obtained strict supersets. However, for the limiting normal cone (2.9b), an element of the right–hand side of (2.9b) will lie in $\mathrm{N}_D(\bar{x}, \bar{u})$ if and only if we choose (a, b) such that $(a, b) \in \Gamma_1 \cup \Gamma_2$. When we apply a similar procedure to the implicit programming approach, we will consider only (a, b) such that $(a, b) \in \Gamma_1$ (see the second step of the algorithm at the end of Section 5.3), and thus arrive at a correct element of $\mathrm{N}_D(\bar{x}, \bar{u})$. On the other hand, this is not true for the Clarke normal cone (2.9c). This gives a strong motivation for working with limiting objects instead of the Clarke ones. △

Example 2.1.13. Consider a single–valued locally Lipschitz function

$$f(x) = \begin{cases} x^2 & \text{if } x < 0, \\ -x & \text{if } x \geq 0 \end{cases}$$

and a point $\bar{x} = (0, 0)$. We would like to compute the limiting and Clarke normal cones to $\operatorname{gph} f$ at $\bar{x}$. The situation is depicted in Figure 2.4. We see that the limiting normal cone consists of half of a quadrant and two halflines. This means that the Clarke normal cone amount to the whole space $\mathbb{R}^2$.

This precisely corresponds to [109, Theorem 3.2], which states that the Clarke normal to the graph of a Lipschitz function is a linear subspace. △

More information about the presented objects can be found either in [110] for finite–dimensional setting or in [28, 87] for the infinite–dimensional one.

2.2 Implicit programming approach

We have mentioned in the previous chapter several techniques for solving (1.4) and noted that most of them do not exploit the structural difference between the control and state variables. This changes when S is at least locally single–valued around the solution point and possibly has some additional nice properties. The main idea of the implicit programming approach is to plug $x = S(u)$ into the

objective function to obtain the problem

$$\begin{aligned} &\underset{u}{\text{minimize}}\ J(u, S(u)) \\ &\text{subject to } u \in U_{ad}. \end{aligned} \tag{2.10}$$

Note that the state variable x has been eliminated.

Assume now that both J and S are continuously differentiable. Such situation arises for example when the solution mapping S is governed by an equation

$$e(u, x) = 0,$$

where $e : U \times X \to Z$ is continuously differentiable around $(\bar{u}, \bar{x})$ and the partial derivative $e_x(\bar{u}, \bar{x}) \in \mathcal{L}(X, Z)$ has a bounded inverse, where $\bar{x} = S(\bar{u})$. This statement is nothing less than the famous implicit function theorem. This often happens in optimal control or in optimization with PDE constraints where this approach is known as *direct approach* or *reduced approach*. Multiple examples of such behavior can be found in [58, 132].

Defining $\tilde{J} : U \to \mathbb{R}$ as

$$\tilde{J}(u) := J(u, S(u)),$$

from the chain rule and implicit function theorem we obtain

$$\tilde{J}'(\bar{u}) = J_u(\bar{u}, \bar{x}) + e_u(\bar{u}, \bar{x})^* p,$$

where $p \in Z^*$ is the solution to the *adjoint equation*

$$e_x(\bar{u}, \bar{x})^* p = -J_x(\bar{u}, \bar{x}).$$

For a precise derivation of these formulas, see [58, Section 1.6.2]. In the finite–dimensional case, this corresponds to formulas

$$\begin{aligned} \nabla \tilde{J}(\bar{u}) &= \nabla_u J(\bar{u}, \bar{x}) + \nabla_u e(\bar{u}, \bar{x})^\top p, \\ \nabla_x e(\bar{u}, \bar{x})^\top p &= -\nabla_x J(\bar{u}, \bar{x}), \end{aligned}$$

where gradients are understood as column vectors. Having formula for derivative of $\tilde{J}$, problem (2.10) can be solved by using standard techniques of constrained optimization.

Note that even though we refer to this problem as an MPEC because it is of our abstract form (1.1), thus the lower level $e(u, x) = 0$ describes an equilibrium, most authors would probably not use such terminology in this case.

It often happens that S is not continuously differentiable but only Lipschitz continuous. In the following several paragraphs we will consider such case. We will present first the implicit programming approach in finite dimension and then comment on infinite–dimensional results as well.

We intend to use a subgradient method, which in every step requires to compute function value $\tilde{J}(u)$ and at least one element of Clarke subdifferential $\overline{\partial}\tilde{J}(u)$. After that, we may use a nonsmooth Newton–like method, such as nonsmooth BFGS [74], or we may use a bundle method [123] where the bundle method of storing a bundle of subgradients is combined with the standard idea of trust region.

The computation of $\tilde{J}(u)$ is usually simple. However, the computation of $\overline{\partial}\tilde{J}(u)$ is not so straightforward. There are two main approaches how to compute a Clarke subgradient of this function. Either we will compute precisely a Clarke subgradient or an element of some smaller set than the Clarke subdifferential, for example a regular or limiting subgradient.

A formula for direct computation of the Clarke subdifferential of $\tilde{J}$ would contain the Clarke normal cone to $\operatorname{gph} S$ or the Clarke subdifferential to S. The first object is typically too large as shown in Example 2.1.13. The second object may be too large as well, as shown in the same example. Moreover, since the limiting subdifferential possesses a better developed calculus than the Clarke subdifferential, we have decided to take the second approach. As shown in Example 2.1.8, regular subdifferential may be an empty set and thus we cannot rely on using it. From scheme (2.7) we see that for the computation of $\overline{\partial}\tilde{J}(\bar{u})$ it is advantageous to compute $D^*S(\bar{u})$. Assuming that

$$\operatorname{gph} S = \{(u,x) |\ f(u,x) \in \Lambda\}$$

and that a constraint qualification is satisfied for this system, we have by schemes (2.6) and (2.7) and the definition of coderivative that

$$\partial\tilde{J}(\bar{u}) \subset \nabla_u J(\bar{u},\bar{x}) + \left\{ \nabla_u f(\bar{u},\bar{x})^\top a \;\middle|\; \begin{array}{r} \nabla_x f(\bar{u},\bar{x})^\top a = -\nabla J(S(\bar{u})) \\ a \in \mathrm{N}_\Lambda(f(\bar{u},\bar{x})) \end{array} \right\}. \tag{2.11}$$

Moreover, we obtain equality in the previous inclusion provided at least one of the following two conditions is satisfied. Either $\nabla f(\bar{u},\bar{x})$ has full row rank (and the constraint qualification can be omitted), see [110, Exercise 6.7] or Λ is regular at $f(\bar{u},\bar{x})$, see [110, Theorem 6.14].

From formula (2.11) we see that elements of its right–hand side generally do not have to belong to $\partial\tilde{J}(\bar{u})$. However, from the commentary below it, we see that they will lie in it whenever Λ is regular at $f(\bar{u},\bar{x})$. Even though we will usually consider $\Lambda = \operatorname{gph} \mathrm{N}_C$, which is a nonregular set, computational experience shows that most iterations of implicit programming approach are regular points of Λ and the nonregular points are encountered only when a solution has already been almost reached. For this reason, we may conclude that any solution of the right–hand side of (2.11) usually provides an element of $\partial\tilde{J}(\bar{u})$.

The infinite–dimensional case is much more demanding. The usual way of dealing with such problem is to smoothen it, derive necessary optimality conditions for the smoothened problem and finally obtain optimality conditions for the original problem by passing to the limit. Unfortunately, this often leads to usually weak stationary concepts, see classical results [18, 84] or some newer results [56, 61]. If we again consider $\Lambda = \operatorname{gph} \mathrm{N}_C$ for some set C, then the implicit programming approach requires us to compute $\mathrm{N}_{\operatorname{gph} \mathrm{N}_C}$, which is not a simple task in infinite dimension. Some progress in this direction has been achieved in [57, 93, 140]. Note that all these papers deal with a case of time–independent problems.

For more information about the implicit programming approach, see [80, 94].

2.3 Notation

Our notation is basically standard. We use $\mathbb{R}_+, \mathbb{R}_-, \mathbb{R}_{++}$ and $\mathbb{R}_{--}$ to denote nonnegative, nonpositive, positive and negative real number, respectively. For a set Ω, $\operatorname{cl}\Omega$, $\operatorname{rint}\Omega$, $\operatorname{co}\Omega$, $\operatorname{cone}\Omega$, $\operatorname{span}\Omega$ denote its closure, relative interior, convex hull, conic hull and linear hull, respectively, where relative interior is defined as interior with respect to the smallest affine subspace which contains Ω. We say that Ω is relatively open if $\Omega = \operatorname{rint}\Omega$. For scalar product of x and y we use both $x^\top y$ and $\langle x, y\rangle$. The inverse function is considered as a multifunction so that

$$g^{-1}(K) = \{x|\ g(x) \in K\}.$$

For a function f, by $\operatorname{dom} f$ we understand its domain and by $\operatorname{epi} f$ its epigraph.

From Chapter 4 onward, we will often omit the arguments of f. Partial derivatives $\nabla_u f$, $\nabla_x f$ and $\nabla_v f$ are taken with respect to u, x and $\dot{x}$, respectively. Upper index K denoting the discretization level will often be omitted, especially in cases when K is fixed and no convergence analysis comes into play. For discretized problem we will often use shortened notation such as $x = (x_1, \ldots, x_K)$ or $u = (u_1, \ldots, u_K)$.

Concerning the used function spaces and norms, by L^p we understand the Lebesgue space $L^p([0,T], \mathbb{R}^n)$ and $W^{1,p}$ denotes the Sobolev spaces of functions with (weak) derivative in L^p. Time derivative will be denoted by a dot and the norms in the above spaces by $\|\cdot\|_p$ and $\|\cdot\|_{1,p}$, respectively. Further, we define $|\cdot| := \|\cdot\|_2$. However, sometimes it will be almost obligatory to emphasize which norm has been used, especially when both finite– and infinite–dimensional ones are present. In this case, we use $|\cdot|_{l^2}$ and $|\cdot|_{L^2}$, respectively. On product spaces we consider the standard Euclidean norm. For the weak convergence we use the notation $x_n \rightharpoonup x$, for the uniform one $x_n \rightrightarrows x$ and $x_n \overset{A}{\to} x$ signifies the usual convergence $x_n \to x$ with the condition that $x_n \in A$.

Apart from Chapter 6, u will denote the control variable and x the state variable. Because Chapter 6 contains integrals with respect to space variables, we have used the standard notation from partial differential equations and x denotes the space variable and u displacement. In this last chapter, π stands for the control variable and (u, z) for the state variable.

3. Computation of limiting normal cone to union of polyhedral sets

3.1 Introduction

In the past few decades, applied mathematicians have paid a lot of attention to optimization and optimal control problems with various types of nonconvex constraints. In the variational geometry of nonconvex sets, the so-called tangent (Bouligand-Severi, contingent) cone, regular (Fréchet) normal cone and limiting (Mordukhovich) normal cone play important role in study of optimization and optimal control, such as optimality conditions, related constraint qualifications, stability analysis etc., see [110] for theory in finite dimensions and [87, 88] for analysis in infinite-dimensional spaces. All cones mentioned above enjoy calculus rules that may simplify their calculations. However, in many cases, calculus provides only approximation (inclusion) which may not be useful for further analysis. Thus, exact computation for even trivial nonconvex set may become a very technical and lengthy procedure.

In this chapter we focus on computation of normal cones to a finite–dimensional set Γ, which is a union of finitely many (convex) polyhedra. By polyhedron we understand a finite intersection of halfspaces, which is always closed and convex. Such sets naturally arise whenever a parameterized generalized equation

$$0 \in F(u, x) + G(u, x) \tag{3.1}$$

is considered with a continuously differentiable function $F : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$, a polyhedral multifunction $G : \mathbb{R}^d \times \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, a parameter or control variable u and a state variable x. Computation of a generalized derivative of a solution map $S : u \mapsto x$ associated with (3.1) is often connected with evaluation of some of the above mentioned cones to $\Gamma := \operatorname{gph} G$. Since G is a polyhedral multifunction, Γ is indeed a union of a finite number of polyhedra.

More specifically, the computation of such cones is essential for sensitivity analysis of generalized equations with polyhedral multifunction (3.1). Moreover, it plays an important role in the so-called disjunctive programs [45]. The class of disjunctive programs include, e.g., bilevel problems with linear constraints on the lower level [33], mathematical programs with complementarity constraints (MPCCs) [80, 94] and related problems such as mathematical programs with vanishing constraints [2], etc. Considering a polyhedral set C, the graph of the normal cone mapping $\mathrm{N}_C(\cdot)$ in the sense of convex analysis also enjoys this special structure, as already observed in [105], having importance in many aspects of variational analysis.

There has already been some attempts to provide formulas for normal cones to such sets Γ. In [37], the authors provide formula for the limiting normal cone to $\operatorname{gph} \mathrm{N}_C$, with C polyhedral, in terms of the so-called critical cones and their polars. This special case of a union of polyhedra has also been studied in [55]. In [54], the formula for the fully general case of a union of polyhedra has been provided utilizing the Motzkin's Theorem of the Alternative. There, the authors already build upon the well-known fact that the tangent and normal cones are constant

on relative interior of a face of a polyhedral set, result that goes back to Robinson [105]. Additionally to simplified formulas for several special cases, a formula for normal cone to a particular case of a union of non-polyhedral sets is provided in [54]. In all the above mentioned papers, however, the resulting formulas are non-trivial with highly growing complexity with respect to the number of faces.

In this chapter, we describe an alternative procedure for computation of full graph of normal cone mappings to Γ along with normal cones at a specific point. For this, we introduce the so-called *normally admissible stratification* of a union of polyhedra in order to generalize the observation of constant-valuedness of tangent and normal cone mappings on certain subsets of a polyhedra. Our results can be considered as a natural generalization of [25] where formulas for tangent and normal cones were derived for a special case of a union of polyhedra with each polyhedral set being a subset of $\{\mathbb{R}, \mathbb{R}_+, \mathbb{R}_-, \{0\}\}^n$. We obtain formulas which hold as equalities without any constraint qualification. This seems to be natural for the considered polyhedral setting. However, to the best of our knowledge, such a result cannot be achieved by applying general calculus rules without any additional information.

The chapter is organized as follows. In Section 3.2 we provide the definition of a normally admissible stratification of Γ and show that such stratification always exists. Further, we derive formulas for graphs of regular and limiting normal cones to Γ. In Section 3.3 we compare our procedure to those of Dontchev and Rockafellar [37] and Henrion and Outrata [54]. Finally, in Section 3.4 we consider an application arising in discretized time–dependent problems [1, 19]. We provide a theoretical background, specifying the form of normally admissible stratifications in this particular class of problems, and illustrate the benefits of our procedure on a special case arising in delamination modeling [114], a topic which will be later studied in Chapter 6.

3.2 Computation of normal cone

The main goal of this section is to compute $\hat{\mathrm{N}}_\Gamma$ and N_Γ, where $\Gamma \subset \mathbb{R}^n$ is a finite union of polyhedral sets Ω_r for $r = 1, \ldots, R$, that is

$$\Gamma = \bigcup_{r=1}^{R} \Omega_r. \tag{3.2}$$

In order to compute these normal cones, we will first introduce a convenient partition of Γ which satisfies certain suitable conditions. Next, we show existence of such partition. Finally, we derive formulas for both Fréchet and limiting normal cones to Γ.

Definition 3.2.1. We say that $\{\Gamma_s | \, s = 1, \ldots, S\}$ forms a *partition* of Γ if Γ_s are nonempty and pairwise disjoint for all $s = 1, \ldots, S$ and $\cup_{s=1}^{S} \Gamma_s = \Gamma$.

The following definition of normally admissible stratification is based on the strata theory [49, 102] which was developed for general manifolds. In the polyhedral case, we may add additional assumptions such as that stratas Γ_s are relatively open. Note that condition (3.3) is well–known as the so–called *frontier condition*. Similar partition was proposed in [121] under the term *polyhedral subdivision* with all the partition elements being closed polyhedra of the same dimension as Γ.

Definition 3.2.2. We say that $\{\Gamma_s | \, s = 1, \dots, S\}$ forms a *normally admissible stratification* of Γ if it is a partition of Γ with Γ_s, $s = 1, \dots, S$ relatively open, convex and $\operatorname{cl} \Gamma_s$ polyhedral such that the following property holds true for all $i, s = 1, \dots, S$

$$\Gamma_s \cap \operatorname{cl} \Gamma_i \neq \emptyset \implies \Gamma_s \subset \operatorname{cl} \Gamma_i. \tag{3.3}$$

The term normally admissible stratification is coined in order to reflect the forthcoming Theorem 3.2.5 saying that normal cones are constant with respect to this stratification in a particular sense. Next, for a normally admissible stratification of Γ denoted by $\{\Gamma_s | \, s = 1, \dots, S\}$ we define two index sets which are extensively used throughout this chapter

$$I(s) := \{i \in \{1, \dots, S\} | \; \Gamma_s \cap \operatorname{cl} \Gamma_i \neq \emptyset\}, \tag{3.4a}$$

$$\tilde{I}(s) := \{i \in I(s) | \; \nexists j \in I(s) : \; \operatorname{cl} \Gamma_i \subsetneq \operatorname{cl} \Gamma_j\} \subset I(s). \tag{3.4b}$$

Clearly, $I(s)$ has a close connection with (3.3) and $\tilde{I}(s)$ is composed of such indices of $I(s)$ that correspond to maximal elements of $\{\operatorname{cl} \Gamma_i | \, i \in I(s)\}$ in the sense of subsets. We will often work with the following alternative representations of $\tilde{I}(s)$

$$\tilde{I}(s) = \{i \in I(s) | \; \forall j \in I(s) : \; \operatorname{cl} \Gamma_i \subset \operatorname{cl} \Gamma_j \implies i = j\} \tag{3.4c}$$

$$= \{i \in I(s) | \; j \in I(s) \cap I(i) \implies i = j\}. \tag{3.4d}$$

For a normally admissible stratification, formula (3.4b) is equivalent to (3.4c) due to [108, Theorem 6.3]. The equivalence of (3.4c) and (3.4d) follows from the fact that $j \in I(i)$ is equivalent to $\Gamma_i \subset \operatorname{cl} \Gamma_j$.

Next, we provide a constructive proof of existence of a normally admissible stratification to Γ.

Lemma 3.2.3. *Let $\Gamma \subset \mathbb{R}^n$ be a finite union of polyhedral sets. Then there exists a normally admissible stratification of Γ.*

Proof. Consider Γ in the form (3.2) with Ω_r defined as

$$\Omega_r = \{x | \, \langle c_t^r, x \rangle \leq b_t^r, \; t = 1, \dots, T(r)\}.$$

We now relabel all c_t^r to c_u, $u = 1, \dots, U$ with $U = \sum_{r=1}^{R} T(r)$ and similarly for b_u. For $I, J \subset \{1, \dots, U\}$ define the following sets

$$\Omega_{I,J} := \left\{ x \;\middle|\; \begin{array}{ll} \langle c_u, x \rangle < b_u & \text{for } u \in I \\ \langle c_u, x \rangle > b_u & \text{for } u \in J \\ \langle c_u, x \rangle = b_u & \text{for } u \in \{1, \dots, U\} \setminus (I \cup J) \end{array} \right\}, \tag{3.5}$$

$$\Theta := \{(I, J) \mid \Omega_{I,J} \neq \emptyset, \; \Omega_{I,J} \subset \Gamma\}. \tag{3.6}$$

We claim that $\{\Omega_{I,J} | \, (I, J) \in \Theta\}$ is a normally admissible stratification of Γ.

First, we show that $\{\Omega_{I,J} | \, (I, J) \in \Theta\}$ is a stratification of Γ. Indeed, if we restrict ourselves to $(I, J) \in \Theta$, then $\Omega_{I,J}$ are nonempty and pairwise disjoint by construction. Moreover, since $\Omega_{I,J} \subset \Gamma$, we have

$$\bigcup_{(I,J) \in \Theta} \Omega_{I,J} \subset \Gamma.$$

To show that the equality holds in the previous relation, choose any $x \in \Gamma$. By construction of sets $\Omega_{I,J}$, there exists exactly one couple (I, J) such that $x \in \Omega_{I,J}$. To show that $(I, J) \in \Theta$, it remains to realize that

$$\Omega_{I,J} \subset \bigcap_{\{r|\, x \in \Omega_r\}} \Omega_r \subset \Gamma.$$

Hence, we have shown that $\{\Omega_{I,J} |\, (I, J) \in \Theta\}$ is indeed a stratification of Γ.

To prove that $\{\Omega_{I,J} |\, (I, J) \in \Theta\}$ is a normally admissible stratification of Γ, recall that for all $(I, J) \in \Theta$ we have $\Omega_{I,J}$ nonempty, which allows us to apply Lemma A.1.1 to obtain that $\Omega_{I,J}$ is relatively open and

$$\operatorname{cl} \Omega_{I,J} = \left\{ x \;\middle|\; \begin{array}{ll} \langle c_u, x \rangle \leq b_u & \text{for } u \in I \\ \langle c_u, x \rangle \geq b_u & \text{for } u \in J \\ \langle c_u, x \rangle = b_u & \text{for } u \in \{1, \ldots, U\} \setminus (I \cup J) \end{array} \right\}.$$

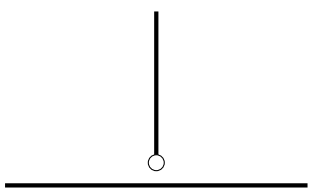
Clearly, $\Omega_{I,J}$ is convex and $\operatorname{cl} \Omega_{I,J}$ polyhedral. Thus, it remains to show that property (3.3) holds. Assume that there is some $x \in \Omega_{I_1,J_1} \cap \operatorname{cl} \Omega_{I_2,J_2}$. This immediately means $I_1 \subset I_2$ and $J_1 \subset J_2$. But this implies that $\Omega_{I_1,J_1} \subset \operatorname{cl} \Omega_{I_2,J_2}$, which concludes the proof. □

Next we show a simple example with several possible partitions of a given set, where only some are normally admissible stratifications.

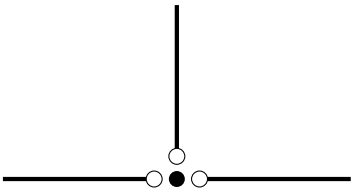
Example 3.2.4. Consider the following union of two polyhedral sets $\Gamma = \big(\mathbb{R} \times \{0\}\big) \cup \big(\{0\} \times \mathbb{R}_+\big)$. One possible partition of Γ to relatively open sets is $\Gamma = \Gamma_1 \cup \Gamma_2$ with

$$\Gamma_1 = \mathbb{R} \times \{0\},\ \Gamma_2 = \{0\} \times \mathbb{R}_{++}.$$

Since $(0, 0) \in \Gamma_1 \cap \operatorname{cl} \Gamma_2$, we have $I(1) = \{1, 2\}$. However, as $(1, 0) \in \Gamma_1$ and $(1, 0) \notin \operatorname{cl} \Gamma_2$ condition (3.3) is not satisfied for $s = 1$ and $i = 2$ and hence this partition is not normally admissible stratification. This situation is depicted in Figure 3.1a.



(a) Not a normally admissible partition



(b) A normally admissible partition

Figure 3.1: Possible partitions of the set from Example 3.2.4

To remedy the situation, one may consider the following partition $\Gamma = \bigcup_{s=1}^{4} \tilde{\Gamma}_s$ with

$$\tilde{\Gamma}_1 = \mathbb{R}_{--} \times \{0\},\ \tilde{\Gamma}_2 = \{0\} \times \{0\},\ \tilde{\Gamma}_3 = \mathbb{R}_{++} \times \{0\},\ \tilde{\Gamma}_4 = \{0\} \times \mathbb{R}_{++},$$

see Figure 3.1b. It is simple to verify that this is indeed a normally admissible stratification of Γ.

Now we present the main motivation for considering normally admissible stratification which states that the tangent and normal cone mappings are constant with respect to a particular component of this stratification.

Theorem 3.2.5. *Consider a finite union of polyhedral sets Γ and its normally admissible stratification $\{\Gamma_s | \, s = 1, \dots, S\}$. Then for any $s \in \{1, \dots, S\}$, $i \in I(s)$ and $x, y \in \Gamma_s$ we have*

$$\mathrm{T}_{\mathrm{cl}\,\Gamma_i}(x) = \mathrm{T}_{\mathrm{cl}\,\Gamma_i}(y) \qquad \textit{and} \qquad \hat{\mathrm{N}}_{\mathrm{cl}\,\Gamma_i}(x) = \hat{\mathrm{N}}_{\mathrm{cl}\,\Gamma_i}(y). \tag{3.7}$$

Proof. From [108, Theorem 18.2] we know that Γ_s is contained in a relatively open face of $\mathrm{cl}\,\Gamma_i$, and so the statement follows from [43, Chapter 1, Lemma 4.11]. □

From Theorem 3.2.5 we know that for any s and $i \in I(s)$, tangent cone $\mathrm{T}_{\mathrm{cl}\,\Gamma_i}(x)$ does not depend on a choice of $x \in \Gamma_s$. To simplify notation, we denote this constant value by

$$\mathrm{T}_{\mathrm{cl}\,\Gamma_i}(\Gamma_s) := \mathrm{T}_{\mathrm{cl}\,\Gamma_i}(x_0) \qquad \text{for arbitrary} \qquad x_0 \in \Gamma_s.$$

In a similar way, we will use notation $\hat{\mathrm{N}}_{\mathrm{cl}\,\Gamma_i}(\Gamma_s)$ and $\mathrm{N}_{\mathrm{cl}\,\Gamma_i}(\Gamma_s)$. In the sequel it will become clear that formula (3.7) is the cornerstone of this chapter.

In the following example we present a set and its several possible partitions. The first partition satisfies formula (3.7) even though one of its components is nonconvex, meaning that this partition is not normally admissible stratification. For the other two partitions considered, we show that neither condition (3.3) nor convexity can be dropped from Definition 3.2.2 in order to satisfy Theorem 3.2.5.

Example 3.2.6. Consider $\Gamma = \Omega_1 \cup \Omega_2$ to be union of $\Omega_1 = [0, 3] \times [0, 1]$ and $\Omega_2 = [0, 2] \times [1, 2]$. Then, one of the possible partitions of Γ, elements of which are relatively open and satisfy condition (3.3), contains a nonconvex plane segment

$$\Gamma_1 = \Big((0, 3) \times (0, 1)\Big) \cup \Big((0, 2) \times (0, 2)\Big),$$

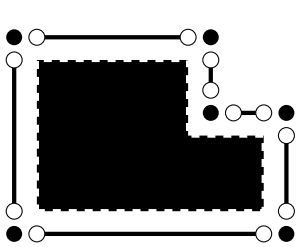
six points and six line segments, see Figure 3.2a. Since $\mathrm{cl}\,\Gamma_1$ is nonconvex, this partition is not normally admissible stratification. However, it is not difficult to verify that the statement of Theorem 3.2.5 holds true. To show an example, consider $s = 1$. Clearly, $I(1) = \{1\}$ and for all $x \in \Gamma_1$ we observe that $\mathrm{T}_{\mathrm{cl}\,\Gamma_1}(x) = \mathbb{R}^2$ and thus $\mathrm{T}_{\mathrm{cl}\,\Gamma_i}(\Gamma_1)$ is indeed well-defined for all $i \in I(1)$.

It is simple to find a normally admissible stratification of Γ. For example, it may consists of two rectangles, eight line segments and seven points as depicted in Figure 3.2b. Now we illustrate the role of condition (3.3) in Theorem 3.2.5. Consider any normally admissible stratification of Γ containing the following sets

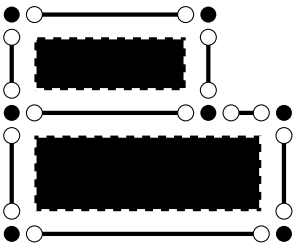
$$\tilde{\Gamma}_1 = (0, 3) \times \{1\}, \; \tilde{\Gamma}_2 = (0, 2) \times (1, 2).$$

Since $(1, 1) \in \tilde{\Gamma}_1 \cap \mathrm{cl}\,\tilde{\Gamma}_2$, we have $2 \in I(1)$. However, it is clear that $\tilde{\Gamma}_1 \not\subset \mathrm{cl}\,\tilde{\Gamma}_2$ and thus (3.3) is violated. Moreover, we have

$$\mathrm{T}_{\mathrm{cl}\,\tilde{\Gamma}_2}((2, 1)) = \mathbb{R}_- \times \mathbb{R}_+,$$
$$\mathrm{T}_{\mathrm{cl}\,\tilde{\Gamma}_2}((1, 1)) = \mathbb{R} \times \mathbb{R}_+,$$



(a) Partition satisfying the result of Theorem 3.2.5 but not being normally admissible.



(b) A partition showing the need of convexity. Note that the rectangles are considered as one set.

Figure 3.2: Possible partitions of the set from Example 3.2.6

even though $(2,1) \in \tilde{\Gamma}_1$ and $(1,1) \in \tilde{\Gamma}_1$. Thus, formula (3.7) does not hold for $s = 1$ and $i = 2$.

Next, consider a partition of Γ with

$$\hat{\Gamma}_1 = [(0,2) \times \{1\}] \cup [(2,3) \times \{1\}],$$
$$\hat{\Gamma}_2 = [(0,3) \times (0,1)] \cup [(0,2) \times (1,2)],$$

and seven points and six line segments, see Figure 3.2b. Then all the conditions for normally admissible stratification with the exception of convexity of $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ and the polyhedrality of $\operatorname{cl}\hat{\Gamma}_2$ are satisfied but Theorem 3.2.5 does not hold true. Finally, observe that indeed $\tilde{\Gamma}_1 \subset \operatorname{cl}\hat{\Gamma}_2$.

We are now ready to provide the main result of this section which concerns the computation of normal cones to finite union of polyhedra.

Theorem 3.2.7. *Let Γ be a finite union of polyhedral sets and $\{\Gamma_s | \, s = 1, \dots, S\}$ be its normally admissible stratification. Then for any $x \in \Gamma_s$ we have $\hat{\mathrm{N}}_\Gamma(x) = \hat{\mathrm{N}}_\Gamma(\Gamma_s)$ and further*

$$\hat{\mathrm{N}}_\Gamma(\Gamma_s) = \bigcap_{i \in I(s)} \hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(\Gamma_s) = \bigcap_{i \in \tilde{I}(s)} \hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(\Gamma_s). \tag{3.8}$$

Moreover, for graphs of Fréchet and limiting normal cones we have the following formulas

$$\operatorname{gph}\hat{\mathrm{N}}_\Gamma = \bigcup_{s=1}^{S} \left(\Gamma_s \times \hat{\mathrm{N}}_\Gamma(\Gamma_s)\right), \tag{3.9}$$

$$\operatorname{gph}\mathrm{N}_\Gamma = \bigcup_{s=1}^{S} \left(\operatorname{cl}\Gamma_s \times \hat{\mathrm{N}}_\Gamma(\Gamma_s)\right). \tag{3.10}$$

Proof. Fix any $x \in \Gamma_s$. Then by simple calculus we obtain

$$\mathrm{T}_\Gamma(x) = \mathrm{T}_{\bigcup_{i \in I(s)} \operatorname{cl}\Gamma_i}(x) = \bigcup_{i \in I(s)} \mathrm{T}_{\operatorname{cl}\Gamma_i}(x),$$
$$\hat{\mathrm{N}}_\Gamma(x) = \bigcap_{i \in I(s)} \hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(x).$$

With regards to Theorem 3.2.5 we obtain the first equality in (3.8). The second equality in (3.8) follows from the fact that $\Gamma_s \subset \operatorname{cl}\Gamma_i \subset \operatorname{cl}\Gamma_j$ implies $\hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(\Gamma_s) \supset \hat{\mathrm{N}}_{\operatorname{cl}\Gamma_j}(\Gamma_s)$.

Formula (3.9) is a direct consequence of (3.8). Since $\operatorname{gph}\mathrm{N}_\Gamma$ is a closure of $\operatorname{gph}\hat{\mathrm{N}}_\Gamma$ by definition, equation (3.10) follows as well. □

In some situations, computation of normal cone $\mathrm{N}_\Gamma(\bar{x})$ only at one particular point $\bar{x} \in \Gamma$ is required instead of computation of the whole graph of the normal cone mapping. The following corollary concerns such a case.

Corollary 3.2.8. *Under assumptions of Theorem 3.2.7, for any $\bar{x} \in \Gamma$ denote by $\bar{s}$ the index of the unique component $\Gamma_{\bar{s}}$ such that $\bar{x} \in \Gamma_{\bar{s}}$. Then*

$$\hat{\mathrm{N}}_\Gamma(\bar{x}) = \hat{\mathrm{N}}_\Gamma(\Gamma_{\bar{s}}) = \bigcap_{i\in I(\bar{s})} \hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(\Gamma_{\bar{s}}) = \bigcap_{i\in \tilde{I}(\bar{s})} \hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(\Gamma_{\bar{s}}), \tag{3.11}$$

$$\mathrm{N}_\Gamma(\bar{x}) = \bigcup_{s\in I(\bar{s})} \hat{\mathrm{N}}_\Gamma(\Gamma_s) = \bigcup_{s\in I(\bar{s})} \bigcap_{i\in I(s)} \hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(\Gamma_s) = \bigcup_{s\in I(\bar{s})} \bigcap_{i\in \tilde{I}(s)} \hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(\Gamma_s). \tag{3.12}$$

Remark 3.2.9. Relations similar to (3.11) and (3.12), see (3.13) and (3.14) below, can be obtained by simpler means. We present them to show the possible advantages of our approach. First, defining $J(x) := \{s|\, x \in \operatorname{cl}\Gamma_s\}$ we observe that $J(x) = I(t)$ where t is the unique index such that $x \in \Gamma_t$. Indeed, if $s \in J(x)$, then $x \in \operatorname{cl}\Gamma_s$, which together with assumed $x \in \Gamma_t$ implies $x \in \Gamma_t \cap \operatorname{cl}\Gamma_s$ and thus $s \in I(t)$. On the other hand, if $s \in I(t)$, then as the considered partition is normally admissible stratification, we have $x \in \Gamma_t \subset \operatorname{cl}\Gamma_s$ and thus $s \in J(x)$, which implies the desired equality.

Formula (3.11) may then be derived in the following way

$$\hat{\mathrm{N}}_\Gamma(\bar{x}) = (\mathrm{T}_{\bigcup_{i\in J(\bar{x})} \operatorname{cl}\Gamma_i}(\bar{x}))^* = \left(\bigcup_{i\in J(\bar{x})} \mathrm{T}_{\operatorname{cl}\Gamma_i}(\bar{x})\right)^* = \bigcap_{i\in J(\bar{x})} \hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(\bar{x}) = \bigcap_{i\in I(\bar{s})} \hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(\bar{x}), \tag{3.13}$$

Similarly, for a sufficiently small neighborhood $\mathcal{X}$ of $\bar{x}$, one may obtain formula for limiting normal cone directly from (3.13) as

$$\mathrm{N}_\Gamma(\bar{x}) = \bigcup_{x\in\mathcal{X}} \bigcap_{i\in J(x)} \hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(x). \tag{3.14}$$

Although it is obvious that the union with respect to $x \in \mathcal{X}$ will reduce to a union with respect to a finite number of elements, it is not entirely clear how to obtain this reduction without the concept of normally admissible stratification.

We conclude this section with a note that the computation of normal cones can be performed repeatedly, by which we mean that formula (3.10) provides a good background for computation of $\operatorname{gph}\mathrm{N}_{\operatorname{gph}\mathrm{N}_\Gamma}$.

Remark 3.2.10. Consider a normally admissible stratification $\{\Gamma_s|\, s = 1, \ldots, S\}$ of Γ. It follows from Lemma A.1.3 that $\{\Gamma_t|\, s \in I(t)\}$ is a normally admissible stratification of $\operatorname{cl}\Gamma_s$ for any s.

Moreover, it is possible to show that

$$\{\Gamma_s \times D_{st}|\, s = 1, \ldots, S,\ t = 1, \ldots, T(s)\}$$

is a normally admissible stratification of $\operatorname{gph} \mathrm{N}_\Gamma$, where $\{D_{st} | \, t = 1, \ldots, T(s)\}$ are suitable normally admissible stratifications of $\mathrm{N}_\Gamma(\Gamma_s)$ for $s = 1, \ldots, S$. However, since the construction of D_{st} is not entirely simple and it is not used later in the text, we omit it here.

3.3 Relation to known results

This section revisits some notable results of other authors on computation of the limiting normal cone to a union of polyhedral sets and exploits the relationship between their results and those presented in the previous section. We firstly recall the result of Dontchev and Rockafellar in [37], where formula for the limiting normal cone to a special case of a union of polyhedral sets was given in terms of critical cones and then show that formulas from Corollary 3.2.8 coincides with those of Dontchev and Rockafellar. Secondly, we summarize the results of Henrion and Outrata in [54] who also considered a general union of polyhedral sets. Direct comparison yields that the explicit formula derived by Henrion and Outrata can be considered as a special case of our approach. We omit a detailed comparison with results of Červinka, Outrata and Pištěk in [25] due to the fact that their results are special case of Theorem 3.2.7.

3.3.1 Normal cones to graph of a normal cone to a polyhedral set

To our knowledge, the first attempt to provide explicit formulas for computation of the limiting normal cone to a union of polyhedral sets can be found in [37]. It concerns a rather special case where $\Gamma = \operatorname{gph} \mathrm{N}_C \subset \mathbb{R}^{2n}$ with $C \subset \mathbb{R}^n$ being polyhedral. Due to polyhedrality of C, Γ is indeed a union of finitely many polyhedral sets. Interestingly, the formula for $\mathrm{N}_\Gamma(\bar{x}, \bar{y})$ was not given in [37] as a separate result but as a part of a proof of another result. We first define the critical cone and then state the result in a proposition.

Definition 3.3.1. For a polyhedral set C and some $\bar{x} \in C$ and $\bar{y} \in \mathrm{N}_C(\bar{x})$, the critical cone to C at $\bar{x}$ for $\bar{y}$ is defined as

$$K_C(\bar{x}, \bar{y}) := \{w \in \mathrm{T}_C(\bar{x}) | \; w^\top \bar{y} = 0\}. \tag{3.15}$$

Proposition 3.3.2 ([37], part of the proof of Theorem 2)**.** *Consider a polyhedral set C and some $\bar{x} \in C$ and $\bar{y} \in \mathrm{N}_C(\bar{x})$. Then*

$$\begin{aligned} \hat{\mathrm{N}}_{\operatorname{gph} \mathrm{N}_C}(\bar{x}, \bar{y}) &= K_C(\bar{x}, \bar{y})^* \times K_C(\bar{x}, \bar{y}), \\ \mathrm{N}_{\operatorname{gph} \mathrm{N}_C}(\bar{x}, \bar{y}) &= \bigcup_{(x,y) \in \mathcal{U}} K_C(x, y)^* \times K_C(x, y), \end{aligned} \tag{3.16}$$

for some sufficiently small neighborhood $\mathcal{U}$ of $(\bar{x}, \bar{y})$.

The original proof of Proposition 3.3.2 by Dontchev and Rockafellar is based on application of the so-called Reduction Lemma, cf. [38, Lemma 2E.4]. To illuminate the relation between Proposition 3.3.2 and Corollary 3.2.8, we provide an alternative proof exploiting the properties of relatively open faces forming a partition of a polyhedral set, see [108, Theorem 18.2]. To this end, we recall the definition of faces of a convex set, see [79].

Definition 3.3.3. A subset F of a convex set P is called a face of P provided the following implication holds true: if x_1 and x_2 belong to P and $\lambda x_1+(1-\lambda)x_2 \in F$ for some $\lambda \in (0,1)$, then x_1 and x_2 belong to F as well. We say that $\tilde{F}$ is a relatively open face of P if there exists a face F of P such that $\tilde{F} = \operatorname{rint} F$.

Consider all nonempty faces of a polyhedral set C and let us denote them $\tilde{C}_s$ with $s = 1, \dots, S$. We shall call $C_s := \operatorname{rint} \tilde{C}_s$ relatively open faces of C. By virtue of Lemma A.1.4 we obtain that $\{C_s | \, s = 1, \dots, S\}$ form a normally admissible stratification of C. Thus, Theorem 3.2.5 implies that $\mathrm{N}_C(x)$ has the same value for all $x \in C_s$. Following the notation developed in previous sections, let us denote it by $\mathrm{N}_C(C_s)$. Since $\mathrm{N}_C(C_s)$ is also a polyhedral set, we can as well find its relatively open faces D_{st}. Again, let $\{D_{st} | \, t = 1, \dots, T(s)\}$ form a normally admissible stratification of $\mathrm{N}_C(C_s)$. This results in the following representation of Γ:

$$\Gamma := \operatorname{gph} \mathrm{N}_C = \bigcup_{s=1}^{S} \bigcup_{t=1}^{T(s)} C_s \times D_{st}.$$

It follows from Lemma A.1.4 that $\{C_s \times D_{st} | \, s = 1, \dots, S, \ t = 1, \dots, T(s)\}$ forms a normally admissible stratification of Γ.

