University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Grzegorz Dudziuk

Optimization of closed-loop controls by thermostats for a class of nonlinear reaction-diffusion processes

PhD dissertation

Supervisor

prof. dr hab. Marek Niezgódka

Interdisciplinary Centre for Mathematical and Computational Modelling
University of Warsaw

May 2014

Author's declaration:

Aware of legal responsibility, I hereby declare that I have written this dissertation myself and all the contents of the dissertation have been obtained by legal means.

. . . .

date   *Grzegorz Dudziuk*

Supervisors' declaration:

The dissertation is ready to be reviewed.

. . . .

date   *prof. dr hab. Marek Niezgódka*

Abstract

In this dissertation, both qualitative and numerical analysis for an optimization problem is performed for a feedback control law applied to a class of nonlinear reaction-diffusion processes. A finite number of control and measurement devices target their actions inside the process domain. The measurement devices collect data on the process evolution, while the control devices obtain those data and activate an appropriate reaction. The aim of this control system is to keep the process evolution close to a user-defined reference state. The above optimization problem consists in choosing geometrical targeting of the control and measurement devices actions according to a suitable optimality criterion.

Such an idea of the closed-loop control of reaction-diffusion processes is implemented by a system of equations with a semilinear PDE coupled to several nonlinear ODEs. The cost functional utilized for a precise definition of the announced problem of optimal targeting is constructed as an integral of the difference between the process and reference states.

The present work is divided into two main parts. The first of them focuses on analysis of the PDE-ODE model under consideration. The second one concerns the problem of optimal targeting, exploiting some of the results of the first part.

In the analysis of the PDE-ODE model we focus on questions concerning existence, uniqueness and stability of solutions as well as on the efficiency of the closed-loop control mechanism implemented there. By efficiency we mean here an ability of moving the process close to the reference state. The existence, uniqueness and stability proofs are provided. The efficiency of the closed-loop control is validated by results of numerical simulations for the investigated PDE-ODE model. The numerical results suggest that the efficiency of the considered closed-loop control depends on changes of the model parameters. Moreover, the long-time behavior visible in the subject simulations also is examined. In all simulations, the process appeared to tend to some time-invariant state, after sufficiently long time. In some cases, that time-invariant state seemed to be, at some rate, independent of the initial condition of the PDE-ODE model.

In the part on the optimal targeting problem, we first focus on analytical questions. We prove there the existence of minimizers and characterize the differential of the cost functional too. Then, we describe numerical optimization experiments, utilizing three gradient optimization algorithms (the steepest descent and two variants of the nonlinear conjugate gradient) and compare their performance. Here, the aforementioned characterization of the cost functional differential is used to implement the formula for the gradient. The results show how the performance of the optimization algorithms varies with changes of the parameters entering the cost functional. It is also shown that modifications of the subject parameters can result in independence of the optimization output on the initial condition of the PDE-ODE model.

Streszczenie

W niniejszej rozprawie przeprowadzone są zarówno jakościowa, jak i numeryczna analiza problemu optymalizacji sterowania ze sprzężeniem zwrotnym zastosowanego do pewnej klasy nieliniowych procesów reakcji-dyfuzji. Skończona liczba urządzeń sterujących i pomiarowych skupia swoje działania wewnątrz obszaru procesu. Urządzenia pomiarowe zbierają dane o ewolucji procesu, następnie urządzenia sterujące otrzymują zebrane dane i uruchamiają odpowiednią reakcję. Celem sterowania jest utrzymać ewolucję procesu blisko zdefiniowanego przez użytkownika stanu referencyjnego. Powyżej wspomniany problem optymalizacji polega na ustaleniu geometrycznego wycelowania działań urządzeń sterujących i pomiarowych w odniesieniu do odpowiedniego kryterium optymalności.

Przedstawiona idea sterowania w układzie zamkniętym procesem reakcji-dyfuzji jest zaimplementowana poprzez układ równań z semiliniowym równaniem różniczkowym cząstkowym sprzężonym z wieloma nieliowymi równaniami różniczkowymi zwyczajnymi. Funkcjonał kosztu wykorzystany na potrzeby precyzyjnej definicji zapowiedzianego problemu optymalnego wycelowania jest skonstruowany jako całka z różnicy między stanem procesu a stanem referencyjnym.

Niniejsza praca podzielona jest na dwie główne części. Pierwsza z nich skupia się na analizie wspomnianego układu równań. Druga część dotyczy problemu optymalnego wycelowania, wykorzystując pewne rezultaty z części pierwszej.

W części dotyczącej analizy wspomnianego układu równań skupiam się na pytaniach dotyczących istnienia, jednoznaczności oraz stabilności rozwiązań, jak również na skuteczności sterowania w układzie zamkniętym zaimplementowanego w rozważanym układzie. Przez skuteczność rozumiem zdolność do sprowadzania procesu w pobliże stanu referencyjnego. Zaprezentowane są dowody istnienia, jednoznaczności oraz stabilności. Skuteczność rozważanego sterowania w układzie zamkniętym jest zilustrowana za pomocą rezultatów symulacji numerycznych dotyczących badanego układu równań. Rezultaty numeryczne sugerują, że skuteczność rozważanego sterowania w układzie zamkniętym zależy od parametrów układu równań. Dodatkowo, poczynione są obserwacje dotyczące zachowania dla dużych czasów widocznego w przedmiotowych symulacjach. We wszystkich symulacjach proces zdawał się dążyć, po upływie odpowiedniego czasu, do pewnego stanu niezmienniczego w czasie. W niektórych przypadkach zaobserwowany stan niezmienniczy wydawał się być w pewnym stopniu niezależny od stanu początkowego dla rozważanego układu równań.

W części dotyczącej problemu optymalnego wycelowania najpierw skupiam się na pytaniach analitycznych. Dowodzę instnienia minimizerów oraz charakteryzuję różniczkę funkcjonału kosztu. Następnie opisuję eksperymenty dotyczące numerycznej optymalizacji, wykorzystujące trzy gradientowe algorytmy optymalizacji (największy spadek oraz dwa warianty nieliniowego gradientu sprzężonego) oraz porównuję ich wydajność. Wspomniana przed chwilą charakterzacja różniczki funkcjonału kosztu wykorzystana jest do implementacji formuły na gradient. Rezultaty pokazują, że wydajność algorytmów optymalizacji zmienia się wraz ze zmianami parametrów funkcjonału kosztu. Pokazane jest również, że modyfikacje przedmiotowych parametrów mogą skutkować niezależnością wyników optymalizacji od warunku początkowego rozważanego układu równań.

Keywords

closed-loop control, thermostats, optimal targeting, existence and uniqueness, stability analysis, differentiation in Banach spaces, numerical simulations, numerical optimization, steepest descent method, nonlinear conjugate gradient method

AMS Mathematics Subject Classification

35-04, 35A05, 35B20, 35K10, 35K55, 35B37, 49-04, 49K20, 90-04, 90C30, 90C90

Słowa kluczowe

sterowanie w układzie zamkniętym, termostaty, optymalne wycelowanie, istnienie i jednoznaczność, analiza stabilności, różniczkowanie w przestrzeniach Banacha, symulacje numeryczne, optymalizacja numeryczna, metoda największego spadku, metoda nieliniowego gradientu sprzężonego

Klasyfikacja tematyczna według AMS

35-04, 35A05, 35B20, 35K10, 35K55, 35B37, 49-04, 49K20, 90-04, 90C30, 90C90

Acknowledgements

I would like to express my gratitude towards prof. Marek Niezgódka (Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw), who supported me during my research with several valuable ides and comments. Conversations with prof. Marek Niezgódka allowed me to get deeper insight into possible applications of the mathematical problems treated in this dissertation. Prof. Marek Niezgódka supervised my work during all stages of preparing this dissertation.