As a consequence, for a given pair $\bar{x} \in C$ and $\bar{y} \in \mathrm{N}_C(\bar{x})$ there is a unique couple of indices $(\bar{s}, \bar{t})$ such that $(\bar{x}, \bar{y}) \in C_{\bar{s}} \times D_{\bar{s}\bar{t}}$. By application of Corollary 3.2.8 to $(\bar{x}, \bar{y}) \in \operatorname{gph} \mathrm{N}_C$, we immediately obtain

$$\begin{aligned} \hat{\mathrm{N}}_{\operatorname{gph} \mathrm{N}_C}(\bar{x}, \bar{y}) &= \bigcap_{(i,j) \in I(\bar{s},\bar{t})} \mathrm{N}_{\operatorname{cl}(C_i \times D_{ij})}(C_{\bar{s}} \times D_{\bar{s}\bar{t}}), \\ \mathrm{N}_{\operatorname{gph} \mathrm{N}_C}(\bar{x}, \bar{y}) &= \bigcup_{(s,t) \in I(\bar{s},\bar{t})} \bigcap_{(i,j) \in I(s,t)} \mathrm{N}_{\operatorname{cl}(C_i \times D_{ij})}(C_s \times D_{st}). \end{aligned} \tag{3.17}$$

Since Γ is the union of finitely many polyhedral sets, only finitely many cones can be manifested as $\hat{\mathrm{N}}_\Gamma(x, y)$ at points $(x, y) \in \Gamma$ near $(\bar{x}, \bar{y})$. It is not difficult to see that each of such cones corresponds to $\hat{\mathrm{N}}_\Gamma(C_s, D_{st})$ with $(s, t) \in I(\bar{s}, \bar{t})$. Invoking Remark 3.2.9, this establishes the correspondence of union in (3.17) with union in (3.16). In order to show equivalence of (3.16) and (3.17), consider a fixed pair of indices $(s, t) \in I(\bar{s}, \bar{t})$ and let us simplify the intersection in (3.17). By elementary operations and [110, Proposition 6.41] we obtain

$$\bigcap_{(i,j) \in I(s,t)} \mathrm{N}_{\operatorname{cl}(C_i \times D_{ij})}(C_s \times D_{st}) = \bigcap_{\{(i,j) |\ C_s \subset \operatorname{cl} C_i,\ D_{st} \subset \operatorname{cl} D_{ij}\}} \left[\mathrm{N}_{\operatorname{cl} C_i}(C_s) \times \mathrm{N}_{\operatorname{cl} D_{ij}}(D_{st})\right]. \tag{3.18}$$

Note that for any i there exists an index $l \in \{1, \dots, T(i)\}$ such that $\operatorname{cl} D_{il} = \mathrm{N}_C(C_i)$. This means that for every $j \in \{1, \dots, T(i)\}$ such that $D_{st} \subset \operatorname{cl} D_{ij}$ we have $\operatorname{cl} D_{ij} \subset \operatorname{cl} D_{il} = \mathrm{N}_C(C_i)$. This, in turn, implies that $\mathrm{N}_{\operatorname{cl} D_{ij}}(D_{st}) \supset \mathrm{N}_{\mathrm{N}_C(C_i)}(D_{st})$. In particular, we have

$$\begin{aligned} \bigcap_{(i,j) \in I(s,t)} \mathrm{N}_{\operatorname{cl}(C_i \times D_{ij})}(C_s \times D_{st}) &= \bigcap_{\{i |\ C_s \subset \operatorname{cl} C_i,\ D_{st} \subset \mathrm{N}_C(C_i)\}} \left[\mathrm{N}_{\operatorname{cl} C_i}(C_s) \times \mathrm{N}_{\mathrm{N}_C(C_i)}(D_{st})\right] \\ = \left[\bigcap_{\{i |\ C_s \subset \operatorname{cl} C_i,\ D_{st} \subset \mathrm{N}_C(C_i)\}} \mathrm{N}_{\operatorname{cl} C_i}(C_s)\right] &\times \left[\bigcap_{\{i |\ C_s \subset \operatorname{cl} C_i,\ D_{st} \subset \mathrm{N}_C(C_i)\}} \mathrm{N}_{\mathrm{N}_C(C_i)}(D_{st})\right]. \end{aligned} \tag{3.19}$$

It suffices to show that both parts of the Cartesian product in (3.16) correspond to those of (3.19). To verify that, we present the following two lemmas. Note that a result similar to the first lemma was proved in [78, Theorem 5.2].

Lemma 3.3.4. *For any $x \in C_s$ and $y \in D_{st}$ the following equality holds*

$$K(x,y) = \bigcap_{\{i|\ C_s \subset \operatorname{cl} C_i,\ D_{st} \subset \mathrm{N}_C(C_i)\}} \mathrm{N}_{\mathrm{N}_C(C_i)}(D_{st}). \tag{3.20}$$

Proof. In order to verify (3.20), note first that for any i such that $C_s \subset \operatorname{cl} C_i$ and $D_{st} \subset \mathrm{N}_C(C_i)$ we have $\mathrm{N}_C(C_i) \subset \mathrm{N}_C(C_s)$. This, in turn, yields $\mathrm{N}_{\mathrm{N}_C(C_i)}(D_{st}) \supset \mathrm{N}_{\mathrm{N}_C(C_s)}(D_{st})$. This implies that

$$\bigcap_{\{i|\ C_s \subset \operatorname{cl} C_i,\ D_{st} \subset \mathrm{N}_C(C_i)\}} \mathrm{N}_{\mathrm{N}_C(C_i)}(D_{st}) = \mathrm{N}_{\mathrm{N}_C(C_s)}(D_{st}). \tag{3.21}$$

Since the set $\mathrm{N}_C(C_s)$ is a cone, from Theorem 3.2.5 and [110, Example 11.4 (b)] we obtain

$$\mathrm{N}_{\mathrm{N}_C(C_s)}(D_{st}) = \mathrm{N}_{\mathrm{N}_C(x)}(y) = \left\{u \in (\mathrm{N}_C(C_s))^* \;\middle|\; u^\top y = 0\right\} = K(x,y), \tag{3.22}$$

which concludes the proof. □

Lemma 3.3.5. *For any $x \in C_s$ and $y \in D_{st}$ the following equality holds*

$$K(x,y)^* = \bigcap_{\{i|\ C_s \subset \operatorname{cl} C_i,\ D_{st} \subset \mathrm{N}_C(C_i)\}} \mathrm{N}_{\operatorname{cl} C_i}(C_s). \tag{3.23}$$

Proof. Recall first that due to [79, relation (42)] one has $\mathrm{T}_P(x_0) = \operatorname{co}\operatorname{cone}(P - x_0)$ for any polyhedral set P and any $x_0 \in P$. This, by virtue of Theorem 3.2.5 implies

$$\mathrm{T}_C(C_s) = \operatorname{co}\operatorname{cone}(C - \operatorname{cl} C_s). \tag{3.24}$$

Similarly, from the definition of normal cone and Theorem 3.2.5 one has

$$\mathrm{N}_{\operatorname{cl} C_i}(C_s) = \{y |\, y^\top(\operatorname{cl} C_i - \operatorname{cl} C_s) \le 0\} = (\operatorname{co}\operatorname{cone}(\operatorname{cl} C_i - \operatorname{cl} C_s))^*.$$

Since the equality of two sets implies equality of their polars, to prove the desired equality (3.23) it is enough to show that

$$K(x,y) = \bigcup_{\{i|\ C_s \subset \operatorname{cl} C_i,\ D_{st} \subset \mathrm{N}_C(C_i)\}} \operatorname{co}\operatorname{cone}(\operatorname{cl} C_i - \operatorname{cl} C_s).$$

Suppose that $u \in \operatorname{co}\operatorname{cone}(\operatorname{cl} C_i - \operatorname{cl} C_s)$ for some i such that $C_s \subset \operatorname{cl} C_i,\ D_{st} \subset \mathrm{N}_C(C_i)$. To show that $u \in K(x,y)$ we need to prove that $u \in \mathrm{T}_C(C_s)$ and that $y^\top u = 0$. The first relation follows immediately from (3.24) and the second one from the following chain of implications

$$\begin{array}{rcl} y \in D_{st} \subset \mathrm{N}_C(C_s) & \Longrightarrow & y^\top(C - \operatorname{cl} C_s) \le 0 \\ y \in D_{st} \subset \mathrm{N}_C(C_i) & \Longrightarrow & y^\top(C - \operatorname{cl} C_i) \le 0 \end{array} \quad \Longrightarrow\ y^\top(\operatorname{cl} C_s - \operatorname{cl} C_i) \le 0.$$

To show the opposite inclusion, we obtain first from [79, Lemma 4] and [79, relation (44)] that there exists an index i such that $C_s \subset \operatorname{cl} C_i$ and such that

$$K(x,y) = \mathrm{T}_{\operatorname{cl} C_i}(C_s) = \operatorname{co}\operatorname{cone}(\operatorname{cl} C_i - \operatorname{cl} C_s). \tag{3.25}$$

To finish the proof, it remains to show that $D_{st} \subset \mathrm{N}_C(C_i)$. From (3.25) we immediately obtain $y^\top(\operatorname{cl} C_i - \operatorname{cl} C_s) = 0$. Due to Theorem 3.2.5 $K(x,y)$ does not depend on the particular choice of $y \in D_{st}$ and thus we obtain $D_{st}^\top(\operatorname{cl} C_i - \operatorname{cl} C_s) = 0$. As already stated above, $D_{st} \subset \mathrm{N}_C(C_s)$ implies $D_{st}^\top(C - \operatorname{cl} C_s) \le 0$. Together, this shows that $D_{st}^\top(C - \operatorname{cl} C_i) \le 0$, which in turn implies $D_{st} \subset \mathrm{N}_C(C_i)$. This concludes the proof. □

Summarizing this special case, the relatively open faces of polyhedral sets appear to be a suitable choice for normally admissible stratifications. In such a case one can enjoy special properties of faces of polyhedral sets and relations to tangent an critical cones.

In the following subsection, we revisit another previously developed representation of normal cones for the general case considered in Section 3.2.

3.3.2 Relation to a union of polyhedral sets

In [54], the authors studied the case of a union of general polyhedral sets. Apart from providing explicit formulas for values of limiting normal cone at a point, the authors in [54] also focused on several special cases of polyhedral sets, such as finite union of halfspaces and finite union of orthants. In this subsection, we briefly summarize their main result concerning the case of a union of R polyhedral sets, for details see [54, Section 6].

Consider Γ as in (3.2). For $x \in \Gamma$ denote the set of active components by

$$\mathbb{I}(x) = \{r \in \{1, \dots, R\} | \, x \in \Omega_r\}.$$

Fix any $\bar{x} \in \Gamma$ and let us denote by Δ_r the polyhedral cones $\Delta_r := \mathrm{T}_{\Omega_r}(\bar{x})$. Then for $\Delta := \bigcup_{r \in \mathbb{I}(\bar{x})} \Delta_r$ one has

$$\mathrm{N}_\Gamma(\bar{x}) = \mathrm{N}_\Delta(0).$$

Now, for all $r \in \mathbb{I}(x)$, consider the explicit description of the polyhedral cones Δ_r

$$\Delta_r = \{x \,|\, \langle c_t^r, x\rangle \le 0, \ t = 1, \dots, T(r)\} \, .$$

Note that we will work with tangent and normal cones to Δ_r at 0 and that all constraints are active at this point. For $\mathbb{I} \subset \mathbb{I}(\bar{x})$ define the following index set

$$\mathcal{J}_{\mathbb{I}} = \begin{cases} \times_{r \in \mathbb{I}} \{1, \dots, T(r)\} & \text{if } \mathbb{I} \neq \emptyset, \\ \{\emptyset\} & \text{if } \mathbb{I} = \emptyset, \end{cases}$$

which adopts the convention that $\mathcal{J}_\emptyset$ contains one element, an empty (zero-dimensional) vector.

For any integer vectors $\mathbb{I}^c = (i_{n_1}, \dots, i_{n_L})$ and $J = (J_{n_1}, \dots, J_{n_L}) \in \mathcal{J}_{\mathbb{I}^c}$ put

$$\Gamma_{\mathbb{I}}^J = \left\{ x \;\middle|\; \begin{array}{l} \langle c_t^r, x\rangle \le 0, \ t = 1, \dots, T(r), \ r \in \mathbb{I} \\ \left\langle c_{J_r}^r, x \right\rangle > 0, \ r \in \mathbb{I}^c \end{array} \right\} .$$

Then

$$\mathrm{N}_\Gamma(\bar{x}) = \bigcup_{\emptyset \neq \mathbb{I} \subset \mathbb{I}(\bar{x})} \bigcup_{J \in \mathcal{J}_{\mathbb{I}^c}} \bigcup_{x \in \Gamma_{\mathbb{I}}^J} \bigcap_{k \in \mathbb{I}} \hat{\mathrm{N}}_{\Delta_k}(x), \tag{3.26}$$

and for each $x \in \Gamma_{\mathbb{I}}^J$ and $r \in \mathbb{I}$ there exist exactly one subsets $\mathcal{J}_{x,r} \subset \{1, \dots, T(r)\}$ such that

$$\begin{aligned} &\langle c_t^r, x\rangle = 0 \ \forall t \in \mathcal{J}_{x,r}, r \in \mathbb{I}, \\ &\langle c_t^r, x\rangle < 0 \ \forall t \in \{1, \dots, T(r)\} \setminus \mathcal{J}_{x,r}, r \in \mathbb{I}, \\ &\langle c_{J_r}^r, x\rangle > 0 \ \forall r \in \mathbb{I}^c. \end{aligned} \tag{3.27}$$

For such x and fixed k we have $\hat{\mathrm{N}}_{\Delta_k}(x) = \operatorname{co}\operatorname{cone}\{c_t^k | \, t \in \mathcal{J}_{x,k}\}$. For any subset $\mathcal{J} = \times_{r\in\mathbb{I}} \mathcal{J}_r \subset \mathcal{J}_{\mathbb{I}}$, put

$$\begin{aligned} R_{\mathbb{I}}^{\mathcal{J},J} &:= \operatorname{co}\operatorname{cone}\{\{c_{J_r}^r | \, r \in \mathbb{I}^c\} \cup \{-c_t^r | \, r \in \mathbb{I}, t \in \{1,\dots,n_r\} \setminus \mathcal{J}_r\}\}, \\ S_{\mathbb{I}}^{\mathcal{J}} &:= \operatorname{span}\{c_t^r | \, r \in \mathbb{I}, t \in \mathcal{J}_r\}, \end{aligned}$$

and

$$\mathcal{A}_{\mathbb{I}}^J := \{\mathcal{J} \subset \mathcal{J}_{\mathbb{I}} | \, R_{\mathbb{I}}^{\mathcal{J},J} \cap S_{\mathbb{I}}^{\mathcal{J}} = \{0\}\}. \tag{3.28}$$

Applying Motzkin's Theorem, solvability of systems of conditions (3.27) can be represented by elements of $\mathcal{A}_{\mathbb{I}}^J$.

Proposition 3.3.6. *Under the notation above, the limiting normal cone to a finite union of polyhedral sets calculates as*

$$\mathrm{N}_\Gamma(\bar{x}) = \bigcup_{\emptyset\neq\mathbb{I}\subset\mathbb{I}(\bar{x})} \bigcup_{J\in\mathcal{J}_{\mathbb{I}^c}} \bigcup_{\mathcal{J}\in\mathcal{A}_{\mathbb{I}}^J} \bigcap_{k\in\mathbb{I}} \operatorname{co}\operatorname{cone}\{c_j^k | \, j \in \mathcal{J}_k\}. \tag{3.29}$$

We will now compare the results of Proposition 3.3.6 to our results in Theorem 3.2.7. From direct comparison of sets defined by conditions (3.27) with sets $\Omega_{I,J}$ defined in (3.5), it follows that elements of $\mathcal{A}_{\mathbb{I}}^J$, which represent only the nonempty sets given by conditions (3.27), correspond to relatively open sets that form one particular normally admissible stratification of Γ. In fact, this is exactly the partition constructed in the proof of Lemma 3.2.3. Thus, it is not difficult to see that $\bigcap_{k\in\mathbb{I}} \operatorname{co}\operatorname{cone}\{c_j^k | j \in \mathcal{J}_k\}$ in (3.29) corresponds to $\bigcap_{i\in\tilde{I}(s)} \hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(\Gamma_s)$ in (3.12) via (3.8). Similarly $\bigcup_{\emptyset\neq\mathbb{I}\subset\mathbb{I}(\bar{x})} \bigcup_{J\in\mathcal{J}_{\mathbb{I}^c}} \bigcup_{\mathcal{J}\in\mathcal{A}_{\mathbb{I}}^J}$ in (3.29) corresponds to $\bigcup_{i\in I(\bar{s})}$ in (3.12).

Taking into account that there might exist other normally admissible stratifications of Γ with less components, we have managed to generalize the approach from [54] by considering a larger family of possible partitions instead of the particular one considered in [54]. On top of that, we are able to provide the corresponding result for the whole graph of N_Γ.

By means of the following example we show the differences in both approaches. These differences will become even clearer in Section 3.4 where we present an example in which a suitable choice of a normally admissible stratification plays a crucial role.

Example 3.3.7. Consider $\Gamma \subset \mathbb{R}^2$ to be a union of R different rays emanating from a common point $\bar{x} \in \mathbb{R}^2$. One can easily find a normally admissible stratification of Γ which consists of $R+1$ sets. For such a normally admissible stratification, the application of Corollary 3.2.8 is straightforward and the number of elements in union (3.12) grows linearly in R. On the other hand, it is clear that direct application of Proposition 3.3.6 results in exponential growth of the number of elements in union (3.29).

3.4 Application to time–dependent problems

In this section we will investigate a special structure of set Γ, which may arise during a discretization of time–dependent problems [1, 19]. To give a short intro-

duction, consider the following differential inclusion with given initial condition

$$\begin{aligned} &\dot{x}(t) \in \Lambda(t, x(t)),\ t \in [0,T] \text{ a.e.} \\ &x(0) = x_0, \end{aligned} \tag{3.30}$$

where $[0,T]$ is time interval, $x : [0,T] \to \mathbb{R}^n$ is the state variable, $\Lambda : [0,T] \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a multifunction and $x_0 \in \mathbb{R}^n$ is an initial point.

After performing a discretization of (3.30), we may obtain the following set of discretized feasible solutions to problem (3.30)

$$\Gamma := \left\{ x \in \mathbb{R}^{Kn} \,\middle|\, x_k \in \Lambda^k(x_{k-1}),\ k = 1, \dots, K \right\}. \tag{3.31}$$

Here, we consider $x = (x_1, \dots, x_K) \in \mathbb{R}^{Kn}$ to be the discretization of the state variable $x(\cdot)$ and for notational simplicity, we identify the initial point x_0 from (3.30) with x_0 from (3.31). Moreover, $n \in \mathbb{N}$ is the dimension of the state variable x_k and $K \in \mathbb{N}$ denotes the number of time discretization steps. Finally, $\Lambda^k : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ for $k = 1, \dots, K$ are multifunctions.

The main goal of this section is to use particular structure of Γ defined by (3.31) and simplify the formula for $\operatorname{gph} \mathrm{N}_\Gamma$ from Theorem 3.2.7. To be able to do so, we will need the following assumption

$$\Lambda^k \text{ is a polyhedral multifunction for } k = 1, \dots, K, \tag{3.32}$$

where a polyhedral multifunction is a multifunction which graph is a finite union of polyhedral sets. We recall that there is a unique correspondence between multifunctions $S : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ and sets $A \subset \mathbb{R}^{p+q}$ via graph operator

$$A = \operatorname{gph} S := \{ (x, y) \in \mathbb{R}^p \times \mathbb{R}^q \,|\, y \in S(x) \}.$$

Moreover, in this section, we will often work with a closure of multifunction $S : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$, which is denoted by $\operatorname{cl} S : \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ and defined via its graph by $\operatorname{gph} \operatorname{cl} S = \operatorname{cl} \operatorname{gph} S$.

3.4.1 Theoretical background

In this subsection, we will provide a theoretical background for computation of $\operatorname{gph} \mathrm{N}_\Gamma$ where Γ is given by (3.31). In particular, we will express normally admissible stratification of Γ in terms of normally admissible stratifications of $\operatorname{gph} \Lambda^k$ and based on these partitions, we will provide a formula for computation of a normal cone to Γ based on normal cones to elements of partitions of $\operatorname{gph} \Lambda^k$.

Observe that under assumption (3.32), application of Lemma 3.2.3 yields a normally admissible stratification $\{A_i^k \subset \mathbb{R}^{2n} | \, i = 1, \dots, M(k)\}$ of $\operatorname{gph} \Lambda^k$ for all $k = 1, \dots, K$. Due to unique correspondence between multifunctions and their graphs, this is equivalent to existence of multifunctions $\Lambda_i^k : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ with $\operatorname{gph} \Lambda_i^k = A_i^k$ such that $\{\operatorname{gph} \Lambda_i^k | \, i = 1, \dots, M(k)\}$ is a normally admissible stratification of $\operatorname{gph} \Lambda^k$. Further, for $s \in \{1, \dots, M(k)\}$ we denote by $I^k(s) \subset \{1, \dots, M(k)\}$ and $\tilde{I}^k(s) \subset I^k(s)$ index sets (3.4) associated with this stratification.

Now, we consider the following sets

$$\Gamma_i := \Gamma_{i_1 \dots i_K} := \left\{ x \in \mathbb{R}^{Kn} \,\middle|\, x_k \in \Lambda_{i_k}^k(x_{k-1}),\ k = 1, \dots, K \right\} \tag{3.33}$$

for $i := (i_1, \ldots, i_K)$ with $i_k \in \{1, \ldots, M(k)\}$. Defining

$$\Theta := \left\{ i \in \bigtimes_{k=1}^{K} \{1, \ldots, M(k)\} \,\middle|\, \Gamma_i \neq \emptyset \right\} \tag{3.34}$$

we show that $\{\Gamma_i | \, i \in \Theta\}$ forms a normally admissible stratification of Γ. To this end we develop a series of lemmas which allow us to express properties of Γ in terms of properties of Λ^k.

Lemma 3.4.1. *For $i \in \Theta$ we have*

$$\operatorname{cl} \Gamma_i = \left\{ x \in \mathbb{R}^{Kn} \,\middle|\, x_k \in (\operatorname{cl} \Lambda^k_{i_k})(x_{k-1}),\ k = 1, \ldots, K \right\}. \tag{3.35}$$

Proof. Denote the right–hand side of (3.35) by G. Directly from the definition of closure of a multifunction we have $\operatorname{cl} \Gamma_i \subset G$. To prove the opposite inclusion, consider some $x \in G$. Since $i \in \Theta$, there exists some $y \in \Gamma_i$, which means that $y_0 = x_0$ and $(y_{k-1}, y_k) \in \operatorname{gph} \Lambda^k_{i_k}$ for $k = 1, \ldots, K$. Since $\operatorname{gph} \Lambda^k_{i_k}$ is convex and relatively open due to definition of normally admissible stratification, by virtue of Lemma A.1.2 we obtain for $c \in (0, 1)$ and $k = 1, \ldots, K$ the following formula

$$(cy_{k-1} + (1-c)x_{k-1}, cy_k + (1-c)x_k) \in \operatorname{gph} \Lambda^k_{i_k}.$$

Defining $z^c_k := cy_k + (1-c)x_k$ and $z^c := (z^c_1, \ldots, z^c_K)$ we have $z^c \in \Gamma_i$ and with $c \to 0$ we have $z^c \to x$, which finishes the proof. □

Lemma 3.4.2. *For $s \in \Theta$ and index sets $I(s)$ and $\tilde{I}(s)$ defined by (3.4), it holds that*

$$I(s) = \left\{ i \in \Theta \,\middle|\, i_k \in I^k(s_k),\ k = 1, \ldots, K \right\}, \tag{3.36a}$$

$$\tilde{I}(s) = \left\{ i \in I(s) \,\middle|\, \forall j \in I(s) : j_k \in I^k(i_k),\ k = 1, \ldots, K \implies i = j \right\}, \tag{3.36b}$$

where index sets $I^k(s_k)$ are associated to a normally admissible stratifications of $\operatorname{gph} \Lambda^k$ for $k = 1, \ldots, K$. Moreover, for any $i \in I(s)$ condition (3.3) holds true.

Proof. First, take any $i \in I(s)$. From the definition of $I(s)$ this is equivalent to $\Gamma_s \cap \operatorname{cl} \Gamma_i \neq \emptyset$, which implies $i \in \Theta$. For contradiction assume that there is some k such that $i_k \notin I^k(s_k)$. This means that $\operatorname{gph} \Lambda^k_{s_k} \cap \operatorname{gph} \operatorname{cl} \Lambda^k_{i_k} = \emptyset$. Using Lemma 3.4.1 this further implies that $\Gamma_s \cap \operatorname{cl} \Gamma_i = \emptyset$, which concludes the contradiction.

Now, take any $i \in \Theta$ such that $i_k \in I^k(s_k)$ for all $k = 1, \ldots, K$. Due to definition of $I^k(s)$ this implies $\operatorname{gph} \Lambda^k_{s_k} \cap \operatorname{gph} \operatorname{cl} \Lambda^k_{i_k} \neq \emptyset$ for all k. By condition (3.3) for stratification of $\operatorname{gph} \Lambda^k$ this implies $\operatorname{gph} \Lambda^k_{s_k} \subset \operatorname{gph} \operatorname{cl} \Lambda^k_{i_k}$ for all k. Invoking Lemma 3.4.1, we have $\Gamma_s \subset \operatorname{cl} \Gamma_i$. Firstly, this implies that $\Gamma_s \cap \operatorname{cl} \Gamma_i = \Gamma_s \neq \emptyset$ proving (3.36a), and secondly it also means that property (3.3) holds true as well.

Formula (3.36b) then follows directly from (3.36a) and (3.4d). □

Lemma 3.4.3. $\{\Gamma_i | \, i \in \Theta\}$ *forms a normally admissible stratification of* Γ.

Proof. Observe first that due to definition of Θ we have $\Gamma = \cup_{i \in \Theta} \Gamma_i$ and that all Γ_i are nonempty. Since $\{\operatorname{gph} \Lambda^k_j | \, j \in \{1, \ldots, M(k)\}\}$ is a normally admissible stratification of $\operatorname{gph} \Lambda^k$, it follows that Γ_i are pairwise disjoint. Hence we have shown that $\{\Gamma_i | \, i \in \Theta\}$ is indeed a partition of Γ.

To prove that this partition is a normally admissible stratification of Γ, it remains to show that Γ_i are relatively open and convex, $\operatorname{cl}\Gamma_i$ are polyhedral and that property (3.3) holds. Since Γ_i can be written as an intersection of relatively open convex sets, it is relatively open and convex as well. Similarly, as $\operatorname{cl}\Gamma_i$ is an intersection of polyhedral sets due to Lemma 3.4.1, it is polyhedral. Finally, condition (3.3) follows directly from Lemma 3.4.2 and so the proof has been finished. □

The following theorem proposes a convenient formula for computation of $\hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(\Gamma_s)$. This formula is presented purely in terms of individual Λ^k and not the original Γ.

Theorem 3.4.4. *Assume that Γ is defined via* (3.31) *and that assumption* (3.32) *is satisfied. Assume moreover that $\{\operatorname{gph}\Lambda_i^k | \, i = 1, \dots, M(k)\}$ forms a normally admissible stratification of $\operatorname{gph}\Lambda^k$ for all $k = 1, \dots, K$. Then for any $s \in \Theta$ and $i \in I(s)$ we have*

$$\hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(\Gamma_s) = \left\{ \begin{pmatrix} p_1 + q_1 \\ \vdots \\ p_K + q_K \end{pmatrix} \in \mathbb{R}^{Kn} \;\middle|\; \begin{array}{c} \begin{pmatrix} p_{k-1} \\ q_k \end{pmatrix} \in \hat{\mathrm{N}}_{\operatorname{cl}\operatorname{gph}\Lambda^k_{i_k}}(\operatorname{gph}\Lambda^k_{s_k}), \;\; k = 1, \dots, K \\ p_K = 0 \end{array} \right\}.$$

Proof. The set $\operatorname{cl}\Gamma_i$ can be by virtue of Lemma 3.4.1 written as multivalued inverse $F^{-1}(\Omega_i)$, where

$$F(x) := \left(\begin{array}{c} x_0 \\ x_1 \\ \hline x_1 \\ x_2 \\ \hline \vdots \\ \hline x_{K-1} \\ x_K \end{array}\right), \;\; \Omega_i := \left(\begin{array}{c} \operatorname{cl}\operatorname{gph}\Lambda^1_{i_1} \\ \hline \operatorname{cl}\operatorname{gph}\Lambda^2_{i_2} \\ \hline \vdots \\ \hline \operatorname{cl}\operatorname{gph}\Lambda^K_{i_K} \end{array}\right).$$

Now, consider some $\bar{x} \in \Gamma_s \subset \operatorname{cl}\Gamma_i$ and define $\bar{x}_0 = x_0$. Since F is affine linear function and Ω_i is a polyhedral set, multifunction $S_i(p) := \{x | \, p + F(x) \in \Omega_i\}$ is calm at $(0, \bar{x})$. Then [60, Proposition 3.4] implies that

$$\mathrm{N}_{\operatorname{cl}\Gamma_i}(\bar{x}) \subset (\nabla F(\bar{x}))^\top \mathrm{N}_{\Omega_i}(F(\bar{x})).$$

But since Ω_i is convex, it is regular, and thus [110, Theorem 6.14] implies that

$$\hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(\bar{x}) = (\nabla F(\bar{x}))^\top \hat{\mathrm{N}}_{\Omega_i}(F(\bar{x})), \tag{3.37}$$

Plugging in the original data, we observe that $x^* \in \hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}(\bar{x})$ if and only if for every $k = 1, \dots, K$ there exist some multipliers $p_{k-1}, q_k \in \mathbb{R}^K$ with

$$\begin{pmatrix} p_{k-1} \\ q_k \end{pmatrix} \in \hat{\mathrm{N}}_{\operatorname{cl}\operatorname{gph}\Lambda^k_{i_k}}(\bar{x}_{k-1}, \bar{x}_k), \;\; k = 1, \dots, K,$$

such that equations $x_k^* = p_k + q_k$ hold for $k = 1, \dots, K$ with $p_K := 0$. But this is equivalent to the stated result by virtue of Lemma 3.4.2, Lemma 3.4.3 and Theorem 3.2.5. □

The previous result may be used directly to calculate $\operatorname{gph}\hat{\mathrm{N}}_\Gamma$ and $\operatorname{gph}\mathrm{N}_\Gamma$, and $\hat{\mathrm{N}}_\Gamma(\bar{x})$ for $\bar{x}\in\Gamma$, using Theorem 3.2.7 and Corollary 3.2.8, respectively. We note that $I(s)$ can be computed in a convenient way due to Lemma 3.4.2.

Remark 3.4.5. Even though we were able to express $I(s)$ in terms of $I^k(s_k)$ in Lemma 3.4.2 and similarly $\hat{\mathrm{N}}_{\operatorname{cl}\Gamma_i}$ in terms of $\hat{\mathrm{N}}_{\operatorname{cl}\operatorname{gph}\Lambda^k_{i_k}}$ in Theorem 3.4.4, we are convinced that it is not possible to derive a similar formula for $\hat{\mathrm{N}}_\Gamma$. In this remark we show that the following intuitive formula

$$\hat{\mathrm{N}}_\Gamma(\Gamma_s)=\left\{\begin{pmatrix}p_1+q_1\\ \vdots\\ p_K+q_K\end{pmatrix}\in\mathbb{R}^{Kn}\;\middle|\;\begin{matrix}\begin{pmatrix}p_{k-1}\\ q_k\end{pmatrix}\in\hat{\mathrm{N}}_{\operatorname{gph}\Lambda^k}(\operatorname{gph}\Lambda^k_{s_k}),\ k=1,\dots,K\\ p_K=0\end{matrix}\right\},\tag{3.38}$$

does not hold true. This is closely connected with violation of the so–called intersection property [45, Definition 9] for (3.37), which says that

$$\bigcap_{i\in I(s)}(\nabla F(\bar{x}))^\top\hat{\mathrm{N}}_{\Omega_i}(F(\bar{x}))=(\nabla F(\bar{x}))^\top\bigcap_{i\in I(s)}\hat{\mathrm{N}}_{\Omega_i}(F(\bar{x})).$$

Indeed, consider the following example with $n=2$, $K=2$,

$$\operatorname{gph}\Lambda^1=[\mathbb{R}\times\mathbb{R}\times\{0\}\times\mathbb{R}_{--}]\cup[\mathbb{R}\times\mathbb{R}\times\{0\}\times\{0\}]\cup\{(a,b,c,d)\in\mathbb{R}^4|\,c\in\mathbb{R}_{--},d=-c\},$$
$$\operatorname{gph}\Lambda^2=[\mathbb{R}_{--}\times\{0\}\times\mathbb{R}\times\mathbb{R}]\cup[\{0\}\times\{0\}\times\mathbb{R}\times\mathbb{R}]\cup\left\{(a,b,c,d)\in\mathbb{R}^4\,|\,a\in\mathbb{R}_{++},b=-a\right\}$$

and initial point $x_0=(0,0)$. Then one observes that $\Gamma=\{0\}\times\{0\}\times\mathbb{R}\times\mathbb{R}$ and thus for any $\bar{x}\in\Gamma$ we have $\mathrm{N}_\Gamma(\bar{x})=\mathbb{R}\times\mathbb{R}\times\{0\}\times\{0\}$. On the other hand, the right–hand side of formula (3.38) results in $\mathbb{R}_+\times\mathbb{R}_+\times\{0\}\times\{0\}$ and thus (3.38) does not hold true.

3.4.2 Example

Consider set

$$\Gamma:=\left\{(y,z)\in\mathbb{R}^K\times\mathbb{R}^K\;\middle|\;z_k\in\mathrm{N}_{[0,y_{k-1}]}(y_k),\ k=1,\dots,K\right\}\tag{3.39}$$

with $y_0=1$. Such set arises in delamination modeling [114] where variable $y_k\in[0,1]$ signifies the delamination level of an adhesive. Specifically, $y_k=1$ corresponds to a situation where the adhesive is not damaged while $y_k=0$ corresponds to a complete delamination. Due to the definition of normal cone, we see that (3.39) contains a hidden constraint $0\le y_k\le y_{k-1}$, meaning that a glue cannot heal back to its original state y_0. When considering optimal control or parameter identification in such model, it is advantageous to compute $\operatorname{gph}\mathrm{N}_\Gamma$, see Chapter 6. Such parameter identification in a delamination model will be also thoroughly investigated in Chapter 6. For simplicity, we will show the result only for the case of $y_k\in\mathbb{R}$ and $z_k\in\mathbb{R}$. However, the generalization to a more–dimensional space is straightforward and can be conducted in componentwise way.

We are not able to use the standard results of variational analysis to compute $\mathrm{N}_\Gamma(\bar{y},\bar{z})$. Since the set $[0,y_{k-1}]$ depends on y, we would have to introduce first additional variables. For example, it is possible to rewrite

$$z_k\in\mathrm{N}_{[0,y_{k-1}]}(y_k)$$

into the following system

$$\begin{aligned} z_k &= z_k^+ + z_k^-, \\ z_k^+ &\in \mathrm{N}_{(-\infty,0]}(y_k - y_{k-1}), \\ z_k^- &\in \mathrm{N}_{[0,\infty)}(y_k). \end{aligned}$$

However, Mangasarian–Fromovitz constraint qualification is not satisfied for this case if $\bar{y}_{k-1} = \bar{y}_k = 0$, and thus results such [110, Theorem 6.14] or [90] cannot be used. Considering this reformulation, it would be possible to use calculus rules with calmness constraint qualification [60] leading only to an inclusion instead of equality.

For these reasons, we will compute $\operatorname{gph} \mathrm{N}_\Gamma$ with Γ defined in (3.39) using Theorem 3.2.7 and Theorem 3.4.4. We consider $x = (y, z)$ and rewrite $z_k \in \mathrm{N}_{[0,y_{k-1}]}(y_k)$ equivalently as $(y_k, z_k) \in \Lambda^k(y_{k-1}, z_{k-1}) = \bigcup_{j=1}^8 \Lambda_j^k(y_{k-1}, z_{k-1})$ with initial condition $(y_0, z_0) = (1, 0)$ and Λ_i^k, $i = 1, \dots, 8$, being defined via respective graphs as follows

$$\begin{aligned} \operatorname{gph} \Lambda_1^k &= \left\{ (\tilde{y}, \tilde{z}, y, z) \in \mathbb{R}^4 \,|\, \tilde{y} \in \mathbb{R}_{++}, \tilde{z} \in \mathbb{R}, y = \tilde{y}, z \in \mathbb{R}_{++} \right\}, \\ \operatorname{gph} \Lambda_2^k &= \left\{ (\tilde{y}, \tilde{z}, y, 0) \in \mathbb{R}^4 \,|\, \tilde{y} \in \mathbb{R}_{++}, \tilde{z} \in \mathbb{R}, y = \tilde{y} \right\}, \\ \operatorname{gph} \Lambda_3^k &= \left\{ (\tilde{y}, \tilde{z}, y, 0) \in \mathbb{R}^4 \,|\, \tilde{y} \in \mathbb{R}_{++}, \tilde{z} \in \mathbb{R}, y \in (0, \tilde{y}) \right\}, \\ \operatorname{gph} \Lambda_4^k &= \mathbb{R}_{++} \times \mathbb{R} \times \{0\} \times \{0\}, \\ \operatorname{gph} \Lambda_5^k &= \mathbb{R}_{++} \times \mathbb{R} \times \{0\} \times \mathbb{R}_{--}, \\ \operatorname{gph} \Lambda_6^k &= \{0\} \times \mathbb{R} \times \{0\} \times \mathbb{R}_{--}, \\ \operatorname{gph} \Lambda_7^k &= \{0\} \times \mathbb{R} \times \{0\} \times \{0\}, \\ \operatorname{gph} \Lambda_8^k &= \{0\} \times \mathbb{R} \times \{0\} \times \mathbb{R}_{++}. \end{aligned} \tag{3.40}$$

Set $\operatorname{gph} \Lambda^k$ is visualized in Figure 3.3 where we omit the coordinate $\tilde{z}$ because there are no constraints on it. Its partition $\operatorname{gph} \Lambda_i^k$ for $i = 1, \dots, 8$ is depicted in Figure 3.4. It is not difficult to show that $\{\operatorname{gph} \Lambda_j^k | \, j = 1, \dots, 8\}$ forms a normally admissible stratification of $\operatorname{gph} \Lambda^k$ for all $k = 1, \dots, K$.

Next, directly from (3.4) by the help of Figure 3.4, we obtain for all $k = 1, \dots, K$

$$\begin{aligned} I^k(1) &= \{1\} & I^k(5) &= \{5\}, \\ , I^k(2) &= \{1, 2, 3\}, & I^k(6) &= \{5, 6\}, \\ I^k(3) &= \{3\}, & I^k(7) &= \{1, \dots, 8\}, \\ I^k(4) &= \{3, 4, 5\}, & I^k(8) &= \{1, 8\}. \end{aligned} \tag{3.41}$$

To construct normally admissible stratification of Γ, we need to characterize Θ given by (3.34).

Lemma 3.4.6. *Setting $i_0 = 1$, it holds that*

$$\Theta = \left\{ (i_1, \dots, i_K) \in \{1, \dots, 8\}^K \;\middle|\; \begin{array}{rcl} i_{k-1} \in \{1, 2, 3\} & \Longrightarrow & i_k \in \{1, 2, 3, 4, 5\} \\ i_{k-1} \in \{4, 5, 6, 7, 8\} & \Longrightarrow & i_k \in \{6, 7, 8\} \end{array} \right\}. \tag{3.42}$$

Proof. Denote the right–hand side of (3.42) by A. If $i \in \Theta$, then there exists some $(y, z) \in \Gamma_i$. If $i_{k-1} \in \{1, 2, 3\}$, then we have $y_{k-1} > 0$, which immediately implies

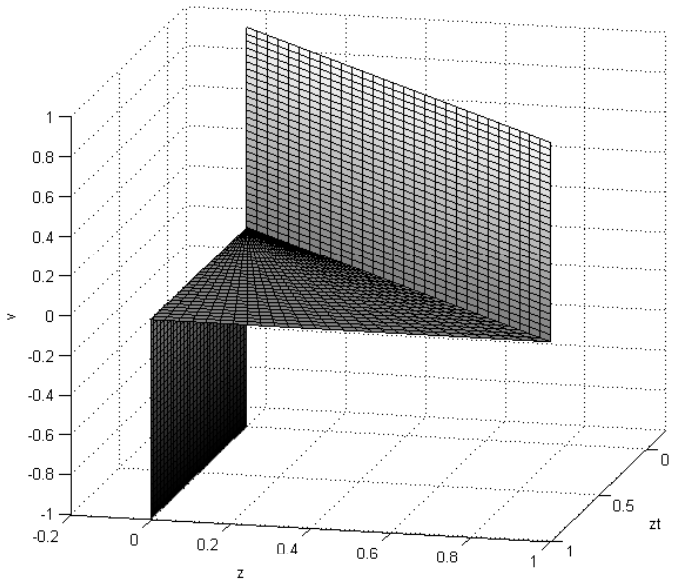


Figure 3.3: Visualization of $\operatorname{gph} \Lambda^k$.

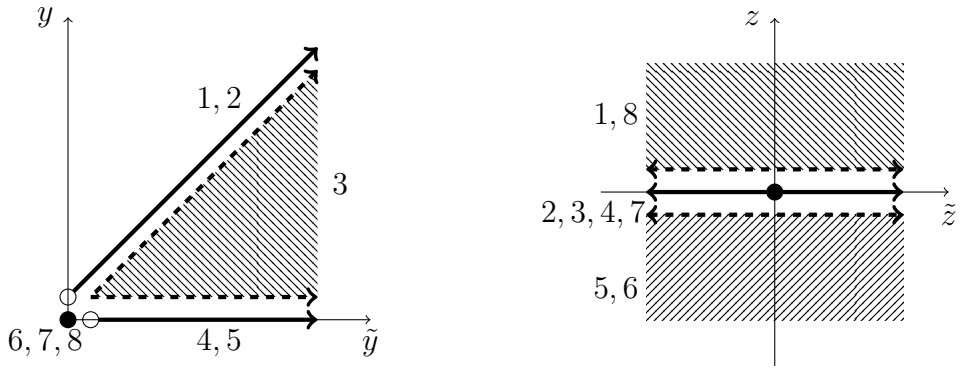


Figure 3.4: Visualization of $\operatorname{gph} \Lambda_i^k$

$i_k \in \{1,2,3,4,5\}$. If $i_{k-1} \in \{4,5,6,7,8\}$, then $y_k = 0$ and thus $i_k \in \{6,7,8\}$. Hence $\Theta \subset A$.

To finish the proof, consider now any $i \in A$ and define y coordinatewise as follows

$$y_k = \begin{cases} y_{k-1} & \text{if } i_k \in \{1,2\}, \\ \frac{1}{2} y_{k-1} & \text{if } i_k = 3, \\ 0 & \text{if } i_k \in \{4,5,6,7,8\}, \end{cases}$$

with $y_0 = 1$. Then it is not difficult to find z such that $(y,z) \in \Gamma_i$, and thus $i \in \Theta$, which completes the proof. □

Based on formula (3.41) we define

$$I(s) = \left\{ i \in \Theta \;\middle|\; \begin{array}{llll} s_k = 1 \implies i_k = 1, & & s_k = 5 \implies i_k = 5 \\ s_k = 2 \implies i_k \in \{1,2,3\}, & & s_k = 6 \implies i_k \in \{5,6\} \\ s_k = 3 \implies i_k = 3, & & s_k = 7 \implies i_k \in \{1,\ldots,8\} \\ s_k = 4 \implies i_k \in \{3,4,5\}, & & s_k = 8 \implies i_k \in \{1,8\} \end{array} \right\},$$

where we assume that $i = (i_1, \dots, i_K), s = (s_1, \dots, s_K) \in \{1, \dots, 8\}^K$ and all relations are required to hold for all $k = 1, \dots, K$.

Now we provide a summary of this part in the following theorem. Note that we have managed to simplify the computation of $\mathrm{N}_\Gamma(\bar{y}, \bar{z})$ into multiple computation of $\mathrm{N}_{\mathrm{cl}\,\mathrm{gph}\,\Lambda^k_{i_k}}(\bar{y}_k, \bar{z}_k)$. While set $\Gamma \subset \mathbb{R}^{2N}$ is nonconvex and depends on time, set $\mathrm{cl}\,\mathrm{gph}\,\Lambda^k_{i_k} \subset \mathbb{R}^4$ is convex and its dimension is independent of time.

Theorem 3.4.7. *Consider Γ as defined in (3.39), fix any $(\bar{y}, \bar{z}) \in \Gamma$ and denote by $\bar{s}$ the index such that $(\bar{y}_{k-1}, \bar{z}_{k-1}, \bar{y}_k, \bar{z}_k) \in \mathrm{gph}\,\Lambda^k_{\bar{s}_k}$ for all $k = 1, \dots, K$, where we set $\bar{y}_0 = 0$ and $\bar{z}_0 = 1$. Then*

$$\mathrm{N}_\Gamma(\bar{y}, \bar{z}) = \bigcup_{s \in I(\bar{s})} \bigcap_{i \in I(s)} \left\{ \begin{pmatrix} \mu_1 + \tilde{\mu}_1 \\ \vdots \\ \mu_K + \tilde{\mu}_K \\ \nu_1 \\ \cdots \\ \nu_K \end{pmatrix} \in \mathbb{R}^{2K} \;\middle|\; \begin{pmatrix} \tilde{\mu}_{k-1} \\ 0 \\ \mu_k \\ \nu_k \end{pmatrix} \in \hat{\mathrm{N}}_{\mathrm{cl}\,\mathrm{gph}\,\Lambda^k_{i_k}}(\mathrm{gph}\,\Lambda^k_{s_k}) \right\}, \quad \tilde{\mu}_K = 0 \tag{3.43}$$

where the inclusion is required to hold for all $k = 1, \dots, K$.

We will conclude this chapter by several instructive applications of Theorem 3.4.7 and compute $\mathrm{N}_\Gamma(\bar{y}, \bar{z})$ for two given points $(\bar{y}, \bar{z})$. The first one is rather simple and will be computed thoroughly, while for the second one we show only the first stage of the computation.

Example 3.4.8. Consider Γ defined in (3.39) with $K = 5$, $\bar{y} = (1, 0.5, 0, 0, 0)$ and $\bar{z} = (1, 0, 0, 1, -1)$. First, we realize that $\bar{s} = (1, 3, 4, 8, 6)$, where $\bar{s} \in \Theta$ is the unique index such that $(\bar{y}, \bar{z}) \in \Gamma_{\bar{s}}$. Employing (3.41), we realize that

$$I^1(\bar{s}_1) = \{1\},\ I^2(\bar{s}_2) = \{3\},\ I^3(\bar{s}_3) = \{3, 4, 5\},\ I^4(\bar{s}_4) = \{1, 8\} \text{ and } I^5(\bar{s}_5) = \{5, 6\}.$$

Then, denoting $i = (1, 3, 3, 1, 5)$, $j = (1, 3, 4, 8, 6)$ and $l = (1, 3, 5, 8, 6)$, Lemma 3.4.2 together with formula (3.42) yields

$$\begin{aligned} I(\bar{s}) &= \{i, j, l\}, \\ I(i) &= \tilde{I}(i) = \{i\}, \\ I(j) &= \{i, j, l\},\ \ \tilde{I}(j) = \{i, l\}, \\ I(l) &= \tilde{I}(l) = \{l\}. \end{aligned}$$

Thus, invoking formula (3.12) we have

$$\mathrm{N}_\Gamma(\bar{y}, \bar{z}) = \hat{\mathrm{N}}_{\mathrm{cl}\,\Gamma_i}(\Gamma_i) \cup \left[\hat{\mathrm{N}}_{\mathrm{cl}\,\Gamma_i}(\Gamma_j) \cap \hat{\mathrm{N}}_{\mathrm{cl}\,\Gamma_l}(\Gamma_j)\right] \cup \hat{\mathrm{N}}_{\mathrm{cl}\,\Gamma_l}(\Gamma_l).$$

Each of the regular normal cones in this formula can be computed as the term in

brackets in (3.43) with the use of the following regular normal cones, $k = 1, \dots, 5$,

$$
\begin{aligned}
&\hat{\mathrm{N}}_{\mathrm{cl\,gph}\,\Lambda_1^k}(\mathrm{gph}\,\Lambda_1^k) = \left\{(\tilde{\alpha}, 0, \alpha, 0) \in \mathbb{R}^4 \,|\, \tilde{\alpha} \in \mathbb{R}, \alpha = -\tilde{\alpha}\right\},\\
&\hat{\mathrm{N}}_{\mathrm{cl\,gph}\,\Lambda_3^k}(\mathrm{gph}\,\Lambda_3^k) = \{0\} \times \{0\} \times \{0\} \times \mathbb{R},\\
&\hat{\mathrm{N}}_{\mathrm{cl\,gph}\,\Lambda_3^k}(\mathrm{gph}\,\Lambda_4^k) = \{0\} \times \{0\} \times \mathbb{R}_- \times \mathbb{R},\\
&\hat{\mathrm{N}}_{\mathrm{cl\,gph}\,\Lambda_5^k}(\mathrm{gph}\,\Lambda_4^k) = \{0\} \times \{0\} \times \mathbb{R} \times \mathbb{R}_+,\\
&\hat{\mathrm{N}}_{\mathrm{cl\,gph}\,\Lambda_5^k}(\mathrm{gph}\,\Lambda_5^k) = \{0\} \times \{0\} \times \mathbb{R} \times \{0\},\\
&\hat{\mathrm{N}}_{\mathrm{cl\,gph}\,\Lambda_5^k}(\mathrm{gph}\,\Lambda_6^k) = \mathbb{R}_- \times \{0\} \times \mathbb{R} \times \{0\},\\
&\hat{\mathrm{N}}_{\mathrm{cl\,gph}\,\Lambda_6^k}(\mathrm{gph}\,\Lambda_6^k) = \mathbb{R} \times \{0\} \times \mathbb{R} \times \{0\},\\
&\hat{\mathrm{N}}_{\mathrm{cl\,gph}\,\Lambda_1^k}(\mathrm{gph}\,\Lambda_8^k) = \left\{(\tilde{\alpha}, 0, \alpha, 0) \in \mathbb{R}^4 \,|\, \tilde{\alpha} \in \mathbb{R}, \alpha \le -\tilde{\alpha}\right\},\\
&\hat{\mathrm{N}}_{\mathrm{cl\,gph}\,\Lambda_8^k}(\mathrm{gph}\,\Lambda_8^k) = \mathbb{R} \times \{0\} \times \mathbb{R} \times \{0\}.
\end{aligned}
$$

This results in

$$
\begin{aligned}
\mathrm{N}_\Gamma(\bar{y}, \bar{z}) &= \cup_{t\in\mathbb{R}} \begin{bmatrix} \mathbb{R}\times\{0\} \\ \{0\}\times\mathbb{R} \\ \{t\}\times\mathbb{R} \\ \{-t\}\times\{0\} \\ \mathbb{R}\times\{0\} \end{bmatrix} \bigcup \left[\cup_{s\in\mathbb{R}} \begin{bmatrix} \mathbb{R}\times\{0\} \\ \{0\}\times\mathbb{R} \\ (-\infty, s]\times\mathbb{R} \\ (-\infty, -s]\times\{0\} \\ \mathbb{R}\times\{0\} \end{bmatrix} \bigcap \begin{bmatrix} \mathbb{R}\times\{0\} \\ \{0\}\times\mathbb{R} \\ \mathbb{R}\times\mathbb{R}_+ \\ \mathbb{R}\times\{0\} \\ \mathbb{R}\times\{0\} \end{bmatrix} \right] \bigcup \begin{bmatrix} \mathbb{R}\times\{0\} \\ \{0\}\times\mathbb{R} \\ \mathbb{R}\times\{0\} \\ \mathbb{R}\times\{0\} \\ \mathbb{R}\times\{0\} \end{bmatrix} \\
&= \cup_{t\in\mathbb{R}} \begin{bmatrix} \mathbb{R}\times\{0\} \\ \{0\}\times\mathbb{R} \\ \{t\}\times\mathbb{R} \\ \{-t\}\times\{0\} \\ \mathbb{R}\times\{0\} \end{bmatrix} \bigcup \cup_{s\in\mathbb{R}} \begin{bmatrix} \mathbb{R}\times\{0\} \\ \{0\}\times\mathbb{R} \\ (-\infty, s]\times\mathbb{R}_+ \\ (-\infty, -s]\times\{0\} \\ \mathbb{R}\times\{0\} \end{bmatrix} \bigcup \begin{bmatrix} \mathbb{R}\times\{0\} \\ \{0\}\times\mathbb{R} \\ \mathbb{R}\times\{0\} \\ \mathbb{R}\times\{0\} \\ \mathbb{R}\times\{0\} \end{bmatrix}. \qquad (3.44)
\end{aligned}
$$

Example 3.4.9. In the setting of Example 3.4.8 we consider $\bar{y} = (1, 0.5, 0, 0, 0)$ and $\bar{z} = (1, 0, 0, 0, 1)$. Then we have $\bar{s} = (1, 3, 4, 7, 8)$ and

$$
I(\bar{s}) = \left\{ i \in \{1, \dots, 8\}^K \;\middle|\; \begin{array}{l} i_1 = 1,\ i_2 = 3,\ i_3 \in \{3, 4, 5\} \\ i_3 = 3 \implies i_4 \in \{1, 2, 3, 4, 5\} \\ i_3 \in \{4, 5\} \implies i_4 \in \{4, 5, 6, 7, 8\} \\ i_4 \in \{1, 2, 3\} \implies i_5 = 1 \\ i_4 \in \{4, 5, 6, 7, 8\} \implies i_5 = 8 \end{array} \right\}.
$$

It is not difficult to verify that $I(\bar{s})$ contains 15 elements and hence we will have to consider a union with respect to 15 elements in formula (3.43). Then it would be necessary to compute $\tilde{I}(s)$ for every $s \in I(\bar{s})$ using Lemma 3.4.2, which would, however, in most cases amount to only one or two elements.

Finally, in the light of Example 3.4.8 and especially Example 3.4.9 we present another comparison of our approach with the theory developed in [54]; a comparison which was already slightly touched in Example 3.3.7.

Remark 3.4.10. Consider set Γ defined in (3.39) and let us show that even though our approach is not simple, it could be more applicable than the approach developed in [54]. There it is necessary to compute $\mathrm{T}_\Gamma(\bar{y}, \bar{z})$ first, which, due to our best knowledge, cannot be tackled by standard calculus rules because of the same reasons as described earlier in this subsection. Even though it is possible to derive formula for $\mathrm{T}_\Gamma(\bar{y}, \bar{z})$ directly from the definition, it is not a simple task.

Consider now the same point $(\bar{y}, \bar{z})$ as in Example 3.4.8. With the notation of Subsection 3.3.2 it is possible to show that $|\mathbb{I}(\bar{y}, \bar{z})| = 2$ with

$$\Delta_1 = \bigcup_{t \in \mathbb{R}_+} \begin{bmatrix} \{0\} \times \mathbb{R} \\ \mathbb{R} \times \{0\} \\ \{t\} \times \{0\} \\ \{t\} \times \mathbb{R} \\ \{0\} \times \mathbb{R} \end{bmatrix}, \quad \Delta_2 = \begin{bmatrix} \{0\} \times \mathbb{R} \\ \mathbb{R} \times \{0\} \\ \{0\} \times \mathbb{R}_- \\ \{0\} \times \mathbb{R} \\ \{0\} \times \mathbb{R} \end{bmatrix}.$$

Now, we show that a direct application of Proposition 3.3.6 can be rather cumbersome. It is clear that the first union in (3.29) will be performed with respect to three elements. Since each Δ_i can be described as an intersection of 11 halfspaces, the any fixed $\mathbb{I}$ for expressing the second and third union in (3.29), one has to check 121 combinations of sets $R_{\mathbb{I}}^{\mathcal{J},J}$ and $S_{\mathbb{I}}^{\mathcal{J}}$, leading together to necessity of solving 363 systems of linear equations (3.28). The number is so high because the majority of this systems will have some solution apart from 0 and thus the set $A_{\mathbb{I}}^{J}$ will contain lesser number of elements. Note that in Example 3.4.8 we need to compute only union of 3 elements. The situation would become more difficult, or possibly intractable should we consider $(\bar{y}, \bar{z})$ as in Example 3.4.9.

Another approach to compute the desired normal cone is to realize that Δ_1 and Δ_2 differ only at components y_3, z_3 and y_4, and so we obtain from [54, Proposition 3.1] that

$$\mathrm{N}_\Gamma(\bar{x}, \bar{y}) = \mathbb{R} \times \{0\} \times \{0\} \times \mathbb{R} \times \Omega \times \{0\} \times \mathbb{R} \times \{0\},$$

where

$$\begin{aligned} \Omega &= \mathrm{bd}\,\Theta_1^* \bigcup (\Theta_1^* \cap \Theta_2^*) \bigcup \mathrm{bd}\,\Theta_2^*, \\ \Theta_1 &= \{(y_3, 0, y_4) |\, y_3 \in \mathbb{R}_+,\ y_4 = y_3\}, \\ \Theta_2 &= \{0\} \times \mathbb{R}_- \times \{0\}. \end{aligned} \tag{3.45}$$

After computing the polars to Θ_1 and Θ_2, it becomes clear that the three elements of unions in (3.44) and (3.45) do correspond.

Finally, we would like to point out that regular points (such as $\bar{y}_k = \bar{y}_{k-1} > 0$ and $\bar{z}_k > 0$) fit well into our approach, while in [54] these points considerably increase the number of halfplanes defining Δ_i.

4. Sensitivity of parameterized differential inclusions

4.1 Introduction

We consider a general differential inclusion with known initial value

$$\begin{aligned} f(t,u,x(t),\dot{x}(t)) &\in \Omega(t,u,x(t),\dot{x}(t)), \quad t\in[0,T] \text{ a.e.} \\ x(0) &= a. \end{aligned} \tag{4.1}$$

This inclusion is parameterized by time independent control variable/parameter u and is to be solved for state variable x. Function f is single–valued and continuously differentiable in all but the time variable while multifunction Ω is only continuous in the time variable. The main goal is to investigate the stability properties of the so–called solution mapping $S : u \mapsto x$ which assigns an infinite–dimensional solution x of (4.1) to a finite–dimensional parameter u.

Even though it is simple to obtain local Lipschitzian continuity of $S : U \to W^{1,\infty}([0,T],\mathbb{R}^n)$ in case of an ordinary differential equation with sufficiently smooth data, which corresponds to $\Omega \equiv 0$ and f having a special form, to our best knowledge, there are not many results for differential inclusions with parameters entering the inclusion. However, for models with parameterized initial condition, numerous results exist, see [72, 77, 97, 126, 134].

Similar dependence was studied in a series of papers by N. S. Papageorgiou, see [64, 98, 99, 100], in which the following infinite–dimensional problem was considered

$$\begin{aligned} -\dot{x}(t) &\in \partial_x f(t,u,x(t)) + \Omega(t,u,x(t)), \quad t\in[0,T] \text{ a.e.} \\ x(0) &= a(u) \end{aligned} \tag{4.2}$$

with $f(t,u,\cdot)$ being a convex lower semicontinuous function. In this series the continuity of $S : U \rightrightarrows \mathcal{C}([0,T],H)$ was proved for a Hilbert space H. This result was obtained for both Vietoris and Hausdorff topologies on the power set of $\mathcal{C}([0,T],H)$.

In the context of rate–independent processes and hysteresis models, the following differential inclusion was studied

$$\begin{aligned} -\dot{x}(t) + \dot{u}(t) &\in \mathrm{N}_Z(x(t)), \quad t\in[0,T] \text{ a.e.} \\ x(0) &= a \end{aligned} \tag{4.3}$$

by P. Krejčí, see [68, 69, 70]. In this case, the global Lipschitzian continuity of $S : \mathcal{C}([0,T],H) \to \mathcal{C}([0,T],H)$ and $S : W^{1,1}([0,T],H) \to W^{1,1}([0,T],H)$ was shown.

In the case of (4.2), a general model was considered but only the state variable and not its derivative entered the estimates. On the other hand, model (4.3) provided opposite results, rather specific model was considered but the estimates were sharper. We try to combine the strengths of both papers. Thus, we consider both general models and are able to obtain local Lipschitzian continuity

of $S : U \rightrightarrows W^{1,2}([0,T],\mathbb{R}^n)$, which is significantly stronger result than the continuity of $S : U \rightrightarrows \mathcal{C}([0,T],H)$ for $H = \mathbb{R}^n$ obtained for (4.2). However, this approach introduces some deficiencies as well. We cannot handle time dependent perturbation and are able to work only with finite dimensional values $x(t)$.

Instead of considering a fixed model, we consider rather a general one, perform its discretization and derive necessary conditions for the local Lipschitzian continuity of the discretized solution map. If the corresponding Lipschitzian modulus exhibits uniform behaviour in some sense, we are able to deduce the Lipschitzian behavior of the original solution map as well. The Lipschitzian modulus must exhibit some boundedness feature both in a neighborhood of the investigated parameter and upon the decrease of the time step. Even though the conditions for this boundedness may seem to be difficult to verify, we propose a class of models, for which these results are verifiable.

Similar stability results can be found in [50, 51], in which a differential variational inequality, which is a differential inclusion of special form coupled with algebraic equation, is considered. Data in such models are approximated and if the approximated solutions converge in some sense, then this limit is a solution to the original problem.

Concerning the possible applicability of the obtained results, differential inclusion are nowadays an established field of research, see monographs [12, 27, 126]. We believe that our results can be used in postoptimal analysis of time–dependent models where some parameters are not known exactly or in Mathematical Programs with Evolutionary Equilibrium Constraints (MPEECs) where inclusion (4.1) is part of the constraint system. We also derive an estimate for the Lipschitzian modulus, providing not only a qualitative but also quantitative estimate. We have in mind one particular application, namely nonregular electrical circuits with ideal diodes [7] where the parameter u plays the role of parameters of various components in the circuit and the state variable x is the charge in the circuit.

The chapter is organized as follows. It Section 4.2 a discretization of (4.1) is considered. First an upper estimate of a generalized derivative of the discretized solution map S^K is found. This estimate is stated via adjoint equations. Then we show that having some bound on the adjoint variables results immediately in the local Lipschitzian continuity of S^K and if this bound is uniform in a certain sense, then we can deduce the local Lipschitzian continuity for the original solution map S as well.

Since it may not be immediately clear how to use these results, in Section 4.3 we present two examples of their possible applications. In the first one, we apply these results to an ordinary differential equation and in the second one to a model arising in modeling of electrical circuits with ideal diodes [7]. We are fully aware that it is possible to obtain stronger results by simpler means for the first case, however, we decided to present this example because the remaining example uses very similar ideas as the first one, only its implementation is much more technically difficult.

4.2 Stability result

In this section we will consider problem (4.1) and derive conditions under which the solution map $S : u \mapsto x$ is locally Lipschitz. To ease the notational burden,

instead of problem (4.1) we will also consider the following problem

$$\begin{gathered} g(t,u,x(t),\dot{x}(t)) \in \Lambda(t), \quad t \in [0,T] \text{ a.e.} \\ x(0) = a, \end{gathered} \tag{4.4}$$

in which $u \in \mathbb{R}^d$ plays the role of a parameter or a control variable and the systems are to be solved for almost any time instant for $x(t) \in \mathbb{R}^n$ for which the initial value is given. Concerning the data, $g : [0,T] \times \mathbb{R}^d \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ is a single–valued function which is continuously differentiable in all but the time variable and $\Lambda(t) \subset \mathbb{R}^m$ is a closed set.

Even though it seems that (4.1) is more general than (4.4), it is not entirely true. Having problem (4.1), we may state it in the form (4.4) by setting

$$\begin{gathered} g(t,u,x(t),\dot{x}(t)) := \begin{pmatrix} u \\ x(t) \\ \dot{x}(t) \\ f(t,u,x(t),\dot{x}(t)) \end{pmatrix} \\ \Lambda(t) := \operatorname{gph} \Omega(t,\cdot,\cdot,\cdot). \end{gathered} \tag{4.5}$$

By doing so, we add artificial functions, which means that the used constraint qualifications will become more difficult to verify. As an example, consider the implicit function theorem or its numerous generalizations from [38], in which one of the constraint qualifications states that ∇g has to have full row rank. This reformulation will be considered in examples in Section 4.3.

Consider now time discretization $0 = t_0^K < \cdots < t_K^K = T$ and together with the infinite–dimensional problem (4.4) also its finite–dimensional discretized counterpart

$$\begin{gathered} g_{k+1}^K(u^K, x_k^K, x_{k+1}^K) \in \Lambda_{k+1}^K, \ k = 0, \ldots, K-1 \\ x_0^K = a \end{gathered} \tag{4.6}$$

in which $g_k : \mathbb{R}^d \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ is a continuously differentiable function and Λ_k is a closed set. For simplicity, the upper index K denoting discretization level will be often omitted. The exact discretization scheme is not specified but we require that for computation of x_{k+1} only the previous value x_k may be used.

As all the problems depend on a parameter u, we are interested in analysis of the so–called solution map, also known as the control–to–state mapping, which assigns a solution x of a system to a parameter u. This mapping will be denoted by S for continuous case (4.4) and S^K for discretized case (4.6). In this section we first derive an upper estimate for coderivative D^*S^K which is used to present conditions for verification of the Aubin property of S^K, a property which coincides with local Lipschitzian property under single–valuedness of S^K. Finally, we show that under single–valuedness of S^K and S and some boundedness of the Lipschitzian moduli, the Lipschitzian property can be transferred from S^K to S. These results will be used later in Section 4.3 where the local Lipschitzian continuity of

$$S : \mathbb{R}^d \to W^{1,2}([0,T], \mathbb{R}^n)$$

is shown.

Before stating the first lemma, we remind that $u \in \mathbb{R}^d$ and $x := (x_1, \ldots, x_K) \in \mathbb{R}^{Kn}$. The initial value x_0 is omitted because $x_0 = a$ is not a subject to change.

Lemma 4.2.1. *Consider problem* (4.6) *and fix any* $\bar{x} \in S^K(\bar{u})$. *Assume that for all* $k = 1, \dots, K$, g_k *is continuously differentiable around* $(\bar{u}, \bar{x}_k, \bar{x}_{k+1})$ *and* Λ_k *is closed. Assume further that the following constraint qualification holds: if* (4.7)–(4.10) *are satisfied with* $x^* = 0$, *then* $u^* = 0$.

Then for $D^*S^K : \mathbb{R}^{Kn} \to \mathbb{R}^d$ *and for any element*

$$u^* \in D^*S^K(\bar{u}, \bar{x})(x^*)$$

with $x^* = (x_1^*, \dots, x_K^*)$ *there exist for* $k = 0, \dots, K-1$ *multipliers*

$$p_{k+1} \in \mathrm{N}_{\Lambda_{k+1}}(g_{k+1}(\bar{u}, \bar{x}_k, \bar{x}_{k+1})) \tag{4.7}$$

such that

$$u^* = \sum_{k=1}^{K} (\nabla_u g_k(\bar{u}, \bar{x}_{k-1}, \bar{x}_k))^\top p_k. \tag{4.8}$$

Moreover, for $k = 1, \dots, K-1$ *the adjoint equations*

$$-x_k^* = (\nabla_v g_k(\bar{u}, \bar{x}_{k-1}, \bar{x}_k))^\top p_k + (\nabla_x g_{k+1}(\bar{u}, \bar{x}_k, \bar{x}_{k+1}))^\top p_{k+1} \tag{4.9}$$

and the terminal condition

$$-x_K^* = (\nabla_v g_K(\bar{u}, \bar{x}_{K-1}, \bar{x}_K))^\top p_K \tag{4.10}$$

are satisfied.

Proof. It is simple to see that

$$\operatorname{gph} S^K = \left\{ (u,x) \middle| \begin{array}{l} g_1(u, x_0, x_1) \in \Lambda_1 \\ \dots \\ g_K(u, x_{K-1}, x_K) \in \Lambda_K \end{array} \right\} = \{(u,x) |\ G(u,x) \in \Sigma\}$$

where we have defined $\Sigma := \Lambda_1 \times \dots \Lambda_K$ and

$$G(u,x) := \begin{pmatrix} g_1(u, x_0, x_1) \\ \dots \\ g_K(u, x_{K-1}, x_K) \end{pmatrix}.$$

Using the multi–valued inverse, we can write $\operatorname{gph} S^K = G^{-1}(\Sigma)$, which due to the assumed constraint qualification by virtue of [110, Theorem 6.14] implies

$$\mathrm{N}_{\operatorname{gph} S^K}(\bar{u}, \bar{x}) \subset \nabla G(\bar{u}, \bar{x})^\top \mathrm{N}_\Sigma(G(\bar{u}, \bar{x})). \tag{4.11}$$

Using the definition of coderivative and (4.11), we obtain

$$\begin{aligned} D^*S^K(\bar{u}, \bar{x})(x^*) &= \left\{ u^* \middle| \begin{pmatrix} u^* \\ -x^* \end{pmatrix} \in \mathrm{N}_{\operatorname{gph} S^K}(\bar{u}, \bar{x}) \right\} \\ &\subset \left\{ u^* \middle| \begin{pmatrix} u^* \\ -x^* \end{pmatrix} \in \nabla G(\bar{u}, \bar{x})^\top \mathrm{N}_\Sigma(G(\bar{u}, \bar{x}) \right\}. \end{aligned} \tag{4.12}$$

The Jacobian of G has the following form

$$\nabla G(\bar{u}, \bar{x}) = \begin{pmatrix} \nabla_u g_1 & \nabla_v g_1 & & & \\ \nabla_u g_2 & \nabla_x g_2 & \nabla_v g_2 & & \\ \vdots & & \ddots & \ddots & \\ \nabla_u g_K & & & \nabla_x g_K & \nabla_v g_K \end{pmatrix} \tag{4.13}$$

From [110, Proposition 6.41] we know that

$$\mathrm{N}_\Sigma(G(\bar{u},\bar{x})) = \bigtimes_{k=0}^{K-1} \mathrm{N}_{\Lambda_{k+1}}(g_{k+1}(\bar{u},\bar{x}_k,\bar{x}_{k+1})). \tag{4.14}$$

Finally by plugging (4.13) and (4.14) into (4.12) we obtain statement of the lemma. □

In the next few lines we will briefly comment on Lemma 4.2.1. As we have already said, the main goal is to perform the sensitivity analysis of solution map S^K. From Mordukhovich criterion [110, Theorem 9.40] we immediately see that S^K has the Aubin property if $\nabla_v g_{k+1}(\bar{u},\bar{x}_k,\bar{x}_{k+1})$ has full row rank for all $k = 0,\dots,K-1$. Unfortunately, as in our desired application the sets Λ_k will mostly depend on state or control variables and will have to be rewritten via their graphs as shown in (4.5), additional artificial functions will be added and hence, this full row rank property cannot be expected and another approach has to be used.

It is certainly possible to relax the used constraint qualification. According to [60, Proposition 3.3] the constraint qualification from Lemma 4.2.1 corresponds to the Aubin property of multifunction

$$M(q) = \{(u,x)|\ g_{k+1}(u,x_k,x_{k+1}) + q_{k+1} \in \Lambda_{k+1},\ k = 0,\dots,K-1\}$$

around point $(0,\bar{u},\bar{x})$. According to [60, Proposition 3.2], the same result would hold true if we assumed only calmness of M at the same point. However, since the main goal of this chapter lies in the next theorem whose assumptions imply the Aubin property of M around the reference point, and thus the constraint qualification of Lemma 4.2.1 is satisfied, we keep the current state.

Having Lemma 4.2.1 at hand, the condition for fulfillment of the Aubin property of S^K, and thus of the local Lipschitzian property if S^K is single–valued, is well–known. To be able to deduce some some results for S, we need some uniformity boundedness of the Lipschitzian moduli both over time and on some neighborhood of $\bar{u}$. This condition is stated in (4.15).

Together with $x^K = (x_1^K,\dots,x_K^K) \in \mathbb{R}^{Kn}$, we will also consider its piecewise linear and piecewise constant extensions $x^K(\cdot)$. Both will satisfy $x^K(0) = a$ and $x^K(t_k) = x_k^K$ for all $k = 1,\dots,K$. The piecewise linear extension will be obtained by connecting these points while the piecewise constant extension will satisfy $x^K(t) = x^K(t_{k-1}^K)$ whenever $t \in [t_{k-1}^K, t_k^K)$ for all $k = 1,\dots,K$.

Theorem 4.2.2. *If in the setting of Lemma 4.2.1 there exists a constant $L(K,\bar{u})$ such that*

$$|u^{*K}| \le L(K,\bar{u})|x^{*K}|,$$

then S^K has the Aubin property around $(\bar{u},\bar{x})$ with modulus not larger than $L(K,\bar{u})$.

Assume further that there exists a neighborhood V of $\bar{u}$ such that multifunctions S^K and S are single–valued on V. For all $u \in V$ find $L(K,u)$ as above, define

$$M(K,V) := \sup_{u\in V} L(K,u)$$

to be the upper bound for the Lipschitzian modulus of S^K on V and assume that

$$M(V) := \liminf_{K\to\infty} \frac{1}{\sqrt{K}} M(K,V) < \infty. \tag{4.15}$$

Finally, assume that for every $u \in V$ we have $x^K(\cdot) \rightharpoonup x$ in $L^2([0,T],\mathbb{R}^n)$, where $x = S(u)$, $x^K = S^K(u)$ and $x^K(\cdot)$ is the piecewise constant or piecewise linear extension of $S^K(u)$.

Then $S : \mathbb{R}^d \to L^2([0,T],\mathbb{R}^n)$ is locally Lipschitz on V with modulus less or equal to $\sqrt{T}M(V)$.

Proof. The first statement follows immediately from [110, Theorem 9.40].

For the second part recall the already several times mentioned fact that the Aubin property coincides with the locally Lipschitzian property for single–valued mappings. Using the same theorem as in the first part, we obtain that S^K is locally Lipschitzian on V with modulus at most $M(K,V)$. Fix arbitrary parameters $u, \tilde{u} \in V$ and denote the corresponding state variables by x^K and $\tilde{x}^K$. Even though the Lipschitzian modulus is defined as infimum of all constants satisfying (2.4), having uniform bound for this modulus on V, we can deduce that

$$\frac{1}{\sqrt{K}} |\tilde{x}^K - x^K|_{l^2} \le \frac{1}{\sqrt{K}} M(K,V) |\tilde{u} - u|. \tag{4.16}$$

For simplicity denote $z^K := \tilde{x}^K - x^K$ and consider first its piecewise constant extension denoted by $z^K(\cdot)$. For $k = 1,\dots,K$ and $j = 1,\dots,n$ we denote by $z^K_{k,j}$ the j–th component of z^K_k and similarly by $z^K_j(t)$ we understand the j–th component of its piecewise constant extension. Further recall that, as mentioned in the introduction, we consider the Euclidean norm on product spaces and hence

$$|z^K(\cdot)|_{L^2} = \sqrt{\sum_{j=1}^n |z^K_j(\cdot)|^2_{L^2}} = \sqrt{\sum_{j=1}^n \int_0^\top (z^K_j(t))^2 dt}.$$

For the left–hand side of (4.16) due to $T = Kh$ we obtain

$$\frac{1}{\sqrt{K}} |z^K|_{l^2} = \sqrt{\frac{1}{Kh} \sum_{j=1}^n \sum_{k=1}^K h (z^K_{k,j})^2} = \sqrt{\frac{1}{T} \sum_{j=1}^n \int_0^\top (z^K_j(t))^2 dt} = \frac{1}{\sqrt{T}} |z^K(\cdot)|_{L^2}. \tag{4.17}$$

From the assumptions we know that for the solutions of the continuous problems $\tilde{x} := S(\tilde{u})$ and $x := S(u)$ we have $\tilde{x}^K - x^K \rightharpoonup \tilde{x} - x$ in $L^2([0,T],\mathbb{R}^n)$ and thus

$$\liminf_{K\to\infty} \frac{1}{\sqrt{K}} |\tilde{x}^K - x^K|_{l^2} = \liminf_{K\to\infty} \frac{1}{\sqrt{T}} |\tilde{x}^K(\cdot) - x^K(\cdot)|_{L^2} \ge \frac{1}{\sqrt{T}} |\tilde{x}(\cdot) - x(\cdot)|_{L^2} \tag{4.18}$$

and combination of (4.18), (4.16) and (4.15) yields

$$\begin{aligned} \frac{1}{\sqrt{T}} |\tilde{x}(\cdot) - x(\cdot)|_{L^2} &\le \liminf_{K\to\infty} \frac{1}{\sqrt{K}} |\tilde{x}^K - x^K|_{l^2} \\ &\le \liminf_{K\to\infty} \left[\frac{1}{\sqrt{K}} M(K,V) |\tilde{u} - u| \right] = M(V) |\tilde{u} - u| \end{aligned}$$

and the result for the case of piecewise constant extension has been proven.

If the extension is piecewise linear, the only thing which needs to be modified is (4.17). Again, we keep the same symbols for this extension as in the previous case. By simple computation it can be shown that

$$\begin{aligned}\frac{1}{\sqrt{T}}|z^K(\cdot)|_{L^2} &= \sqrt{\frac{h}{3T}\sum_{j=1}^{n}\sum_{k=1}^{K}(z_{k,j}^2 + z_{k,j}z_{k-1,j} + z_{k-1,j}^2)} \\ &\leq \sqrt{\frac{h}{3T}\sum_{j=1}^{n}\sum_{k=1}^{K}(3z_{k,j}^2 + 2z_{0,j}^2)} = \sqrt{\frac{1}{K}\sum_{j=1}^{n}\sum_{k=1}^{K} z_{k,j}^2} = \frac{1}{\sqrt{K}}|z^K|_{l^2}\end{aligned} \tag{4.19}$$

because $z_{0,j} = 0$. Finally, the theorem statement follows exactly from the same arguments as in the previous case. □

4.3 Applications

In this final section we present applications of Theorem 4.2.2. Even though the main strength of this theorem lies in admitting the multi–valued part, first we set this part to zero and consider an ordinary differential equation. We are fully aware that for this case it is possible to obtain stronger results by simpler means, however, we have decided to include this example to illustrate the basic estimates on which the next example is based. In the second part we consider a differential inclusion called the sweeping process motivated by electrical circuits with ideal diodes [7] and show the local Lipschitzian continuity of the solution map.

In all cases we will need the following two lemmas, the first one being a discrete version of Gronwall's lemma. The first one will be later used in Chapter 5 while the second one both in Chapters 5 and 6.

Lemma 4.3.1 (Gronwall discrete)**.** *Let the following inequality*

$$\frac{a_{j+1} - a_j}{h_j} \leq g_j + (1-\theta)\lambda_j a_j + \theta\lambda_{j+1}a_{j+1},\ j = 0, \ldots, k-1$$

hold true and assume that

$$1 - \theta\lambda_{j+1}h_j > 0,\ 1 + (1-\theta)\lambda_j h_j > 0,\ j = 0, \ldots, k-1.$$

Then for $j = 1, \ldots, k$ the following estimate holds true

$$a_j \leq a_0 \prod_{l=1}^{j} \frac{1 + (1-\theta)\lambda_{l-1}h_{l-1}}{1 - \theta\lambda_l h_{l-1}} + \sum_{n=0}^{j-1} \frac{h_n g_n}{1 + (1-\theta)\lambda_n h_n} \prod_{l=n+1}^{j} \frac{1 + (1-\theta)\lambda_{l-1}h_{l-1}}{1 - \theta\lambda_l h_{l-1}}.$$

Proof. Taken from [42, Proposition 3.3]. □

Corollary 4.3.2. *Suppose that for $j = 0, \ldots, k-1$ one has*

$$\frac{a_{j+1} - a_j}{h_j} \leq g_j + \lambda a_j$$

with $\lambda > 0$, $h_j > 0$ and $\sum_{j=0}^{k-1} h_j = T$. Then there exists constants C_1 and C_2 depending only on λ, a_0 and T such that for all $j = 1, \dots, k$ the following estimate holds true

$$a_j \leq C_1 + C_2 \sum_{l=0}^{j-1} h_l g_l.$$

Proof. The imposed assumptions correspond to Lemma 4.3.1 with $\theta = 0$ and $\lambda_j = \lambda$. Then the estimate takes the from

$$a_j \leq a_0 \prod_{l=1}^{j} (1 + \lambda h_{l-1}) + \sum_{n=0}^{j-1} \frac{h_n g_n}{1 + \lambda h_n} \prod_{l=n+1}^{j} (1 + \lambda h_{l-1}). \tag{4.20}$$

From the inequality of algebraic and geometric mean we obtain

$$\prod_{l=1}^{j} (1 + \lambda h_{j-1}) \leq \left(\frac{1}{j} \sum_{l=1}^{j} (1 + \lambda h_{j-1}) \right)^j \leq \left(1 + \frac{\lambda T}{j} \right)^j \leq e^{\lambda T}.$$

Plugging this into (4.20) we obtain

$$a_j \leq a_0 e^{\lambda T} + \sum_{n=0}^{j-1} \frac{h_n g_n}{1} e^{\lambda T},$$

which was to be proven. □

Lemma 4.3.3. *Let A be a positive definite matrix. Fix any r and find any p and q solving the following system*

$$\begin{aligned} p - Aq &= r \\ p^\top q &\leq 0. \end{aligned} \tag{4.21}$$

Denoting

$$d := \min_{\|x\|=1} x^\top A x,$$

then one has

$$\|q\| \leq \frac{1}{d} \|r\|.$$

Finally, for $p^\top q \leq 0$ it is sufficient that

$$\begin{pmatrix} p \\ q \end{pmatrix} \in \mathrm{N}_{\operatorname{gph} \mathrm{N}_\Gamma}(x, y)$$

for any $y \in \mathrm{N}_\Gamma(x)$ and convex Γ.

Proof. Constant d is positive because A is positive definite. For $q = 0$ the statement is obvious. In the opposite case, multiply equation (4.21) by q and perform simple algebraic operations to obtain

$$q^\top r = q^\top p - q^\top A q \leq -d \|q\|^2,$$

which implies

$$\|q\|^2 \leq -\frac{1}{d} q^\top r \leq \frac{1}{d} \|q\| \|r\|,$$

which in turn amounts to the first part of the lemma statement.

If Γ is convex, then by virtue of [107, Theorem 4] we know that $x \mapsto \mathrm{N}_\Gamma(x)$ is a maximal monotone operator, and thus [104, Theorem 2.1] implies $p^\top q \leq 0$. □

4.3.1 Ordinary differential equation

Consider an ordinary differential equation

$$\dot{y}(t) = f(t, u, y(t)) \tag{4.22}$$

with initial condition $y(0) = a$. As we have already said, this is only an illustrative example with not satisfactory results and thus we impose rather strong assumptions on f to keep this subsection as short as possible. Specifically, we will assume that f is bounded, continuous in the time variable and continuously differentiable in last two variables.

First, we need to discretize (4.22). Since we are not interested in convergence analysis, we keep the discretization as simple as possible and use the forward Euler method. It is possible to use

$$y_{k+1}^K - y_k^K - h^K f(t_k^K, u, y_k^K) = 0, \tag{4.23}$$

however, in this case, we would obtain only local Lipschitzian continuity of the solution map of (4.22) when considered as $S : \mathbb{R}^d \to L^2([0,T], \mathbb{R}^n)$. Thus, we introduce an artificial variable and perform the following discretization

$$y_{k+1}^K - y_k^K - h^K z_{k+1}^K = 0 \tag{4.24a}$$
$$z_{k+1}^K - f(t_k^K, u, y_k^K) = 0 \tag{4.24b}$$

with $y_0 = a$ fixed. As we will see later, in this case, we will be able to prove local Lipschitzian continuity of the solution map of (4.22) when considered as $S : \mathbb{R}^d \to W^{1,2}([0,T], \mathbb{R}^n)$.

For discretized problem (4.24) we consider slightly redefined solution map, specifically we will consider $S^K : u \mapsto (y^K, z^K)$, hence to the parameter not only the state variable but also its derivate is assigned. It is clear that S^K is single–valued. Define now $f_k^K(u, y_k^K) := f(t_k^K, u, y_k^K)$, fix any $\bar{u}$ and some its bounded neighborhood V. Then the following constants are finite

$$\begin{aligned} c_1 &:= \sup\{\nabla_y f_k(u, y_k)|\ u \in V, (y^K, z^K) \in S^K(u), K \in \mathbb{N}, k = 1, \dots, K\} \\ c_2 &:= \sup\{\nabla_u f_k(u, y_k)|\ u \in V, (y^K, z^K) \in S^K(u), K \in \mathbb{N}, k = 1, \dots, K\}. \end{aligned} \tag{4.25}$$

Fix now any $u \in V$, compute $(y^K, z^K) = S^K(u)$ and set $(y^K(\cdot), z^K(\cdot)) \in L^2([0,T], \mathbb{R}^n) \times L^2([0,T], \mathbb{R}^n)$ to be the extension of $(y^K, z^K) \in \mathbb{R}^{Kn} \times \mathbb{R}^{Kn}$ which is piecewise linear for $y^K(\cdot)$ and piecewise constant for $z^K(\cdot)$. From the assumptions on f it is simple to deduce that $y^K(\cdot)$ and $z^K(\cdot)$ are uniformly bounded in $L^\infty([0,T], \mathbb{R}^n)$ and thus we may extract a subsequence such that $y^K(\cdot) \rightharpoonup y(\cdot)$ and $z^K(\cdot) \rightharpoonup z(\cdot)$, both in $L^2([0,T], \mathbb{R}^n)$. Moreover, it can be shown that $\dot{y}(\cdot) = z(\cdot)$ and that $y(\cdot)$ is the unique solution to (4.22). In the next several lines, we will consider fixed discretization level K and thus we omit it.

Denote the multiplier corresponding to (4.24a) by p_k and to (4.24b) by q_k. Then Lemma 4.2.1, with the not yet verified constraint qualification, yields that if $u^* \in D^* S^K(u, y, z)(y^*, z^*)$, then

$$u^* = -\sum_{k=1}^{K} (\nabla_u f_k(u, y_k))^\top q_k \tag{4.26}$$

and the terminal condition

$$-y_K^* = p_K \tag{4.27a}$$
$$-z_K^* = -hp_K + q_K \tag{4.27b}$$

and for $k = 1, \ldots, K-1$ the adjoint equations

$$-y_k^* = p_k - p_{k+1} - (\nabla_y f_{k+1}(u, y_{k+1}))^\top q_{k+1} \tag{4.28a}$$
$$-z_k^* = -hp_k + q_k. \tag{4.28b}$$

are satisfied. Moreover, from these expressions, it is simple to see that the constraint qualification of Lemma 4.2.1 indeed holds.