I would also thank prof. Willi Jäger (Interdisciplinary Center for Scientific Computing, University of Heidelberg) for his support during the part of my work concerning the optimization problem which is the subject of Chapter 3 and Chapter 4 of the present dissertation. The comments of prof. Willi Jäger gave me better understanding of various optimization strategies and indicated interesting directions of further research.

During preparation of this dissertation, a lot of kindness towards my questions was demonstrated also by Maria Gokieli, Anna Trykozko and Krzysztof Nowiński (Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw), who spared their time for discussions concerning my research.

During the research described in this dissertation, I obtained a financial support as a participant of *Ph.D. Programme: Mathematical Methods in Natural Sciences*, which is a project financed by *International Ph.D. Projects* program of Foundation for Polish Science financed within the *Innovative Economy Operational Programme* 2007-2013 funded by European Regional Development Fund.

The numerical optimization simulations described in Chapter 4 were performed on computer cluster Halo2 (Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw), in frames of computational grant G51-0 of Interdisciplinary Centre for Mathematical and Computational Modelling.

The numerical simulations described in Chapter 2 and Chapter 4 were performed with use of the GNU Octave software.

Except Figure 0.1, pictures presented in the present work were generated with use of the GNU Octave software. The picture presented in Figure 0.1 was created with Inkscape.

Contents

Introduction

This question addresses a range of questions on closed-loop control of nonlinear distributed systems governed by a combination of partial and ordinary differential equations. The control system set-up comprises a finite number of measurement devices and a finite number of control devices. We analyze such a class of control systems, addressing the existence and uniqueness of solutions, the efficiency of the closed-loop controls and their optimization.

Mathematical models applied in science suffer from inaccuracies originating due to at least two sources:

1) First, the models represent only approximations of real phenomenas.

2) Second, also the values of model parameters frequently only approximate the values which are, in some sense, the best (optimal).

In the thesis, we consider a control system imposed on a process governed by the reaction-diffusion equation:

$$y_t(x,t) \;-\; \Delta y(x,t) \;=\; f\big(y(x,t)\big) \;+\; \hat{\mathbf{u}}(x,t) \tag{0.A}$$

with a control term $\hat{\mathbf{u}}$. The control term is a model parameter, selected according to a particular aim of the control, for instance, to reach a given state $y^* = y^*(x)$ at a given time T.

In the above context, first, one faces the question to what extend the semilinear reaction-diffusion equation is a precise representation of the underlying process. But even though the above semilinear equation, with certain concrete f, were considered to be satisfactory in this connection, a second question, concerning the choice of the control term $\hat{\mathbf{u}}$ (the model parameter), would be faced. The choice of $\hat{\mathbf{u}}$ should provide a „sufficiently accurate" approximation of the optimal $\hat{\mathbf{u}}$. Here, the meaning of optimality is determined by the above mentioned aim of the control.

The direct approach concerning the issue 2) as above consists in improving the approximation of optimal values of the model parameters. However, this approach has several limitations:

- In general, the only way to approximate those optimal values is based on numerical approaches. As often the numerical optimization is computationally of high complexity, such a treatment proves time consuming.

- The results of numerical optimization usually remain different from actual optimal values. This produces a next obstacle for models of instable nature, where even small perturbation of model parameters can result in big changes in the solution of the model.

- In (0.A), an optimal parameter $\hat{\mathbf{u}}$, being the control variable, depends not only on the control objective (to achieve a state y^* at time T) but also on the initial condition of the model. Thus, a change of the initial condition results in a necessity of computing the model parameter again.

All above considerations refer actually to the open-loop set-up of the control problem. The latter shows, as discussed, its obvious limitations. As an alternative concept, a closed-loop set-up can be developed. In this context, our approach shall be to accept the parameters inaccuracies in the model and extend the model of an additional mechanism of automatic real-time parameters corrections, basing on the observed actual evolution of the model solution. In the context of (0.A), this idea can be implemented by allowing the parameter $\hat{\mathbf{u}}$ to depend on the solution of the model itself:

$$\hat{\mathbf{u}}(x,t) \;=\; \hat{\mathbf{u}}\big(x,t,y(\,.\,,t)\big)$$

or more generally

$$\hat{\mathbf{u}}(x,t) \;=\; \hat{\mathbf{u}}\big(x,\,t,\,y|_{\mathrm{space}\times[0,t)}(\,.\,,\,.\,)\big) \tag{0.B}$$

The latter formula stresses that the values of $\hat{\mathbf{u}}$ in a given moment of time t can be computed using the whole information about the past behavior of the solution y (not only the information on the present time t). The above idea of *automatic correction mechanisms* assumes that a computational algorithm for the values of the model parameters is given. Such an algorithm will be called *a feedback law* in our thesis. In our control theory model (0.A), under assumption that the term $\hat{\mathbf{u}}$ is of form (0.B), the feedback law can be identified with the definition of $\hat{\mathbf{u}}$.

The above approach involving the idea of automatic correction mechanisms, potentially, may be a way to overcome the aforementioned difficulties, because:

- With such an approach, a computationally expensive procedure of searching for the optimal values of the parameters can be unnecessary.
- Since the basic idea of the discussed approach is not to predict the behavior of the solution of the model *a priori*, but to react to the behavior of the solution in real time, the following consequences, hypothetically, can be faced:
  a) The approach can be effective in the case of the models exhibiting unstable nature. Here, by the effectiveness we mean the result of making the behavior of the solutions of the model close, in a suitable sense, to a desired reference.
  b) In (0.A), with the objective to reach a state y^* at time T, a control involving the automatic correction idea (i.e. a parameter $\hat{\mathbf{u}}$ of form (0.B)) can occur to preserve the control effectiveness under changes of the initial condition. In consequence, the proposed approach may help to avoid computing the model parameter every time when the initial condition is changed.
  c) In control systems, a control based on the automatic corrections idea may prove effective even if the utilized description of the underlying process (i.e. the equation $y_t - \Delta y = f(y)$, in the case of the model (0.A)) is inaccurate. In other words, closed-loop controls of the considered type can preserve the effectiveness under changes of the model. Thus, in the control theory context, the automatic correction idea can also offer a solution to the issue 1) as formulated above.

The aim of this thesis is to demonstrate the concept of automatic correction mechanism in the control theory model (0.A) in a specific implementation, *a control by thermostats*. The control by thermostats assumes that the feedback law, built into the control term $\hat{\mathbf{u}}$, relies on a finite system measurement devices and control devices. The measurement devices gather the information on the current state of the process. The control devices influence the process, basing on the information provided by the measurement devices.

Questions on the models with an automatic correction mechanism that we shall address include:

I) Whether a given model with the latter type of mechanism is mathematically well posed, i.e. its solutions exist and are unique, further are stable subject to the data perturbations. We investigate this questions for a model with control by thermostats in Chapter 1.

II) Whether the automatic correction mechanism applied in a given model indeed ensures an effectiveness and insensitivity to changes of the initial condition. We focus on this questions in Chapter 2, where results of numerical simulations for a model with control by thermostats are exposed.