Plugging (4.28b) into (4.28a) yields

$$\frac{p_k - p_{k+1}}{h} = (\nabla_y f_{k+1}(u, y_{k+1}))^\top p_{k+1} - \frac{1}{h}(y_k^* + (\nabla_y f_{k+1}(u, y_{k+1}))^\top z_{k+1}^*), \tag{4.29}$$

which due to Lemma 4.3.1 results in

$$\|p_k\|_1 \leq e^{c_1 T}(\|y_K^*\|_1 + \sum_{j=k}^{K-1}(\|y_j^*\|_1 + c_1\|z_{j+1}^*\|_1)) \leq e^{c_1 T}\sum_{j=1}^{K}(\|y_j^*\|_1 + c_1\|z_j^*\|_1).$$

By plugging this estimate back to (4.28b) we have

$$\|q_k\|_1 \leq \|z_k^*\|_1 + he^{c_1 T}\sum_{j=1}^{K}(\|y_j^*\|_1 + c_1\|z_j^*\|_1) \tag{4.30}$$

and noting that $y^{*K} \in \mathbb{R}^{Kn}$ stands for vector $(y_1^*, \ldots, y_K^*)$, we have

$$\|y^{*K}\|_1 = \sum_{k=1}^{K}\|y_k^*\|_1 \tag{4.31}$$

and thus from (4.26) due to (4.30) we have estimate

$$\begin{aligned}\|u^*\|_2 \leq \|u^*\|_1 &\leq c_2\sum_{k=1}^{K}\|q_k\|_1 \leq c_2\sum_{k=1}^{K}\left(\|z_k^*\|_1 + he^{c_1 T}\sum_{j=1}^{K}(\|y_j^*\|_1 + c_1\|z_j^*\|_1)\right)\\ &= c_2\left(\|z^{K*}\|_1 + Te^{c_1 T}(\|y^{K*}\|_1 + c_1\|z^{K*}\|_1)\right)\\ &\leq c_2\sqrt{Kn}\left(\|z^{K*}\|_2 + Te^{c_1 T}(\|y^{K*}\|_2 + c_1\|z^{K*}\|_2)\right).\end{aligned} \tag{4.32}$$

Since c_1 and c_2 are not dependent on the choice of $u \in V$, we may apply Theorem 4.3.4 to obtain that mapping $u \mapsto (y, \dot{y})$ solving (4.22) is locally Lipschitz continuous as $V \to L^2([0,T], \mathbb{R}^n) \times L^2([0,T], \mathbb{R}^n)$. This is equivalent to local Lipschitzian continuity of solution map $u \mapsto y$ to (4.22) when considered as $V \to W^{1,2}([0,T], \mathbb{R}^n)$. We again emphasize that this is only a toy example with simplified assumptions and not entirely satisfactory results.

4.3.2 Sweeping process

The purpose of the previous example was to give an insight of what needs to be done. In this second example we consider a proper differential inclusion with discontinuous multifunction and derive similar results as in the first example. The considered model is a version of the sweeping process from [7] which takes the following form

$$\begin{gathered} -A_1\dot{y}(t) - A_0 y(t) + f(t) \in \mathrm{N}_{C(t)}(\dot{y}(t)), \quad t \in [0,T] \text{ a.e.} \\ y(0) = a. \end{gathered} \tag{4.33}$$

Such systems arise in modeling electrical circuits with ideal diodes, matrices A_0 and A_1 accumulate information about components of the circuit and y stands for the charge in circuit. We add perturbation u to data and consider the following model

$$\begin{gathered} -A_1(u)\dot{y}(t) - A_0(u) y(t) + f(t,u) \in \mathrm{N}_{C(t)}(\dot{y}(t)), \quad t \in [0,T] \text{ a.e.} \\ y(0) = a. \end{gathered} \tag{4.34}$$

The goal is to perform sensitivity analysis of $S : u \mapsto x(\cdot)$ with $x(\cdot) := (y(\cdot), \dot{y}(\cdot))$ solving (4.34). Thus, we investigate how the change of various parameters of the model, such as resistances, influences the change of the charge in the circuit.

On the contrary to the original paper [7], instead of an arbitrary Hilbert space we restrict ourselves only to $y(t) \in \mathbb{R}^n$. Under certain assumptions, in [7, Theorem 5.6] the convergence of the discretized solutions is proved using a modified version of the catching up algorithm, in which one solves for $k = 0, \dots, K-1$ defines doubles $x_{k+1} := (y_{k+1}, z_{k+1})$ and solves iteratively the following system starting with $y_0 = a$

$$g_{k+1}(u, x_k, x_{k+1}) := \begin{pmatrix} z_{k+1} \\ -A_1(u) z_{k+1} - A_0(u) y_{k+1} + f_{k+1}(u) \\ y_{k+1} - y_k - h z_{k+1} \end{pmatrix} \in \begin{pmatrix} \mathrm{gph}\, \mathrm{N}_{C_{k+1}} \\ 0 \end{pmatrix} =: \Lambda_{k+1} \tag{4.35}$$

where $f_k^K(u) := f(t_k^K, u)$ and $C_k^K := C(t_k^K)$ are closed convex sets. It is shown in [7, Theorem 5.7] that S is single–valued under additional assumptions and similar result is shown for S^K in [7, Remark 2]. In the rest of this section we assume that this single–valuedness is satisfied.

From now on, we assume that the constraint qualification of Lemma 4.2.1 is satisfied. This constraint qualification will be verified later on. Fix any $\bar{u}$ and some its neighborhood V and assume further that for $i = 0, 1$ and $t \in [0,T]$ the mappings $u \mapsto A_i(u)$ and $u \mapsto f(t,u)$ are continuously differentiable for all $u \in V$.

First, fix any $u \in V$ and apply Lemma 4.2.1. It is well–known that $\mathrm{gph}\, \mathrm{N}_{C_k}$ is closed for any set C_k and hence, with the not–yet–verified constraint qualification, all the assumptions of this lemma are satisfied. For simplicity, for the corresponding multipliers from this lemma we omit their dependence on u, and hence write e. g. only p_k instead of $p_k(u)$. However, for various constants used as upper bounds, their dependence on u is emphasized.

Computing the partial derivatives of g_k, we obtain

$$\nabla_u g_k = \begin{pmatrix} 0 \\ B_k(u) \\ 0 \end{pmatrix}, \nabla_x g_k = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ -I & 0 \end{pmatrix}, \nabla_v g_k = \begin{pmatrix} 0 & I \\ -A_0(u) & -A_1(u) \\ I & -hI \end{pmatrix}$$

with

$$B_{k+1}(u) := -\nabla_u A_1(u) z_{k+1} - \nabla_u A_0(u) y_{k+1} + \nabla_u f_{k+1}(u).$$

Lemma 4.2.1 then states that if $u^* \in D^* S(u,x)(x^*)$, then this element can be expressed as

$$u^* = \sum_{k=1}^{K} B_k^\top(u) q_k \tag{4.36}$$

where there are multipliers (p_k, q_k, r_k) satisfying for $k = 1, \dots, K$

$$\begin{pmatrix} p_k \\ q_k \end{pmatrix} \in \mathrm{N}_{\mathrm{gph}\, \mathrm{N}_{C_k}}(z_k, -A_1(u) z_k - A_0(u) y_k + f_k(u)). \tag{4.37}$$

Assuming that matrices $A_0(u)$ and $A_1(u)$ are symmetric, then the adjoint equations (4.9) read for $k = 1, \dots, K-1$

$$r_k = r_{k+1} + A_0(u) q_k - y_k^* \tag{4.38a}$$
$$p_k - A_1(u) q_k = h r_k - z_k^* \tag{4.38b}$$

and the terminal condition (4.10) is equal to

$$r_K = A_0(u) q_K - y_K^* \tag{4.39a}$$
$$p_K - A_1(u) q_K = h r_K - z_K^*. \tag{4.39b}$$

Define now the following constants

$$c(u) := \frac{1}{\min_{\|x\|_1 = 1} x^\top A_1(u) x}$$
$$d(K,u) := \frac{1}{1 - h c(u) \|A_0(u)\|_\infty} \left(1 - \frac{T c(u) \|A(u)\|_\infty}{K}\right)^{-K}$$

If $A_1(u)$ is positive definite, we may employ Lemma 4.3.3 to (4.38b) and (4.39b) to obtain for $k = 1, \dots, K$

$$\|q_k\|_1 \le c(u) \|h r_k - z_k^*\|_1. \tag{4.40}$$

Then by plugging estimate (4.40) into (4.39a) and (4.38a) we obtain

$$\begin{aligned} \|r_K\|_1 &\le c(u) \|A_0(u)\|_\infty (h \|r_K\|_1 + \|z_K^*\|_1) + \|y_K^*\|_1 \\ \|r_K\|_1 &\le \frac{1}{1 - h c(u) \|A_0(u)\|_\infty} (c(u) \|A_0(u)\|_\infty \|z_K^*\|_1 + \|y_K^*\|_1) \end{aligned} \tag{4.41}$$

and for $k = 1, \dots, K-1$

$$\|r_k\|_1 = \|r_{k+1} + A_0(u) q_k - y_k^*\|_1 \le \|r_{k+1}\|_1 + c(u) \|A_0(u)\|_\infty (h \|r_k\|_1 + \|z_k^*\|_1) + \|y_k^*\|_1$$

or equivalently

$$\frac{\|r_k\|_1 - \|r_{k+1}\|_1}{h} \le c(u) \|A_0(u)\|_\infty \|r_k\|_1 + \frac{1}{h} \|y_k^*\|_1 + \frac{1}{h} c(u) \|A_0(u)\|_\infty \|z_k^*\|_1.$$

This enables us to use Lemma 4.3.1 with, together with (4.41) and (4.31), to obtain

$$\begin{aligned}\|r_k\|_1 &\le \left(1-\frac{Tc(u)\|A(u)\|_\infty}{K}\right)^{-K}\left(\|r_K\|_1+\sum_{l=k}^{K-1}[\|y_l^*\|_1+c(u)\|A_0(u)\|_\infty\|z_l^*\|_1]\right)\\ &\le d(K,u)\sum_{l=1}^{K}[\|y_l^*\|_1+c(u)\|A_0(u)\|_\infty\|z_l^*\|_1]\\ &= d(K,u)[\|y^{K*}\|_1+c(u)\|A_0(u)\|_\infty\|z^{K*}\|_1].\end{aligned} \tag{4.42}$$

Assume further that there exist constants ρ_y and ρ_z such that for all $u \in V$, for all K and for the corresponding $(y^K, z^K) = S^K(u)$ we have

$$|y_k^K| \le \rho_y,\ |z_k^K| \le \rho_z$$

for all K and $k = 1, \dots, K$. Due to the structure of the considered model, this assumption is satisfied when all sets $C(t)$ are uniformly bounded. Then if $\|\nabla_u f(\cdot, u)\|_\infty$ is bounded on $[0, T]$, then we get the following estimate

$$\|B_k(u)\|_\infty \le \|\nabla_u A_1(u)\|_\infty\rho_z + \|\nabla_u A_0(u)\|_\infty\rho_u + \sup_{t\in[0,T]}\|\nabla_u f(t,u)\|_\infty =: b(u). \tag{4.43}$$

Formulas (4.36), (4.43), (4.40) and (4.42) then imply

$$\begin{aligned}|u^*| \le \|u^*\|_1 &\le b(u)\sum_{k=1}^{K}\|q_k\|_1 \le b(u)c(u)\sum_{k=1}^{K}(h\|r_k\|_1+\|z_k^*\|_1)\\ &\le b(u)c(u)[Td_1(K,u)d_2(K,u)[\|y^{K*}\|_1+c(u)\|A_0(u)\|_\infty\|z^{K*}\|_1]+\|z^{K*}\|_1]\\ &\le b(u)c(u)\max\{Td(K,u), Td(K,u)c(u)\|A_0(u)\|_\infty+1\}\|(y^{K*},z^{K*})\|_1\\ &\le \sqrt{2Kn}b(u)c(u)\max\{Td(K,u), Td(K,u)c(u)\|A_0(u)\|_\infty+1\}|(y^{K*},z^{K*})|.\end{aligned}$$

Moreover, from this formula it is clear that the constraint qualification of Lemma 4.2.1 is satisfied.

Theorem 4.2.2 then implies that S^K has the local Lipschitzian property around u with modulus less or equal to $L(K, u)$ where

$$L(K,u) := \sqrt{2Kn}b(u)c(u)\max\{Td(K,u), Td(K,u)c(u)\|A_0(u)\|_\infty+1\}. \tag{4.44}$$

To use the second part of Theorem 4.2.2, we need to find $\sup_{u\in V} L(K, u)$. Assume furthermore that $u \mapsto \nabla_u f(t, u)$ is continuous on V uniformly in t and fix any $\varepsilon > 0$ such that $\varepsilon c(\bar{u}) < 1$. This, together with previously imposed assumptions, allows us to shrink the neighborhood V of $\bar{u}$ such that for all $u \in V$ and for all $t \in [0, T]$ we have

$$\begin{aligned}\|A_i(u)-A_i(\bar{u})\|_\infty &\le \varepsilon\\ \|\nabla_u A_i(u)-\nabla_u A_i(\bar{u})\|_\infty &\le \varepsilon\\ \|\nabla_u f(t,u)-\nabla_u f(t,\bar{u})\|_\infty &\le \varepsilon.\end{aligned} \tag{4.45}$$

Hence we obtain

$$\begin{aligned}b(u) &= \|\nabla_u A_1(u)\|_\infty\rho_z + \|\nabla_u A_0(u)\|_\infty\rho_u + \sup_{t\in[0,T]}\|\nabla_u f(t,u)\|_\infty\\ &\le \|\nabla_u A_1(\bar{u})\|_\infty\rho_z + \|\nabla_u A_0(\bar{u})\|_\infty\rho_u + \sup_{t\in[0,T]}\|\nabla_u f(t,\bar{u})\|_\infty + \varepsilon\rho_z + \varepsilon\rho_u + \varepsilon\end{aligned} \tag{4.46}$$

For an estimate for $c(u)$, fix first any $x \in \mathbb{R}^n$ with $\|x\|_1 = 1$. Then by (4.45) we have

$$|x^\top A_1(u)x - x^\top A_1(\bar{u})x| \leq \|x\|_1 \|A_1(u) - A_1(\bar{u})\|_\infty \|x\|_1 \leq \varepsilon,$$

which implies

$$c(u) = \frac{1}{\min_{\|x\|_1=1} x^\top A_1(u)x} \leq \frac{1}{\min_{\|x\|_1=1} x^\top A_1(\bar{u})x - \varepsilon} = \frac{c(\bar{u})}{1 - \varepsilon c(\bar{u})}. \tag{4.47}$$

Similarly, for K large enough, the following estimate, which is independent of the choice of $u \in V$, can be deduced

$$\begin{aligned} d(K,u) &\leq \left(\frac{1}{1 - h(c(\bar{u}) + \varepsilon)(\|A(\bar{u})\|_\infty + \varepsilon)}\right)\left(1 - \frac{Tc(\bar{u})(\|A_0(\bar{u})\|_\infty + \varepsilon)}{K(1 - \varepsilon c(\bar{u}))}\right)^{-K} \\ &\overset{K\to\infty}{\longrightarrow} \exp\left(\frac{Tc(\bar{u})(\|A_0(\bar{u})\|_\infty + \varepsilon)}{1 - \varepsilon c(\bar{u})}\right). \end{aligned} \tag{4.48}$$

By plugging (4.46), (4.47) and (4.48) into (4.44) we find upper estimate of $\sup_{u\in V} L(K,u)$, which in accordance with Theorem 4.2.2 will be denoted by $M(K,V)$ and similarly for

$$M(V) := \limsup_{K\to\infty} \frac{1}{\sqrt{K}} M(K,V)$$

we have

$$M(V) \leq \sqrt{2n}\hat{b}\hat{c}\max\{T\hat{d}, T\hat{d}\hat{c}\hat{A} + 1\}.$$

where

$$\begin{aligned} \hat{A} &:= \|A_0(\bar{u})\|_\infty + \varepsilon \\ \hat{b} &:= \|\nabla_u A_1(\bar{u})\|_\infty \rho_z + \|\nabla_u A_0(\bar{u})\|_\infty \rho_u + \sup_{t\in[0,T]} \|\nabla_u f(t,\bar{u})\|_\infty + \varepsilon\rho_u + \varepsilon\rho_z + \varepsilon \\ \hat{c} &:= \frac{c(\bar{u})}{1 - \varepsilon c(\bar{u})} \\ \hat{d} &:= \exp\left(\frac{Tc(\bar{u})(\|A_0(\bar{u})\|_\infty + \varepsilon)}{1 - \varepsilon c(\bar{u})}\right). \end{aligned}$$

If $y^K(\cdot) \rightharpoonup y(\cdot)$ and $z^K(\cdot) \rightharpoonup \dot{y}(\cdot)$ both in $L^2([0,T],\mathbb{R}^n)$ with $y(\cdot)$ being the unique solution of (4.34) corresponding to u, then considering that ε may be arbitrarily small, Theorem 4.2.2 tells us that S is locally Lipschitz around $\bar{u}$ with modulus no more than

$$\sqrt{2n}bc\max\{Te^{Tc\|A_0(\bar{u})\|_\infty}, Te^{Tc\|A_0(\bar{u})\|_\infty}c\|A_0(\bar{u})\|_\infty + 1\}. \tag{4.49}$$

where

$$\begin{aligned} b &:= \|\nabla_u A_1(\bar{u})\|_\infty \rho_z + \|\nabla_u A_0(\bar{u})\|_\infty \rho_u + \sup_{t\in[0,T]} \|\nabla_u f(t,\bar{u})\|_\infty \\ c &:= \frac{1}{\min_{\|x\|_1=1} x^\top A_1(\bar{u})x}. \end{aligned}$$

We summarize the result in the following theorem.

Theorem 4.3.4. *For $u \in \mathbb{R}^d$ define the set of solutions y of problem (4.34) by $S(u)$. Assume that for all $t \in [0,T]$ the sets $C(t)$ are closed and convex. Moreover, let $C(0)$ be a bounded set and $C(\cdot)$ have a continuous variation, which means that there exists a nondecreasing continuous function $\nu : [0,T] \to \mathbb{R}$ with $\nu(0) = 0$ such that*

$$|d(v, C(t)) - d(v, C(s)| \leq |\nu(t) - \nu(s)| \tag{4.50}$$

for all $v \in \mathbb{R}^n$ and $s, t \in [0,T]$.

Fix any $\bar{u}$ and assume that there exists its neighborhood V with the following properties

- *f is a continuous function on $[0,T] \times V$*
- *$f(t,\cdot)$ is differentiable on V for every $t \in [0,T]$ and $\nabla_u f(t,\cdot)$ is continuous at $\bar{u}$ uniformly in t by which we understand that for every $\varepsilon > 0$ there exists a neighborhood $\tilde{V}$ of $\bar{u}$ such that*

$$\sup_{t\in[0,T]} \sup_{u\in\tilde{V}} |\nabla_u f(t,u) - \nabla_u f(t,\bar{u})| \leq \varepsilon.$$

- *$A_i(u)$ is symmetric positive definite matrices such that $A_i(\cdot)$ is continuously differentiable at $\bar{u}$ for $i = 0, 1$ and $u \in V$.*

Then S^K and S are single–valued and there exist constants ρ_y and ρ_z such that for all $u \in V$ and all the corresponding $(y^K, z^K) = S^K(u)$ we have $|y_k^K| \leq \rho_y$, $|z_k^K| \leq \rho_z$ for all K and all $k = 1, \ldots, K$. Finally, $S : \mathbb{R}^d \to W^{1,2}([0,T], \mathbb{R}^n)$ is locally Lipschitz at $\bar{u}$ with modulus no more than (4.49).

Proof. The proof has been basically performed in this section. We only fit the remaining blank spots concerning assumptions. Constant ρ_z exists due to the boundedness of sets $C(t)$. From this the existence of ρ_y follows immediately. The single–valuedness of S^K and of S follows from [7, Remark 2] and [7, Theorem 5.7]. The required convergence $y^K(\cdot) \rightharpoonup y(\cdot)$ and $z^K(\cdot) \rightharpoonup \dot{y}(\cdot)$ in $L^2([0,T], \mathbb{R}^n)$ is a result of the proof of [7, Theorem 5.6]. □

5. Optimal control of a dynamical system

5.1 Introduction

In this chapter, we study an optimal control problem with a variational inequality in the constraint system. If we write down this variational inequality in the equivalent form of a generalized equation, it becomes

$$-\dot{z}(t) + R\dot{y}(t) \in \mathrm{N}_{Z(t)}(z(t)), \quad t \in [0,T] \text{ a.e.} \tag{5.1}$$

where $Z(t)$ is a closed convex set, $\mathrm{N}_{Z(t)}(z(t))$ denotes the normal cone to $Z(t)$ at $z(t)$, $\dot{y}(t)$ and $\dot{z}(t)$ stand for the time derivatives and R is a matrix with appropriate dimensions.

Differential inclusion (5.1) can be understood as a special form of the maximal dissipation principle in evolution systems with convex constraints. Models of this type play a central role in modeling nonequilibrium processes with rate–independent memory in mechanics of elastoplastic and thermoelastoplastic materials as well as in ferromagnetism, piezoelectricity or phase transitions, see [15], [23], [71] or [138]. Furthermore, differential inclusion (5.1) is a special case of a sweeping process which was introduced (in its basic form) in [91] and then gradually generalized in a number of papers, e. g. [40] or [81]. In most of these works, however, the authors do not consider any control and concentrate solely on the existence and regularity of its solution. More references to and applications of (5.1) can be found in [68] or [131].

Another view at this model is provided by [96] where, under certain condition, inclusion (5.1) is reformulated as a differential variational inequality, for which some theoretical results and numerous applications exist. For applications to game theory and Nash equilibria see [26], [85] or [128]. If the aforementioned conditions are not fulfilled, inclusion (5.1) can be reformulated as a mixture of a differential variational inequality and a differential algebraic equation [20]. For these reformulations see again [96].

By adding a differential equation

$$\dot{y}(t) = f(t, y(y), z(t)) + Bu(t), \quad t \in [0,T] \text{ a.e.}, \tag{5.2}$$

and an objective function in variables u, y and z, we obtain a system introduced in [22] where necessary optimality conditions for this problem have been derived. The authors considered only a fixed set Z and made use of the rate–independence of the solution map $y \mapsto z$ defined by (5.1). We utilize some of their results and demonstrate the usefulness of this model by means of a simple example from the area of queuing theory.

The incurred optimization problem is a special difficult optimal control problem which cannot be tackled by standard optimal control theory via maximum principle or dynamic programming. Indeed, when dealing with a differential inclusion, the right–hand side is usually required to be Lipschitz continuous, such as in [27] or [87] or this requirement is somehow relaxed, such as the one–sided

Lipschitzian property from [36]. Unfortunately, neither of these assumptions is satisfied in our setting.

From another point of view, system (5.1)–(5.2) models an evolutionary equilibrium and hence its optimization is a special type of Mathematical program with evolutionary equilibrium constraints (MPEEC) which has been defined and studied in [65].

In [29] the authors also investigated an optimization problem with a sweeping process among the constraints. They use a similar technique, based on time discretization and the coderivative calculus to derive necessary optimality conditions for the original problem. However, their control enters the system via the sweeping set and so the results of [29] cannot be compared with those presented in this chapter in a straightforward way.

The aim of this chapter is twofold. First we perform a simple discretization of the problem and show that from any sequence of solutions to the discretized problems one can select a subsequence converging in a certain sense to a solution of the original infinite–dimensional problem. This part depends on some results from [22]. Then, in the second part of the chapter, we apply the so–called implicit programming approach, developed for the numerical solution of non–evolutionary MPECs, to the numerical solution of individual discretized problems. Thereby, we make use of some advanced tools of modern variational analysis, which enables us to compute the so–called "subgradient information" required by most numerical methods of nonsmooth optimization. This part follows essentially the scheme developed in [94]. In fact, this approach has been already successfully used in a different MPEEC describing the magnetization of a piece of ferromagnet [65] where, however, the underlying evolutionary equilibrium has been governed by a completely different infinite–dimensional model.

Hence, our technique differs substantially from the smoothing approach which is usually used in differential inclusions with hysteresis operators [21].

The organization of the chapter is as follows: in Section 5.2 we present the problem and state the assumptions used in the chapter. Discretization of the time interval is performed and discretized optimization problems are stated. Furthermore, the topic of convergence of the optimal solutions to the discretized problems is discussed.

In Section 5.3 we ensure first existence of optimal solutions, computes an upper estimate of the coderivative of the solution map and on its basis an upper estimate of the subdifferential of the composite objective function. Moreover, several cases in which this estimate is exact are discussed. Section 5.4 concludes this chapter by the already–mentioned example from the area of queuing theory.

5.2 Problem statement and its approximation

The main goal of this chapter is to propose a way of the numerical solution of the following optimization problem:

$$\begin{aligned} \text{minimize } & \int_0^T [L_1(t, y(t), z(t)) + L_2(u(t))]dt + L_3(y(T), z(T)) \\ \text{subject to } & -\dot{z}(t) + R\dot{y}(t) \in \mathrm{N}_{Z(t)}(z(t)), \quad t \in [0,T] \text{ a.e.} \\ & \dot{y}(t) = f(t, y(t), z(t)) + Bu(t), \quad t \in [0,T] \text{ a.e.} \\ & u(t) \in \Omega \\ & y(0) = a,\ z(0) = b, \end{aligned} \tag{5.3}$$

where $u : [0,T] \to \mathbb{R}^d$, $y : [0,T] \to \mathbb{R}^n$, $z : [0,T] \to \mathbb{R}^m$, E, R and B are fixed matrices with corresponding dimensions, $Z(t) \subset \mathbb{R}^m$ is a moving convex set and $\mathrm{N}_{Z(t)}(z(t))$ denotes the normal cone to this set at point $z(t)$. The exact assumptions will be given below. The constraint system for problem (5.3) was taken from paper [22], in which necessary optimality conditions were derived. In this chapter we consider more general objective function.

Although the main result of paper [22] was the development of necessary optimality conditions, several of the proven auxiliary results can be also used to derive their discretized–version counterparts. Among the obtained results, a modification of the following facts will be used in this chapter. In particular, we learn that under some assumptions, for any $u \in L^1$ there exists exactly one pair $(y, z) \in W^{1,p}$ satisfying the constraints of (5.3). For this reason u will be referred to as the control variable while y and z will be called the state variables.

Moreover, the following a priori estimate was derived for every $p \in [1, \infty]$:

$$\|y(u)\|_{1,p} + \|z(u)\|_{1,p} \leq C(p)(1 + \|u\|_p)$$

with the constant C independent of the choice of u. Furthermore, $u^K \rightharpoonup u$ in L^2 implies $y(u^K) \rightrightarrows y(u)$ and $z(u^K) \rightrightarrows z(u)$, which under additional assumptions implies the existence of an optimal solution to (5.3). In this section, we show similar statements for the solutions of discretized problems and use these results to prove the convergence of the solutions of the discretized problems to a solution of (5.3).

For any K, we consider a rather general discretization scheme $0 = t_0^K < t_1^K < \cdots < t_K^K = T$; hence we do not require the division to be equidistant. Denoting the distance between two consecutive time instants $h_k^K = t_{k+1}^K - t_k^K$, problem (5.3) may be discretized as follows:

$$\begin{aligned} \text{minimize } & \sum_{k=0}^{K-1} h_k^K [L_1(t_k^K, y_k^K, z_k^K) + L_2(u_k^K)] + L_3(y_K^K, z_K^K) \\ \text{subject to } & -z_{k+1}^K + z_k^K + Ry_{k+1}^K - Ry_k^K \in \mathrm{N}_{Z(t_{k+1}^K)}(z_{k+1}^K),\ k = 0, \ldots, K-1 \\ & y_{k+1}^K - y_k^K = h_k^K f(t_k^K, y_k^K, z_k^K) + h_k^K B u_k^K,\ k = 0, \ldots, K-1 \\ & u_k^K \in \Omega,\ k = 0, \ldots, K-1 \\ & y_0^K = a,\ z_0^K = b.\text{`} \end{aligned} \tag{5.4}$$

To simplify the notation, we will use the notation $f_k^K := f(t_k^K, y_k^K, z_k^K)$ and similarly $Z_k^K := Z(t_k^K)$. Usually, the upper discretization index K will be fixed and hence, it will often be omitted. However, in some cases and especially in the convergence analysis, it will be helpful to emphasize the discretization level by keeping it. If all the values u_k, y_k and z_k are known, functions y^K and z^K denote the piecewise linear functions obtained by connecting the known points. Function u^K will be constructed in a similar way, with the difference that it will be piecewise constant. For the sake of simplicity, the discretization of the differential equation was performed in a forward way.

As the normal cone is a cone by definition, it is not necessary to include term h_k in the inclusion in (5.4). The reason for considering $\mathrm{N}_{Z_{k+1}}(z_{k+1})$ instead of $\mathrm{N}_{Z_k}(z_k)$ in (5.4) is that inclusion

$$-z_{k+1} + z_k + Ry_{k+1} - Ry_k \in \mathrm{N}_{Z_{k+1}}(z_{k+1}) \tag{5.5}$$

may be for convex sets Z_{k+1} equivalently rewritten as equation

$$z_{k+1} = P_{Z_{k+1}}(z_k + Ry_{k+1} - Ry_k). \tag{5.6}$$

This immediately implies that if we know all the values u_k and the initial points a and b, we are able to unambiguously compute the values y_k and z_k for all $k = 0, \ldots, K$. Similarly as in the continuous case, u^K can be seen as the control variable and it is sufficient to optimize only with respect to this variable. Because of this, y^K and z^K will be referred to as state variables.

Having a brief knowledge about the model, we now state the assumptions, which will be used later in the text. Unfortunately, we do not work with Carathéodory functions; on the other hand, we admit some possible discontinuities in the time variable. First define the set

$$\Gamma := \{t \in [0,T] |\ \exists\ y, z : f(\cdot, y, z) \text{ or } L_1(\cdot, y, z) \text{ is not continuous at } t\}.$$

In the rest of this chapter, we require that Γ is a set of zero measure and that for all discretization levels K we have

$$\{t_0^K, \ldots, t_K^K\} \cap \Gamma = \emptyset.$$

This also implies that we have to consider nonequidistant divisions. So we suppose that

- $(H1)$: $Z(t)$ is closed and convex for all t
- $(H2)$: $Z(t)$ moves in a Lipschitz continuous way with constant K_1 with respect to the Hausdorff distance, hence $d_H(Z(t), Z(s)) \leq K_1|t-s|$ where d_H denotes the Hausdorff distance between two sets
- $(H3)$: Ω is closed and convex
- $(A1)$: f is continuous on $\Gamma \times \mathbb{R}^n \times \mathbb{R}^m$
- $(A2)$: $|f(t,y,z)| \leq \alpha_0 + \alpha_1(|y| + |z|)$ for all $t \in [0,T]$, $y \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$
- $(A3)$: L_1 is continuous on $\Gamma \times \mathbb{R}^n \times \mathbb{R}^m$ and bounded below
- $(A4)$: there exists convex $\tilde{L}_2 : \mathbb{R}^d \to \mathbb{R}$ such that $\tilde{L}_2|_\Omega = L_2$ and $L_2(u) \geq \alpha_3 + u^T E_2 u$ for all $u \in \Omega$ with E_2 being positive definite
- $(A5)$: Ω is bounded or $\tilde{L}_2(u) \leq \alpha_2 + u^T E_1 u$ for all $u \in \mathbb{R}^d$
- $(A6)$: L_3 is continuous and bounded below

The imposed assumptions are followed by two remarks. In the first one, we discuss the difference of the model and assumptions from the core chapter and in the second one, possible assumption relaxation is considered.

Remark 5.2.1. We have made several extensions of the model considered in [22]. To the objective function we have added a term depending on the terminal state via L_3 and we consider a more general L_2 which was restricted only to a quadratic function in [22]. As seen from $(A5)$, L_2 still possesses some features of a quadratic function; however, if Ω is bounded, then L_2 can be any bounded function.

For the sake of simplicity, we still insist on the separability of the objective function in the state and control variables. It is certainly possible to overcome this restriction by strengthening several assumptions to hold uniformly in the control or state variables.

Furthermore, we allow $Z(\cdot)$ to depend on time. On the other hand, we require several estimates to hold uniformly in t. Specifically, α_0 should be time–independent and L_1 should be bounded on compact sets in all three variables, not only in the state ones. The time uniformity is used only two times in this text, namely in (5.9) and (5.21). In both cases the Lebesgue integral has to be approximated by a limit of Riemann sums. Although this is not possible for arbitrary choice of time instants, according to [32, Lemma 4.12], it is possible to choose time instants in such a way that Riemann sums converge to the Lebesgue integral.

It is certainly possible to consider $\alpha_0 \in L^1$ instead of the assumed $\alpha_0 \in L^\infty$ but in this case Γ would have to be restricted to admit only time divisions mentioned above. The same consideration applies also to L_1.

Remark 5.2.2. In the assumptions we require that the discontinuities in the data may occur only on a subset of time instants with zero measure; moreover, these instants should not depend on the state variables y and z. The reason for this requirement is simple. For the convergence analysis, it could possibly suffice to choose any point t in the interval $[t_k, t_{k+1})$, at which $f(\cdot, y_k, z_k)$ is continuous or a Lebesgue point. On the other hand, for the optimization and the computation of coderivative of the solution map we need to have fixed time division and not change it every time when the state variables y_k and z_k change.

We will discuss only the case of f but of course this remark is valid for L_1 as well. It is certainly possible to weaken assumption $(A1)$ into f being only a Carathéodory function. On the other hand, then the discretized equation

$$y_{k+1} - y_k = h_k f(t_k, y_k, z_k) + h_k B u_k$$

would have to be replaced by

$$y_{k+1} - y_k = \int_{t_k}^{t_{k+1}} f(t, y_k, z_k) dt + h_k B u_k. \tag{5.7}$$

By considering (5.7) it is possible to prove the convergence. However, this would cause trouble in the numerical computation since instead of evaluation of f at one point, an integral would have to be computed.

It is also possible to weaken the convexity assumption in $(H1)$ by assuming that $Z(t)$ is a prox–regular set with a specific structure. This follows from the

discretization scheme used in [137]. The main argument is that the points z_k will not be far away from the set Z_k and thus the projection will retain its single–valued Lipschitzian character, even though the set may be nonconvex.

Having posed and discussed the assumptions, we start now developing a chain of lemmas analyzing the properties of solutions to the discretized problems. First of all, we show the uniform boundedness of the state variables.

Lemma 5.2.3. *Assume that (H1), (H2) and (A2) are valid. Then for any feasible solution of the discretized problems u^K, y^K, z^K the following estimate holds true for all $k = 0, \dots, K-1$*

$$\left|\frac{y_{k+1}-y_k}{h_k}\right| + \left|\frac{z_{k+1}-z_k}{h_k}\right| \le (1+\|R\|)[\alpha_0 + \alpha_1|y_k| + \alpha_1|z_k| + \|B\||u_k|] + K_1.$$

Proof. At first, we derive an estimate for the derivative of z^K. Due to the construction of z^K we know that $z_k \in Z_k$, and thus $P_{Z_k}(z_k) = z_k$. Furthermore

$$\begin{aligned}|z_{k+1}-z_k| &= |P_{Z_{k+1}}(z_k + Ry_{k+1} - Ry_k) - z_k| \\ &= |P_{Z_{k+1}}(z_k + Ry_{k+1} - Ry_k) - P_{Z_{k+1}}(z_k) + P_{Z_{k+1}}(z_k) - P_{Z_j}(z_k)| \\ &\le |Ry_{k+1} - Ry_k| + |P_{Z_{k+1}}(z_k) - P_{Z_j}(z_k)| \le \|R\||y_{k+1}-y_k| + K_1 h_k.\end{aligned}$$

Dividing the previous inequality by h_k, we obtain

$$\left|\frac{z_{k+1}-z_k}{h_k}\right| \le \|R\| \left|\frac{y_{k+1}-y_k}{h_k}\right| + K_1.$$

Altogether, we obtain that

$$\begin{aligned}\left|\frac{y_{k+1}-y_k}{h_k}\right| + \left|\frac{z_{k+1}-z_k}{h_k}\right| &\le (1+\|R\|)\left|\frac{y_{k+1}-y_k}{h_k}\right| + K_1 \\ &\le (1+\|R\|)[|f_k| + \|B\||u_k|] + K_1 \\ &\le (1+\|R\|)[\alpha_0 + \alpha_1|y_k| + \alpha_1|z_k| + \|B\||u_k|] + K_1.\end{aligned} \tag{5.8}$$

□

In the previous lemma we derived the estimate needed in the discrete version of the Gronwall's Lemma, which allows us to estimate the Sobolev norm of the state variables by the Lebesgue norm of the control variable.

Lemma 5.2.4. *Let assumptions (H1), (H2) and (A2) be fulfilled. Then for any $p \in [1, \infty]$ there exists a constant C such that for any K and for any feasible solution of the discretized problem $u^K \in L^p$ and the corresponding y^K and z^K the following estimate is valid*

$$\|y^K\|_{1,p} + \|z^K\|_{1,p} \le C(1 + \|u^K\|_p).$$

Proof. Since

$$|a| - |b| \le |a-b|,$$

due to Lemma 5.2.3 we obtain

$$\frac{|y_{k+1}| - |y_k| + |z_{k+1}| - |z_k|}{h_k} \le (1+\|R\|)(\alpha_0 + \|B\||u_k|) + K_1 + (1+\|R\|)\alpha_1(|y_k| + |z_k|),$$

and hence the Corollary 4.3.2 to the discrete Gronwall's Lemma can be applied to $a_k := |y_k| + |z_k|$. This infers that

$$|y_k| + |z_k| \leq C_1 + C_2 \sum_{i=0}^{k-1} h_i |u_i| \leq C_1 + C_2 \sum_{i=0}^{K-1} h_i |u_i| = C_1 + C_2 \|u^K\|_1, \qquad (5.9)$$

where the last equality holds true since u^K is a piecewise constant function. This leads to the first part of the estimate

$$\|y^K\|_\infty + \|z^K\|_\infty \leq C_1 + C_2 \|u^K\|_1.$$

For the estimate of derivatives combine Lemma 5.2.3 with (5.9), which yields

$$\left| \frac{y_{k+1} - y_k}{h_k} \right| + \left| \frac{z_{k+1} - z_k}{h_k} \right| \leq C_3 + C_4 \|u^K\|_1 + C_5 |u_k|. \qquad (5.10)$$

Temporarily ignoring one of the terms on the left–hand side of the previous estimate, integrating the p–th power of the rest, using the well–known equivalence of norms on any finite–dimensional space and the estimate of $\|\cdot\|_1 \leq T^{\frac{p-1}{p}} \|\cdot\|_p$ we obtain

$$\|\dot{y}^K\|_p^p \leq C_6 (T C_3^p + T C_4^p \|u^K\|_1^p + C_5^p \|u^K\|_p^p) \leq C_6 (T C_3^p + (C_4^p T^{2p} + C_5^p) \|u^K\|_p^p)$$

To finish the proof it is sufficient to take the p–th root and again use the equivalence of norms on $\mathbb{R}^2$. The estimate for $\|\dot{z}^K\|_p$ follows from the symmetry of (5.10). □

Insofar, we have been working only with the constraint system and not with the objective function. In the next theorem, we prove the existence of a convergent subsequence to an arbitrary sequence of feasible solutions to the discretized problems. For notational simplicity, any subsequence will be denoted by the same indices as the original sequence. Denote further the composite objective function of the original problem (5.3) by $J(u)$ and the composite objective function of the discretized problems (5.4) by $J^K(u^K)$.

Theorem 5.2.5. *Let assumptions* $(H1)$–$(H3)$, $(A1)$–$(A4)$ *and* $(A6)$ *hold true and let* u^K *be a sequence of any feasible solutions to the discretized problems which is bounded in* L^2. *Then there exists a subsequence, denoted without relabeling, and some* $u \in L^2$ *such that* $u^K \rightharpoonup u$ *in* L^2. *Moreover, for the corresponding* y^K *and* z^K *there are Lipschitz continuous functions* y *and* z *such that* $y^K \rightrightarrows y$ *and* $z^K \rightrightarrows z$. *Finally, the triple* (u, y, z) *is a feasible solution to the original problem* (5.3).

Furthermore, we obtain

$$J(u) \leq \liminf J^K(u^K). \qquad (5.11)$$

If u^K *converges to* u *strongly in* L^2 *and assumption* $(A5)$ *is valid, then*

$$J(u) = \lim J^K(u^K). \qquad (5.12)$$

Proof. Since u^K is bounded in L^2, a weakly convergent subsequence may be chosen. The convergent subsequences of y^K and z^K may be chosen due to their boundedness in $W^{1,2}$ from Lemma 5.2.4 and the compact embedding of $W^{1,2}$ into $\mathcal{C}(0,T)$. As a byproduct we obtain the weak convergence of the derivatives $\dot{y}^K \rightharpoonup \dot{y}$ and $\dot{z}^K \rightharpoonup \dot{z}$ in L^2.

The feasibility of (u,y,z) will be proven in several steps. Since the proof that the triple (u,y,z) satisfies the differential inclusion and equation is not straightforward, these proofs are postponed to two following Lemmas 5.2.6 and 5.2.7. Here we will prove only that $u(t) \in \Omega$. As $u^K \rightharpoonup u$, due to Mazur's lemma we may choose integers $n(K)$ and a convex hull of $\{u^i |\ i = K, \dots, n(K)\}$ to define new functions, say

$$\tilde{u}^K(\cdot) := \sum_{i=k}^{n(K)} \lambda_{ik} u^i(\cdot),$$

such that $\tilde{u}^K \to u$ in L^2. This implies the pointwise convergence for a subsequence. As Ω is convex and does not depend on t, the relation $u^K(t) \in \Omega$ implies $\tilde{u}^K(t) \in \Omega$ and the closedness of Ω implies $u(t) \in \Omega$.

To show inequality (5.11), we consider first the last term of the objective function. Due to the uniform convergence of the state variables and the continuity of L_3 from $(A6)$, we obtain

$$L_3(y^K(T), z^K(T)) \to L_3(y(T), z(T)).$$

Concerning the first term, define the piecewise constant function

$$\tilde{L}_1^K(t) := L_1(t_k^K, y_k^K, z_k^K),\ t \in [t_k^K, t_{k+1}^K).$$

By virtue of the uniform convergence $y^K \rightrightarrows y$, $z^K \rightrightarrows z$ and the continuity of L_1, we get for all $t \notin \Gamma$

$$\tilde{L}_1^K(t) = L_1(t_k^K, y_k^K, z_k^K) \to L_1(t, y(t), z(t)).$$

Since y^K and z^K are uniformly bounded in L^∞ due to Lemma 5.2.4, assumption $(A3)$ implies that $\tilde{L}_1^K$ are uniformly bounded in the same space as well and hence the dominated convergence theorem may be used. This leads to

$$\sum_{k=0}^{K-1} h_k^K L_1(t_k^K, y_k^K, z_k^K) = \sum_{k=0}^{K-1} h_k^K \tilde{L}_1^K(t_k^K) = \int_0^T \tilde{L}_1^K(t)dt \to \int_0^T L_1(t, y(t), z(t))dt, \tag{5.13}$$

thus the first part of the objective function converges.

As to the second term, define

$$J_2(u) := \int_0^T L_2(u(t))dt.$$

According to [31, Proposition 2.32] assumption $(A5)$ implies that there exists a constant β such that for $\tilde{u}_1, \tilde{u}_2 \in \Omega$ one has

$$|L_2(\tilde{u}_1) - L_2(\tilde{u}_2)| \le \beta(1 + |\tilde{u}_1| + |\tilde{u}_2|)|\tilde{u}_1 - \tilde{u}_2|.$$

Integrating the previous inequality and applying the Hölder inequality, we obtain that $J_2 : L^2 \to \mathbb{R}$ is continuous, which together with the previous parts implies (5.12). To prove (5.11) it is sufficient to show that J_2 is weakly lower semicontinuous. But this follows from the fact that for convex functions the lower semicontinuity and the weak lower semicontinuity coincide.