III) It is natural to ask a question concerning possibilities of refining the effectiveness of a given model with an automatic correction mechanism. This leads to the problem of optimization of the feedback law, constituting the automatic correction mechanism. Investigating the latter problem for the model with control by thermostats is the main aim of the present work and is the subject of Chapter 3 and Chapter 4.

The set-up of the optimal feedback problem may seem in dissonance with the former remarks, as one of the highlighted advantages of automatic correction mechanisms was their low computational cost due to avoiding computationally expensive optimization procedures. However, the latter dissonance is only virtual. First, some of the numerical prototypes described in Chapter 2 show that an effective feedback law can be defined heuristically, without optimization procedures involved. Still, even if one is able to intuitively construct *a good* feedback law, searching for *a better* one remains natural and hence our interest in the related optimization problem. Second, as mentioned, in the context of the control theory, other possible advantage of automatic correction mechanisms is its insensitivity to the changes the initial condition (for the control by thermostats, it also seems to be the case in certain situations, as the results described in Chapter 2 indicate). In consequence, for a given model with a given aim of the control, re-optimizing the feedback law every time the initial condition is changed may be unnecessary. In such cases, it is sufficient to perform the optimization procedure just once.

To formulate the subject optimization problem precisely, a parametrization of the feedback law is necessary. To this end, we assume that the thermostat feedback law is parametrized by the localization of the actions of control and measurement devices. In other words, in Chapter 3 and Chapter 4, we will focus on the problem of choosing optimal localizations of the actions of those devices.

For the optimization problem, a number of related questions will be explored. In Chapter 3, theoretical aspects as the existence of minimizers of a suitable cost functional and the analysis of its differentiability will be examined. Chapter 4 outlines the results of related numerical simulations. There, the problem of choice of an appropriate optimization method and the question concerning independence of the optimal feedback law on the initial condition of the model are discussed.

In the remaining part of *Introduction* we set the framework for the thesis. The precise definition of the model with control by thermostats addressed in this work is given in §1. In §2, we formulate the optimization problem that will be considered throughout the thesis. Some possible applications of the control by thermostats, as well as bibliographical information concerning the latter control concept, are exposed in §3. Finally, §4 provides bibliographical notes concerning the present dissertation, a summary of its results in the above mentioned fields of interest, as well as a more details on the content of the subsequent chapters.

§1 Model with the control by thermostats

In the present work, we take into consideration the following mathematical model, realizing the concept of control by thermostats:

$$\begin{cases} y_t(x,t) - D\Delta y(x,t) = f(y(x,t)) + \sum_{j=1}^{J} g_j(x)\kappa_j(t) & \text{on } Q_T \\ \dfrac{\partial y}{\partial n} = 0 & \text{on } \partial\Omega \times (0,T) \\ y(x,0) = y_0(x) & \text{for } x \in \Omega \end{cases} \tag{0.1}$$

together with

$$\begin{cases} \beta_1\kappa_1'(t) + \kappa_1(t) = W_1\big(y(\,.\,,t), y^*(x,t)\big) & \text{on } [0,T] \\ \vdots & \vdots \\ \beta_J\kappa_J'(t) + \kappa_J(t) = W_J\big(y(\,.\,,t), y^*(x,t)\big) & \text{on } [0,T] \\ \kappa_j(0) = \kappa_{j0} \in \mathbb{R} & \text{for } j = 1,\dots,J \end{cases} \tag{0.2}$$

where $Q_T = \Omega \times (0,T)$, $T > 0$ and $\Omega \subset \mathbb{R}^{\mathbf{d}}$ is a bounded domain with sufficiently regular boundary. The unknown in the above equations is $(y, \kappa_1, \dots, \kappa_J)$, where $y\colon Q_T \to \mathbb{R}$ and $\kappa_j\colon [0,T] \to \mathbb{R}$. The term $f\colon \mathbb{R} \to \mathbb{R}$ represents a given nonlinearity. The diffusion coefficient $D > 0$ is given, as well as coefficients $\beta_1, \dots, \beta_J > 0$. Functions $y^*\colon Q_T \to \mathbb{R}$ and $g_j\colon \Omega \to \mathbb{R}$ also are known. The functionals W_j are defined as follows, for $j = 1, \dots, J$:

$$W_j(y(\,.\,,t), y^*(\,.\,,t)) = \sum_{k=1}^{K} \alpha_{jk} w_k\Big(\int_\Omega h_k(x)\big(y(x,t) - y^*(x,t)dx\big)\Big) \tag{0.3}$$

where $\alpha_{jk} \in \mathbb{R}$, $w_k\colon \mathbb{R} \to \mathbb{R}$ and $h_k\colon \Omega \to \mathbb{R}$.

In (0.1) - (0.3), y^* describes *a reference trajectory* — the purpose of the introduced model is to stabilize the reaction-diffusion process possibly close to the reference trajectory y^*. If y^* is independent of the time variable, we will call it *a reference state*. Functions g_j are constant in time, characterizing the actions of *control devices* in space. The actions of control devices alternate in time according to the values of functions κ_j, called *response functions* or *power functions*. The response functions depend on the process evolution, described by variable y. This dependence can be described as follows. *Measurement devices*, whose actions are characterized by functions h_k, acquire the data on the current state of the process. Each measurement device is responsible for computing *the measurement value*, represented by the term $\int_\Omega h_k(y - y^*)\,dx$, entering the right hand side of (0.3). The measurement values returned by the measurement devices are processed by functions w_k. The processed measurement data are synthesized by *the signal generator* associated with j-th control device, with *weights* α_{jk}, $k = 1, \dots, K$. The function $W_j(y(\,.\,,t), y^*(\,.\,,t))$, as a function of time, can be interpreted as *the signal* generated by the signal generator for the j-th control device. Next, the j-th control device responses to the input signal. The response of the j-th control device is described by the response function κ_j.

Figure 0.1 illustrates a functional structure of the control mechanism that we have described. The below remarks can be helpful for understanding of the system (0.1) - (0.3):

- Functions κ_j are modeled with ODEs in (0.2), meaning that the changes of the response are continuous in time.

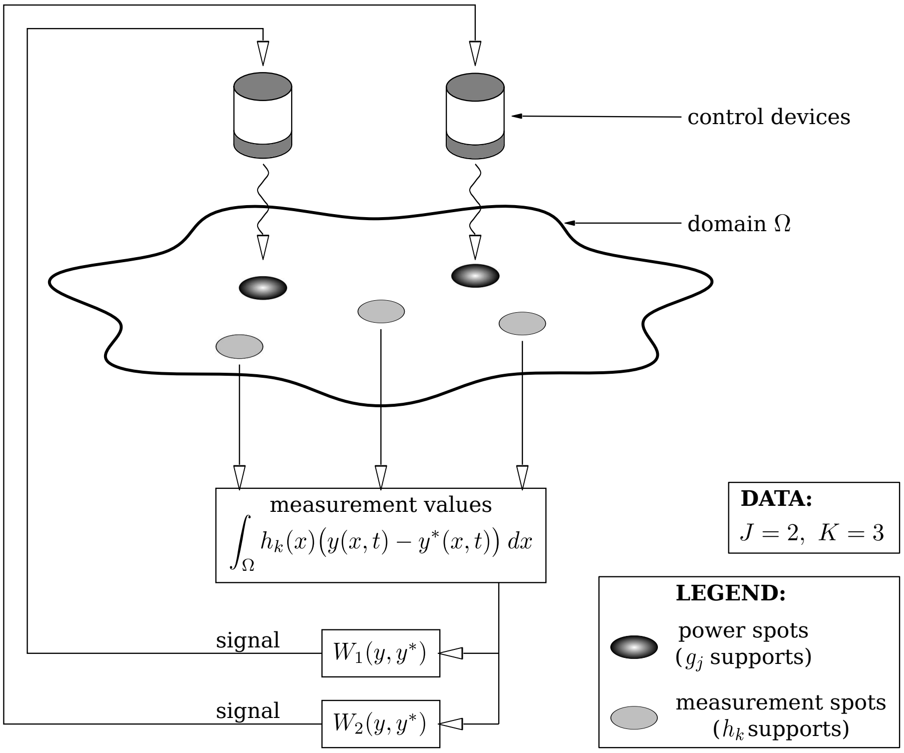

control devices
domain Ω
measurement values
$\int_\Omega h_k(x)\big(y(x,t)-y^*(x,t)\big)\,dx$
DATA:
$J=2,\ K=3$
LEGEND:
power spots (g_j supports)
measurement spots (h_k supports)
signal
$W_1(y,y^*)$
signal
$W_2(y,y^*)$

Figure 0.1: Schematic presentation of the closed-loop control concept, implemented in the system (0.1) - (0.3), for the case of two control devices and three measurement devices.

- A natural example of the functions g_j and h_k is a characteristic function of a small ball, being a subset of Ω, times a constant. If this is the case for h_k, then the measurement devices return measurement values representing the mean difference between the current process state and the reference trajectory in a neighborhood (the ball supporting h_k) of certain point (the center of the ball). If g_j are functions as above, then the control devices deliver the energy uniformly over the balls being the supports of g_j.

- For the functions w_k, a natural example is $w_k(s) = -sgn(s)$. In this case, the function w_k returns simple information understood by the signal generators as „cool down" or „heat up", depending on whether the k-th measurement value indicates that the process values exceed the reference values or are below them. Hence, functions w_k can be understood as functions describing *a switching mechanism* implemented in the system. We will call w_k *the switching functions.*

- The assumption that $\beta_j > 0$ has a practical interpretation. If, for certain $j \in \{1,\ldots,J\}$, $\beta_j > 0$ and the signal W_j in the RHS of (0.2) is zero, then it follows straight by the basic properties of the ODE (0.2) that the power function κ_j tends to zero. This is the behavior which one can intuitively expect, meaning „no signal — no power". And the opposite, if one assumed that $\beta_j < 0$, then the power function κ_j would tend to infinity for the signal W_j equal zero and nonzero initial condition κ_{j0}, what is a less natural behavior.

- Functions h_k in the system (0.1) - (0.3) describe measurement abilities of specific measurement devices, not just measurement devices understood as physical units. Similarly, functions g_j describe power spots created in the process domain by control devices rather than physical devices itself. Putting the latter in another way, functions g_j and h_k do not describe the mechanism of work of the control and measurement devices, but only the effect of the work of the devices.

Note that the control devices can be placed outside the domain of the controlled process. For example, the control devices can be electromagnetic transmitting antennas, placed outside the domain and focusing the electromagnetic waves at some spot placed inside the domain. Then, the function g_j describe the spatial distribution of the intensity of the electromagnetic effects generated in the domain by the j-th antenna.

- The above interpretation of the role of g_j and h_k has quite essential consequences. If one assumed that g_j describe physical units, then one could expect some additional no-collision restrictions, as e.g. the condition of disjoint supports of all functions g_j or the condition that the supports of g_j are contained in Ω. Instead, we only assume that g_j describe some immaterial energy injections, hence there is no reason to forbid intersections of the supports of g_j or to forbid the supports of g_j to intersect with the exterior of Ω. An analogous remark holds for functions h_k.

- In many situations it is natural that the control devices act through the boundary of the domain. Even if the control devices are physically located in the process domain, then the volume they occupy should not be the influenced by the control action. To achieve this, for example, one could modify the domain of the process and exclude the volume occupied by the control devices from the domain, what in fact leads to a model with control acting through some part of the boundary (i.e. the part being the boundary of the volumes occupied by the devices).

  Hence, if one interpreted functions g_j in the model (0.1) - (0.3) as physical units, then the model might seem not quite realistic. But, as mentioned, functions g_j do not describe the physical units and can be understood e.g. as functions describing electromagnetic effects in some volume of the domain, generated by electromagnetic antennas placed outside the domain. With this interpretation, the model (0.1) - (0.3) becomes coherent.

Throughout this thesis, we will keep the above interpretation of the system (0.1) - (0.3), assuming that functions g_j and h_k do not represent physical objects. Instead, g_j and h_k will be assumed to characterize the actions of the control and measurement devices actions.

In the present work, we will use the term *the control by thermostats* or *the thermostat control mechanism* to refer to the control concept applied in the system of equations (0.1) - (0.3) for controlling the reaction-diffusion process. In the literature, some variants of the above described control concept were already considered. We will briefly comment on those variants in §3. For further convenience, we will call the mentioned variants thermostat control mechanisms or controls by thermostats, as well. Thus, in the present work, the notion of „the thermostat control mechanism" or „the control by thermostats" refers to a family of closed-loop control concepts, to which the control concept applied in (0.1) - (0.3) belongs.

Remark. For $D = 1$, the system (0.1) - (0.3) can be understood as a particular case of the equation (0.A) with the control term $\hat{\mathbf{u}}$ of form (0.B). Indeed, it suffices to set $\hat{\mathbf{u}} := \sum_{j=1}^{J} g_j \kappa_j$ in (0.1). Equations (0.2) and (0.3) can be understood as conditions describing the feedback law for computing functions κ_j and hence the term $\hat{\mathbf{u}}$. It follows by (0.2) and (0.3) that functions κ_j depend on y, or more precisely, that $\kappa_j(t)$, for given $t \in (0, T)$, depends on the past values of y, earlier than t. Thus, $\hat{\mathbf{u}}$ defined as proposed above, is a realization of (0.B). ▲

The properties of the system (0.1) - (0.3), such as the existence and uniqueness of solutions, stability of the system or efficiency of the thermostat control mechanism will be the discussed in Chapter 1 and Chapter 2. In Chapter 3 and Chapter 4, the system (0.1) - (0.3) will be considered in the context of optimization of the feedback law implemented by the thermostat control mechanism.

§2 Formulation of the optimal targeting problem

Below, we introduce the optimization problem which will be investigated in Chapter 3 and Chapter 4. The optimality criterion will refer to bringing the state of the controlled process possibly close to a given reference state at time T. In the problem, a feedback control law in (0.1) - (0.3) (i.e. the algorithm for computing the response functions κ_j) will be optimized so as to meet such a requirement. The feedback law will be optimized with respect to the choice of geometrical targeting of control and measurement devices actions.

To this purpose, we will assume that the pattern of energy distributed in the domain by a given control device is fixed and that the user can adjust the energy distribution only by translations of the latter pattern. For instance, the situation can be considered where a control device can produce a uniform energy distribution in a small ball-shaped volume and the user is expected just to choose the center of the volume. An analogous assumption will be made for the measurement devices, stating that the measurement abilities of the measurement devices are described by fixed patterns and can be adjusted only by spatial translations of the subject patterns.

We pursue the above concept by the following mathematical assumptions.