Functions y and z are Lipschitz continuous because they are uniform limits of sequences of functions with uniform Lipschitz moduli. □

To finish the proof of Theorem 5.2.5 it remains to show that (u, y, z) satisfies the differential inclusion and the differential equation. This will be conducted in the following two lemmas.

Lemma 5.2.6. *Under assumptions of Theorem 5.2.5, the pair (y, z) satisfies almost everywhere the differential inclusion*

$$-\dot{z}(t) + R\dot{y}(t) \in \mathrm{N}_{Z(t)}(z(t)).$$

Proof. This proof is a modification of a similar proof presented in [137], the first part goes in the same way, yet for the second part we depart from finite–dimensional setting into the infinite–dimensional one and then return back.

Define first

$$S := \{x \in L^2 |\ x(t) \in Z(t)\ a.e.\}.$$

Since $Z(t)$ is convex for every t, it is easy to see that S is convex as well. Further for fixed K define $\theta^K : \mathbb{R} \to \mathbb{R}$ as a function satisfying $\theta^K(t) := t^K_{k+1}$ on the interval $t \in [t^K_k, t^K_{k+1})$. Then for almost every t it holds true that

$$-\dot{z}^K(t) + R\dot{y}^K(t) \in \mathrm{N}_{Z(\theta^K(t))}(z^K(\theta^K(t))). \tag{5.14}$$

Consequently, from (5.14) and from the equivalence between statements 1 and 2 in Lemma A.2.2 we obtain that for every $\xi(t)$ we have

$$\langle -\dot{z}^K(t) + R\dot{y}^K(t), \xi(t)\rangle \le |-\dot{z}^K(t) + R\dot{y}^K(t)| d_{Z(\theta^K(t))}(z^K(\theta^K(t)) + \xi(t)). \tag{5.15}$$

Since $\xi(t)$ may be chosen in an arbitrary way, we can also restrict our attention to all $\xi \in L^2$. From Lemma 5.2.4 and imposed assumptions it follows that

$$\| -\dot{z}^K + R\dot{y}^K\|_2 \le C(1 + \|u^K\|_2) \le \tilde{C}.$$

Integrating (5.15) over $[0, T]$ and using Hölder inequality implies that

$$\begin{aligned}
&\int_0^T \langle -\dot{z}^K(t) + R\dot{y}^K(t), \xi(t)\rangle dt \\
&\qquad \le \| -\dot{z}^K + R\dot{y}^K\|_2 \left[\int_0^T d^2_{Z(\theta^K(t))}(z^K(\theta^K(t)) + \xi(t))dt\right]^{\frac{1}{2}} \qquad (5.16)\\
&\qquad \le \tilde{C}\left[\int_0^T d^2_{Z(\theta^K(t))}(z^K(\theta^K(t)) + \xi(t))dt\right]^{\frac{1}{2}}.
\end{aligned}$$

Passing to the limit, the left–hand side converges in L^2 due to the weak convergence of the time derivatives to

$$\int_0^T \langle -\dot{z}(t) + R\dot{y}(t), \xi(t)\rangle dt.$$

For the right–hand side we intend to use the Lebesgue dominated convergence theorem. To find the dominating function, we employ the relation

$$z^K(\theta^K(t)) \in Z(\theta^K(t)),$$

which implies

$$d^2_{Z(\theta^K(t))}(z^K(\theta^K(t)) + \xi(t)) \leq \xi^2(t).$$

Since $\xi \in L^2$, the Lebesgue dominated convergence theorem may be used to obtain

$$\lim_K \int_0^T d^2_{Z(\theta^K(t))}(z^K(\theta^K(t)) + \xi(t))dt = \int_0^T \lim_K d^2_{Z(\theta^K(t))}(z^K(\theta^K(t)) + \xi(t))dt \tag{5.17}$$

Since $Z(t)$ moves in a Lipschitzian way and $z^K(\theta^K(t)) \to z(t)$ due to $z^K \rightrightarrows z$, for any fixed t we obtain

$$\begin{aligned}
&|d_{Z(\theta^K(t))}(z^K(\theta^K(t)) + \xi(t)) - d_{Z(t)}(z(t) + \xi(t))| \\
&\quad \leq |d_{Z(\theta^K(t))}(z^K(\theta^K(t)) + \xi(t)) - d_{Z(t)}(z^K(\theta^K(t)) + \xi(t))| \\
&\quad + |d_{Z(t)}(z^K(\theta^K(t)) + \xi(t)) - d_{Z(t)}(z(t) + \xi(t))| \to 0.
\end{aligned}$$

This implies

$$d_{Z(\theta^K(t))}(z^K(\theta^K(t)) + \xi(t)) \to d_{Z(t)}(z(t) + \xi(t))$$

for every fixed t and $\xi(t)$.

Putting all the previous results together, we conclude that for all $\xi \in L^2$ we have

$$\int_0^T \langle -\dot{z}(t) + R\dot{y}(t), \xi(t)\rangle dt \leq \tilde{C} \left[\int_0^T d^2_{Z(t)}(z(t) + \xi(t))dt\right]^{\frac{1}{2}}. \tag{5.18}$$

Next we claim that this relation implies the following relation in L^2 (compare with the implication 3 to 1 in Lemma A.2.2)

$$-\dot{z} + R\dot{y} \in \mathrm{N}_S(z). \tag{5.19}$$

Assume that this is not the case. Since S is convex, by the definition of the normal cone there exists some $x \in S$ such that

$$\int_0^T \langle -\dot{z}(t) + R\dot{y}(t), x(t) - z(t)\rangle dt > 0. \tag{5.20}$$

Putting $\xi = x - z$ into (5.18), we get

$$\int_0^T \langle -\dot{z}(t) + R\dot{y}(t), x(t) - z(t)\rangle dt \leq \tilde{C} \left[\int_0^T d^2_{Z(t)}(x(t))dt\right]^{\frac{1}{2}}.$$

As $x(t) \in Z(t)$ by $x \in S$, the right–hand side is equal to zero, which is a contradiction with (5.20), and hence the validity of (5.19) is proven. But (5.19) is almost everywhere equivalent to

$$-\dot{z}(t) + R\dot{y}(t) \in \mathrm{N}_{Z(t)}(z(t))$$

by virtue of [28, Proposition 3.5.7]. □

Lemma 5.2.7. *Under assumptions of Theorem 5.2.5, the triple (u, y, z) satisfies almost everywhere the differential equation*

$$\dot{y}(t) = f(t, y(t), z(t)) + Bu(t).$$

Proof. Similarly as in the previous lemma define for fixed K the function $\nu^K : \mathbb{R} \to \mathbb{R}$ satisfying $\nu^K(t) := t_k^K$ on interval $t \in [t_k^K, t_{k+1}^K)$.

Fix any $t \in [0, T]$. Then we have

$$\dot{y}^K(t) = f(\nu(t), y^K(\nu(t)), z^K(\nu(t)) + Bu^K(t).$$

In the last term it is not necessary to insert ν because u^K is piecewise constant. Integrating the previous equation, we obtain

$$y^K(t) = a + \int_0^t [f(\nu(s), y^K(\nu(s)), z^K(\nu(s))) + Bu^K(s)]ds.$$

To obtain the desired result we intend to pass to the limit and compute the derivative afterwards. The left–hand side converges to $y(t)$ and the convergence

$$\int_0^t Bu^K(s)ds \to \int_0^t Bu(s)ds$$

is evident from the definition of weak convergence.

For the missing term, we use again the Lebesgue dominated convergence theorem. Observe that a large enough constant C' can be chosen such that

$$|f(\nu(s), y^K(\nu(s)), z^K(\nu(s)))| \leq \alpha_0 + \alpha_1(|y^K(\nu(s))| + |z^K(\nu(s))|) \leq C'. \quad (5.21)$$

Thus

$$\begin{aligned} &\lim \int_0^t f(\nu(s), y^K(\nu(s)), z^K(\nu(s)))ds \\ &\qquad = \int_0^t \lim f(\nu(s), y^K(\nu(s)), z^K(\nu(s)))ds = \int_0^t f(s, y(s), z(s))ds, \end{aligned} \quad (5.22)$$

where in the last equality we employed the continuity of f on $\Gamma \times \mathbb{R}^n \times \mathbb{R}^m$ and the fact that Γ has zero measure.

All in all, we receive

$$y(t) = a + \int_0^t [f(s, y(s), z(s)) + Bu(s)]ds,$$

which almost everywhere amounts to

$$\dot{y}(t) = f(t, y(t), z(t)) + Bu(t)$$

and $y(0) = a$. □

In Theorem 5.2.5 we have shown the convergence of a sequence of any feasible solutions. In the next theorem, we make use of it and prove similar result for optimal solutions. This theorem is further utilized by the end of the chapter where a numerical solution scheme is presented.

Theorem 5.2.8. *Let assumptions (H1)–(H3) and (A1)–(A6) be fulfilled and let $(\bar{u}^K, \bar{y}^K, \bar{z}^K)$ be optimal solutions of discretized problems (5.4). Then there exists their subsequence such that, without relabeling, $\bar{u}^K \rightharpoonup \bar{u}$ in L^2, $\bar{y}^K \rightrightarrows \bar{y}$ and $\bar{z}^K \rightrightarrows \bar{z}$, where the triple $(\bar{u}, \bar{y}, \bar{z})$ is an optimal solution of the original problem (5.3). Moreover, $\bar{y}$ and $\bar{z}$ are Lipschitz continuous and the values of the objective functions converge as well, i. e., $J^K(\bar{u}^K) \to J(\bar{u})$.*

Proof. Since E_2 from (A5) induces an equivalent norm on $\mathbb{R}^d$, we obtain for some $\tilde{C}$ the estimate:

$$\begin{aligned}\sum_{k=0}^{K-1} h_k L_2(\bar{u}_k^K) &\geq \sum_{k=0}^{K-1} h_k(\alpha_3 + (\bar{u}_k^K)^T E_2 \bar{u}_k^K) \\ &\geq \alpha_3 T + \sum_{k=0}^{K-1} h_k \tilde{C}(\bar{u}_k^K)^T \bar{u}_k^K = \alpha_3 T + \tilde{C}\|\bar{u}^K\|_2^2.\end{aligned}$$

We claim now that $\|\bar{u}^K\|_2$ is bounded. If this is not the case, then since L_1 and L_3 are bounded below, we obtain $J^K(\bar{u}^K) \to \infty$. But taking any u^K constant and feasible we obtain a contradiction with optimality of $\bar{u}^K$.

This implies that the assumptions of Theorem 5.2.5 are fulfilled and there exists a subsequence converging to some $(\bar{u}, \bar{y}, \bar{z})$, which is a feasible solution to the original problem (5.3). To finish the proof it is necessary to show that $(\bar{u}, \bar{y}, \bar{z})$ is indeed the optimal solution to (5.3). Consider arbitrary feasible solution (u, y, z) to (5.3). From Lemma A.2.1 we obtain that u can be approximated by simple functions u^K such that the intervals, on which $\bar{u}^K$ and u^K are constant, coincide. To u^K we find the corresponding y^K and z^K.

Due to Theorem 5.2.5, we obtain that

$$J(\bar{u}) \leq \liminf J^K(\bar{u}^K) \leq \liminf J^K(u^K) = J(u), \tag{5.23}$$

where in the second inequality we used the optimality of $\bar{u}^K$ and the property that u^K is constant on the same intervals as $\bar{u}^K$. This implies that $\bar{u}$ is indeed an optimal solution to (5.3). As (5.23) holds true with limsup instead of liminf as well, choosing $u = \bar{u}$ implies $J^K(\bar{u}^K) \to J(\bar{u})$. □

5.3 Numerical solution of discretized problems

In Section 5.2 we performed the convergence analysis. This section is devoted to the numerical solution of discretized problems. Hence, we fix the index K and discretization time instants t_k^K and suggest a solution method for the optimization problem

$$\begin{aligned}\text{minimize } & \sum_{k=0}^{K-1} h_k[L_1(t_k, y_k, z_k) + L_2(u_k)] + L_3(y_K, z_K) \\ \text{subject to } & z_{k+1} = P_{Z(t_{k+1})}(z_k + Ry_{k+1} - Ry_k),\ k = 0, \ldots, K-1 \\ & y_{k+1} = y_k + h_k f(t_k, y_k, z_k) + h_k B u_k,\ k = 0, \ldots, K-1 \\ & u_k \in \Omega,\ k = 0, \ldots, K-1 \\ & y_0 = a,\ z_0 = b,\end{aligned} \tag{5.24}$$

which is clearly equivalent to (5.4). Often the abbreviated notation will be used such as $u = (u_0, \ldots, u_{k-1}) = (u_0^K, \ldots, u_{k-1}^K)$.

Even though problem (5.24) is a finite–dimensional optimization problem and the only nonsmoothness may appear in the objective function, the projection operator and possibly in f, the number of equalities may be large and the computation may not be simple. It is also not entirely clear how to handle the fact that the optimization is performed only over u and not over y and z.

To handle this problem, we will make use of the implicit programming approach which was thoroughly described in Section 2.2. Thus, define the solution mapping $S : \mathbb{R}^{Kd} \rightrightarrows \mathbb{R}^{(K+1)n} \times \mathbb{R}^{(K+1)m}$ as

$$\begin{aligned} S(u_0, \ldots, u_{k-1}) := \{(y_0, \ldots, y_K, z_0, \ldots, z_K) | \\ z_{k+1} = P_{Z(t_{k+1})}(z_k + Ry_{k+1} - Ry_k),\ k = 0, \ldots, K-1 \\ y_{k+1} = y_k + h_k f(t_k, y_k, z_k) + h_k B u_k,\ k = 0, \ldots, K-1 \\ y_0 = a,\ z_0 = b\} \end{aligned} \tag{5.25}$$

and introduce the function

$$\tilde{L}(y_0, \ldots, y_K, z_0, \ldots, z_K) := \sum_{k=0}^{K-1} h_k L_1(t_k, y_k, z_k) + L_3(y_K, z_K).$$

Then the discretized problem (5.4) attains the form

$$\begin{aligned} &\text{minimize } (\tilde{L} \circ S)(u_0, \ldots, u_{k-1}) + \sum_{k=0}^{K-1} h_k L_2(u_k) \\ &\text{subject to } u_k \in \Omega. \end{aligned} \tag{5.26}$$

For the numerical solution of (5.26), we intend to use a numerical method which employs the so–called subgradient information as described in Section 2.2. To do so, we need to be able to evaluate the values of $\tilde{L} \circ S$ and to be able to compute at least one element of its Clarke subdifferential $\overline{\partial}(\tilde{L} \circ S)$. The evaluation of the values will be simple but possibly time consuming for large K. It is sufficient to evaluate $S(u)$ and then plug it into $\tilde{L}$. The computation of the subdifferential will be performed in the subsequent text.

We give additional assumptions strengthening the previous ones:

- $(A7)$: $f(t_k, \cdot, \cdot)$ is $\mathcal{C}^1$ for every k
- $(A8)$: $L_1(t_k, \cdot, \cdot)$ is locally Lipschitz at for every k and L_3 is locally Lipschitz
- $(A8')$: $L_1(t_k, \cdot, \cdot)$ is $\mathcal{C}^1$ for every k and L_3 is $\mathcal{C}^1$.

First, we prove that S is single–valued and locally Lipschitz continuous and then we state a result about existence of optimal solutions to the discretized problems.

Lemma 5.3.1. *Under assumptions $(H1)$, $(H2)$, $(A2)$ and $(A7)$, the mapping S is single–valued and locally Lipschitz.*

Proof. Single–valuedness of S is evident from (5.25). To prove the local Lipschitzian property, we will make use of the well–known fact that a function is

locally Lipschitz on an open set if and only if it is globally Lipschitz on its any compact subset. Choosing two control variables u and $\tilde{u}$ with $u_k, \tilde{u}_k \in \Omega \cap B(0,r)$, we obtain from Lemma 5.2.4 that all y_k, $\tilde{y}_k$, z_k and $\tilde{z}_k$ are contained in a compact set. Due to $(A7)$, $f(t_k,\cdot,\cdot)$ is globally Lipschitz on this set, with, say, constant K_2.

Let us make the following estimates:

$$\begin{aligned}
|z_{k+1} - \tilde{z}_{k+1}| &= |P_{Z_{k+1}}(z_k + Ry_{k+1} - Ry_k) - P_{Z_{k+1}}(\tilde{z}_k + R\tilde{y}_{k+1} - R\tilde{y}_k)| \\
&\leq |z_k + Ry_{k+1} - Ry_k - \tilde{z}_k - R\tilde{y}_{k+1} + R\tilde{y}_k)| \\
&\leq |z_k - \tilde{z}_k| + \|R\||y_{k+1} - \tilde{y}_{k+1}| + \|R\||y_k - \tilde{y}_k| \\
|y_{k+1} - \tilde{y}_{k+1}| &= |y_k + h_k f(t_k, y_k, z_k) + h_k B u_k - \tilde{y}_k - h_k f(t_k, \tilde{y}_k, \tilde{z}_k) - h_k B\tilde{u}_k| \\
&\leq (h_k K_2 + 1)|y_k - \tilde{y}_k| + h_k K_2 |z_k - \tilde{z}_k| + h_k \|B\||u_k - \tilde{u}_k|.
\end{aligned}$$

From the previous relations we obtain the existence of a constant C dependent only on K_2, R, B and the fixed time division which satisfies

$$\begin{aligned}
|z_{k+1} - \tilde{z}_{k+1}| &\leq C|z_k - \tilde{z}_k| + C|y_k - \tilde{y}_k| + C|u_k - \tilde{u}_k| \\
|y_{k+1} - \tilde{y}_{k+1}| &\leq C|z_k - \tilde{z}_k| + C|y_k - \tilde{y}_k| + C|u_k - \tilde{u}_k|.
\end{aligned}$$

Summing these inequalities and chaining them k times, we obtain for some constant $\tilde{C}$ that

$$|z_k - \tilde{z}_k| + |y_k - \tilde{y}_k| \leq \tilde{C}|z_0 - \tilde{z}_0| + \tilde{C}|y_0 - \tilde{y}_0| + \tilde{C}\sum_{i=0}^{k-1} |u_i - \tilde{u}_i| = \tilde{C}\sum_{i=0}^{k-1} |u_i - \tilde{u}_i|.$$

The Lipschitzian property of S follows now directly from the previous estimate. □

Lemma 5.3.2. *Under assumptions* $(H1)$*–*$(H3)$ *and* $(A2)$*–*$(A7)$ *discretized problems* (5.4) *have an optimal solution.*

Proof. A sufficient condition for the equivalence of (5.5) and (5.6) is convexity and closedness of Z_k. Due to assumptions and Lemma 5.3.1, the objective function of (5.26) is continuous, coercive and bounded below. Since Ω is closed, the statement follows. □

After proving that S is locally Lipschitz, we will compute its coderivative in the following lemma and then discuss the used constraint qualification. We will learn later from Lemma 5.3.4 that the constraint qualification is always satisfied. However, we keep it in the lemma statement and omit it only in the final Theorem 5.3.7. Before stating the lemma we will need yet another representation of the solution map.

Define function $g : \mathbb{R}^{Kd} \times \mathbb{R}^{(K+1)n} \times \mathbb{R}^{(K+1)m} \to \mathbb{R}^{(K+1)n} \times \mathbb{R}^{(K+1)m} \times \mathbb{R}^{Km}$

and set Q as follows

$$g(u,y,z) := \begin{pmatrix} -y_0 + a \\ h_0 B u_0 - y_1 + y_0 + h_0 f(t_0, y_0, z_0) \\ \dots \\ h_{k-1} B u_{k-1} - y_K + y_{k-1} + h_{k-1} f(t_{k-1}, y_{k-1}, z_{k-1}) \\ z_0 - b \\ z_1 - z_0 - R y_1 + R y_0 \\ \dots \\ z_K - z_{k-1} - R y_K + R y_{k-1} \\ -z_1 \\ \dots \\ -z_K \end{pmatrix}$$

$$Q := \left\{ \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \\ 0 \\ \alpha_1 \\ \dots \\ \alpha_K \\ \beta_1 \\ \dots \\ \beta_K \end{pmatrix} \mid \alpha_k \in \mathrm{N}_{Z_k}(\beta_k) \right\} \subset \mathbb{R}^{(K+1)n} \times \mathbb{R}^{(K+1)m} \times \mathbb{R}^{Km}.$$

Then the solution map can be equivalently written as

$$\begin{aligned} S(u) &= \{(y,z) |\ 0 \in g(u,y,z) + Q)\} \\ \operatorname{gph} S &= \{(u,y,z) |\ -g(u,y,z) \in Q)\} = (-g)^{-1}(Q). \end{aligned} \tag{5.27}$$

On the basis of this reformulation we are able to compute D^*S.

Lemma 5.3.3. *Let $(\bar{y}, \bar{z}) = S(\bar{u})$ and let assumptions $(H1)$, $(H2)$, $(A2)$ and $(A7)$ be fulfilled. Further, let the multifunction*

$$\Xi(\eta) := \{(u,y,z) |\ \eta \in g(u,u,z) + \operatorname{gph} Q)\}$$

be calm at $(0, \bar{u}, \bar{y}, \bar{z})$. Then for $\mu \in \mathbb{R}^{(K+1)n}$, $\nu \in \mathbb{R}^{(K+1)m}$ and for each element

$$u^* = \begin{pmatrix} u_0^* \\ \dots \\ u_{k-1}^* \end{pmatrix} \in D^*S(u,y,z)(\mu,\nu), \tag{5.28}$$

one has the representation

$$u^* = \begin{pmatrix} h_0 B^T p_1 \\ \dots \\ h_{k-1} B^T p_K \end{pmatrix}, \tag{5.29}$$

where $p_1, \dots, p_K$ are solutions of the adjoint equations with $k = 1, \dots, K-1$

$$p_k = p_{k+1} + h_k (\nabla_y f_k)^T p_{k+1} - R^T q_{k+1} + R^T q_k + \mu_k \tag{5.30a}$$

$$q_k = q_{k+1} + h_k (\nabla_z f_k)^T p_{k+1} + r_k + \nu_k \tag{5.30b}$$

with the terminal conditions

$$p_K = R^T q_K + \mu_K \tag{5.31a}$$
$$q_K = r_K + \nu_K \tag{5.31b}$$

and the multipliers r_k and q_k satisfy for $k = 1, \dots, K-1$ the relations

$$\begin{pmatrix} r_k \\ q_k \end{pmatrix} \in \mathrm{N}_{\mathrm{gph}\, \mathrm{N}_{Z_k}} \begin{pmatrix} \bar{z}_k \\ -\bar{z}_k + \bar{z}_{k-1} + R\bar{y}_k - R\bar{y}_{k-1} \end{pmatrix}. \tag{5.32}$$

Proof. Since the continuous differentiability of f implies the continuous differentiability of g, [60, Proposition 3.8] together with the imposed constraint qualification implies

$$\mathrm{N}_{\mathrm{gph}\, S}(\bar{u}, \bar{y}, \bar{z}) \subset -(\nabla g(\bar{u}, \bar{y}, \bar{z}))^T \, \mathrm{N}_Q(-g(\bar{u}, \bar{y}, \bar{z})). \tag{5.33}$$

Observe that for $\tilde{p} \in \mathbb{R}^{(K+1)n}$, $q \in \mathbb{R}^{(K+1)m}$ and $r \in \mathbb{R}^{Km}$ the relation

$$\begin{pmatrix} \tilde{p} \\ q \\ r \end{pmatrix} \in \mathrm{N}_Q(-g(\bar{u}, \bar{y}, \bar{z}))$$

amounts to

$$\begin{pmatrix} r_k \\ q_k \end{pmatrix} \in \mathrm{N}_{\mathrm{gph}\, \mathrm{N}_{Z_k}} \begin{pmatrix} \bar{z}_k \\ -\bar{z}_k + \bar{z}_{k-1} + R\bar{y}_k - R\bar{y}_{k-1} \end{pmatrix}, \ k = 1, \dots, K \tag{5.34}$$

and the variables $\tilde{p}_k$ and q_0 being unconstrained.

From (5.33) we obtain

$$\begin{aligned} D^*S(u, y, z)(\mu, \nu) &= \left\{ \tilde{u}^* \middle| \ \begin{pmatrix} \tilde{u}^* \\ -\mu \\ -\nu \end{pmatrix} \in \mathrm{N}_{\mathrm{gph}\, S}(\bar{u}, \bar{y}, \bar{z}) \right\} \\ &\subset \left\{ \tilde{u}^* \middle| \ \begin{pmatrix} -\tilde{u}^* \\ \mu \\ \nu \end{pmatrix} = (\nabla g(\bar{u}, \bar{y}, \bar{z}))^T \begin{pmatrix} \tilde{p} \\ q \\ r \end{pmatrix}; (5.34) \text{ holds true} \right\}. \end{aligned} \tag{5.35}$$

As the matrix $\nabla g(\bar{u}, \bar{y}, \bar{z})$ is rather large, we do not write it down. However, it can be retrospectively computed from the following relation, which follows from (5.35)

$$\left(\begin{array}{c} -\tilde{u}_0^* \\ \dots \\ -\tilde{u}_{k-1}^* \\ \hline \mu_0 \\ \mu_1 \\ \dots \\ \mu_{k-1} \\ \mu_K \\ \hline \nu_0 \\ \nu_1 \\ \dots \\ \nu_{k-1} \\ \nu_K \end{array}\right) = \left(\begin{array}{c} h_0 B^T \tilde{p}_1 \\ \dots \\ h_{k-1} B^T \tilde{p}_K \\ \hline -\tilde{p}_0 + \tilde{p}_1 + h_0 (\nabla_y f_0)^T \tilde{p}_1 + R^T q_1 \\ -\tilde{p}_1 + \tilde{p}_2 + h_1 (\nabla_y f_1)^T \tilde{p}_2 + R^T q_2 - R^T q_1 \\ \dots \\ -\tilde{p}_{k-1} + \tilde{p}_K + h_{k-1} (\nabla_y f_{k-1})^T \tilde{p}_K + R^T q_K - R^T q_{k-1} \\ -\tilde{p}_K - R^T q_K \\ \hline h_0 (\nabla_z f_0)^T \tilde{p}_1 + q_0 - q_1 \\ h_1 (\nabla_z f_1)^T \tilde{p}_2 + q_1 - q_2 - r_1 \\ \dots \\ h_{k-1} (\nabla_z f_{k-1})^T \tilde{p}_K + q_{k-1} - q_K - r_{k-1} \\ q_K - r_K \end{array}\right)$$

Now, observe that $\tilde{p}_0$ and q_0 appear both in one equation only. Due to this fact, it is not necessary to consider these two equations because a suitable choice of $\tilde{p}_0$ and q_0 can make these equations fulfilled under any circumstances. The lower two thirds can be rewritten in a compact form as

$$\begin{aligned}\mu_k &= -\tilde{p}_k + \tilde{p}_{k+1} + h_k(\nabla_y f_k)^T \tilde{p}_{k+1} + R^T q_{k+1} - R^T q_k \\ \nu_k &= h_k(\nabla_z f_k)^T \tilde{p}_{k+1} + q_k - q_{k+1} - r_k,\end{aligned}$$

where $k = 1, \ldots, K-1$ and the terminal conditions attain the form

$$\begin{aligned}\mu_K &= -\tilde{p}_K - R^T q_K \\ \nu_K &= q_K - r_K.\end{aligned}$$

To simplify the formulas, it remains to set $p = -\tilde{p}$ and formula (5.29) follows. □

In the next lemma we will show that the mapping Ξ has even the Aubin property around any feasible (u, y, z) and thus is also calm at that point.

Lemma 5.3.4. *Under assumptions (H1), (H2), (A2) and (A7) the mapping Ξ from Lemma 5.3.3 has the Aubin property at any feasible (u, y, z).*

Proof. The mapping S, defined in (5.25), is single–valued and locally Lipschitz from Lemma 5.3.1 and admits the representation (5.27). Assume for a while that the associated partially linearized mapping $\Delta : \mathbb{R}^{(K+1)n} \times \mathbb{R}^{(K+1)m} \times \mathbb{R}^{Km} \to \mathbb{R}^{(K+1)n} \times \mathbb{R}^{(K+1)m}$, defined by

$$\Delta(\eta) := \{(y, z) |\ \eta \in g(\bar{u}, \bar{y}, \bar{z}) + \nabla_y g(\bar{u}, \bar{y}, \bar{z})(y - \bar{y}) + \nabla_z g(\bar{u}, \bar{y}, \bar{z})(z - \bar{z}) + Q\}, \tag{5.36}$$

has the same properties. Since

$$D^*\Delta(0, \bar{y}, \bar{z})(0, 0) = \{\eta^* \in \mathrm{N}_Q(-g(\bar{u}, \bar{y}, \bar{z}))|\ \begin{pmatrix}\nabla_y g(\bar{u}, \bar{y}, \bar{z})^T \\ \nabla_z g(\bar{u}, \bar{y}, \bar{z})^T\end{pmatrix} \eta^* = 0\},$$

the Mordukhovich criterion [110, Theorem 9.40] would imply that

$$\mathrm{Ker} \begin{bmatrix}\nabla_y g(\bar{u}, \bar{y}, \bar{z})^T \\ \nabla_z g(\bar{u}, \bar{y}, \bar{z})^T\end{bmatrix} \cap \mathrm{N}_Q(-g(\bar{u}, \bar{y}, \bar{z})) = \{0\}.$$

The last condition ensures, however, directly the Aubin property of Ξ around $(0, \bar{u}, \bar{y}, \bar{z})$ and so, its calmness at this point. Since $(\bar{u}, \bar{y}, \bar{z})$ was an arbitrary point of $\mathrm{gph}\, S$, the statement holds provided we verify the required properties of Δ.

Setting $\eta = (\xi, \zeta, \chi)$, from (5.36) we obtain for $k = 0, \ldots, K-1$

$$\begin{aligned}\xi_0 &= -y_0 + a \\ \xi_{k+1} &= h_k B\bar{u}_k - y_{k+1} + y_k + h_k f_k + h_k \nabla_y f_k(y_k - \bar{y}_k) + h_k(\nabla_z f_k)(z_k - \bar{z}_k) \\ \zeta_0 &= z_0 - b \\ \zeta_{k+1} &= R(y_k - y_{k+1}) - z_k + z_{k+1} + \alpha_{k+1} \\ \chi_{k+1} &= -z_{k+1} + \beta_{k+1},\end{aligned}$$

where $\alpha_{k+1} \in \mathrm{N}_{Z_{k+1}}(\beta_{k+1})$.

Using again the equivalent representation of normal cone via the projection operator, this system can be rewritten as

$$\begin{aligned}
y_0 &= -\xi_0 + a \\
y_{k+1} &= h_k B\bar{u}_k + y_k + h_k f_k + h_k \nabla_y f_k(y_k - \bar{y}_k) + h_k(\nabla_z f_k)(z_k - \bar{z}_k) - \xi_{k+1} \\
z_0 &= \zeta_0 + b \\
z_{k+1} &= P_{Z_{k+1}}(z_k - R(y_k - y_{k+1}) + \zeta_{k+1} + \chi_{k+1}) - \chi_{k+1}
\end{aligned}$$

From this representation the single–valuedness of Δ follows immediately.

To prove the Lipschitz continuity of Δ, choose any η, $\tilde{\eta}$ and corresponding variables y, z, $\tilde{y}$, $\tilde{z}$ and compute

$$\begin{aligned}
y_0 - \tilde{y}_0 &= \tilde{\xi}_0 - \xi_0 \\
y_{k+1} - \tilde{y}_{k+1} &= (I + h_k \nabla_y f_k)(y_k - \tilde{y}_k) + h_k(\nabla_z f_k)(z_k - \tilde{z}_k) - \xi_{k+1} + \tilde{\xi}_{k+1} \\
z_0 - \tilde{z}_0 &= \zeta_0 - \tilde{\zeta}_0 \\
|z_{k+1} - \tilde{z}_{k+1}| &\le |z_k - \tilde{z}_k| + \|R\||y_k - \tilde{y}_k| + \|R\||y_{k+1} - \tilde{y}_{k+1}| \\
&\qquad + |\zeta_{k+1} - \tilde{\zeta}_{k+1}| + 2|\chi_{k+1} - \tilde{\chi}_{k+1}|.
\end{aligned}$$

Since $\nabla_y f_k$ and $\nabla_z f_k$ are fixed matrices, the Lipschitz continuity can be obtained in the same way as in the proof of Lemma 5.3.1. □

Under the surjectivity of ∇g, we obtain equality in the coderivative estimate from Lemma 5.3.3. As we will see later, this surjectivity is ensured if matrices B and R have full row ranks. Hence, a necessary condition for fulfillment of this requirement is that B and R do not have less columns than rows. This implies in particular that the dimension of the control variable u must be greater than or equal to the dimensions of both the state variables y and z.

Lemma 5.3.5. *If matrices B and R have full row ranks, then the statement of Lemma 5.3.3 holds true with equality in the sense that if conditions* (5.29)–(5.32) *are fulfilled, then* (5.28) *holds true as well.*

Proof. If ∇g is surjective, we can use [110, Theorem 6.7] to obtain

$$\mathrm{N}_{\mathrm{gph}\, S}(\bar{u}, \bar{y}, \bar{z}) = -(\nabla g(\bar{u}, \bar{y}, \bar{z}))^T \, \mathrm{N}_Q(g(\bar{u}, \bar{y}, \bar{z}))$$

instead of (5.33). So, it remains to verify that the matrix ∇g computed in the proof of Lemma 5.3.3 has full row rank. By removing rows with only one nonzero cell occupied by an identity matrix and the corresponding columns, it can be easily shown that the full row rank requirement on ∇g is equivalent to the following matrix having full row rank:

$$\left(\begin{array}{cccc|cccc}
h_0 B & 0 & \cdots & 0 & -I & 0 & \cdots & 0 \\
0 & h_1 B & \cdots & 0 & I + h_1 \nabla_y f_1 & -I & \ddots & \vdots \\
0 & \vdots & \ddots & \vdots & \ddots & \ddots & \ddots & 0 \\
0 & 0 & \cdots & h_{k-1} B & \cdots & 0 & I + h_{k-1} \nabla_y f_{k-1} & -I \\
\hline
0 & 0 & \cdots & 0 & -R & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & R & -R & \ddots & 0 \\
\vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & R & -R
\end{array}\right)$$

However, the full rank of this matrix is equivalent to the full rank of matrices B and R. □

Remark 5.3.6. In Lemma 5.3.3 we have worked with the limiting normal cone. It is not difficult to show similar results for the Fréchet normal cone. As in the proof of Lemma 5.3.5, the only change would concern inclusion (5.33). This change would be based on [110, Theorem 6.14] which gives the opposite estimate

$$\hat{\mathrm{N}}_{\mathrm{gph}\, S}(\bar{u},\bar{y},\bar{z}) \supset -(\nabla g(\bar{u},\bar{y},\bar{z}))^T \hat{\mathrm{N}}_Q(-g(\bar{u},\bar{y},\bar{z})).$$

In the lemma statement, two changes would occur. Firstly, the upper estimate would be replaced by the lower one and secondly, relation (5.34) would become

$$\begin{pmatrix} r_k \\ q_k \end{pmatrix} \in \hat{\mathrm{N}}_{\mathrm{gph}\, \mathrm{N}_{Z_k}} \begin{pmatrix} \bar{z}_k \\ -\bar{z}_k + \bar{z}_{k-1} + R\bar{y}_k - R\bar{y}_{k-1} \end{pmatrix}. \tag{5.37}$$

However, it may happen that adjoint system (5.30) has no solution if coupled with (5.37) instead of (5.32).

After the preparatory work, we can now provide a workable description of $\partial(\tilde{L} \circ S)$ in the following form.

Theorem 5.3.7. *Let $(\bar{y},\bar{z}) = S(\bar{u})$ and let assumptions $(H1)$, $(H2)$, $(A2)$, $(A7)$ and $(A8)$ be fulfilled. Then for any element*

$$u^* = \begin{pmatrix} u_0^* \\ \dots \\ u_{k-1}^* \end{pmatrix} \in \partial(\tilde{L} \circ S)(\bar{u}_0, \dots, \bar{u}_{k-1}) \tag{5.38}$$

one has the representation (5.29) *such that adjoint equations* (5.30) *with terminal conditions* (5.31) *and multipliers* (5.32) *are satisfied. Moreover, for the additional multipliers the following conditions are satisfied as well*

$$\begin{pmatrix} \mu_k \\ \nu_k \end{pmatrix} \in h_k \partial L_1(t_k, \bar{y}_k, \bar{z}_k), \ k = 1, \dots, K-1 \tag{5.39}$$

$$\begin{pmatrix} \mu_K \\ \nu_K \end{pmatrix} \in \partial L_3(\bar{y}_K, \bar{z}_K). \tag{5.40}$$

Conversely, if relations (5.29), (5.30), (5.31), (5.32), (5.39) *and* (5.40) *hold true and the set* $\mathrm{gph}\, \mathrm{N}_{Z_k}$ *is regular at*

$$\begin{pmatrix} \bar{z}_k \\ -\bar{z}_k + \bar{z}_{k-1} + R\bar{y}_k - R\bar{y}_{k-1} \end{pmatrix},$$

$L_1(t_k,\cdot,\cdot)$ *is regular at* $(\bar{y}_k,\bar{z}_k)$ *for all* $k = 1,\dots,K-1$ *and* L_3 *is regular at* $(\bar{y}_K,\bar{z}_K)$*, then we obtain* (5.38).

Similarly, under assumption $(A8')$ *one obtains*

$$u^* \in \overline{\partial}(\tilde{L} \circ S)(\bar{u}_0, \dots, \bar{u}_{k-1}) \tag{5.41}$$

provided relations (5.29), (5.30), (5.31), (5.32), (5.39) *and* (5.40) *hold true and matrices B and R have full row rank. For this result, no additional regularity requirements are needed.*

Proof. From Lemma 5.3.1 it follows that S is single–valued and Lipschitz. As the Lipschitz continuity of $L_1(t_k, \cdot, \cdot)$ implies that $\partial^\infty L_1(t_k, \bar{y}_k, \bar{z}_k) = \{0\}$, the assumptions of [110, Theorem 10.49] are satisfied, which implies

$$\partial(\tilde{L} \circ S)(\bar{u}) \subset D^* S(\bar{u})[\partial \tilde{L}(S(\bar{u}))].$$

Due to $\tilde{L}(y, z) = \sum_{k=0}^{K-1} h_k L_1(t_k, y_k, z_k) + L_3(y_K, z_K)$, the relation

$$\begin{pmatrix} \mu \\ \nu \end{pmatrix} \in \partial \tilde{L}(\bar{y}, \bar{z})$$

is equivalent with (5.39) and (5.40). The upper estimate of the coderivative is computed in Lemma 5.3.3, which gives the first statement of the theorem.

For the second part of the theorem statement, we find upper and lower estimates which coincide. The upper estimate has been already computed in the first part. For the lower one, we obtain

$$\hat{\partial}(\tilde{L} \circ S)(\bar{u}) \supset \hat{D}^* S(\bar{u})[\hat{\partial} \tilde{L}(S(\bar{u}))].$$

According to Remark 5.3.6 and the comparison of both estimates, we conclude that the equality is ensured by the regularity of set $\operatorname{gph} \mathrm{N}_{Z_k}$ and functions $L_1(t_k, \cdot, \cdot)$ and $L_3(\cdot, \cdot)$ at the points in question.

For the final part of the theorem, note that

$$\overline{\partial}(\tilde{L} \circ S)(\bar{u}) = \overline{\partial} S(\bar{u})^T \nabla \tilde{L}(S(\bar{u})) = \operatorname{co} D^* S(\bar{u})(\nabla \tilde{L}(S(\bar{u}))) \supset D^* S(\bar{u})(\nabla \tilde{L}(S(\bar{u}))),$$

where the first equality comes from [27, Theorem 2.6.6] and the second one from [86, Proposition 2.11]. As the full row ranks of B and R imply by virtue of Lemma 5.3.5 that we are able to compute $D^* S$ exactly, the proof has been finished. □

When solving numerically problem (5.4), or equivalently (5.26), we have to be able to compute the function value and one subgradient from $\overline{\partial}(\tilde{L} \circ C)$. The function value can easily be computed from (5.25) and for the computation of a subgradient, Theorem 5.3.7 can be used. We are able to compute such subgradient under full row ranks of matrices B and R or under the regularity assumptions from the previous theorem. We are fully aware that these regularity assumptions cannot be simply checked on the basis of given data. However, computational experience indicates that points evaluated during the solution process of (5.26) are almost exclusively regular and nonregularity may typically occur only in several last iterations without disturbing the convergence.

In the rest of the chapter, we will address two issues. The first one is the choice of r_k and q_k which would satisfy both (5.32) and (5.30), the other one are more precise conditions under which the regularity conditions needed in the second part of Theorem 5.3.7 are satisfied.

If Z_k is a polyhedral set, then from Proposititon 3.3.2 we have the following characterization of Fréchet and limiting normal cones to the graph of the normal cone mapping:

$$\begin{aligned} \hat{\mathrm{N}}_{\operatorname{gph} \mathrm{N}_{Z_k}}(\bar{z}_k, \bar{v}_k) &= K(\bar{z}_k, \bar{v}_k)^* \times K(\bar{z}_k, \bar{v}_k) \\ \mathrm{N}_{\operatorname{gph} \mathrm{N}_{Z_k}}(\bar{z}_k, \bar{v}_k) &= \bigcup_{(z_k, v_k) \in U \cap \operatorname{gph} \mathrm{N}_{Z_k}} K(z_k, v_k)^* \times K(z_k, v_k), \end{aligned} \tag{5.42}$$

where U is some sufficiently small neighborhood of (z_k, v_k) and $K(z_k, v_k)$ is the critical cone to Z_k at z_k with respect to v_k as defined in (3.15). To be able to exploit this result, we will assume that Z_k are polyhedral sets.

The next lemma provides a basis, on which an algorithm for choosing the variables r_k and q_k will be derived. Moreover, it suggests a case in which the graph of the normal cone mapping is regular. One part of the proof is a specific case of [109, Theorem 3.5]. However, to use this part it would be necessary to define new objects and thus, we decided to provide a more technical but self–contained proof. Note that the last part can be easily deduced from the concept of normally admissible stratification from Subsection 3.3.1.