We will understand *the control* as the set of all functions characterizing control and measurement devices along with weights entering to (0.1) - (0.3), i.e. the control is $(g_j, h_k, \alpha_{jk})_{j=1,\ldots,J}^{k=1,\ldots,K}$. The choice of control determines the feedback law in (0.1) - (0.3), assuming that functions w_k and coefficients β_j are prescribed. Let functions $\sigma_g, \sigma_h : \mathbb{R}^d \to \mathbb{R}$ and points $x_1, \ldots, x_J$ and $z_1, \ldots, z_K$ in $\mathbb{R}^{\mathbf{d}}$ be given. We assume that the functions describing the control and measurement devices actions are given by

$$g_j(x) := \sigma_g(x - x_j)|_\Omega, \qquad h_j(x) := \sigma_h(x - z_k)|_\Omega \tag{0.4}$$

for $j = 1, \ldots, J$, $k = 1, \ldots, K$. Functions σ_g and σ_h will be called *the pattern functions*. For example, in the case of control devices distributing energy uniformly in a ball-shaped volume, one can set $\sigma_g := C_g \mathbf{1}_{B(0,r_g)}$, with parameters C_g and r_g chosen accordingly. Points x_j and z_k characterize *targeting* of specific control and measurement devices actions.

Under the above assumptions, for prescribed pattern functions σ_g and σ_h, the control is determined by a choice of targetings $x_1, \ldots, x_J$ and $z_1, \ldots, z_K$ as well as weights $\alpha_{1,1}, \ldots, \alpha_{J,K}$. However, we do not plan to address the problem of optimal choice of weights in the termostats control system. In the thesis we focus on the problem of optimal targeting of the devices actions. To this end, we make the following simplifying assumptions. We postulate that

$$K = J \tag{0.5}$$

and that

$$z_j = x_j \quad \text{for } j = 1, \ldots, J \tag{0.6}$$

In addition, we set

$$\alpha_{j,k} := \delta_{j,k} \quad \text{for } j, k = 1, \ldots, J \tag{0.7}$$

As a result, the problem of choice of the weights disappears.

Now, with assumptions (0.4), (0.5), (0.6) and (0.7), for fixed pattern functions σ_g and σ_h, the choice of targetings $x_1, \dots, x_J$ determines the control and hence the feedback law in the system (0.1) - (0.3). For this reason, the sequence $(x_1, \dots, x_J)$ will be called *the control parameter.*

Assumptions (0.4), (0.5), (0.6) and (0.7) together can be interpreted as a set conditions that the control and measurement devices are pairwise coupled in the thermostat control mechanism.

We are now ready to formulate the complete optimization problem to be studied in Chapter 3 and Chapter 4. Let the pattern functions σ_g and σ_h be given. The problem is to choose the control parameter in an optimal manner, with respect to the criterion of minimizing the following cost functional:

$$(x_1, \dots, x_J) \;\mapsto\; \widetilde{\lambda} \int_{T_0}^{T} \int_{\Omega} \left| y(x,t) - y^*(x,t) \right|^2 dxdt \tag{0.8}$$

for certain $\widetilde{\lambda} > 0$, $T_0 \in [0, T)$, where y^* is a reference trajectory entering the system (0.1) - (0.3) and y is the first component of solution $(y, \kappa_1, \dots, \kappa_J)$ of the system (0.1) - (0.3) with conditions (0.4), (0.5), (0.6) and (0.7), corresponding to the control parameter $(x_1, \dots, x_J)$.

The minimization problem for the cost functional (0.8) can be referred to as *the problem of optimal targeting of control and measurement devices actions.* However, it will be convenient to have a shorter name, thus in this thesis we shall refer to it as *the optimal targeting problem.*

Remark. The cost functional (0.8) reflects the idea of measuring the gap between the process evolution and the reference state. In particular, setting T_0 close to T and $\widetilde{\lambda} = (T - T_0)^{-1}$, the above cost functional approximates the gap at time T of the experiment. As such, the subject cost functional is appropriate to describe the idea of bringing the process state close to the reference state at the terminal time T, mentioned in the beginning of §2. ▲

Remark. Due to our interpretation of the system (0.1) - (0.3), which allows intersections of the supports of functions g_j and h_j with each other and with the exterior of Ω (see §1), we do not impose any control parameter restrictions for preventing the subject intersections. Thus, we will view the optimal targeting problem as an unconstrained optimization problem, consisting in minimization of the cost functional (0.8) over whole $\left(\mathbb{R}^{\mathbf{d}}\right)^J$. ▲

Remark. It will be convenient for the reader to remember the terminology introduced in §1 and §2 of the present chapter (reference trajectory, switching functions, control parameter, optimal targeting problem e.t.c.) because we will use it frequently in this work. ▲

§3 Control by thermostats in the literature and possible applications

We will now give some comments on the history and variants of the concept of control by thermostats. We also remark on possible applications.

In the mathematical literature, the idea of control by thermostats of processes governed by evolutionary PDEs was probably introduced first in [26], [25]. There, a parabolic linear heat flow was controlled by thermostats. A model of control by thermostat of a parabolic linear heat flow was considered also in [10]. However, the applications of thermostat control mechanisms were not limited to control of linear parabolic PDEs. The work [11] addressed the control by thermostats of a thermodynamical process modeled by the telegraph equation. In [30] and [19], the authors focused on models with processes described by a semilinear equation controlled by thermostats, in [12] a system of semilinear equations with an additional convolution term was considered in the context of control by thermostats. A lot of attention was directed toward control by thermostats of phase transition processes modeled by various versions of the Stefan model, see e.g. [23], [33], [28], [15]. The strain and temperature in a viscoelastic body subject to a thermodynamical process were controlled by thermostats in the model presented in [29]. A problem of control of saturation in a model of filtration of a porous medium was considered in [5], with the control involving the thermostat concept. In more recent works [31] and [32], a model for control by thermostats of a linear heat flow was considered.

Not only the controlled process varies in the models considered in the mentioned works. The thermostat control mechanism also has its variants. One of the point where the differences in the thermostat control mechanism can occur is the placement of actions of the control devices. In all indicated references, except for [19] and [30], the control devices are acting through the boundary of the process. In [19] and the present work the control devices create a power spot distributed in the domain of the controlled process. In [30], the control acts both through the boundary and as a quantity distributed in the domain.

Also, various versions of the switching mechanism, being a part of the thermostat control mechanism, can be found in the literature. A frequently encountered case is that hysteresis in the work of the switching mechanism is assumed to be present. See [33], [28], [29], [11], [12], [31], [32] for applications of the so-called relay switch hysteresis or [23], [10], [15], [28], [29], [12], [5], [30] for the Preisach hysteresis model. In [33], [15] or [19], the case of no hysteresis effects in the switching mechanism was addressed. In the present work, we also do not assume hysteresis effects.

The version of the thermostat control mechanism investigated in this work is very similar to that in [19] or one of the cases taken into account in [33].

Certain potential applications of the thermostat control mechanisms have been already indicated above, in the description of mathematical literature. They cover control of thermodynamical processes, strain in viscoelastic bodies, saturation of porous media and phase transition processes. Besides, the control concepts similar to the concept of the thermostat control mechanism were present also in technical literature.

In this context, we mention the application of thermostat control mechanism mechanisms in the hypertermia cancer therapy. Roughly speaking, hyperthermia consists in heating the body of a patient to influence the cancer tissue. See [48], [46] for general overview of the latter therapy method, its variants and limitations. According to those references, one of the variants of hyperthermia assumes ultrasounds or electromagnetic waves to be the heating medium, delivering energy directly to the deep tissues of the body of the patient. A typical strategy in this hyperthermia variant is to heat the cancer tissue area to a possibly high temperature without rising the temperature in the neighboring tissues above certain critical level. A feedback information concerning the heating results is necessary. The measurement actions can be carried out by interstitial heat probes or the magnetic resonance imaging.

The model (0.1) - (0.3) can be understood as describing the above situation, assuming that the domain Ω represents the heated tissue. Note that the subject variant of the hyperthermia, consisting in the deep heating, is coherent with our interpretation of the functions g_j in the model (0.1) - (0.3), describing the control effects in a certain volume of the domain of the controlled process. In the model (0.1) - (0.3), the strategy of selective heating the tumor can be implemented by a proper choice of the reference state y^*, describing a desired temperature distribution.

In many publications addressing hyperthermia, the feedback information obtained by magnetic resonance is utilized to control the actions of the heating medium transmitters. Control mechanisms which share control concepts in certain way related to the concept of thermostat control mechanisms are described (examples can be found in [42], [8]). However, methods basing on other control concepts also were introduced in the hypertermia-related publications (for instance, see [16], [35], [47]).

In the context of hyperthermia, an interesting hybrid control concept is presented in [36], combining a thermostat-like control concept for controlling the power of the control devices in time with other kind of control strategy for the control of energy delivery in space. The latter strategy consists in optimization of the control devices settings, and hence, indirectly, in optimization of the patterns of the spatial distribution of the delivered energy. Thus, at the level of general concepts, the aims of the control mechanism in [36] are similar to the aims of both our thermostat control mechanism and our optimal targeting problem, introduced in §1 and §2. Nonetheless, comparing to our work, many differences occur there. In particular, the control mechanism in [36] assumes other feedback law in the thermostat-like mechanism used there and there considered optimization problem is formulated in significantly other way.

§4 Summary of the results and bibliographical notes

Below, we sketch the plan of the present work, summarize the main results and provide bibliographical notes. Chapter 1 and Chapter 2 are focused purely on the properties of the system (0.1) - (0.3) and do not touch the optimal targeting problem. The optimal targeting problem, associated with the cost functional (0.8), is the subject of Chapter 3 and Chapter 4.

In Chapter 1, we focus on analytical properties of the system (0.1) - (0.3). Two main problems are addressed in this chapter. The first one is: what can be proven if we decide to put discontinuous switching functions in the system (0.1) - (0.3), e.g. if we put $w_k = -sgn$. Unfortunately, in this case we prove only existence of solutions, without any uniqueness results. Moreover, we prove the existence result not for the system (0.1) - (0.3) directly, but for its modification (see comments below). The second problem consists in proving existence, uniqueness and stability w.r.t. perturbations of control for solutions of the system (0.1) - (0.3), under sufficiently strong assumptions. These sufficiently strong assumptions exclude the possibility of discontinuous switching functions. Knowledge on the existence, uniqueness and stability w.r.t. control for the system (0.1) - (0.3) is essential also in further parts of the thesis, concerning directly the optimal targeting problem formulated in §2. Hence, investigating the above properties is necessary prior to proceed up to this optimization problem. For both problems, Lipschitz continuity of the reactive term f in the system (0.1) - (0.3) is assumed.

The first of the problems, concerning discontinuous switching functions in the system (0.1) - (0.3), is treated in Section 1.1. Our approach is the following one. For a given discontinuous switching function w_k, we replace it with a multivalued upper semicontinuous mapping $\widetilde{w}_k$ whose graph contains the graph of w_k. This means that the right hand side of (0.2) becomes a multivalued mapping. Hence, in Section 1.1, we temporarily replace the differential equation (0.2) with a differential inclusion, obtaining a modified version of the system (0.1) - (0.3). As mentioned, we prove only the existence of solutions for the postulated modification of the system (0.1) - (0.3). The proof of the existence theorem exploits the generalized Kakutani fixed-point theorem.

The second problem, concerning existence, uniqueness and stability topics for the system (0.1) - (0.3), is considered in Section 1.2. Here, we conduct our reasoning under the assumption of Lipschitz continuity of the switching functions. This means that (0.2) becomes equality again rather than inclusion, what brings us back to analysis of the system (0.1) - (0.3). In Section 1.2, stability of solutions of the system (0.1) - (0.3) under perturbations of control is proven, with the mentioned assumption on Lipschitz continuity of w_k and with the assumption that $y^* \in L^2(0,T;L^2(\Omega))$. Under the same assumptions, stability w.r.t. perturbations of the initial condition is shown, what proves the uniqueness of solutions of (0.1) - (0.3). The existence result also is shown, with additional restriction for y^* and w_k, namely that one of the following hypotheses is fulfilled: 1) $y^* \in L^2(0,T;L^2(\Omega))$ and w_k are bounded or 2) $y^* \in L^\infty(0,T;L^2(\Omega))$. Eventually, as a complementary result, we prove also weak stability of solutions of the system (0.1) - (0.3), under the same assumptions under that the stability and uniqueness are proven. In Section 1.2, we provide also generalization of some of the above mentioned results for the case of f only locally Lipschitz with certain growth condition and y_0 essentially bounded.

In Chapter 2, we present results of numerical simulations for the thermostat control mechanism, involved in the system (0.1) - (0.3). These simulations were intended to give an insight into the properties of the system in some aspects not touched in Chapter 1.

In particular, Chapter 1 does not concern the efficiency of the thermostat control mechanism in any sense, i.e. does not give an information whether the thermostat control mechanism, described by (0.1) - (0.3), brings the process close to the reference state y^* or not. Thus, in Chapter 2, we describe numerical results illustrating efficiency of the thermostat control mechanism, in the above sense.

As a second focus of our attention in the analysis of the numerical results, we take into account the problem of dependence on the initial state y_0 of the large time behavior of the process controlled by thermostats (i.e. of solution component y in the system (0.1) - (0.3)). The information on independence of the process state at the terminal time T on the initial state are important for the optimal targeting problem, considered in Chapter 3 and Chapter 4. To be precise, if the process state at the terminal time T is independent of the initial state then, perhaps, the cost functional (0.8) also becomes independent of the initial state, assuming T_0 close to T. In consequence, the local minimums of the cost functional become independent of the initial state.

In our simulations, two-dimensional square domain was considered and a triangulation of triangular elements was used. To obtain the results, the system of equations was treated with finite element method combined with the implicit Euler scheme. The finite element space was the space of continuous functions, linear on each element of the triangulation. The nonlinear terms entering (0.1) - (0.3) were treated with the use of Picard iterations.

The simulations addressed the cases of various reference states y^*, various initial states y_0 and various configurations of the control and measurement devices in the thermostat control mechanism, described by (0.1) - (0.3).

The simulation results suggest that the efficiency of thermostat control mechanism differs with changes of the model parameters. As a general rule, greater number of the control and measurement devices, not surprisingly, results in better efficiency. Moreover, in all simulations, stabilization of the process near to some time-invariant state was observed. The independence of the subject time-invariant states on the initial state was observed in some, but not in all, of the simulations.

In Chapter 3, we report an analysis of the optimal targeting problem, announced in §2. The main objective of Chapter 3 is to derive a formula characterizing the gradient of the cost functional (0.8). The gradient formula will be necessary further, in Chapter 4, to perform optimization procedures for approximation of local solutions of the subject optimization problem. Chapter 3 is split into two parts: 1) part concerning the properties of the operator assigning the solution of the system (0.1) - (0.3) to a given control parameter, let us call this operator *the state operator* and 2) part concerning properties of the mentioned cost functional, including the formula for its gradient.