Lemma 5.3.8. *Assume that Z_k is a polyhedral set and fix any $\bar{v}_k \in \operatorname{rint} \mathrm{N}_{Z_k}(\bar{z}_k)$. Then the critical cone $K(\bar{z}_k, \bar{v}_k)$ is a linear subspace and $\operatorname{gph} \mathrm{N}_{Z_k}$ is regular at $(\bar{z}_k, \bar{v}_k)$. Moreover, the cones $K(\bar{z}_k, v_k)$ coincide for all $v_k \in \operatorname{rint} \mathrm{N}_{Z_k}(\bar{z}_k)$.*

Proof. Since the critical cone is an intersection of a convex cone and a linear subspace, it is also a convex cone. To prove that $K(\bar{z}_k, \bar{v}_k)$ is a linear subspace, it is sufficient to show that if $s \in K(\bar{z}_k, \bar{v}_k)$, then $-s \in K(\bar{z}_k, \bar{v}_k)$, or equivalently $-s \in \mathrm{T}_{Z_k}(\bar{z}_k)$. Because Z_k is convex, we have full duality between the normal and tangent cone, hence not only the standard relation $\mathrm{N}_{Z_k}(\bar{z}_k) = \mathrm{T}^*_{Z_k}(\bar{z}_k)$ but also the relationship $\mathrm{N}^*_{Z_k}(\bar{z}_k) = \mathrm{T}_{Z_k}(\bar{z}_k)$. For contradiction assume that $-s \notin \mathrm{T}_{Z_k}(\bar{z}_k)$, which implies that there exists some $r \in \mathrm{N}_{Z_k}(\bar{z}_k)$ such that $\langle -s, r\rangle > 0$.

As $s \in K(\bar{z}_k, \bar{v}_k)$, from the definition of the critical cone we know that $\langle s, \bar{v}_k\rangle = 0$. Define $\tilde{s} := \bar{v}_k + \lambda(\bar{v}_k - r)$ with a positive λ. Since $\bar{v}_k \in \operatorname{rint} \mathrm{N}_{Z_k}(\bar{z}_k)$ and $r \in \mathrm{N}_{Z_k}(\bar{z}_k)$, for small $\lambda > 0$ we have $\tilde{s} \in \mathrm{N}_{Z_k}(\bar{z}_k)$ and

$$\langle s, \tilde{s}\rangle = \langle s, \bar{v}_k + \lambda(\bar{v}_k - r)\rangle = 0 + \lambda\langle s, -r\rangle > 0,$$

which is a contradiction with $\tilde{s} \in \mathrm{N}_{Z_k}(\bar{z}_k)$ because $s \in \mathrm{T}_{Z_k}(\bar{z}_k)$.

To check the regularity of $\operatorname{gph} \mathrm{N}_{Z_k}$, we need to prove that

$$\hat{\mathrm{N}}_{\operatorname{gph} \mathrm{N}_{Z_k}}(\bar{z}_k, \bar{v}_k) \supset \mathrm{N}_{\operatorname{gph} \mathrm{N}_{Z_k}}(\bar{z}_k, \bar{v}_k),$$

which is by virtue of (5.42) equivalent to

$$K(\bar{z}_k, \bar{v}_k)^* \times K(\bar{z}_k, \bar{v}_k) \supset \bigcup_{(z_k, v_k) \in U \cap \operatorname{gph} \mathrm{N}_{Z_k}} K(z_k, v_k)^* \times K(z_k, v_k), \tag{5.43}$$

where U is some small neighborhood of $(\bar{z}_k, \bar{v}_k)$.

We make use of the concept of a normal fan $\mathcal{N}$ to a set A, which is defined as a collection of all normal cones $\mathrm{N}_A(x)$ over $x \in A$. By [79, Corollary 1] we know that if A is polyhedral, then for any two normal cones $N_1, N_2 \in \mathcal{N}$ one has

$$(\operatorname{rint} N_1) \cap N_2 \neq \emptyset \implies N_1 \subset N_2. \tag{5.44}$$

Take any $z_k \in Z_k$ sufficiently close to $\bar{z}_k$. Choosing $N_1 = \mathrm{N}_{Z_k}(\bar{z}_k)$ and $N_2 = \mathrm{N}_{Z_k}(z_k)$ in (5.44), we obtain either

$$(\operatorname{rint} \mathrm{N}_{Z_k}(\bar{z}_k)) \cap \mathrm{N}_{Z_k}(z_k) = \emptyset \tag{5.45}$$

or $\mathrm{N}_{Z_k}(\bar{z}_k) \subset \mathrm{N}_{Z_k}(z_k)$, which, due to the polyhedrality of Z_k, implies that

$$\mathrm{N}_{Z_k}(\bar{z}_k) = \mathrm{N}_{Z_k}(z_k). \tag{5.46}$$

Because $v_k \to \bar{v}_k \in \operatorname{rint} \mathrm{N}_{Z_k}(\bar{z}_k)$, the case (5.45) does not need to be considered in the computation of right–hand side of (5.43). Then (5.46) implies that in the union on the right–hand side of (5.43) it suffices to set $z_k = \bar{z}_k$. To finish the proof, it remains to prove that the critical cone does not depend on the choice of v_k. Choose any $v_k, \tilde{v}_k \in \operatorname{rint} \mathrm{N}_{Z_k}(\bar{z}_k)$. We need to prove that if $s \in K(\bar{z}_k, v_k)$, then $s \in K(\bar{z}_k, \tilde{v}_k)$ or equivalently $\langle s, \tilde{v}_k \rangle = 0$. As $K(\bar{z}_k, v_k)$ is a linear subspace by the previous statement, we also have $-s \in K(\bar{z}_k, v_k) \subset \mathrm{T}_{Z_k}(\bar{z}_k)$, which implies that $\langle -s, \tilde{v}_k \rangle \leq 0$. This, together with $\langle s, \tilde{v}_k \rangle \leq 0$, implies the desired equality $\langle s, \tilde{v}_k \rangle = 0$. □

The previous lemma tells us that for any $\bar{v}_k \in \operatorname{rint} \mathrm{N}_{Z_k}(\bar{z}_k)$ we have

$$\mathbb{R}^m = K(\bar{z}_k, \bar{v}_k)^* \oplus K(\bar{z}_k, \bar{v}_k).$$

This allows us to solve find q_k and r_k which satisfy conditions (5.32), (5.30b) and (5.31b). Indeed, the adjoint and terminal equations may be written in a compact form as

$$q_k - r_k = s_k \tag{5.47}$$

for some s_k. Due to Lemma 5.3.8 we can choose exactly one $q_k \in K(\bar{z}_k, \bar{v}_k)$ and exactly one $-r_k \in -K(\bar{z}_k, \bar{v}_k)^* = K(\bar{z}_k, \bar{v}_k)^*$ such that (5.32) and (5.47) are satisfied.

We may summarize the algorithm for computing an element of the upper estimate of the subdifferential of $\partial(\tilde{L} \circ S)$ as follows:

1. Given $\bar{u}$, compute the values of $\bar{y}$ and $\bar{z}$. Set $s_K = \nu_K$, $k = K$, find any element
$$\begin{pmatrix} \mu_K \\ \nu_K \end{pmatrix} \in \partial L_3(\bar{y}_K, \bar{z}_K)$$
and continue to step 2.

2. If $\bar{z}_k \in \operatorname{int} Z_k$, then set $q_k = s_k$. Otherwise find any $\bar{v}_k \in \operatorname{rint} \mathrm{N}_{Z_k}(\bar{z}_k)$, compute $K(\bar{z}_k, \bar{v}_k)$ and set q_k to be the projection of s_k onto $K(\bar{z}_k, \bar{v}_k)$. In both cases proceed to step 3.

3. Compute
$$p_k = p_{k+1} + h_k(\nabla_y f_k)^T p_{k+1} - R^T q_{k+1} + R^T q_k + \mu_k$$
with the convention that $p_{k+1} = q_{k+1} = 0$. If $k = 1$, go to step 5, otherwise decrease k by one and proceed to step 4.

4. Find any element
$$\begin{pmatrix} \mu_k \\ \nu_k \end{pmatrix} \in h_k \partial L_1(t_k, \bar{y}_k, \bar{z}_k),$$
set
$$s_k = q_{k+1} + h_k(\nabla_z f_k)^T p_{k+1} + \nu_k$$
and return to step 2.

5. Having $p_1, \ldots, p_K$, compute the estimate of a subgradient in the form

$$\begin{pmatrix} h_0 B^T p_1 \\ \ldots \\ h_{k-1} B^T p_K \end{pmatrix}.$$

On the basis of Theorem 5.3.7 and the discussion onward we can now provide a final statement about the computation of subgradients from $\partial(\tilde{L} \circ S)$. Since the proof has been basically completed in the previous text, we omit it.

Theorem 5.3.9. *Let assumptions $(H1)$, $(H2)$, $(A2)$, $(A7)$ and $(A8)$ be fulfilled and assume that all the sets Z_k are polyhedral. Then the previous algorithm finds an element of the upper estimate of the subdifferential $\partial(\tilde{L} \circ S)(\bar{u})$.*

Moreover, if for all $k = 1, \ldots, K-1$ we have

$$-\bar{z}_k + \bar{z}_{k-1} + R\bar{y}_k - R\bar{y}_{k-1} \in \operatorname{rint} \mathrm{N}_{Z_k}(\bar{z}_k),$$

functions $L_1(t_k, \cdot, \cdot)$ are regular at $(\bar{y}_k, \bar{z}_k)$ and $L_3(\cdot, \cdot)$ is regular at $(\bar{y}_K, \bar{z}_K)$, then the computed point is actually an element of the subdifferential $\partial(\tilde{L} \circ S)(\bar{u})$.

Similarly, if $(A8')$ is fulfilled and if matrices B and R have full row rank, then the computed point is an element of $\overline{\partial}(\tilde{L} \circ S)(\bar{u})$ without the regularity requirements from the previous part.

5.4 Numerical examples

We have already mentioned in the Introduction that most applications of the investigated model lie in various fields of physics. However, in this section we investigate a different type of application, namely optimization of the efficiency of a service point handling a queue. A similar model was studied in [67] where, however, no optimization was considered.

To solve the respective problem (5.26), we used two algorithms, both employing the subgradient information computed in Theorem 5.3.7. The first one was BT (Bundle Trust) algorithm proposed in [123], the other one a version of BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm. The BT algorithm was designed especially for nonsmooth optimization while for BFGS we have used its modification from [74] which, under a weakening of the Wolfe condition, works well even for larger nonsmooth optimization problems. While for small problems, BT exhibited slightly better results, it often collapsed for larger problems and then BFGS has been used. Furthermore, since BT solves a quadratic optimization problem in every iteration, its time consumption was higher. Due to these reasons, we present here only the results computed by the BFGS algorithm. The implementation of BFGS was taken from Master thesis [125].

Example 5.4.1. Let us consider a service point handling a deterministic queue. Assuming that we know the customer inflow, we find an optimal rate, at which the service point should operate. Both the arrival and service rates depend on time and are considered as processes with continuous values. This means that even though it is possible to use this model for people behavior and round off the optimal values, we prefer to think about it as a model for product processing.

Since agreements for such processing are often negotiated beforehand, even the deterministic inflow and outflow make sense in this case. This model can be also applied to control a water reservoir where the water inflow and outflow form a process with continuous values as well.

We denote the known number of customers who entered the system prior to time t by $v(t)$, the rate at which the service point operates by $u(t)$ and finally the queue length by $z(t)$. We assume that customers may leave the queue, with the rate of leaving equal to $g(z(t))$ depending purely on the length of the queue. Further we assume that the maximum waiting capacity of the service point is unbounded, which leads to the definition $Z := [0, \infty)$.

Under the above assumptions, the service point operates according to the model

$$-\dot{z}(t) + \dot{v}(t) - g(z(t)) - u(t) \in \mathrm{N}_Z(z(t)).$$

Due to the definition of the normal cone mapping, the condition $z(t) \in Z$ is always satisfied. Moreover, if $z(t) > 0$, then the change in the queue length corresponds to the difference between the arrival and departure rates, hence

$$\dot{z}(t) = \dot{v}(t) - g(z(t)) - u(t). \tag{5.48}$$

However, if the queue is empty, or equivalently $z(t) = 0$, then instead of (5.48) we obtain

$$\dot{z}(t) \geq \dot{v}(t) - g(z(t)) - u(t). \tag{5.49}$$

If $\dot{v}(t) - g(z(t)) - u(t) < 0$, then due to $z(t) \geq 0$ we obtain $\dot{z}(t) \geq 0$, which indicates that some of the service point working capacity is idle.

As the penalization of large queues we have chosen the function

$$g(z) := \begin{cases} 0 & \text{if } z \in [0, 5] \\ (z-5)^2 & \text{if } z > 5. \end{cases}$$

This indicates that if there are more than five customers in the queue, some of them start leaving.

Setting $\dot{y}(t) = \dot{v}(t) - g(z(t)) - u(t)$, we obtain the constraint system from problem (5.3) with $n = m = d = 1$, $R = 1$ and $B = -1$. The objective function is parameterized by $\lambda_1, \ldots, \lambda_4$ and equals to

$$\int_0^T [L_1(z(t)) + L_2(u(t))]dt + \lambda_3 z(T)^2$$

with

$$L_1(z) := \lambda_1 z^2 + \lambda_2[(z-5)^+]^2$$

and

$$L_2(u) := \begin{cases} u & \text{if } u \in [0, 30] \\ \lambda_4 u^2 + (1 - 60\lambda_4)u + 900\lambda_4 & \text{if } u > 30. \end{cases}$$

Function L_1 penalizes the number of customers which are in the queue or have left it prematurely. The definition of L_2 models the situation where only a certain service rate is reachable by simple means and to exceed this rate it is necessary to pay additional costs. This rate was chosen as 30, which is slightly lower than the average customer arrival rate. The initial conditions naturally read $y(0) = z(0) = 0$ and the time interval was chosen $[0, 10]$.

We decided not to penalize short queues and to set $\lambda_1 = 0.01$. It is not equal to zero, if it was, multiple global optima would arise. To ensure that the queue is cleared at the end of the time interval, we chose $\lambda_3 = 1000$. Fixing the previous two parameters, the problem can be understood as a multiobjective problem where one minimizes a tradeoff between the running costs and the queue length. The last two parameters were set to $\lambda_2 = 50$ and $\lambda_4 = 0.2$.

Figure 5.1 summarizes the results. In the graph on the left–hand side, the full line represents the arrival rate $\dot{v}(t)$ and the dotted line represents the optimal service rate $u(t)$. For the sake of completeness, we write the formula for the arrival rate

$$\dot{v}(t) := \begin{cases} 10\,\mathrm{arctg}(t-1.5)+30 & \text{if } u \in [0,3] \\ 10\,\mathrm{arctg}(1.5)+30-(t-3)^2 & \text{if } t \in [3, \sqrt{10\,\mathrm{arctg}(1.5)+2}+3] \\ 28 & \text{if } t \in [\sqrt{10\,\mathrm{arctg}(1.5)+2}+3, 10]. \end{cases}$$

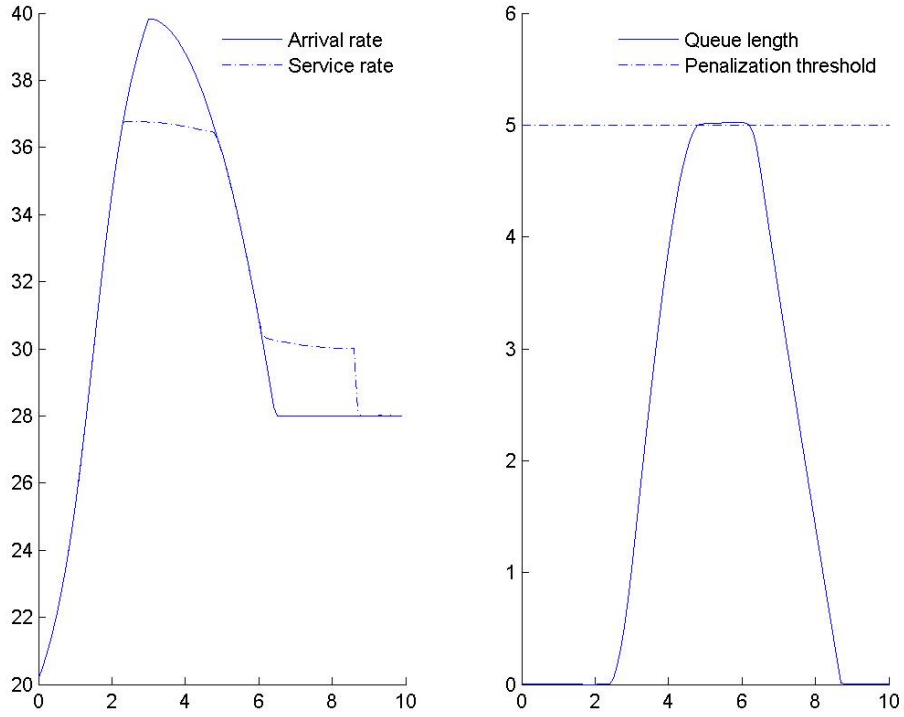


Figure 5.1: The arrival and service rates and the development of the queue length over time.

We see that during customer arrival peak, the optimal service rate does not handle all arrivals and the queue begins to grow. As soon as the arrival rate decreases below 30, which is the maximum rate at which the service rate is not further penalized, the service rate will be higher than the arrival rate and the queue starts to disband until it is disbanded completely.

In the graph on the right–hand side of Figure 5.1 the development of the queue length is shown. We see that its length keeps below or slightly over the penalization threshold, which is in accordance with expectations.

Remark 5.4.2. When testing the convergence, we have created numerous academic test problems. In some of them, we achieved a good convergence, some examples caused minor faults and some even big difficulties or kdowns during the optimization. In this short remark we provide several possible reasons for this unpleasant behavior:

1. Rounding errors resulting in a wrong subgradient. Even though S is Lipschitz, the algorithm presented after Lemma 5.3.8 still heavily depends on the non–Lipschitzian behavior of the normal cone mapping. If z_{k+1} is close to the boundary of Z_{k+1}, then the algorithm may fail to compute q_k in a correct way. Moreover, if $z_k + Ry_{k+1} - Ry_k$ is close to zero, then the used solution method may lie in a nonregular point and either the computed subgradient does not have to be in the Clarke subdifferential or, if it is, its opposite direction does not have to be a descent direction as described in [17].

2. Error accumulation due to the iterative computing in time–dependent problems.

3. Ill–conditioned approximation of the inverted Hessian matrix. During the optimization we encountered cases, in which the condition number of the approximation was of order -13. As most of the scripting languages are able to work only with 16 digits on default, there was a very small margin for error.

4. Ill–posedness of the data which often happens when the penalization of the control variable is too low compared to the penalization of the state variables. This occurs especially when the time mesh is too fine because, in this case, big changes of the control variable cause only small changes of the state variables.

6. Parameter identification in delamination model

6.1 Introduction

Many evolution systems have the structure of the *generalized gradient flow*

$$\dot{q} \in \partial_\xi \mathscr{R}^*(q; -\partial_q \mathscr{E}(t,q))$$

with functionals $\mathscr{E}(t,q)$ and $\mathscr{R}^*(q;\xi)$. Here q is an abstract state of the system and $\dot{q}$ denotes its time derivative. Quite typically, $\mathscr{R}^*(q;\cdot)$ is convex and, making the conjugate of $\mathscr{R}^*(q;\xi)$ with respect to the "driving force" variable ξ, i.e. $\mathscr{R}(q;v) = \sup_\xi[\langle v,\xi\rangle - \mathscr{R}^*(q;\xi)]$, the generalized gradient flow can equivalently be written in the Biot-equation form

$$\partial_{\dot{q}} \mathscr{R}(q;\dot{q}) + \partial_q \mathscr{E}(t,q) \ni 0. \tag{6.1}$$

In many cases, the problem is nonsmooth due to a nonsmoothness of $\mathscr{R}(q;\cdot)$ or $\mathscr{E}(t,q)$, which is why we wrote inclusion in (6.1) rather than equality. Ansatz (6.1) is very general and covers great variety of problems in particular in nonsmooth continuum mechanics. The state variable q may involve displacements and various internal parameters (but also various concentrations of some constituents subjected to diffusive processes). In this chapter, we focus on a subclass of such problems where the state has the structure

$$q = (u,z) \tag{6.2}$$

for each time instance t in a Banach space $U \times Z$. In this way, a quasistatic *plasticity*, or *damage*, or various *phase transformations* at small strains can be modelled, and also various problems in *contact mechanics* like friction or adhesion, together with various combinations of these phenomena.

After a suitable time discretization, (6.1) gives rise to recursive optimization problems. Often (or, in applications in continuum mechanics we have in mind, rather typically) q ranges over an infinite-dimensional Banach space and, after a possible "spatial" discretization, these minimization problems have a structure of strictly convex *Quadratic Programming* problems. It is then relatively easy to use such a discretized evolution problem as a governing system for some optimization problem, e.g., optimal control or identification of parameters. This leads to *Mathematical Programs with Evolution Equilibrium Constraints* (MPEEC) which have been studied e.g. in [65, 66, 92].

The functionals in (6.1) depend also on an abstract parameter π and have a special form

$$\begin{aligned}\mathscr{R}(\pi,q;\dot{q}) &= \mathscr{R}_1(\pi,u,z;\dot{u}) + \mathscr{R}_2(\pi,u,z;\dot{z}),\\ \mathscr{E}(t,\pi,q) &= \mathscr{E}(t,\pi,u,z).\end{aligned}$$

We then consider an optimal-control or an identification problem on a fixed time

interval $[0,T]$:

$$\begin{aligned}
&\text{minimize } \int_0^T j(u,z)\,\mathrm{d}t + H(\pi)\\
&\text{subject to } \partial_{\dot u}\mathscr{R}_1(\pi,u,z;\dot u) + \partial_u\mathscr{E}(t,\pi,u,z) \ni 0, \quad t\in[0,T] \text{ a.e., } u(0)=u_0,\\
&\qquad\qquad \partial_{\dot z}\mathscr{R}_2(\pi,u,z;\dot z) + \partial_z\mathscr{E}(t,\pi,u,z) \ni 0, \quad t\in[0,T] \text{ a.e., } z(0)=z_0,\\
&\qquad\qquad u\in L^\infty(0,T;U), \quad z\in L^\infty(0,T;Z), \quad \pi\in\Pi
\end{aligned} \tag{6.3}$$

with some $j : U\times Z\to\mathbb{R}$ and $H:\Pi\to\mathbb{R}$ specified later; here Π is a closed convex set of a Banach space where π lives. In some models, the flow rule for u in (6.3) is purely static, i.e. $\mathscr{R}_1 = 0$. In this case, if there is no dissipation in this part, then only z_0 but not u_0 is decisive when considering $\partial_u\mathscr{E}(t,\pi,u_0,z_0)\ni 0$. We use the standard notation for Bochner space $L^\infty(0,T;\cdot)$ of Banach-space-valued Bochner measurable functions on $[0,T]$.

The main aim of the chapter is a deep analysis of a discretized version of MPEEC (6.3) leading both to sharp necessary optimality conditions as well as to an efficient numerical procedure based on the so-called *implicit Programming* approach (ImP), cf. [80, 94]. In particular, on the basis of the subdifferential calculus of B. Mordukhovich [87, 110] we will show that the *solution map* S : $\pi\mapsto(u,z)$ defined via the discretized equilibrium relations in (6.3), is single-valued and locally Lipschitz and satisfies henceforth the basic ImP hypothesis. In the respective proof one has to deal with a difficult multifunction arising in connection with our *evolving* constraint sets. The application of standard tools of generalized differential calculus provides us in this case only with an upper estimate of the coderivative of the normal cone mapping to the *overall* constraint set, which could be a substantial drawback both in the optimality conditions as well as in the used numerical approach. To overcome this hurdle, we have employed the results from Chapter 3 which enables us to compute the limiting coderivative of the mentioned normal cone mapping *exactly*.

The plan of the chapter is the following. In Section 6.2, we briefly introduce a suitable discretization of the identification problem (6.3) that yields a unique response (u,z) of the constraint system of (6.3) for a given π and allows for efficient optimization technique. In Section 6.3 we formulate first-order necessary optimality condition for the discrete version of MPEEC (6.3) and derive a subgradient formula for the composite objective function of the discretized problem. In Section 6.4, we formulate a specific application-motivated identification problem from contact continuum mechanics that fits (and illustrates) the system (6.3). Eventually, in Section 6.5, we illustrate the usage of the subgradient evaluation procedure via an adhesive contact problem in a nontrivial two-dimensional example involving a spatial discretization by *Finite-Element-Method* (FEM).

6.2 Discretization of the identification problem

The natural procedure is to discretize the problem (6.3) in time. This might be a rather delicate problem especially when the controlled system (6.3b,c) exhibits responses of different time scales, and in particular with tendencies for jumping, which quite typically happens in rate-independent systems governed by noncon-

vex potentials $\mathscr{E}(t,\pi,\cdot,\cdot)$.

We consider (for simplicity) an equidistant partition of the time interval $[0,T]$ with a time step $\tau>0$ such that $T/\tau =: K \in \mathbb{N}$ and then a fractional-step-type semi-implicit time discretization of (6.3b,c). Moreover, if U, Z or Π is infinite-dimensional, the time-discrete problem still remains infinite-dimensional, and to implement it on computers, we need to apply also an abstract space discretization controlled by a parameter, let us denote it by $h>0$. Such an approximation can be considered by replacing U, Z, and Π in (6.4) by their finite-dimensional subsets U_h, Z_h, and Π_h. Counting also a possible numerical approximation of $\mathscr{E}$, denoted by $\mathscr{E}_h$, altogether, (6.3) turns into the problem

$$\begin{aligned}
&\text{minimize } \tau\sum_{k=1}^{K} j(u_{\tau h}^k, z_{\tau h}^k) + H(\pi)\\
&\text{subject to } \partial_{\dot u}\mathscr{R}_1\Big(\pi, u_{\tau h}^{k-1}, z_{\tau h}^{k-1}; \frac{u_{\tau h}^k - u_{\tau h}^{k-1}}{\tau}\Big) + \partial_u \mathscr{E}_h(k\tau,\pi,u_{\tau h}^k, z_{\tau h}^{k-1}) \ni 0, \quad u_{\tau h}^0 = u_{0h},\\
&\qquad\qquad \partial_{\dot z}\mathscr{R}_2\Big(\pi, u_{\tau h}^{k-1}, z_{\tau h}^{k-1}; \frac{z_{\tau h}^k - z_{\tau h}^{k-1}}{\tau}\Big) + \partial_z \mathscr{E}_h(k\tau,\pi,u_{\tau h}^k, z_{\tau h}^{k}) \ni 0, \qquad z_{\tau h}^0 = z_{0h},\\
&\qquad\qquad u_{\tau h}^k \in U_h \ \text{ and } \ z_{\tau h}^k \in Z_h \ \text{ for } k=1,...,K, \quad \pi\in\Pi_h,
\end{aligned} \tag{6.4}$$

where $(u_{0h}, z_{0h}) \in U_h \times Z_h$ is an approximation of the initial condition (u_0, z_0). Let us note that the controlled system (6.4b,c) decouples so that, for a given π, one is to solve alternating optimization problems

$$\text{minimize } \tau\mathscr{R}_1\Big(\pi, u_{\tau h}^{k-1}, z_{\tau h}^{k-1}; \frac{u - u_{\tau h}^{k-1}}{\tau}\Big) + \mathscr{E}_h(k\tau,\pi,u,z_{\tau h}^{k-1}) \quad \text{subject to } u\in U_h \tag{6.5a}$$

and, taking (one of) its solution for $u_{\tau h}^k$, further

$$\text{minimize } \tau\mathscr{R}_2\Big(\pi, u_{\tau h}^{k-1}, z_{\tau h}^{k-1}; \frac{z - z_{\tau h}^{k-1}}{\tau}\Big) + \mathscr{E}_h(k\tau,\pi,u_{\tau h}^k,z) \qquad \text{subject to } z\in Z_h \tag{6.5b}$$

which yields $z_{\tau h}^k$ as (one of) its solution. Assuming $\mathscr{E}(t,\pi,\cdot,\cdot)$ as well as its approximation $\mathscr{E}_h(t,\pi,\cdot,\cdot)$ separately strictly convex (and, of course, coercive with compact level sets) and $\mathscr{R}_i(\pi,u,z;\cdot)$ convex, $i=1,2$, both problems in (6.5) have unique solutions $u_{\tau h}^k$ and $z_{\tau h}^k$, respectively, and thus the whole recursive problem in the constraint system (6.4) has a unique response for a given π as well. This allows us to reformulate (6.4) as a minimization problem for a functional depending on π only, cf. (6.9) below. This will be exactly the situation we will consider in the rest of this text. The fully discretized system (6.4) can thus be understood as an MPEC for which a developed theory exists.

In what follows, we will confine ourselves to problems with a bit more detailed (but nevertheless still fairly general) structure, namely

$$\mathscr{E}(t,\pi,u,z) = \begin{cases} \mathcal{E}(t,\pi,u,z) & \text{if } u\in\Lambda_0^t,\ z\in K_0^t,\\ \infty & \text{otherwise,}\end{cases} \tag{6.6a}$$

$$\mathscr{R}_1(\pi,u,z,\dot u) = \mathcal{R}_1(\pi,u,z,\dot u), \tag{6.6b}$$

$$\mathscr{R}_2(\pi,u,z,\dot z) = \begin{cases} \mathcal{R}_2(\pi,u,z,\dot z) & \text{if } \dot z\in K_1,\\ \infty & \text{otherwise,}\end{cases} \tag{6.6c}$$

where $\mathcal{E}$, $\mathcal{R}_1$, and $\mathcal{R}_2$ are finite and smooth, Λ_0^t, K_0^t, and K_1 are convex closed set, the last one being a cone. We will use $\mathcal{E}_h$ as a possible approximation of $\mathcal{E}$.

Although, in Section 6.5, we will illustrate usage of this model on a rather special inverse adhesive-contact problem, most of the considerations can expectedly be applied (after possible modification) to many other problems from continuum mechanics and physics, as (various combination of) damage, phase-transformations, plasticity, etc.

Remark 6.2.1 (Stability and convergence for $\tau \to 0$ and $h \to 0$)**.** The focus of this text is on the identification of the discrete finite-dimensional problem. Nevertheless, the convergence towards the original continuous problem when $\tau \to 0$ and $h \to 0$ is of interest.

Without going into (usually rather technical) details, let us only mention that under certain qualification of $\mathscr{R}_1$, $\mathscr{R}_2$ and $\mathscr{E}_h$, a boundedness (= numerical stability) and convergence of a solution $(u_{\tau h}, z_{\tau h})$ to the discrete state problem obtained by interpolation from values $(u_{\tau h}^k, z_{\tau h}^k)_{k=1}^K$ towards a weak solution (u, z) to controlled state system for a fixed π can usually be shown at least in terms of subsequences in various situations. A rather simple situations is if $\mathscr{R}_2$, or possibly also $\mathscr{R}_1$, is uniformly convex; this corresponds to some viscosity. In a special fully rate-independent case when $\mathscr{R}_1 = 0$ and $\mathscr{R}_2(\pi, u, z; \cdot)$ is 1-homogeneous and independent of (u, z), such convergence was proved in [111]; in this case the weak solutions are called local solutions. The uniqueness is however not guaranteed in general. If $\mathscr{E}(t, \pi, \cdot, \cdot)$ is jointly uniformly convex, then this uniqueness and even continuous dependence on π holds, cf. [82, 83] for a survey of such situations. This is e.g. the case of linearized rate-independent plasticity with hardening. Sometimes, viscosity can help for this uniqueness. This is the case of frictional normal-compliance contact of visco-elastic bodies which, after a certain algebraic manipulation gets the structure with $\mathscr{E}(t, \pi, \cdot, \cdot)$ separately uniformly quadratic with linear constraints in two-dimensions, cf. [112], or with conical constraints in three-dimensions. The uniqueness of the response of the continuous problems was shown in [53].

As usual, the convergence of solutions to (6.4) towards solutions to (6.3) is much more delicate and it is a well-known fact that it cannot be expected unless the controlled state system in (6.3) has a unique response or at least any solution to (6.1) can be attained by the discretized solutions, which is however usually not granted unless the solution to (6.1) is unique. In any case, one needs to show the continuous convergence of the solution map $S_{\tau h} : \pi \mapsto (u_{\tau h}, z_{\tau h})$, i.e. that $\tau \to 0$ and $h \to 0$ and $\widetilde{\pi} \to \pi$ implies $S_{\tau h}(\widetilde{\pi}) \to S(\pi)$. This is usually a relatively simple modification of the convergence for π fixed.

Having in mind the discrete problem with $\tau > 0$ and $h > 0$, we will use notation

$$p_{\tau h}^k(\pi, \widetilde{u}, u, \widetilde{z}) := \nabla_u \mathcal{R}_1\Big(\pi, \widetilde{u}, \widetilde{z}, \frac{u - \widetilde{u}}{\tau}\Big) + \nabla_u \mathcal{E}_h\big(k\tau, \pi, u, \widetilde{z}\big), \tag{6.7a}$$

$$q_{\tau h}^k(\pi, \widetilde{u}, u, \widetilde{z}, z) := \nabla_z \mathcal{R}_2\Big(\pi, \widetilde{u}, \widetilde{z}, \frac{z - \widetilde{z}}{\tau}\Big) + \nabla_z \mathcal{E}_h\big(k\tau, \pi, u, z\big), \tag{6.7b}$$

$$\mathfrak{K}^k(\tilde{z}) := \big(K_1 + \tilde{z}\big) \cap K_0^{k\tau}, \tag{6.7c}$$

$$J(\pi, \hat{u}, \hat{z}) := \tau \textstyle\sum_{k=1}^K j(u^k, z^k) + H(\pi). \tag{6.7d}$$

with $\hat{u} = (u^1, ..., u^K)$ and $\hat{z} = (z^1, ..., z^K)$. Since problems (6.5) are convex, necessary optimality conditions are also sufficient and thus, taking into account structure (6.6), problem (6.4) can equivalently be written in the form

$$\begin{aligned} &\text{minimize } J(\pi, u_{\tau h}, z_{\tau h}) \\ &\text{subject to } 0 \in p^k_{\tau h}(\pi, u^{k-1}_{\tau h}, u^k_{\tau h}, z^{k-1}_{\tau h}) + \mathrm{N}_{\Lambda^k_\tau}(u^k_{\tau h}), \qquad k = 1, \dots, K, \qquad u^0_{\tau h} = u_{0h}, \\ &\qquad\qquad 0 \in q^k_{\tau h}(\pi, u^{k-1}_{\tau h}, u^k_{\tau h}, z^{k-1}_{\tau h}, z^k_{\tau h}) + \mathrm{N}_{\mathfrak{K}^k(z^{k-1}_{\tau h})}(z^k_{\tau h}),\ k = 1, \dots, K,\ z^0_{\tau h} = z_{0h}, \\ &\qquad\qquad \pi \in \Pi_h \end{aligned} \tag{6.8}$$

with $K = T/\tau$, $u_{\tau h} := (u^1_{\tau h}, ..., u^K_{\tau h})$ and $z_{\tau h} = (z^1_{\tau h}, ..., z^K_{\tau h})$. Defining the solution map $S_{\tau h} : \pi \mapsto (u, z)$ implicitly via system (6.8), we may used the so-called implicit programming approach to rewrite problem (6.8) equivalently into the form

$$\text{minimize } J(\pi, S_{\tau h}(\pi)) \text{ subject to } \pi \in \Pi_h. \tag{6.9}$$

In the rest of the chapter we will make use of the following standing assumption, which imply in particular the single-valuedness of the solution map $S_{\tau h}$:

(A1): $\mathscr{E}_h(t, \pi, u, \cdot)$ and $\mathscr{E}_h(t, \pi, \cdot, z)$ are strictly convex,

(A2): $\mathscr{R}_1(\pi, u, z, \cdot)$ and $\mathscr{R}_2(\pi, u, z, \cdot)$ are convex,

(A3): $p^k_{\tau h}(\pi, \tilde{u}, \cdot, z)$ and $q^k_{\tau h}(\pi, \tilde{u}, u, \tilde{z}, \cdot)$ are continuously differentiable mappings, and

(A4): Λ^k and $\mathfrak{K}^k(\tilde{z})$ are closed convex sets.

Note that (A1)–(A3) implies that $p^k_{\tau h}(\pi, \tilde{u}, \cdot, z)$ and $q^k_{\tau h}(\pi, \tilde{u}, u, \tilde{z}, \cdot)$ have a positive definite Jacobian.

In what follows, we will fix time (and, if any, also space) discretization and thus we will omit τ and h in the following sections. The dimension of U_h, Z_h, and Π_h will be respectively denoted by N, M, and L.

Before devising a (necessarily) quite complicated procedure to evaluate a gradient information for the nonsmooth functional $\pi \mapsto J(\pi, S(\pi))$, let us still briefly present basic notions from variational analysis which are essential for this chapter. More information can be found in [110] for finite-dimensional setting or in [87] and [28] for the general infinite-dimensional case.

6.3 Evaluation of a subgradient of the solution mapping and first–order necessary optimality conditions

To solve problem (6.8) or equivalently (6.9) efficiently, we need to compute a subgradient information for the mapping $\pi \mapsto J(\pi, S(\pi))$. Unfortunately, we cannot expect that S is a differentiable function and thus, we need first to compute some kind of generalized derivative of S.

We will work with the generalized differential calculus of Mordukhovich [87, 110] and compute the limiting subdifferential of the objective in (6.9). To be

able to do so, we first have to compute the so-called coderivative D^*S, which for continuously differentiable functions amounts to the adjoint Jacobian. First we state a lemma which links these two concepts together.

Lemma 6.3.1. *Consider the solution mapping $S : \pi \mapsto (\bar{u}, \bar{z})$ being implicitly defined by system (6.8) and fix some $(\bar{u}, \bar{z}) = S(\bar{\pi})$. Assume that S is Lipschitz continuous on some neighborhood of $\bar{\pi}$ and that J is continuously differentiable on some neighborhood of $(\bar{\pi}, \bar{u}, \bar{z})$. Denoting $\tilde{J}(\pi) := J(\pi, S(\pi))$, we have*

$$\partial \tilde{J}(\bar{\pi}) \subset \nabla_\pi J(\bar{\pi}, \bar{u}, \bar{z}) + D^*S(\bar{\pi}, \bar{u}, \bar{z})(\nabla_u J(\bar{\pi}, \bar{u}, \bar{z}), \nabla_z J(\bar{\pi}, \bar{u}, \bar{z})).$$

Proof. It follows directly from [89, Theorem 7] and [110, Exercise 8.8]. □

To obtain the necessary optimality conditions in the form of original data, we need to compute D^*S. This is conducted in the next lemma which will also be the basis for proving the Lipschitzian continuity of S later in Corollary 6.3.3.

Lemma 6.3.2. *Consider the setting of the solution mapping $S : \pi \mapsto (\bar{u}, \bar{z})$ being implicitly defined by system (6.8) and fix some $(\bar{u}, \bar{z}) = S(\bar{\pi})$. Assuming (A1)–(A4), the upper estimate of $D^*S(\bar{\pi}, \bar{u}, \bar{z})(u^*, z^*)$ is the collection of all quantities*

$$-\sum_{k=1}^{K} (\nabla_\pi p^k)^\top \beta^k - \sum_{k=1}^{K} (\nabla_\pi q^k)^\top \delta^k \tag{6.10}$$

such that for $k = 1, \ldots, K$ the adjoint equations

$$-u^{*k} = \alpha^k - (\nabla_u p^k)^\top \beta^k - (\nabla_u q^k)^\top \delta^k - (\nabla_{\tilde{u}} p^{k+1})^\top \beta^{k+1} - (\nabla_{\tilde{u}} q^{k+1})^\top \delta^{k+1}, \tag{6.11a}$$

$$-z^{*k} = \gamma^k - (\nabla_z q^k)^\top \delta^k - (\nabla_{\tilde{z}} p^{k+1})^\top \beta^{k+1} - (\nabla_{\tilde{z}} q^{k+1})^\top \delta^{k+1} \tag{6.11b}$$

with the terminal conditions $\beta^{K+1} = 0$ and $\delta^{K+1} = 0$ are fulfilled. For the multipliers $\alpha, \beta, \gamma, \delta$ we have the relations

$$\begin{pmatrix} \alpha^k \\ \beta^k \end{pmatrix} \in \mathrm{N}_{\mathrm{gph}\, \mathrm{N}_{\Lambda^k}}(\bar{u}^k, -p^k(\bar{\pi}, \bar{u}^{k-1}, \bar{u}^k, \bar{z}^{k-1})), \tag{6.12a}$$

$$\begin{pmatrix} \gamma \\ \delta \end{pmatrix} \in \mathrm{N}_{\mathrm{gph}\, Q}(\bar{z}, -q(\bar{\pi}, \bar{u}, \bar{z})), \tag{6.12b}$$

where $\gamma = (\gamma^1, ..., \gamma^K)$ and $\delta = (\delta^1, ..., \delta^K)$ and where, for $u = (u^1, ..., u^K)$ and $z = (z^1, ..., z^K)$, we have defined

$$\begin{aligned} q(\pi, u, z) &:= \begin{pmatrix} q^1(\pi, u^0, u^1, z^0, z^1) \\ \ldots \\ q^K(\pi, u^{K-1}, u^K, z^{K-1}, z^K) \end{pmatrix} && : \mathbb{R}^L \times \mathbb{R}^{KN} \times \mathbb{R}^{KM} \to \mathbb{R}^{KM}, \\ Q(z) &:= \bigtimes_{k=1}^{K} \mathrm{N}_{\mathfrak{K}^k(z^{k-1})}(z^k) && : \mathbb{R}^{KM} \rightrightarrows \mathbb{R}^{KM}. \end{aligned}$$

Proof. Similarly to q and Q, we define

$$\begin{aligned} p(\pi, u, z) &:= \begin{pmatrix} p^1(\pi, u^0, u^1, z^0) \\ \ldots \\ p^K(\pi, u^{K-1}, u^K, z^{K-1}) \end{pmatrix} && : \mathbb{R}^L \times \mathbb{R}^{KN} \times \mathbb{R}^{KM} \to \mathbb{R}^{KN}, \\ P(u) &:= \bigtimes_{k=1}^{K} \mathrm{N}_{\Lambda^k}(u^k) && : \mathbb{R}^{KN} \rightrightarrows \mathbb{R}^{KN} \end{aligned}$$

We define the following partially linearized mapping

$$M(\mu,\nu) := \left\{ (\pi,u,z) \;\middle|\; \begin{array}{l} \mu \in p(\bar\pi,\bar u,\bar z) + \nabla_u p(\bar\pi,\bar u,\bar z)(u-\bar u) + \nabla_z p(\bar\pi,\bar u,\bar z)(z-\bar z) + P(u) \\ \nu \in q(\bar\pi,\bar u,\bar z) + \nabla_u q(\bar\pi,\bar u,\bar z)(u-\bar u) + \nabla_z q(\bar\pi,\bar u,\bar z)(z-\bar z) + Q(z) \end{array} \right\}$$

and show that it is single–valued and locally Lipschitz around $(0,0)$. Indeed, the relations defining M read for $k = 1,\dots,K$

$$\begin{aligned}
\mu^k &\in p^k(\bar\pi,\bar u^{k-1},\bar u^k,\bar z^{k-1}) + \nabla_u p^k(\bar\pi,\bar u^{k-1},\bar u^k,\bar z^{k-1})(u^k-\bar u^k) \\
&\quad + \nabla_{\bar u} p^k(\bar\pi,\bar u^{k-1},\bar u^k,\bar z^{k-1})(u^{k-1}-\bar u^{k-1}) + \nabla_{\bar z} p^k(\bar\pi,\bar u^{k-1},\bar u^k,\bar z^{k-1})(z^{k-1}-\bar z^{k-1}) + \mathrm{N}_{\Lambda^k}(u^k), \\
\nu^k &\in q^k(\bar\pi,\bar u^{k-1},\bar u^k,\bar z^{k-1},\bar z^k) + \nabla_u q^k(\bar\pi,\bar u^{k-1},\bar u^k,\bar z^{k-1}\bar z^k)(u^k-\bar u^k) \\
&\quad + \nabla_{\bar u} q^k(\bar\pi,\bar u^{k-1},\bar u^k,\bar z^{k-1},\bar z^k)(u^{k-1}-\bar u^{k-1}) + \nabla_z q^k(\bar\pi,\bar u^{k-1},\bar u^k,\bar z^{k-1})(z^k-\bar z^k) \\
&\quad + \nabla_{\bar z} q^k(\bar\pi,\bar u^{k-1},\bar u^k,\bar z^{k-1})(z^{k-1}-\bar z^{k-1}) + \mathrm{N}_{\mathfrak{K}^k(z^{k-1})}(z^k)
\end{aligned}$$

with $u^0 = \bar u^0$ and $z^0 = \bar z^0$. Since the first inclusion is solved for u^k and the second one for z^k, we obtain that M is single-valued due to (A1)–(A4). By virtue of [38, Corollary 3D.5] we further obtain that M is Lipschitz continuous around $\bar\pi$, so that the system defining S is strongly regular (in the sense of Robinson [106]) at $(0,0,\bar\pi,\bar u,\bar z)$.