In Section 3.1, we investigate the properties of the state operator. By the existence and uniqueness results from Chapter 1, in Section 3.1 we easily justify that the state operator is well defined. Moreover, by the stability results from Chapter 1, we show that the state operator is Lipschitz continuous. In comparison to the results stating that the state operator is well defined, its Lipschitz continuity requires additionally stronger assumptions for the pattern functions σ_g and σ_h. Eventually, in Section 3.1 we prove also the weak Gâteaux differentiability of the state operator and characterize its weak Gâteaux differential. This is the main result of Section 3.1, necessary also in further considerations, concerning the properties of the cost functional. As we will see, the Lipschitz continuity of the state operator is essential to prove its weak Gâteaux differentiability. In addition, the proof the weak Gâteaux differentiability of the state operator assumes that both the nonlinear term f and the switching functions w_k, $k = 1, \ldots, K$ in the system (0.1) - (0.3) are everywhere differentiable in the classical sense.

In Section 3.2, we investigate the properties of the cost functional (0.8). First, we introduce a simple criterion for existence of minimizers in the subject optimization problem. This criterion assumes that the pattern functions σ_g and σ_h have compact supports. Next, we focus on the matter of differentiability of the cost functional. We show that it is Gâteaux differentiable under the same conditions under which the state operator is weakly Gâteaux differentiable. Finally, we derive a formula for the gradient of the cost functional, what is the main result of Section 3.2.

In Chapter 4, we present results of numerical optimization experiments concerning the optimal targeting problem. Chapter 4 complements the theoretical material provided in Chapter 3 by presenting attempts to construct concrete solutions of the investigated optimization problem.

The simulations described in Chapter 4 were intended mainly 1) to compare performance of various optimization methods for various parameters of the subject optimization problem and 2) to check whether the optimization output is independent of the initial state y_0, entering the system (0.1) - (0.3), when the parameter T_0 in the cost functional (0.8) is close to T.

The independence of the optimization output on y_0 is related with the independence of the process state at the terminal time on y_0 (see the remarks concerning Chapter 2). Since the latter independence was observed in some cases in the simulations described in Chapter 2, one can expect that the former independence, concerning the optimization output, also is possible. The independence of the optimization output on the initial state y_0, if exists, would mean that it is not necessary to re-optimize the feedback law constituting the thermostat control mechanism each time the initial state is changed (see the expectations expressed in the beginning of *Introduction*).

The numerical optimization experiments were performed with the use of steepest descent method (SD method, in short) and nonlinear conjugate gradient method (CG method). The CG method variant was implemented in the Polak-Ribière mode, with a certain modification. Two subvariants of the CG method were considered: 1) the method with a reset of the search direction every N_{dim} iterations, where N_{dim} stands for dimension of the optimization space (CG+r method) and 2) the method without the latter reset procedure (CG-r method). The stop criterion utilized in the experiments was a short step criterion. To implement the optimization methods, we rely on the gradient characterization derived in Chapter 3.

We have compared performance for the three optimization methods (SD, CG-r, CG+r) for three variants of the initial state y_0, three reference states y^* and two values of the left edge, T_0, of the integration interval in the definition of the cost functional (0.8). Here, by performance of an optimization method we mean the number of iterations necessary to meet the stop criterion. The two considered values of T_0 were 1) zero and 2) a value close to terminal time T for the system (0.1) - (0.3). Thus, in case 2), the value of T_0 corresponded to the idea of measuring the gap between the process and the reference state in neighborhood of the terminal time of the system (0.1) - (0.3).

The results show that the average performance of the SD method was much inferior in the case of the parameter T_0 close to T than in the case of T_0 equal zero. Nevertheless, the difference in the average performance of the SD method for two different values of T_0 was leveled by using the CG+r method instead of SD.

We have also compared the average performance of the CG+r method for a given reference state and T_0 close to T, for varying values of parameter T ($T = 2, 4, 6$) and for two variants of y_0. It occurred that the performance of the CG+r method was better in the case of $T = 2$ than in the case of $T = 4$ or $T = 6$.

Hence a hypothesis that the average performance of optimization methods for our optimization problem changes both with changes of T_0 (when using the SD method)) and with changes of T (when using the CG+r method). For changes of T_0, the use of stronger optimization method (CG+r instead of SD) levels the performance differences, while for changes of T, the performance differences occur despite using CG+r.

Other observation concerning our experimental results with varying T is that the optimization output becomes more independent of y_0 when lengthening time horizon T. This stays in accordance with intuition. Unfortunately, greater T results in higher computational cost. Thus, if our observation was a general rule, the desired effect of the independence of the optimization output on y_0 could be expected for those values of parameter T which result in a computationally more expensive numerical treatment of the optimization problem.

Bibliographical notes. As remarked in §3, the thermostat control mechanism was taken into account in the mathematical literature in different versions. The thermostat control mechanism present in the model (0.1) - (0.3) was inspired by and is similar to the version considered in [19] or one of the versions considered in [33]. However, in comparison to those works, we make additional assumptions for the switching functions in the thermostat control mechanism to get stronger results (except Section 1.1, where the assumptions for the switching functions are as in the given references).

The analytical results presented in Section 1.2 of Chapter 1 and in Chapter 3 are obtained with rather standard mathematical methods. The methods utilized in Section 1.2 are an adaptation of methods presented in many PDE handbooks to the PDE-ODE system (0.1) - (0.3). The approach presented in Section 3.1 of Chapter 3 for the investigation of the differentiability of the state operator was inspired, in particular, by some of the arguments utilized in [39]. Some of the key concepts utilized in in Section 3.2 of Chapter 3 for the characterization of the differential of the cost functional base on the methods broadly described in the handbook [45].

The methods utilized to obtain the main result of Section 1.1 (Theorem 1.1.2) are probably less standard (the generalized Kakutani theorem, the properties of multivalued mappings). The latter methods were applied in a similar fashion to models with a similar version of the control by thermostats in works [33] and [19].

To our knowledge, rigorous mathematical analysis of the problem of optimal targeting of the actions of control and measurement devices in PDE models involving thermostat control mechanisms was not performed so far. The latter remark concerns both the variant of the thermostat control mechanism present in the model (0.1) - (0.3) as well as its other variants, present in the models addressed in the mathematical references given in §3. Many other questions were posed for the subject models, including the existence or uniqueness of solutions (see [26], [25], [33], [28], [15], [29], [12], [5]), the existence, or other properties, of time-periodic solutions (see [28], [30], [31], [32]), convergence to stationary solutions (see [28]) or the existence of a global attractor (see [30]). In the mathematical literature, we have encountered only one type of optimization problems for PDE models involving thermostat control mechanisms. It is the problem of choosing the optimal hysteresis law, for the variant of thermostat control mechanism where a switching mechanism with hysteresis was considered — see e.g. [23], [10], [5]. The optimal targeting problem announced in §2, or similar, seems to be not addressed in the mathematical literature.

However, in non-mathematical literature, not providing rigorous mathematical analysis, the problems in certain fashion related to the optimal targeting problem were addressed. For instance, see the reference [36] (some comments on this reference were given in §3).