This enables us to use [92, Proposition 3.2] and [110, Theorem 6.14] to obtain, with I being the identity matrix, that

$$\mathrm{N}_{\mathrm{gph}\, S}(\bar\pi,\bar u,\bar z) \subset \begin{pmatrix} 0 & I & 0 \\ -\nabla_\pi p(\bar\pi,\bar u,\bar z) & -\nabla_u p(\bar\pi,\bar u,\bar z) & -\nabla_z p(\bar\pi,\bar u,\bar z) \\ 0 & 0 & I \\ -\nabla_\pi q(\bar\pi,\bar u,\bar z) & -\nabla_u q(\bar\pi,\bar u,\bar z) & -\nabla_z q(\bar\pi,\bar u,\bar z) \end{pmatrix}^\top \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix}.$$

with some $\alpha,\beta \in \mathbb{R}^{KN}$ and $\gamma,\delta \in \mathbb{R}^{KM}$ satisfying

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \in \mathrm{N}_{\mathrm{gph}\, P}(\bar u, -p(\bar\pi,\bar u,\bar z)) \quad \text{and} \quad \begin{pmatrix} \gamma \\ \delta \end{pmatrix} \in \mathrm{N}_{\mathrm{gph}\, Q}(\bar z, -q(\bar\pi,\bar u,\bar z)).$$

Applying the product rule for normal cones [110, Proposition 6.41] we obtain the statement of the lemma. □

If Λ^k is a polyhedral set, then $\mathrm{N}_{\mathrm{gph}\,\Lambda^k}(\cdot)$ can be computed via [37, Theorem 2] or [55, Proposition 3.2], see also Chapter 3. For the computation of $\mathrm{N}_{\mathrm{gph}\, Q}(\cdot)$, we will consider two cases of $\mathfrak{K}^k$, specifically

$$\mathfrak{K}^k(z^{k-1}) = \mathbb{R}^M \quad \text{or} \tag{6.13a}$$

$$\mathfrak{K}^k(z^{k-1}) = \{z \in \mathbb{R}^M \mid 0 \le z \le z^{k-1}\} \tag{6.13b}$$

where in (6.13b), the inequality is understood componentwise. The former case (6.13a) corresponds to $K_0^t = K_1 = \mathbb{R}^M$, while the latter case (6.13b) corresponds to $K_0^t = \mathbb{R}^M_+$ and $K_1 = \mathbb{R}^M_-$. The former case is simple because from (6.12b) we immediately obtain that $\gamma^k = 0$ and $\delta^k \in \mathbb{R}^M$. For the analysis of the more complicated case we refer the reader to Subsection 3.4.2. We will also use notation from this part.

We will prove now the Lipschitz continuity of S for both cases in (6.13). Due to assumptions (A1)–(A3) and Lemma 4.3.3 we obtain that if a pair (α^k,β^k) satisfies (6.12a), then we have $\alpha^{k\top}\beta^k \le 0$. However, for (γ,δ) satisfying (6.12b), it may happen that $\gamma^{k\top}\delta^k > 0$ (see formula (6.17) below). Nevertheless, we are able to overcome this problem by making use of the specific structure of $\mathrm{gph}\, Q$.

Corollary 6.3.3. *In the setting of Lemma 6.3.2 assume that $\mathfrak{K}^k$ is defined via* (6.13a) *or* (6.13b). *Fix some $(\bar{u}, \bar{z}) = S(\bar{\pi})$. Then S is Lipschitz continuous around $\bar{\pi}$.*

Proof. Without loss of generality we may assume that $M = 1$. Since S is single–valued, it is locally Lipschitz around $\bar{\pi}$ if and only if it has the so–called Aubin property around $(\bar{\pi}, \bar{u}, \bar{z})$. Moreover, this property is according to [110, Theorem 9.40] equivalent to

$$D^*S(\bar{\pi}, \bar{u}, \bar{z})(0,0) = \{0\}. \tag{6.14}$$

To show this, we plug $u^* = z^* = 0$ into system (6.11)–(6.12) and deduce that $\beta = \delta = 0$, which implies that (6.10) is equal to zero as well, and thus condition (6.14) is fulfilled.

To this end, we first realize that the first case (6.13a) implies $\gamma^k = 0$ and $\delta^k \in \mathbb{R}^M$. In the rest of the proof, we will consider only the second case (6.13b) with a note that case (6.13a) can be shown by a slight modification of the last paragraph. Fix any $(\gamma, \delta) \in \mathrm{N}_{\mathrm{gph}\, Q}(\bar{z}, -q(\bar{\pi}, \bar{u}, \bar{z}))$. From Theorem 3.4.7 we know that there is some $s \in I(\bar{s})$ such that for all $i \in I(s)$ there exist some μ_i^k, $\tilde{\mu}_i^k$ and ν_i^k such that $\gamma^k = \mu_i^k + \tilde{\mu}_i^k$, $\delta^k = \nu_i^k$, $\tilde{\mu}_i^K = 0$ and relation

$$\begin{pmatrix} \tilde{\mu}_i^{k-1} \\ \mu_i^k \\ \nu_i^k \end{pmatrix} \in \mathrm{N}_{\mathrm{cl}\, \tilde{Q}_{i^k}}(\tilde{Q}_{s^k}) \tag{6.15}$$

holds for all $k = 1, \ldots, K$.

We will define now the index set

$$I = \left\{ (i^1, \ldots, i^K) \;\middle|\; \begin{array}{l} s^k = 1 \implies i^k = 1, \quad s^k = 2 \implies i^k \in \{1,3\} \\ s^k = 3 \implies i^k = 3, \quad s^k = 4 \implies i^k \in \{3,5\} \\ s^k = 5 \implies i^k = 5 \\ s^k = 6,\ i^{k-1} \in \{1,3\} \implies i^k = 5 \\ s^k = 6,\ i^{k-1} \in \{5,6,8\} \implies i^k = 6 \\ s^k = 7,\ i^{k-1} \in \{1,3\} \implies i^k \in \{1,5\} \\ s^k = 7,\ i^{k-1} \in \{5,6,8\} \implies i^k \in \{6,8\} \\ s^k = 8,\ i^{k-1} \in \{1,3\} \implies i^k = 1 \\ s^k = 8,\ i^{k-1} \in \{5,6,8\} \implies i^k = 8 \end{array} \right\}$$

and say that property (P^k) holds if

$$s^k \in \{1,2\} \implies \mu_i^k = 0 \text{ for all } i \in I, \text{ and} \tag{6.16a}$$

$$\exists j < k : s^j = 4,\ s^{j+1} = \cdots = s^k = 8 \implies \mu_i^k \geq 0 \text{ for some } i \in I \text{ with } i^j = 3 \text{ and } i^{j+1} = \cdots = i^k = 1. \tag{6.16b}$$

Naturally, this property is satisfied if $s^k \notin \{1,2,8\}$ and it can be shown that $I \subset I(s)$. We will now show that for all $k = 1, \ldots, K-1$ we have the following implication

$$\gamma^{k+1\top}\delta^{k+1} \leq 0 \text{ and } (P^{k+1}) \text{ holds} \implies \gamma^{k\top}\delta^k \leq 0 \text{ and } (P^k) \text{ holds}. \tag{6.17}$$

Thus, we assume $\gamma^{k+1\top}\delta^{k+1} \leq 0$ and that property (P^{k+1}) holds. We will now make use of the fact that $\delta^k = \nu_i^k$, and thus ν_i^k does not depend on i. By

evaluating (6.15), we obtain that there exists $i \in I \subset I(s)$ such that

$$
\begin{array}{lll}
s^{k+1}=1 \implies \tilde{\mu}_i^k = -\mu_i^{k+1}, & s^k=1 \implies & \nu_i^k = 0, \\
s^{k+1}=2 \implies \tilde{\mu}_i^k = -\mu_i^{k+1}, & s^k=2 \implies \mu_i^k \geq 0, & \nu_i^k \leq 0, \\
s^{k+1}=3 \implies \tilde{\mu}_i^k = 0, & s^k=3 \implies \mu_i^k = 0, & \\
s^{k+1}=4 \implies \tilde{\mu}_i^k = 0, & s^k=4 \implies \mu_i^k \leq 0, & \nu_i^k \geq 0, \\
s^{k+1}=5 \implies \tilde{\mu}_i^k = 0, & s^k=5 \implies & \nu_i^k = 0, \\
 & s^k=6 \implies & \nu_i^k = 0, \\
 & s^k=7 \implies & \nu_i^k = 0, \\
 & s^k=8 \implies & \nu_i^k = 0.
\end{array}
$$

The implication $s^k = 7 \implies \nu_i^k = 0$ follows from $I \subset I(s)$, the nondependence of ν_i^k on i and from the possibility to choose either $i^k \in \{1,5\}$ or $i^k \in \{6,8\}$. We observe now that in any case we have $\mu_i^{k\top}\delta^k = \mu_i^{k\top}\nu_i^k \leq 0$. This means that we have managed to prove $\gamma^{k\top}\delta^k \leq 0$ provided $\tilde{\mu}_i^k = 0$ or $\nu_i^k = 0$.

Thus, to prove the first part of (6.17) it remains to investigate cases $s^{k+1} \in \{1,2,6,7,8\}$ and $s^k \in \{2,3,4\}$. We will restrict now to these problematic cases. If $s^{k+1} \in \{1,2\}$, then (P^{k+1}) implies $\tilde{\mu}_i^k = -\mu_i^{k+1} = 0$ and we may apply the previous result. If $s^{k+1} \in \{6,7\}$ and $s^k = 4$, then choosing $i^{k+1} = 5$ and $i^k = 3$ results in $\tilde{\mu}_i^k \leq 0$ and $\mu_i^k \leq 0$, which together with $\nu_i^k \geq 0$ implies $\gamma^{k\top}\delta^k \leq 0$. Due to definition of Θ, it remains to investigate the last case: $s^{k+1} = 8$ and $s^k = 4$. In this case, we choose $i^{k+1} = 1$ and $i^k = 3$, which leads to $\mu_i^{k+1} + \tilde{\mu}_i^k \leq 0$ and $\mu_i^k \leq 0$. But since $\mu_i^{k+1} \geq 0$ due to (P^{k+1}), we have $\tilde{\mu}_i^k \leq 0$, and thus we again obtain $\gamma^{k\top}\delta^k \leq 0$. So far, we have managed to prove that if the left–hand side of (6.17) holds true, then we have $\gamma^{k\top}\delta^k \leq 0$.

To show the validity of formula (6.17), we need to verify that (P^k) holds as well. To do so, we multiply the adjoint equation (6.11b) by δ^k, which due to assumption (A1)–(A2) and the already proven $\gamma^{k\top}\delta^k \leq 0$ results in $\gamma^k = \mu_i^k + \tilde{\mu}_i^k = 0$ and $\delta^k = \nu_i^k = 0$ for all $i \in I$. We will now investigate the cases described on the left–hand side of (6.16).

For (6.16a) we have $s^k \in \{1,2\}$. This by definition of Θ yields $s^{k+1} \in \{1,2,3,4,5\}$. If $s^{k+1} \in \{3,4,5\}$, then $\tilde{\mu}_i^k = 0$ and thus $\mu_i^k = 0$ follows. If on the other hand we have $s^k \in \{1,2\}$, then from assumed (P^{k+1}) we get $\tilde{\mu}_i^k = -\mu_i^{k+1} = 0$, and thus $\mu_i^k = 0$ follows for this case as well. To prove (6.16b) consider some $j < k$ and $s^j = 4$, $s^{j+1} = \cdots = s^k = 8$, $i^j = 3$ and $i^{j+1} = \cdots = i^k = 1$. If $s^{k+1} = 8$, then $i^{k+1} = 1$ and we may apply (P^{k+1}) to obtain $\mu_i^{k+1} \geq 0$, which together with $\tilde{\mu}_i^k + \mu_i^{k+1} \leq 0$ and $\mu_i^k + \tilde{\mu}_i^k = 0$ implies $\mu_i^k \geq 0$. If $s^{k+1} \in \{6,7\}$, then choosing $i^{k+1} = 5$ results in $\tilde{\mu}_i^k \leq 0$, which again implies $\mu_i^k \geq 0$. Since these are all possibilities due to the definition of Θ, we have showed formula (6.17).

Having this formula at hand, the rest of the proof is performed by a finite induction. Since $\tilde{\mu}_i^K = 0$, by similar arguments as in the previous text we obtain that $\gamma^{K\top}\delta^K = \mu_i^{K\top}\nu_i^K \leq 0$, which further yields $\gamma^K = \mu_i^K = \delta^K = 0$, and thus property (P^K) is satisfied. Hence, we have obtained the validity of the first step for finite induction. Plugging this into the first adjoint equation (6.11a) and multiplying it by β^K, we obtain that $\alpha^K = \beta^K = 0$. Since the left–hand side of (6.17) is satisfied, we immediately obtain that $\gamma^{K-1\top}\delta^{K-1} \leq 0$ and that (P^{K-1}) holds. Performing this procedure K times, we obtain that (6.14) indeed holds, which finishes the proof. □

Finally, we summarize the derivation of the necessary optimality conditions in Theorem 6.3.4 below. Thereby, the normal cone $\mathrm{N}_{\mathrm{gph}\, Q}(\cdot)$ is computed in Theorem 3.4.7 and for the computation of $\mathrm{N}_{\mathrm{gph}\, \Lambda^k}(\cdot)$ we refer the reader to [37, Theorem 2] or [55, Proposition 3.2]. Moreover, when solving system (6.12) and (6.19), one may use the procedure described at the end of Section 5.3 to its advantage.

Theorem 6.3.4 (First-order optimality conditions)**.** *Consider the setting of the solution mapping $S : \pi \mapsto (\bar{u}, \bar{z})$ implicitly defined by system* (6.8) *and fix some $(\bar{u}, \bar{z}) = S(\bar{\pi})$. Assume (A1)–(A4) and that J is continuously differentiable at $(\bar{\pi}, \bar{u}, \bar{z})$. If $(\bar{\pi}, \bar{y}, \bar{z})$ is a local minimum of problem* (6.8)*, then there exists multipliers $(\alpha, \beta, \gamma, \delta)$ satisfying* (6.12) *such that the optimality condition*

$$0 \in \nabla_\pi J(\bar{\pi}, \bar{u}, \bar{z}) - \sum_{k=1}^{K} (\nabla_\pi p^k)^\top \beta^k - \sum_{k=1}^{K} (\nabla_\pi q^k)^\top \delta^k + \mathrm{N}_\Pi(\bar{\pi}), \tag{6.18}$$

the adjoint equations with $k = 1, \dots, K$

$$-\nabla_{u^k} J(\bar{\pi}, \bar{u}, \bar{z}) = \alpha^k - (\nabla_u p^k)^\top \beta^k - (\nabla_u q^k)^\top \delta^k - (\nabla_{\tilde{u}} p^{k+1})^\top \beta^{k+1} - (\nabla_{\tilde{u}} q^{k+1})^\top \delta^{k+1}, \tag{6.19a}$$

$$-\nabla_{z^k} J(\bar{\pi}, \bar{u}, \bar{z}) = \gamma^k - (\nabla_z q^k)^\top \delta^k - (\nabla_{\tilde{z}} p^{k+1})^\top \beta^{k+1} - (\nabla_{\tilde{z}} q^{k+1})^\top \delta^{k+1} \tag{6.19b}$$

and terminal conditions $\beta^{K+1} = 0$ and $\delta^{K+1} = 0$ are satisfied.

Remark 6.3.5 (More general dissipation I)**.** In a number of applications $\mathscr{R}_2$ is finite but nonsmooth at 0 and $K_1 = Z$. In this case, in the generalized equation system defining S, one has generally a sum of multifunctions which is typically very difficult to handle, cf. [110, Theorem 10.41]. Sometimes, however, an analytic formula for the behavior of S at the single time instances can be obtained and then D^*S can be computed by applying the (first-order) generalized differential calculus [87, 110].

Another possible approach to this situation is to transform it into the form considered here, i.e. $\mathscr{R}_2$ smooth and a suitable K_1. Let us illustrate this on a one-dimensional case $Z = \mathbb{R}$ with $\mathscr{R}_2(\dot{z}) = a \max(0, \dot{z}) + b \max(0, -\dot{z})$ with some $a, b \geq 0$ and, e.g., $\mathscr{E}(z) = \frac{1}{2} z^2$. Considering artificial variable (z_1, z_2) such that $z_1 + z_2 = z$, we may write

$$\mathscr{E}(z_1, z_2) = \frac{1}{2}(z_1 + z_2)^2 \quad \text{and} \quad \mathscr{R}_2(\dot{z}_1, \dot{z}_2) = \begin{cases} a\dot{z}_1 - b\dot{z}_2 & \text{if } \dot{z}_1 \geq 0 \text{ and } \dot{z}_2 \leq 0, \\ \infty & \text{otherwise.} \end{cases} \tag{6.20}$$

Such a transformation allows to widen the application range towards e.g. damage or delamination problems with healing in arbitrary space dimension. Another application can be frictional contact [139] or adhesive contact with an interfacial plasticity [116] allowing to distinguish less dissipative mode I (opening) from more dissipative mode II (shear) in two-dimensional cases. Another, rather academical, application is the bulk plasticity with kinematic hardening in one dimension. Naturally, all these applications are considered with a suitable space discretization.

Remark 6.3.6 (More general dissipation II)**.** In some applications the cone K_1 could be the 2nd-order (Lorentz, or colloquially also called "ice-cream") cone, defined in $\mathbb{R}^l$ by

$$\{x \in \mathbb{R}^l \mid x_l \geq |(x_1, ..., x_{l-1})|\}.$$

where $|\cdot|$ stands for the Euclidean norm. In this case it is possible to make use of coderivatives of the normal cone mapping associated with second-order cones which have been computed in [95]. Then, however, the special technique of Chapter 3, tailored to polyhedral multifunctions, cannot be used any more and we have to confine ourselves to standard calculus rules, which leads to less selective necessary optimality conditions.

Typical applications of this type with $K_1 = Z$ are a frictional contact in three-dimensional case or plasticity with kinematic hardening in two- or three-dimensional case, again having in mind a suitable space discretization in each case. An example which uses a combination of $K_1 \neq Z$ with a nonsmooth potential $\mathscr{R}_2$, both being of the "ice-cream-type", is plasticity with isotropic hardening, cf. [52, 82, 127] which has the dissipation potential acting on the rate of $z = (p, \eta)$ of the form:

$$\delta_S^*(\dot{p}) + \delta_{K_1}(\dot{p}, \dot{\eta}) \quad \text{with } 0 \in S \subset \mathbb{R}^{d\times d}_{\text{dev}} \text{ and } K_1 := \big\{(\dot{p}, \dot{\eta}) \in \mathbb{R}^{d\times d}_{\text{dev}} \times \mathbb{R};\ \dot{\eta} \geq q_{\text{H}} \delta_S^*(\dot{p})\big\} \tag{6.21}$$

where $q_{\text{H}} > 0$ and $\mathbb{R}^{d\times d}_{\text{dev}} := \{A \in \mathbb{R}^{d\times d};\ A = A^\top,\ \operatorname{tr} A = 0\}$, and δ_A stands for an indicator function of a convex set A and δ_A^* of its conjugate. Typically, S a ball, which makes both δ_S^* and K_1 of the "ice-cream-type".

A combination of the preceding case with a general polyhedral convex set K_0 is also possible. This combination allows for some applications in identification of parameters of some phenomenological models of phase transformations in certain ferroic materials as shape-memory alloys where K_0 forms constraints on an internal variable like p in (6.21) and may be considered polyhedral, cf. the polycrystalic models in [48, 124], possibly also in combination with plasticity like that one in (6.21), cf. [13, 117].

6.4 Adhesive contact problem and its identification

We illustrate the above abstract identification problem (6.3) on an unilateral adhesive-contact problem for a linear elastic body at small strains. We consider $\Omega \subset \mathbb{R}^2$ a Lipschitz domain with $\Gamma_{\text{C}} \subset \partial\Omega$ and $\Gamma_{\text{D}} \subset \partial\Omega$ disjoint parts of the boundary $\partial\Omega$ where the delamination is undergoing and time-varying Dirichlet boundary condition where $w_{\text{D}}(t)$ is prescribed, respectively. Now, $u : \Omega \to \mathbb{R}^2$ is the displacement and $z : \Gamma_{\text{C}} \to [0, 1]$ is a delamination parameter having the meaning of the portion of bonds of the adhesive which are not debonded. With $\mathbb{C}$ the tensor of elastic moduli, $h : [0, 1] \to \mathbb{R}$ a convex adhesive-stored-energy function, and with $e(u)$ denoting the small-strain tensor, i.e. $[e(u)]_{ij} = \frac{1}{2}\frac{\partial u_i}{\partial x_j} + \frac{1}{2}\frac{\partial u_j}{\partial x_i}$,

we will consider the boundary-value problem

$$\begin{aligned}
&\operatorname{div}\mathbb{C}e(u)=0 && \text{in } [0,T]\times\Omega, && (6.22a)\\
&\mathbb{C}e(u)\vec{n}=0 && \text{on } [0,T]\times(\Gamma\backslash(\Gamma_{\mathrm{C}}\cup\Gamma_{\mathrm{D}})), && (6.22b)\\
&u|_{\Gamma_{\mathrm{D}}}=w_{\mathrm{D}}(t,\cdot) && \text{on } [0,T]\times\Gamma_{\mathrm{D}}, && (6.22c)\\
&\left.\begin{aligned}
&u_{\mathrm{N}}\geq 0,\quad z\kappa_{\mathrm{N}}u_{\mathrm{N}}+\vec{n}^{\top}\mathbb{C}e(u)\vec{n}\geq 0,\\
&\big(z\kappa_{\mathrm{N}}u_{\mathrm{N}}+\vec{n}^{\top}\mathbb{C}e(u)\vec{n}\big)u_{\mathrm{N}}=0\\
&\dot z\leq 0,\quad \xi+\alpha_{\mathrm{F}}\geq 0,\quad \dot z(\xi+\alpha_{\mathrm{F}})=0,\\
&\xi+h'(z)+\tfrac12(\kappa_{\mathrm{N}}u_{\mathrm{N}}^2+\kappa_{\mathrm{T}}u_{\mathrm{T}}^2)-\varepsilon\operatorname{div}_{\mathrm{S}}\nabla_{\mathrm{S}}z\geq 0,\quad z\geq 0,\\
&\big(\xi+h'(z)+\tfrac12(\kappa_{\mathrm{N}}u_{\mathrm{N}}^2+\kappa_{\mathrm{T}}u_{\mathrm{T}}^2)-\varepsilon\operatorname{div}_{\mathrm{S}}\nabla_{\mathrm{S}}z\big)\,z=0,\\
&z\kappa_{\mathrm{T}}u_{\mathrm{T}}+\mathbb{C}e(u)\vec{n}-(\vec{n}^{\top}\mathbb{C}e(u)\vec{n})\vec{n}=0,
\end{aligned}\right\} && \text{on } [0,T]\times\Gamma_{\mathrm{C}}, && (6.22d)
\end{aligned}$$

where we used the decomposition of the trace of displacement $u=u_{\mathrm{N}}\vec{n}+u_{\mathrm{T}}$ with u_{N} being the normal displacement defined as $u\cdot\vec{n}$ and u_{T} being the tangential displacement on Γ_{C}, and where ∇_{S} denotes a "surface gradient", i.e. the tangential derivative defined as $\nabla_{\mathrm{S}}z=\nabla z-(\nabla z\cdot\vec{n})\vec{n}$ for z defined around Γ_{C}. Alternatively, pursuing the concept of fields defined exclusively on Γ_{C}, we can consider $z:\Gamma_{\mathrm{C}}\to\mathbb{R}$ and extend it to a neighborhood of Γ_{C} and then again define $\nabla_{\mathrm{S}}z:=(\nabla z)P$ with $P=\mathbb{I}-\vec{n}\otimes\vec{n}$ onto a tangent space, which, in fact, does not depend on the particular extension. Moreover, $\operatorname{div}_{\mathrm{S}}:=\operatorname{tr}\nabla_{\mathrm{S}}$. Then $\operatorname{div}_{\mathrm{S}}\nabla_{\mathrm{S}}$ is the so-called Laplace-Beltrami operator.

Let us remark that (6.22a) is the force equilibrium, (6.22b) prescribes the zero-traction (i.e. free surface) on $\Gamma\backslash(\Gamma_{\mathrm{C}}\cup\Gamma_{\mathrm{D}})$. The condition (6.22d) combines three complementarity problems related respectively to the Signorini unilateral contact for the displacement u, the non-negativity constraint for z, and the unidirectionality constraint (i.e. the non-positivity constraint on $\dot z$), and eventually the equilibrium of tangential stress. More in detail, the last two mentioned complementarity problems write in the classical formulation as the inclusion $\partial\delta^*_{[-\alpha_{\mathrm{F}},\infty)}(\dot z)\ni\xi$ with the admissible driving force fulfilling the inclusion $\xi\in-\partial_z\mathscr{E}(t,\pi,u,z)=\varepsilon\operatorname{div}_{\mathrm{S}}\nabla_{\mathrm{S}}z-h'(z)-\frac12(\kappa_{\mathrm{N}}u_{\mathrm{N}}^2+\kappa_{\mathrm{T}}u_{\mathrm{T}}^2)-\mathrm{N}_{[0,\infty)}(z)$.

Referring to the abstract problem (6.1), the boundary-value problem (6.22) corresponds to the stored and the dissipation energies

$$\mathscr{E}(t,\pi,u,z):=\begin{cases}\displaystyle\int_{\Gamma_{\mathrm{C}}}\frac12 z\big(\kappa_{\mathrm{N}}u_{\mathrm{N}}^2+\kappa_{\mathrm{T}}u_{\mathrm{T}}^2\big)+h(z)+\frac12\varepsilon\nabla_{\mathrm{S}}z\cdot\nabla_{\mathrm{S}}z\,\mathrm{d}S\\ \qquad\displaystyle+\int_\Omega\frac12\mathbb{C}e(u)\colon e(u)\,\mathrm{d}x & \text{if } u|_{\Gamma_{\mathrm{D}}}=w_{\mathrm{D}}(t,\cdot)\text{ on }\Gamma_{\mathrm{D}}\text{ and}\\ & u|_{\Gamma_{\mathrm{C}}}\cdot\vec{n}\geq 0\text{ and }z\geq 0\text{ on }\Gamma_{\mathrm{C}},\\ \infty & \text{otherwise,}\end{cases} \qquad (6.23a)$$

$$\mathscr{R}_1\equiv 0,\qquad \mathscr{R}_2(\dot z):=\begin{cases}\displaystyle\int_{\Gamma_{\mathrm{C}}}\alpha_{\mathrm{F}}|\dot z|\,\mathrm{d}S & \text{if }\dot z\leq 0\text{ a.e. on }\Gamma_{\mathrm{C}},\\ \infty & \text{otherwise,}\end{cases}\qquad\text{with }\pi=(\alpha_{\mathrm{F}},\kappa_{\mathrm{N}},\kappa_{\mathrm{T}}), \qquad (6.23b)$$

Note that $\mathscr{E}(t,\pi,\cdot,\cdot)$ is not convex but it is separately convex and, if Γ_{D} is non-empty and h is strictly convex, it is separately strictly convex, complying

with our assumption (A1)–(A2). Considering h quadratic, this leads, after a suitable spatial discretization of (6.4), to recursive alternating strictly convex *Quadratic-Programming* (QP) which can be solved by efficient prefabricated software packages.

The (distributed) parameters to be identified will be the fracture toughness $\alpha_{\rm F}$ and the elasticity-moduli of the adhesive $\kappa_{\rm N}$ and $\kappa_{\rm T}$, i.e. we have considered simply $\pi = (\alpha_{\rm F}, \kappa_{\rm N}, \kappa_{\rm T})$ as outlined in (6.23b). This choice has a certain motivation in engineering where, in contrast to essentially all the bulk material properties, these parameters are largely unknown and have to be set up in a rather ad-hoc way to fit at least roughly some experiments, cf. e.g. [129, 130] based on experiments from [63]. Actually, the models of adhesive contacts used in engineering may be more complicated; typically they distinguish modes of delamination (opening vs shear) and/or may involve friction. Identification of friction/adhesive contacts may have interesting applications in geophysics where such contact surfaces (called faults) are deep in lithosphere and not accessible to direct investigations although a lot of indirect data from earthquakes are usually available; a popular rate-and-state friction model involves one internal parameter (called ageing) which is analogous to the delamination parameter used here, cf. [35] for a survey or also e.g. [113]. Other models that may lead to a recursive QP have been mentioned in Remark 6.3.5, in contrast to problems from Remark 6.3.6 that would lead to a recursive Second-Order Cone Programming (SOCP) for which efficient codes do exist, cf. [9].

We prescribe some initial conditions $u_0 \in H^1(\Omega)$ and $z_0 \in H^1(\Gamma_{\rm C})$, $0 \le z_0 \le 1$; note that then $0 \le z \le 1$ is satisfied during the whole evolution process. We further consider a fixed time horizon $T > 0$ and assume that we have some given desired response $(u_{\rm d}, z_{\rm d})$ corresponding e.g. to some experimentally obtained measurements, and we want to identify parameters π such that the response $(u, z) = S(\pi)$ is as close to $(u_{\rm d}, z_{\rm d})$ as possible, i.e. we want to minimize the objective

$$\int_0^T \Big[\int_\Omega \frac{\zeta}{2} \big|u - u_{\rm d}\big|^2 \,{\rm d}x + \int_{\Gamma_{\rm C}} \frac{1}{2} \big|z - z_{\rm d}\big|^2 \,{\rm d}S \Big] \,{\rm d}t \tag{6.23c}$$

where ζ is a fixed weight balancing both parts of the objective function.

After the semi–implicit time discretization, the whole problem (6.23) reads as

$$\begin{aligned} &\text{minimize } \tau \sum_{k=1}^{K} \Big[\int_\Omega \frac{\zeta}{2} \big|u^k - u_{\rm d}^k\big|^2 \,{\rm d}x + \int_{\Gamma_{\rm C}} \frac{1}{2} \big|z^k - z_{\rm d}^k\big|^2 \,{\rm d}S \Big] \\ &\text{subject to } (u^k, z^k) = S^k(\pi, u^{k-1}, z^{k-1}),\ k = 1, ..., K, \\ &\qquad \pi = (\alpha_{\rm F}, \kappa_{\rm N}, \kappa_{\rm T}) \in \Pi, \end{aligned} \tag{6.24a}$$

where the solution map $S^k : (\pi, u^{k-1}, z^{k-1}) \mapsto (u^k, z^k)$ for a particular time instant is now defined by the alternating recursive system: given $\pi = (\alpha_{\rm F}, \kappa_{\rm N}, \kappa_{\rm T})$ and previous values (u^{k-1}, z^{k-1}), the first one is solved for u^k and the second one for z^k recursively for $k = 1, ..., K$:

$$\begin{aligned} &\underset{u \in H^1(\Omega, \mathbb{R}^d)}{\text{minimize}} \ \int_\Omega \frac{1}{2} \mathbb{C}e(u) \colon e(u) \,{\rm d}x + \int_{\Gamma_{\rm C}} \frac{1}{2} z^{k-1} \big(\kappa_{\rm N} u_{\rm N}^2 + \kappa_{\rm T} u_{\rm T}^2\big) \,{\rm d}S \\ &\text{subject to } u|_{\Gamma_{\rm D}} = w_{\rm D}^k := w_{\rm D}(k\tau, \cdot) \\ &\qquad u|_{\Gamma_{\rm C}} \cdot \vec{n} \ge 0, \end{aligned} \tag{6.24b}$$

and

$$\begin{aligned}\underset{z\in H^1(\Gamma_{\mathrm{C}})\cap L^\infty(\Gamma_{\mathrm{C}})}{\text{minimize}} \ & \int_{\Gamma_{\mathrm{C}}} \Big[h(z)+\frac{\varepsilon}{2}\nabla_{\mathrm{S}} z\cdot\nabla_{\mathrm{S}} z + \Big(\frac{1}{2}\big(\kappa_{\mathrm{N}}(u_{\mathrm{N}}^k)^2+\kappa_{\mathrm{T}}(u_{\mathrm{T}}^k)^2\big) - \alpha_{\mathrm{F}}\Big)z\Big]\,\mathrm{d}S \\ \text{subject to } & 0\le z\le z^{k-1}.\end{aligned} \tag{6.24c}$$

Discretizing system (6.24b) via finite elements, we obtain

$$\begin{aligned}\underset{u=(u_{\mathrm{C}},u_{\mathrm{F}},u_{\mathrm{D}})}{\text{minimize}} \ & \frac{1}{2}u^\top A(\pi,z^{k-1})u \\ \text{subject to } & u_{\mathrm{C}}\in\Lambda_0 := \{u|\, u\cdot\vec{n}\ge 0\}, \\ & u_{\mathrm{D}} = w_{\mathrm{D}}^k,\end{aligned} \tag{6.25}$$

where the components of $u=(u_{\mathrm{C}},u_{\mathrm{F}},u_{\mathrm{D}})$ correspond to the displacement on contact boundary Γ_{C}, in free nodes (interior and Neumann) in $\bar{\Omega}\setminus(\Gamma_{\mathrm{C}}\cup\Gamma_{\mathrm{D}})$, and on Dirichlet boundary Γ_{D}, respectively. Matrix A has the following form

$$A(\pi,z^{k-1}) = \begin{pmatrix} A_{\mathrm{CC}} & A_{\mathrm{CF}} & A_{\mathrm{CD}} \\ A_{\mathrm{FC}} & A_{\mathrm{CF}} & A_{\mathrm{FD}} \\ A_{\mathrm{DC}} & A_{\mathrm{DF}} & A_{\mathrm{DD}}\end{pmatrix} + \begin{pmatrix} \tilde{A}(\pi,z^{k-1}) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0\end{pmatrix}$$

where the first part corresponds to the discretization of the first part of the objective in (6.24b) and similarly for the second part. Using simple calculus, discretized problem (6.25) can be written as

$$\begin{aligned}\underset{u_{\mathrm{C}}}{\text{minimize}} \ & \frac{1}{2}u_{\mathrm{C}}^\top(A_\alpha+\tilde{A}(\pi,z^{k-1}))u_{\mathrm{C}} + (A_\beta w_{\mathrm{D}}^k)^\top u_{\mathrm{C}} \\ \text{subject to } & u_{\mathrm{C}}\in\Lambda_0,\end{aligned} \tag{6.26a}$$

where we have defined

$$\begin{aligned} A_\alpha &:= A_{\mathrm{CC}} - A_{\mathrm{CF}}A_{\mathrm{FF}}^{-1}A_{\mathrm{FC}}, & A_\gamma &:= -A_{\mathrm{FF}}^{-1}A_{\mathrm{FC}}, \\ A_\beta &:= A_{\mathrm{CD}} - A_{\mathrm{CF}}A_{\mathrm{FF}}^{-1}A_{\mathrm{FD}}, & A_\delta &:= -A_{\mathrm{FF}}^{-1}A_{\mathrm{FD}}.\end{aligned}$$

Similarly, when discretizing (6.24c), we obtain the following problem

$$\begin{aligned}\underset{z}{\text{minimize}} \ & \frac{1}{2}z^\top Bz + b(\pi,u^k)^\top z \\ \text{subject to } & 0\le z\le z^{k-1}.\end{aligned} \tag{6.26b}$$

Since both problems in (6.26) are quadratic, we can pass to their necessary optimality conditions and the whole optimization problem (6.8) reads as

$$\begin{aligned}\underset{\pi,u_{\mathrm{C}},z}{\text{minimize}} \ & \tau\sum_{k=1}^{K}\Big[\frac{\zeta}{2}\big|u_{\mathrm{C}}^k - [u_{\mathrm{d}}]_{\mathrm{C}}^k\big|^2 + \frac{\zeta}{2}\big|A_\gamma u_{\mathrm{C}}^k + A_\delta w_{\mathrm{D}}^k - [u_{\mathrm{d}}]_{\mathrm{F}}^k\big|^2 + \frac{1}{2}\big|z^k - z_{\mathrm{d}}^k\big|^2\Big] \\ \text{subject to } & 0\in(A_\alpha+\tilde{A}(\pi,z^{k-1}))u_{\mathrm{C}}^k + A_\beta w_{\mathrm{D}}^k + \mathrm{N}_{\Lambda_0}(u_{\mathrm{C}}^k),\ k=1,...,K,\ u^0=u_0, \\ & 0\in Bz^k + b(\pi,u^k) + \mathrm{N}_{[0,z^{k-1}]}(z^k),\ k=1,...,K,\ z^0=z_0, \\ & \pi\in\Pi.\end{aligned} \tag{6.27}$$

By passing from u^k to u^k_{C} we have managed to reduce the number of parameters in (6.24b) from the number of all nodes to the number of contact nodes only. This is especially powerful because the first inclusion in (6.27) will be solved many times during the parameter identification procedure while it is sufficient to compute matrices A_α, A_β, A_γ and A_δ only once.

To be able to use Theorem 6.3.4, we need to check whether assumptions (A1)–(A4) are satisfied. But this amount to showing that matrices $A_\alpha + \tilde{A}(\pi, z^{k-1})$ and B are positive definite. Since A_α is Schur complement of A_{CC} in $\hat{A} := \begin{pmatrix} A_{\mathrm{CC}} & A_{\mathrm{CF}} \\ A_{\mathrm{FC}} & A_{\mathrm{FF}} \end{pmatrix}$, it is positive definite if $\hat{A}$ is positive definite. But the positive definiteness of $\hat{A}$ follows from the conformal FEM via positive definiteness of $\mathbb{C}$ together with the Korn inequality using Dirichlet boundary conditions on Γ_{D}. More precisely, the FEM may also involve some numerical integration (which in fact has been used for our implementation, too).

Remark 6.4.1 (Boundary-element method)**.** Note that (6.26a) is the optimization problem on Γ_{C} because we eliminated the values u_{D} and u_{F}. This is the philosophy of the *boundary-integral method* and $(A_\beta w^k_{\mathrm{D}})^\top$ in (6.26a) is in the position of the (discretized) Poincaré-Steklov operator transferring Dirichlet boundary conditions on Γ_{C} to traction forces on Γ_{C}. The discretization then leads to the celebrated *Boundary-Element Method* (BEM). One option for this discretization is FEM, cf. e.g. [73], which is in fact what we used here and such BEM represents a noteworthy interpretation of (6.26a). Other options are based on a direct discretization of the Poincaré-Steklov operator by using the approximate evaluation of the so-called Somigliana identity based on the underlying integral Green operators, cf. e.g. [16, 101, 130, 118].

Remark 6.4.2 (Variants of the adhesive model)**.** The contribution $h(z)$ in (6.23a) has the meaning of a stored energy deposited in the adhesive bonds and, during delamination, this energy naturally increases. If a reversible damage (called healing) were allowed, cf. Remark 6.3.5 above, $h'(z)$ would give a driving force for it. Strict convexity of h represents certain *cohesive effects*: when delamination is tended to be complete, still more and more energy is needed for complete delamination. Cohesive effects can also be modelled by letting κ_{N} and κ_{T} dependent on z so that $z \mapsto z\kappa_{\mathrm{N}}(z)$ and $z \mapsto z\kappa_{\mathrm{T}}(z)$ are convex. This however does not guarantee strict convexity of $\mathscr{E}(t, \pi, u, \cdot)$. Other option complying with a purely adhesive contact (e.g. $h = 0$) would be to consider a small, linear *viscosity* in z, i.e. $\mathscr{R}_2$ strictly convex and quadratic. Then the usual concept of weak solution can be used again together with the semi-implicit fractional-step-type time discretization. Yet, such problem becomes computational difficult if the viscosity is small, as often considered with the goal to approximate so-called vanishing-viscosity solution in the rate-independent inviscid limit, cf. [115].

6.5 Numerical experiments

In this section we illustrate usage and efficiency of the theory developed in Section 6.3 and later specified in Section 6.5 on a two-dimensional problem where an elastic body glued along the x-axis and pulled in the direction of the y-axis by

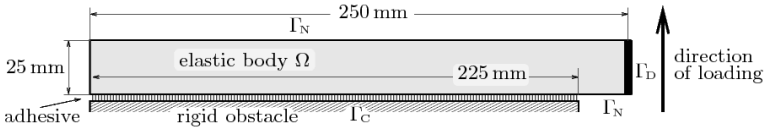


Figure 6.1: Geometry and boundary conditions of the two-dimensional problem used for calculation.

the time-varying loading w_D, cf. Figure 6.1. Considering the parameters α_F, κ_N, and κ_T to be unknown, the main goal is to identify them via an inverse problem. Following the delamination example in [115, 116], we considered the isotropic material in the bulk with the tensor of elastic moduli

$$\mathbb{C}_{ijkl} := \frac{\nu E}{(1+\nu)(1-2\nu)}\delta_{ij}\delta_{kl} + \frac{E}{2(1+\nu)}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$$

with the Young modulus $E = 70$ GPa and the Poisson ratio $\nu = 0.35$. Concerning the adhesive-stored-energy and the gradient terms in (6.23a), we used $h(z) = \frac{1}{2}z^2 - z$ and $\varepsilon = 1\,\mathrm{J}$, while the weight ζ was chosen as $10^{10}\mathrm{m}^{-2}$. For the space discretization we employed a mesh with 14×20 nodes and the equidistant time discretization used 40 time instants. The contact boundary consists of 12 nodes. As already said, there are three parameters to be identified: α_F, κ_N, and κ_T. Moreover, we assume that the values of these parameters are not constant along the contact boundary but it may have different values in every contact node. This leads to a total number of $3 \times 12 = 36$ parameters to be identified.

We fixed these 36 values, to be more specific the mean of α_F, κ_N, and κ_T was 187.5 J/m^2, 150 GPa/m, and 75 GPa/m, respectively. The difference between the smallest and largest value of α_F was approximately 10% and similarly for κ_N and κ_T. Next, we randomly generated some (with time increasing) dragging loading w_D, computed the corresponding (u_d, z_d), and plugged them into the upper level of problem (6.27). Since there was no perturbation of (u_d, z_d) present, the optimal objective value was zero, which allows numerical testing of the efficiency of the optimization algorithm.

The computation of problem (6.27) was performed in Matlab. To compute u^k from the first inclusion in (6.27), we modified and used the already written code [8]. Since a direct application of a gradient algorithm to whole problem (6.27) lead to rather inferior results, we had to find another way to solve (6.27), specifically we used a combination of three optimization algorithms. The first was PSwarm [135], which combines pattern search with genetic algorithm particle swarm, the second one standard Matlab function `fminunc` and the last one a nonsmooth modification of BFGS algorithm [74] with its implementation [125].

The optimization process was run in four phases. For the first phase, we simplified the problem and assumed that the parameters are constant along the contact boundary. This reduced the number of parameters from 36 to 3. To this problem, the algorithm PSwarm was used, however, we did not let it converge to the optimal solution but it was interrupted when the problem reached a priori given threshold or when the optimal value did not improve much in several successive iterations. In other words, the goal of the first phase was to find an estimate of the solution. Since PSwarm works rather with populations instead of single points, multiple initial points had to be chosen. These points were generated

randomly from the following intervals

$$\alpha_{\mathrm{F}} \in \left[100\ \mathrm{J/m^2}, 500\ \mathrm{J/m^2}\right], \qquad \kappa_{\mathrm{N}}, \kappa_{\mathrm{T}} \in \left[10\ \mathrm{GPa/m}, 1000\ \mathrm{GPa/m}\right].$$

In the second phase, the reduced problem was still considered but this time, an algorithm using a gradient information was used. Similarly to the first phase, we did not let the it converge and interrupted it prematurely. Because of this interruption, nonregular points were usually evaded and it was possible to use `fminunc`, even though it is designed for smooth functions.

While in the first two phases, the values of parameters were constant on the contact boundary, this no longer holds true for the last two phases. In the third one, we considered the state in which one parameter corresponds to two nodes on the contact boundary, while in the fourth phase every parameter corresponded to only one node. This means that there were 18 parameters in the third phase and 36 in the last one. The evolution of the optimal value can be seen in Figure 6.2. Note that on the y axis the logarithm of the objective value is depicted and that the vertical lines separate the four phases.

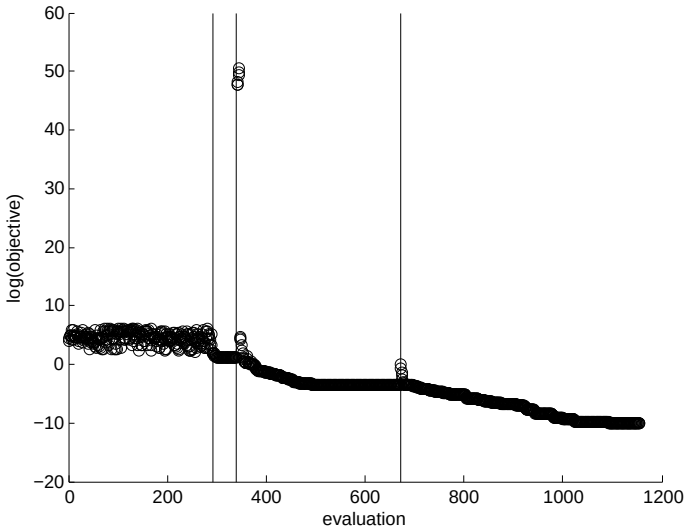


Figure 6.2: Development of the objective value during particular iterations of the optimization algorithms used during the four phases of our optimization: phase 1 used a global optimization algorithm (PSwarm), whereas phases 2–4 used a (sub)gradient algorithm with subsequently refined discretization of Γ_{C}.

The following table summarizes the values of parameters and of the objective function for all phases. The first column presents the best point in the initial population of PSwarm. The next four columns show the optimal solutions and values of all four phases. Finally, the last column corresponds to the actual values of parameters. Since there were multiple values distributed along the boundary for the last three columns, we show only their mean in such cases.

	starting	phase 1	phase 2	phase 3	optimal	desired
α_{F}	203.934	190.405	194.877	187.489	187.512	187.5
κ_{N}	$0.822{\cdot}10^{11}$	$1.586{\cdot}10^{11}$	$1.462{\cdot}10^{11}$	$1.499{\cdot}10^{11}$	$1.500{\cdot}10^{11}$	$1.5{\cdot}10^{11}$
κ_{T}	$47.251{\cdot}10^{10}$	$2.326{\cdot}10^{10}$	$7.317{\cdot}10^{10}$	$7.498{\cdot}10^{10}$	$7.499{\cdot}10^{10}$	$7.5{\cdot}10^{10}$
objective	3138.97	70.503	14.184	$3.573{\cdot}10^{-4}$	$7.538{\cdot}10^{-11}$	0

In Figure 6.3 we show the the displacement u (magnified by factor 50) corresponding to one of the random initial points used for PSwarm and solution of the four phases. A circle on the contact boundary mean that no delamination has taken place yet at the corresponding node while an asterisk means that the corresponding node has been completely delaminated. No symbol being present indicates that only a partial delamination took place. Since the contact boundary is shorter than the length of the body, there are no symbols at the bottom right corner.

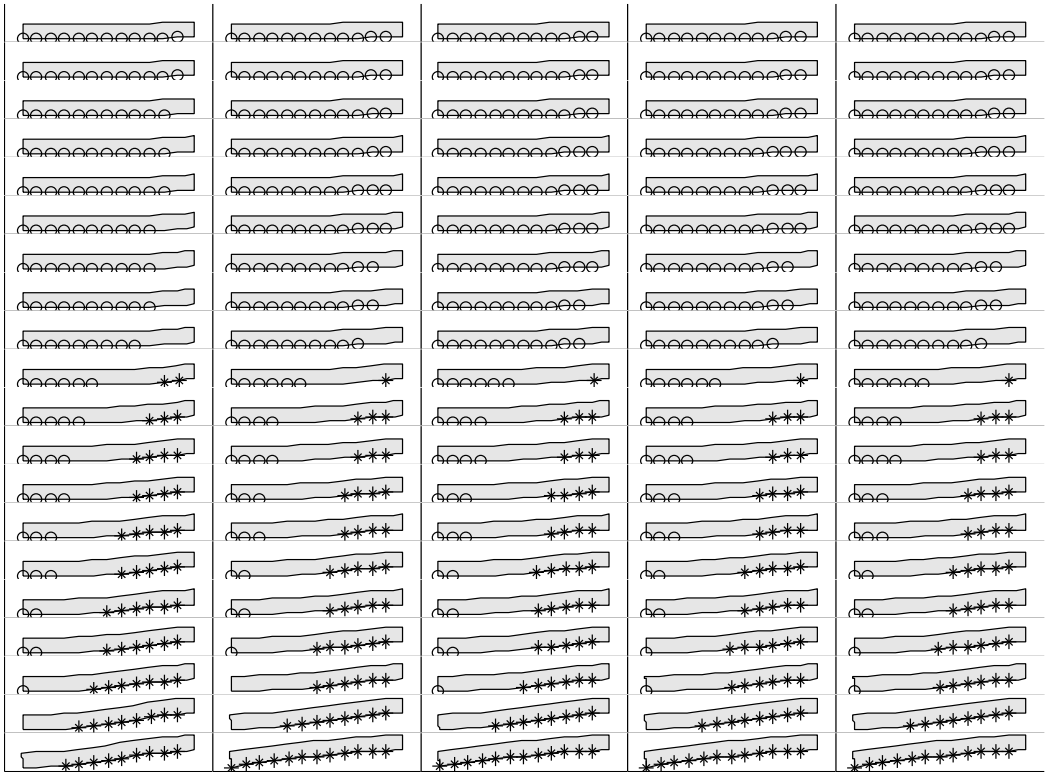


Figure 6.3: Evolution of the deformed specimen with distribution of the delamination parameter z along Γ_{C} (only values 1 or 0 are displayed) at 17 selected time instances. Displacements depicted as magnified by factor 50×.

In Figure 6.4 we show the distribution of the elastic adhesive moduli κ_{N} and κ_{T} as well as the fracture toughness α_{F} along the contact boundary Γ_{C}. Four lines corresponding to the actual parameters and to the terminal points of phases 2, 3 and 4 are depicted. The horizontal line without any symbols corresponds to phase 2, the line with circles corresponds to line 3 and the line with asterisks corresponds to phase 4, which means that it depicts the parameters identified by the algorithm. The last line without any symbols (which coincides with the line with asterisks for κ_{N}) depicts the actual parameters. We see that while the result of phase 3 provided a good estimate for the actual parameters. Phase 4 provided only a slight improvement for κ_{T} while it managed to identify the values of κ_{N} completely.

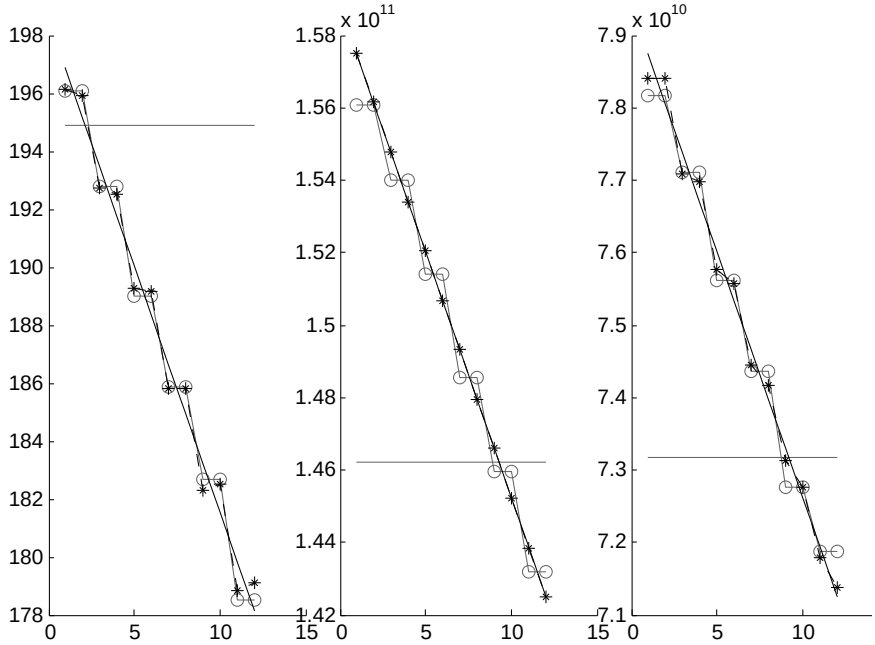


Figure 6.4: Parameter distribution along the contact boundary, graphs depicting form left to right α_{F}, κ_{N} and κ_{T} resulting after particular phases of the optimization algorithm.

7. Conclusion

In this dissertation thesis we have performed a thorough analysis of a class of mathematical programs with equilibrium constraints where the solution mapping is locally single–valued and Lispchitz continuous. To achieve this goal, we have employed the implicit programming approach which requires to compute a subgradient information of the composed objective function. Since this function is usually nondifferentiable, we aimed to find an estimate of the limiting subgradient.

We managed to derive a new approach for an exact computation of limiting normal cone to a set which can be represented as a finite union of polyhedra. Since standard calculus rules would result only in inclusions, we have succeeded to enrich the calculus. This approach was thereafter successfully used in estimating parameters in a delamination problem and in controlling an academic queuing problem. Apart from that, we were able to derive a stability criterion for the Lipschitz continuity for a solution mapping defined by a differential inclusion.

The obtained optimality conditions, such as those stated in Theorem 6.3.4, are in the MPEC literature known as M–stationarity conditions because they are based exclusively on notions from the Mordukhovich subdifferential calculus. They are relatively sharp and can very well be used, e.g., for testing this type of stationarity at points computed via implicit programming approach. It would be a great challenge to derive suitable optimality conditions also for the original continuous problem. Unfortunately, this problem is formulated over non–Asplund spaces (with a possible exception of the space for the control variable) which are not amenable for a treatment via the Mordukhovich calculus.

Concerning the results from Chapter 3, we see possible generalization in two directions. The first one would be to obtain conditions for stability of more general MPECs, where the constraint set on the lower level is governed by a linear system. The second one would be to obtain a similar result about the dependence of the solution of two–stage linear program on the (discretized) probability distribution. This generalization would make use of similar results from [41].

Acknowledgement

The author gratefully acknowledges the support from the EEA/Norwegian Financial Mechanism (project 7F14287 STRADI) and from the Grant Agency of the Czech Republic (projects 15-00735S and P201/12/0671).

Bibliography

[1] V. Acary and B. Brogliato. *Numerical Methods for Nonsmooth Dynamical Systems*, volume 35. Springer, 2008.

[2] W. Achtziger and C. Kanzow. Mathematical programs with vanishing constraints: Optimality conditions and constraint qualifications. *Mathematical Programming*, (114):69–99, 2008.

[3] L. Adam. On the Lipschitz behavior of solution maps of a class of differential inclusions. *Set–Valued and Variational Analysis*, DOI: 10.1007/s11228-015-0323-x, 2015.

[4] L. Adam and J. Outrata. On optimal control of a sweeping process coupled with an ordinary differential equation. *Discret. Contin. Dyn. Syst. - Ser. B*, 19(9):2709–2738, 2014.

[5] L. Adam, J. Outrata, and T. Roubíček. Identification of some rate-independent systems with illustration on cohesive contacts at small strains. *Submitted*.

[6] L. Adam, M. Červinka, and M. Pištěk. Normally admissible partitions and calculation of normal cones to a finite union of polyhedral sets. *Set–Valued and Variational Analysis*, DOI: 10.1007/s11228-015-0325-8, 2015.

[7] S. Adly, T. Haddad, and L. Thibault. Convex sweeping process in the framework of measure differential inclusions and evolution variational inequalities. *Math. Program. Ser. B*, 2014.

[8] J. Alberty, C. Carstensen, S. A. Funken, and R. Klose. Matlab implementation of the finite element method in elasticity. *Computing*, 69(3):239–263, 2002.

[9] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Math. Program., Ser. B*, 95:3–51, 2003.

[10] M. A. Amouzegar and S. E. Jacobsen. A decision support system for regional hazardous waste management alternatives. *Journal of applied mathematics & decision sciences*, 2:23–50, 1998.

[11] J.-P. Aubin. Lipschitz behavior of solutions to convex minimization problems. *Math. Oper. Res.*, 9:87–111, 1984.

[12] J. P. Aubin and A. Cellina. *Differential Inclusions: Set-Valued Maps and Viability Theory*. Springer, 1984.

[13] F. Auricchio, A. Reali, and U. Stefanelli. A three-dimensional model describing stress-induced solid phase transformation with permanent inelasticity. *Int. J. Plasticity*, 23(2):207–226, 2007.

[14] J. F. Bard. Convex two-level optimization. *Mathematical Programming*, 40:15–27, 1988.

[15] A. Bergqvist. Magnetic vector hysteresis model with dry friction–like pinning. *Physica B: Condensed Matter*, 233(4):342–347, 1997.

[16] A. Blázquez, R. Vodička, F. París, and V. Mantič. Comparing the conventional displacement BIE and the BIE formulations of the first and the second kind in frictionless contact problems. *Eng. Anal. Bound. Elem.*, 26:815–826, 2002.

[17] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal. *Numerical optimization: theoretical and practical aspects*. Springer, 2006.

[18] J. F. Bonnans and D. Tiba. Pontryagin's principle in the control of semilinear elliptic variational inequalities. *Appl. Math. Optim.*, 23:299–312, 1991.

[19] M. Bounkhel. *Regularity Concepts in Nonsmooth Analysis: Theory and Applications*. Springer, 2012.

[20] K. E. Brenan, S. L. Campbell, and L. R. Petzold. *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*. SIAM Publications Classics in Applied Mathematics, 1996.

[21] M. Brokate. *Optimale Steuerung von gewöhnlichen Differentialgleichungen mit Nichtlinearitäten vom Hysteresis-Typ*. Peter Lang GmbH, 1987.

[22] M. Brokate and P. Krejčí. Optimal control of ODE systems involving a rate independent variational inequality. *Discret. Contin. Dyn. Syst. - Ser. B*, 18(2):331–348, 2012.

[23] M. Brokate and J. Sprekels. *Hysteresis and Phase Transitions*. Springer, 1996.

[24] L. Brotcorne, M. Labbé, P. Marcotte, and G. Savard. A bilevel model and solution algorithm for a freight tariff setting problem. *Transportation Science*, 34:289–302, 2000.

[25] M. Červinka, J. Outrata, and M. Pištěk. On stability of M-stationary points in MPCCs. *Set-Valued and Variational Analysis*, 22(3):575–595, 2014.

[26] H. S. Chung, R. D. Weaver, and T. L. Friesz. Oligopolies in pollution permit markets: A dynamic game approach. *International Journal of Production Economics*, 140(1):48–56, 2012.

[27] F. H. Clarke. *Optimization and Nonsmooth Analysis*. SIAM Publications Classics in Applied Mathematics, 1983.

[28] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski. *Nonsmooth Analysis and Control Theory*. Springer, 1998.

[29] G. Colombo, R. Henrion, N. D. Hoang, and B. S. Mordukhovich. Optimal control of the sweeping process. *Dynamics of Continuous, Discrete and Impulsive Systems Series B: Applications & Algorithms*, 19:117–159, 2012.

[30] A. A. Cournot. *Recherches sur les Principes Mathematiques de la Theorie des Richesses*. 1838.

[31] B. Dacorogna. *Direct Methods in the Calculus of Variations*, volume 78. Springer, 2008.

[32] G. Dal Maso, G. A. Francfort, and R. Toader. Quasistatic crack growth in finite elasticity. *ArXiv Mathematics e–prints*, 2004.

[33] S. Dempe. *Foundations of Bilevel Programming*. Springer, 2002.

[34] S. Dempe. Annotated Bibliography on Bilevel Programming and Mathematical Programs with Equilibrium Constraints. *Optimization*, 52(3):333–359, 2003.

[35] J. Dieterich. Applications of rate- and state-dependent friction to models of fault slip and earthquake occurence. Chap.4. In *Earthquake Seismology (H. Kanamori, ed.)*, Treatise on Geophys. 4, pages 107–129. Elsevier, 2007.

[36] T. Donchev, E. Farkhi, and B. S. Mordukhovich. Discrete approximations, relaxation, and optimization of one-sided Lipschitzian differential inclusions in Hilbert spaces. *J. Differ. Equ.*, 243(2):301–328, 2007.

[37] A. Dontchev and R. T. Rockafellar. Characterizations of strong regularity for variational inequalities over polyhedral convex sets. *SIAM J. Optim.*, 6(4):1087–1105, 1996.

[38] A. Dontchev and R. T. Rockafellar. *Implicit Functions and Solution Mappings*. Springer, 2009.

[39] D. Drusvyatskiy and A. Ioffe. Quadratic growth and critical point stability of semi-algebraic functions. *Mathematical Programming*, pages 1–19, 2014.

[40] J. F. Edmond and L. Thibault. Relaxation of an optimal control problem involving a perturbed sweeping process. *Math. Program. Ser. B*, 104(2-3):347–373, 2005.

[41] K. Emich, R. Henrion, and W. Römisch. Conditioning of linear-quadratic two-stage stochastic optimization problems. *Mathematical Programming*, 148(1-2):201–221, 2014.

[42] E. Emmrich. Discrete versions of Gronwall's lemma and their application to the numerical analysis of parabolic problems. *Prepr. Ser. Inst. Math. Tech. Univ. Berlin*, 637, 1999.

[43] G. Ewald. *Combinatorial Convexity and Algebraic Geometry*. Springer, 1996.

[44] F. Facchinei, H. Jiang, and L. Qi. A smoothing method for mathematical programs with equilibrium constraints. *Mathematical Programming*, 85A:107–134, 1999.

[45] M. L. Flegel, C. Kanzow, and J. Outrata. Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints. *Set-Valued Anal.*, 15(2):139–162, 2007.

[46] R. Fletcher and S. Leyffer. Numerical experience with solving MPECs as NLPs. Technical Report NA/210, Department of Mathematics, University of Dundee, 2002.

[47] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes. Local convergence of sqp methods for mathematical programs with equilibrium constraints. *SIAM J. on Optimization*, 17(1):259–286, 2006.

[48] M. Frost, B. Benešová, and P. Sedlák. A microscopically motivated constitutive model for shape memory alloys: formulation, analysis and computations. *Math. Mech. of Solids*, 2014. doi:10.1177/1081286514522474.

[49] M. Goresky and R. MacPherson. *Stratified Morse Theory*. Springer, 1988.

[50] J. Gwinner. On differential variational inequalities and projected dynamical systems- equivalence and a stability result. *Discret. Contin. Dyn. Syst. Supl.*, pages 467–476, 2007.

[51] J. Gwinner. On a new class of differential variational inequalities and a stability result. *Math. Program. Ser. B*, 139(1-2):205–221, 2013.

[52] W. Han and B. D. Reddy. *Plasticity (Mathematical Theory and Numerical Analysis)*. Springer, New York, 1999.

[53] W. Han, M. Shillor, and M. Sofonea. Variational and numerical analysis of a quasistatic viscoelastic problem with normal complience, friction and damage. *J. Comput. Appl. Math.*, 137:377–398, 2001.

[54] R. Henrion and J. Outrata. On calculating the normal cone to a finite union of convex polyhedra. *Optim. A J. Math. Program. Oper. Res.*, 57(1):57–78, 2008.

[55] R. Henrion and W. Römisch. On M-stationary points for a stochastic equilibrium problem under equilibrium constraints in electricity spot market modeling. *Appl. Math.*, 52(6):473–494, 2007.

[56] M. Hintermüller and I. Kopacka. Mathematical programs with comple-

mentarity constraints in function space: C- and strong stationarity and a path-following algorithm. *SIAM J. Optim.*, 20(2):868–902, 2009.

[57] M. Hintermüller, B. S. Mordukhovich, and T. M. Surowiec. Several approaches for the derivation of stationarity conditions for elliptic MPECs with upper-level control constraints. *Math. Program.*, 146(1-2):555–582, 2014.

[58] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE Constraints*. Springer, 2012.

[59] T. Hoheisel, C. Kanzow, and A. Schwartz. Convergence of a local regularization approach for mathematical programmes with complementarity or vanishing constraints. *Optimization Methods and Software*, 37(3):483–512, 2012.

[60] A. D. Ioffe and J. Outrata. On metric and calmness qualification conditions in subdifferential calculus. *Set-Valued Anal.*, 16(2-3):199–227, 2008.

[61] K. Ito and K. Kunisch. Optimal Control of Elliptic Variational Inequalities. *Appl. Math. Optim.*, 41(3):343–364, 2000.

[62] H. Jiang and D. Ralph. Extension of quasi-newton methods to mathematical programs with complementarity constraints. *Computational Optimization and Applications*, 25(1-3):123–150, 2003.

[63] M. E. Jiménez, J. Cañas, V. Mantič, and J. E. Ortiz. Numerical and experimental study of the interlaminar fracture test of composite-composite adhesively bonded joints. (in spanish). In *Materiales Compuestos 07*, pages 499–506, Universidad de Valladolid. Asociación Española de Materiales Compuestos.

[64] D. A. Kandilakis and N. S. Papageorgiou. Evolution inclusions of the subdifferential type depending on a parameter. *Comment. Math. Univ. Carolinae*, 33(3):437–449, 1992.

[65] M. Kočvara, M. Kružík, and J. Outrata. On the control of an evolutionary equilibrium in micromagnetics. In S. Dempe and V. Kalashnikov, editors, *Optimization with Multivalued Mappings*, volume 2 of *Springer Optimization and Its Applications*, pages 143–168. Springer, 2006.

[66] M. Kočvara and J. Outrata. On the modeling and control of delamination processes. In J. Cagnol and J.-P. Zolesio, editors, *Control Bound. Anal.*, pages 171–190. Marcel Dekker, New York, 2004.

[67] P. Krejčí and A. Vladimirov. Lipschitz continuity of polyhedral Skorokhod maps. *Zeitschrift für Analysis und Ihre Anwendungen*, 20(4):817–844, 2000.

[68] P. Krejčí. Evolution variational inequalities and multidimensional hysteresis operators. In P. Drabek, P. Krejci, and P. Takac, editors, *Nonlinear Differ. equations*, pages 47–110. 1999.

[69] P. Krejčí. A remark on the local Lipschitz continuity of vector hysteresis operators. *Appl. Math.*, 46(1):1–11, 2001.

[70] P. Krejčí and T. Roche. Lipschitz continuous data dependence of sweeping processes in BV spaces. *Discret. Contin. Dyn. Syst. - Ser. B*, 15(3):637–650, 2011.

[71] P. Krejčí and J. Sprekels. Temperature–dependent hysteresis in one-dimensional thermovisco–elastoplasticity. *Appl. Math*, 43:173–205, 1998.

[72] M. Kunze and M. D. P. M. Marques. An introduction to Moreau's sweeping process. *Impacts Mech. Syst.*, 551:1–60, 2000.

[73] U. Langer and O. Steinbach. Coupled finite and boundary element domain decomposition methods. In M. Schanz and O. Steinbach, editors, *Boundary Element Analysis*, volume 29 of *L. N. in Appl. Comput. Mech.*, pages 61–95, 2007.

[74] A. S. Lewis and M. L. Overton. Nonsmooth optimization via quasi–Newton methods. *Mathematical Programming*, pages 1–29, 2012.

[75] S. Leyffer. MacMPEC: AMPL collection of MPECs, 2005.

[76] S. Leyffer, G. López-Calva, and J. Nocedal. Interior methods for mathematical programs with complementarity constraints. *SIAM J. on Optimization*, 17(1):52–77, 2006.

[77] T.-C. Lim. On fixed point stability for set-valued contractive mappings with applications to generalized differential equations. *J. Math. Anal. Appl.*, 110(2):436–441, 1985.

[78] S. Lu and A. Budhiraja. Confidence regions for stochastic variational inequalities. *Math. Oper. Res.*, 38(3):545–568, 2013.

[79] S. Lu and S. Robinson. Normal fans of polyhedral convex sets. *Set–Valued Analysis*, 16(2):281–305, 2008.

[80] Z. Q. Luo, J. S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996.

[81] B. Maury and J. Venel. A discrete contact model for crowd motion. *ESAIM Math. Model. Numer. Anal.*, 45(1):145–168, 2011.

[82] A. Mielke. Evolution of rate-independent systems. In C. M. Dafermos and E. Feireisl, editors, *Handb. Differ. Equations Evol. Equations*, number October 2004, pages 461–559. 2006.

[83] A. Mielke and T. Roubíček. *Rate-Independent Systems - Theory and Application*. Appl. Math. Sci. Series. Springer, New York, 2015. to appear.

[84] F. Mignot and J. P. Puel. Optimal control in some variational inequalities. *SIAM J. Control Optim.*, 22(3):466–477, 1984.

[85] Y. Moon, T. Yao, and T. L. Friesz. Dynamic pricing and inventory policies: A strategic analysis of dual channel supply chain design. *Service Science*, 2(3):196–215, 2010.

[86] B. S. Mordukhovich. Generalized differential calculus for nonsmooth and set–valued mappings. *Journal of Mathematical Analysis and Applications*, 183(1):250–288, 1994.

[87] B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation I*. Springer, 2006.

[88] B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation II*. Springer, 2006.

[89] B. S. Mordukhovich, N. M. Nam, and N. D. Yen. Subgradients of marginal functions in parametric mathematical programming. *Math. Program.*, 116(1-2):369–396, 2007.

[90] B. S. Mordukhovich and J. Outrata. Coderivative analysis of quasi-variational inequalities with applications to stability and optimization. *SIAM J. Optim.*, 18(2):389–412, 2007.

[91] J. J. Moreau. On unilateral constraints, friction and plasticity. In G. Capriz and G. Stampacchia, editors, *New Variational Techniques in Mathematical Physics*, C.I.M.E. Summer Schools, pages 171–322. Springer Berlin Heidelberg, 1974.

[92] J. Outrata. A generalized mathematical program with equilibrium constraints. *SIAM J. Control Optim.*, 38(5):1623–1638, 2000.
[93] J. Outrata, J. Jarušek, and J. Stará. On optimality conditions in control of elliptic variational inequalities. *Set-Valued Var. Anal.*, 19(1):23–42, 2011.
[94] J. Outrata, M. Kočvara, and J. Zowe. *Nonsmooth approach to Optimization Problems with Equilibrium Constraints.* Kluwer Academic Publishers, Boston, 1998.
[95] J. Outrata and D. Sun. On the coderivative of the projection operator onto the second-order cone. *Set-Valued Anal.*, 16(7-8):999–1014, 2008.
[96] J.-S. Pang and D. E. Stewart. Differential variational inequalities. *Math. Program. Ser. A*, 113(2):345–424, 2008.
[97] J.-S. Pang and D. E. Stewart. Solution dependence on initial conditions in differential variational inequalities. *Math. Program.*, 116(1-2):429–460, 2009.
[98] N. S. Papageorgiou. Continuous dependence results for subdifferential inclusions. *Publ. l'Institut Mathématique. Nouv. Série*, 52(66):47–60, 1992.
[99] N. S. Papageorgiou. On parametric evolution inclusions of the subdifferential type with applications to optimal control problems. *Trans. Am. Math. Soc.*, 347(1):203–231, 1995.
[100] N. S. Papageorgiou. Parametrized relaxation for evolution inclusions of the subdifferential type. *Arch. Math.*, 31(1):9–28, 1995.
[101] F. París and J. Cañas. *Boundary Element Method, Fundamentals and Applications.* Oxford Univ. Press, 1997.
[102] M. Pflaum. *Analytic and Geometric Study of Stratified Spaces.* Springer, 2001.
[103] H. Pieper. Algorithms for Mathematical Programs with Equilibrium Constraint with Applications to Deregulated Electricity markets. *Dissertation thesis*, 2001.
[104] R. A. Poliquin and R. T. Rockafellar. Tilt stability of a local minimum. *SIAM J. Optim.*, 8:287–299, 1998.
[105] S. Robinson. Some continuity properties of polyhedral multifunctions. *Mathematical Programming Study*, (14):206–214, 1981.
[106] S. M. Robinson. Strongly regular generalized equations. *Math. Oper. Res.*, 5(1):43–62, 1980.
[107] R. T. Rockafellar. Convex functions, monotone operators and variational inequalities. *Theory and Applications of Monotone Operators, proc. NATO Institute*, 1968.
[108] R. T. Rockafellar. *Convex Analysis.* Princeton Univ. Press, 1970.
[109] R. T. Rockafellar. Maximal monotone relations and the second derivatives of nonsmooth functions. *Ann. Inst. Henri Poincaré*, 2(3):167–184, 1985.
[110] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis.* Springer, 1998.
[111] T. Roubíček. Maximally-dissipative local solutions to rate-independent systems and application to damage and delamination problems.
[112] T. Roubíček. Adhesive contact of visco-elastic bodies and defect measures arising by vanishing viscosity. *SIAM J. Math. Anal.*, 45(1):101–126, 2013.
[113] T. Roubíček. A note about the rate-and-state-dependent friction model in a thermodynamical framework of the Biot-type equation. *Geophysical J. Intl.*, 199:286–295, 2014.

[114] T. Roubíček, C. G. Panagiotopoulos, and V. Mantic. Quasistatic adhesive contact of visco-elastic bodies and its numerical treatment for very small viscosity. *ZAMM - J. Appl. Math. Mech.*, 93(10-11):823–840, 2013.

[115] T. Roubíček, C. G. Panagiotopoulos, and V. Mantič. Quasistatic adhesive contact of visco-elastic bodies and its numerical treatment for very small viscosity. *Zeits. für angew. Math. u. Mechanik*, 93:823–840, 2013.

[116] T. Roubíček, C. G. Panagiotopoulos, and V. Mantič. Local-solution approach to quasistatic rate-independent mixed-mode delamination. *Math. Meth. Models Appl. Sci.*, submitted.

[117] A. Sadjadpour and K. Bhattacharya. A micromechanics inspired constitutive model for shape-memory alloys. *Smart Mater. Structures*, 16:1751–1765, 2007.

[118] S. Sauter and C. Schwab. *Boundary Element Methods.* Springer, Berlin, 2011.

[119] H. Scheel and S. Scholtes. Mathematical Programs with Complementarity Constraints: Stationarity, Optimality, and Sensitivity. *Math. Oper. Res.*, 25(1):1–22, 2000.

[120] S. Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 11:918–936, 2001.

[121] S. Scholtes. *Introduction to Piecewise Differentiable Equations.* Springer, 2012.

[122] S. Scholtes and M. Stöhr. Exact penalization of mathematical programs with equilibrium constraints. *SIAM J. Control Optim.*, 37(2):617–652, 1999.

[123] H. Schramm and J. Zowe. A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results. *SIAM Journal on Optimization*, 2(1):121–152, 1992.

[124] P. Sedlák, M. Frost, B. Benešová, T. B. Zineb, and P. Šittner. Thermomechanical model for NiTi-based shape memory alloys including R-phase and material anisotropy under multi-axial loadings. *Intl. J. Plasticity*, 39:132–151, 2012.

[125] A. Skajaa. Limited memory BFGS for nonsmooth optimization, 2010.

[126] G. V. Smirnov. *Introduction to the Theory of Differential Inclusions.* American Mathematical Society, 2001.

[127] U. Stefanelli. A variational principle for hardening elastoplasticity. *SIAM J. Math. Anal.*, 40:623–652, 2008.

[128] A. Tasora, M. Anitescu, S. Negrini, and D. Negrut. A compliant viscoplastic particle contact model based on differential variational inequalities. *International Journal of Non-Linear Mechanics*, 53(0):2–12, 2013.

[129] L. Távara, V. Mantič, E. Graciani, J. Cañas, and F. París. Analysis of a crack in a thin adhesive layer between orthotropic materials: an application to composite interlaminar fracture toughness test. *CMES*, 58:247–270, 2010.

[130] L. Távara, V. Mantič, E. Graciani, and F. París. BEM analysis of crack onset and propagation along fiber-matrix interface under transverse tension using a linear elastic-brittle interface model. *Eng. Anal. Bound. Elem.*, 35:207–222, 2011.

[131] L. Thibault. Sweeping process with regular and nonregular sets. *J. Differ. Equ.*, 193(1):1–26, 2003.

[132] M. Ulbrich. *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces.* SIAM Series on Optimization, 2011.

[133] H. v. Stackelberg. *Marktform und Gleichgewicht.* Springer-Verlag, Berlin, 1934. engl. transl.: The Theory of the Market Economy, Oxford University Press, 1952.

[134] A. Vasil'ev. Continuous dependence of the solutions of differential inclusions on the parameter. *Ukrainian Mathematical Journal*, 35(5):520–524, 1983.

[135] A. I. F. Vaz and L. N. Vicente. A particle swarm pattern search method for bound constrained global optimization. *Journal of Global Optimization*, 39(2):197–219, 2007.

[136] S. Veelken. A New Relaxation Scheme for Mathematical Programs with Equilibrium Constraints: Theory and Numerical Experience. *Dissertation thesis*, 2009.

[137] J. Venel. A numerical scheme for a class of sweeping processes. *Numer. Math.*, 118(2):367–400, 2011.

[138] A. Visintin. *Differential Models of Hysteresis.* Springer, 1994.

[139] R. Vodička, V. Mantič, and T. Roubíček. Quasistatic normal-compliance contact problem of visco-elastic bodies with Coulomb friction implemented via SGBEM/QP. In preparation.

[140] G. Wachsmuth. Strong stationarity for optimal control of the obstacle problem with control constraints. *SIAM Journal on Optimization*, 24(4):1914–1932, 2014.

[141] S. Wang and F. A. Lootsma. A hierarchical optimization model of resource allocation. *Optimization*, 28:351–365, 1994.

A. Auxiliary lemmas

In this appendix chapter we present several auxiliary lemmas which were used earlier in the text.

A.1 Lemmas for Chapter 3

Lemma A.1.1. *Consider continuous functions $g_i : \mathbb{R}^n \to \mathbb{R},\ i = 1, \dots, I$ and affine linear $h_j : \mathbb{R}^n \to \mathbb{R},\ j = 1, \dots, J$ and define the following set*

$$A = \{x |\, g_i(x) < 0,\ h_j(x) = 0,\ i = 1, \dots, I,\ j = 1, \dots, J\}.$$

Then A is relatively open. Moreover, if g_i are convex for all $i = 1, \dots, I$ and A is nonempty, then

$$\operatorname{cl} A = \{x |\, g_i(x) \leq 0,\ h_j(x) = 0,\ i = 1, \dots, I,\ j = 1, \dots, J\}. \tag{A.1}$$

Proof. Since g_i are continuous, $A_1 := \{x |\, g_i(x) < 0,\ i = 1, \dots, I\}$ is an open set. As h_j are affine linear, we know that $A_2 := \{x |\, h_j(x) = 0,\ j = 1, \dots, J\}$ is an affine subspace. Thus, $A = A_1 \cap A_2$ is relatively open.

To prove the second result, denote the right–hand side of (A.1) by B. Clearly, we have $\operatorname{cl} A \subset B$ without any additional assumptions. To show the opposite inclusion, consider any $x \in B$. Since A is nonempty, there exists $\bar{x}$ such that $g_i(\bar{x}) < 0$ and $h_j(\bar{x}) = 0$. Due to the assumptions, we know that $x_n := (1 - \frac{1}{n})x + \frac{1}{n}\bar{x} \in A$ and $x_n \to x$, which finishes the proof. □

Lemma A.1.2. *Assume that $A \subset \mathbb{R}^n$ is convex and relatively open and consider some $x \in A$ and $y \in \operatorname{cl} A$. Then for all $\lambda \in (0, 1)$ we have $\lambda x + (1 - \lambda)y \in A$.*

Proof. The statement is a direct consequence of [108, Theorem 6.1]. □

Lemma A.1.3. *Consider a normally admissible stratification $\{\Gamma_s |\, s = 1, \dots, S\}$ of Γ and some $\mathcal{S} \subset \{1, \dots, S\}$. Then*

$$\bigcap_{s \in \mathcal{S}} \operatorname{cl} \Gamma_s = \bigcup_{\{t |\, \mathcal{S} \subset I(t)\}} \Gamma_t. \tag{A.2}$$

Proof. Assume that $x \in \operatorname{cl} \Gamma_s$ for all $s \in \mathcal{S}$. Then there exists some t such that $x \in \Gamma_t$. But this means that $x \in \Gamma_t \cap \operatorname{cl} \Gamma_s$ for all $s \in \mathcal{S}$ and thus $s \in I(t)$ for all $s \in \mathcal{S}$, meaning that $\mathcal{S} \subset I(t)$.

On the other hand, consider any t such that $\mathcal{S} \subset I(t)$. Then for any $s \in \mathcal{S}$, we have $s \in \mathcal{S} \subset I(t)$, and thus $\Gamma_t \subset \operatorname{cl} \Gamma_s$, which finishes the proof. □

Lemma A.1.4. *For a polyhedral set C consider its all nonempty relatively open faces C_s with $s = 1, \dots, S$. Then $\{C_s |\, s = 1, \dots, S\}$ forms a normally admissible stratification of C.*

Proof. Since all properties of Definition 3.2.2 apart from formula (3.3) obviously hold, it remains to verify this formula. Consider thus some C_s and C_i such that $C_s \cap \operatorname{cl} C_i \neq \emptyset$. Since we can write

$$\begin{aligned} C &= \{x |\, \langle c_t, x\rangle \leq b_t,\ t = 1, \ldots, T\}, \\ C_s &= \{x |\, \langle c_t, x\rangle < b_t,\ t \in \mathcal{T}_{11},\ \langle c_t, x\rangle = b_t,\ t \in \mathcal{T}_{12}\}, \\ \operatorname{cl} C_i &= \{x |\, \langle c_t, x\rangle \leq b_t,\ t \in \mathcal{T}_{21},\ \langle c_t, x\rangle = b_t,\ t \in \mathcal{T}_{22}\}, \end{aligned}$$

where $\mathcal{T}_{j1} \cap \mathcal{T}_{j2} = \emptyset$ and $\mathcal{T}_{j1} \cup \mathcal{T}_{j2} = \{1, \ldots, T\}$ for $j = 1, 2$ and since there is some $x \in C_s \cap \operatorname{cl} C_i$, we have $\mathcal{T}_{11} \subset \mathcal{T}_{21}$ and thus $C_s \subset \operatorname{cl} C_i$, which finishes the proof. □

A.2 Lemmas for Chapter 5

The first lemma is a slight modification of the well–known fact that simple functions are dense in L^p for any $p \in [1, \infty)$.

Lemma A.2.1. *Let $u \in L^p([0,T])$ for any $p \in [1,\infty)$ and $\{0 = t_0^k \leq t_1^k \leq \cdots \leq t_k^k = T\}$ be a collection of divisions of the interval $[0,T]$ such that their norms converge to zero. Then there exists sequence $\{u^k\}$ of simple functions, constant on intervals $[t_j^k, t_{j+1}^k)$ for every $j = 0, \ldots, k-1$, which converges to u in L^p.*

Proof. Since continuous functions with compact support are dense in L^p, we can construct a sequence of continuous functions y^k such that $y^k \to u$ in L^p. Define now the function u^k as follows:

$$u^k(t) := \min_{s \in [t_j^k, t_{j+1}^k)} y^k(s)$$

for $t \in [t_j^k, t_{j+1}^k)$. Since any continuous function on a compact set is uniformly continuous by the Heine–Cantor Theorem, we immediately obtain that $\|u^k - y^k\|_\infty \to 0$, which leads to the statement of the lemma.. □

The following lemma is a discrete version of the Gronwall's Lemma. We have taken the full version from the original source [42] and simplified it to the needed form stated in the corollary. Note that $\theta = 0$ corresponds to the forward Euler discretization scheme and $\theta = 1$ to the backward one.

Lemma A.2.2. *Let C be closed convex set. Then for all $x \in C$ and for all w the following conditions are equivalent.*

1. $w \in \mathrm{N}_C(x)$
2. $\forall \xi : \langle w, \xi\rangle \leq |w| d_C(\xi + x)$
3. $\exists k > 0\ \exists \delta > 0\ \forall v, |v| < \delta : \langle w, v\rangle \leq k d_C(v + x)$.

Proof. For the proof, see [137, Lemma 2.4]. □

List of Figures