Some of the results of this thesis were already published in a preliminary form on arXive.org, in the work [18]. This concerns a major part of the content presented in Section 1.2.1, Section 1.2.2 and Chapter 2 of the thesis. Roughly speaking, the content of Section 3 of [18] is included into Section 1.2.1 and Section 1.2.2 of the present dissertation, while the content of Section 4 of [18] is included into Chapter 2. Nevertheless, significant refinements were implemented since the preliminary version in [18]. In Section 1.2.2, the only part imported form [18] is Theorem 1.2.3 and its proof (the latter with certain rearrangements). The rest of Section 1.2.2 is a new content, including the image in Figure 1.3. In Section 1.2.2, the refinements include improved typesetting of mathematical formulas, rearrangements of a big part of the proofs, more precise exposition of certain mathematical arguments and some additional comments. In addition, Section 1.2.2 considers both the case of $y^* \in L^\infty(0,T;L^2(\Omega))$ and $y^* \in L^2(0,T;L^2(\Omega))$, while in [18] we included only the considerations on $y^* \in L^\infty(0,T;L^2(\Omega))$. Chapter 2, in comparison to Section 4 of [18], contains a much more extensive description of the numerical schemes utilized in the simulations and some additional comments. The images in Figures 2.3, 2.4, 2.6 and 2.8 in Chapter 2 represent the same data as some of the images in [18], however they were plotted anew, for better readability. The rest of images in Chapter 2, as well as the tables exposed therein, is the same as corresponding images and tables in [18].

Moreover, some fragments of Section 1 (Introduction) of [18] (text bulk of less than two pages in total) are present in the *Introduction* of the preset dissertation. Section 2 of [18] also is here, splitting its content to *Notation conventions* and the beginning of Chapter 1. To be specific, the list of norms in *Notation conventions*, along with some minor text fragments there, and big parts of the notation remarks in the beginning of Chapter 1 are present in Section 2 in [18].

Notation conventions

In this chapter, we introduce notation which will be binding everywhere else in the present work.

General notation

By „domain" we mean a nonempty open subset of $\mathbb{R}^n$, for some $n \in \mathbb{N} \setminus \{0\}$.

In the present work, $\Omega \subset \mathbb{R}^{\mathbf{d}}$ always denotes the corresponding set appearing in the system (0.1) - (0.3) and is assumed to be a domain. Positive integer $\mathbf{d}$ stands for the dimension of Ω. $T > 0$ is the constant in (0.1) - (0.3) determining the time horizon and $Q_T := \Omega \times (0, T)$.

Unless it is explicitly said to be otherwise, $\mathbb{R}^n$ for an arbitrary $n \in \mathbb{N} \setminus \{0\}$ is always endowed with its standard topology and with Lebesgue measure and so subsets of $\mathbb{R}^n$ are, including Ω.

If F is a function defined on a given set $\mathbb{A}$ and $\widetilde{\mathbb{A}}$ is a subset of $\mathbb{A}$, we denote by $F|_{\widetilde{\mathbb{A}}}$ the restriction of F to $\widetilde{\mathbb{A}}$.

For a given set $\mathbb{A}$ and its subset $\widetilde{\mathbb{A}}$, $\mathbf{1}_{\widetilde{\mathbb{A}}} \colon \mathbb{A} \to \mathbb{R}$ is the indicator function of $\widetilde{\mathbb{A}}$, i.e. $\mathbf{1}_{\widetilde{\mathbb{A}}}(\omega)$ equals 1 for $\omega \in \widetilde{\mathbb{A}}$ and equals 0 for $\omega \notin \widetilde{\mathbb{A}}$.

The function $sgn \colon \mathbb{R} \to \mathbb{R}$ is defined as follows: $sgn(s) = 1$ for $s > 0$, $sgn(s) = -1$ for $s < 0$, $sgn(0) = 0$.

For $j, k \in \mathbb{N}$, we use symbol $\delta_{j,k}$ to denote the Kronecker delta function of j and k, i.e. $\delta_{j,k} = 1$ for $j = k$ and $\delta_{j,k} = 0$ for $j \neq k$.

For vector spaces $\mathbb{X}$, $\mathbb{Y}$ and an operator T acting from $\mathbb{X}$ to $\mathbb{Y}$, we will denote the value of T on an element $x \in \mathbb{X}$ as $T(x)$ or Tx, interchangeably.

Notation for function spaces

Below, any space of scalar functions is understood as a space of real functions and any Banach space is also assumed to be real.

Assume that X is a Banach space. We denote:

X^* — dual of X,

X_w, ${X^*}_{w*}$ — the space X considered with its weak topology and the space X^* considered with its weak-$*$ topology, respectively.

For two Banach spaces X_1 and X_2, $X_1 \hookrightarrow X_2$ means that X_1 can be continuously embedded in X_2. When this notation is used, specification of the embedding operator is necessary. If $X_1 \subseteq X_2$, then we assume that the embedding operator for $X_1 \hookrightarrow X_2$ is the identity operator. If X_1 is a separable, reflexive Banach space, X_2 is a separable Hilbert space and $X_1 \hookrightarrow X_2$ densely, then the embedding operator for $X_2 \hookrightarrow {X_1}^*$ is understood in the standard evolution triples sense (see [51, Chap. 23.4] for explanation of this concept). If none of these two situation takes place, external embedding theorems will be referred in the text to specify the meaning of the embedding operator.

Assume that $k, n \in \mathbb{N} \setminus \{0\}$, $p \in [1, \infty]$, X is a Banach space and let $\mathbb{A}$ be a measure space and $\mathbb{D} \subseteq \mathbb{R}^n$ be a domain. The following notation concerning function spaces will be in use:

$L^p(\mathbb{A})$	—	standard Lebesgue space,
$W^{k,p}(\mathbb{D})$	—	standard Sobolev space,
$H^k(\mathbb{D})$	—	synonym for $W^{k,2}(\mathbb{D})$,
$C(\mathbb{D})$	—	space of real valued continuous functions defined on $\mathbb{D}$ with its standard topology,
$C_c(\mathbb{D})$	—	subspace of $C(\mathbb{D})$ consisting of functions with support that is compact in $\mathbb{D}$,
$L^p(0, T; X)$	—	standard Bochner space,
$C([0, T]; X)$	—	space of continuous functions from $[0, T]$ into X,
$C([0, T]; X_w)$	—	space of weakly continuous functions from $[0, T]$ into X, or in other words, space of continuous functions from $[0, T]$ into X_w,
$C([0, T])$	—	synonym for $C([0, T]; \mathbb{R})$.

Assuming that X is a Banach space, H is a Hilbert space and $\mathbb{E} \subseteq \mathbb{R}^n$ is a measurable set, we denote:

$\lVert \cdot \rVert_X$	—	the norm of X,
$(. , .)_H$	—	the scalar product of H,
$\langle . , . \rangle_{X^*, X}$	—	the natural pairing between X^* and X; the first argument stands for the element of X^*,
$\lVert \cdot \rVert_{p,\mathbb{E}}$	—	the norm of the Lebesgue space $L^p(\mathbb{E})$, $p \in [1, \infty]$,
$\lVert \cdot \rVert_p$	—	the norm of the Lebesgue space $L^p(\Omega)$, $p \in [1, \infty]$,
$\lVert \cdot \rVert_{X,q}$	—	the norm of the Bochner space $L^q(0, T; X)$, $q \in [1, \infty]$,
$\lVert \cdot \rVert_{p,q}$	—	the norm of the Bochner space $L^q(0, T; L^p(\Omega))$,
$\langle . , . \rangle$	—	the natural pairing between $H^1(\Omega)^*$ and $H^1(\Omega)$; the first argument stands for the element of $H^1(\Omega)^*$,
$\left\vert \cdot \right\vert_p$	—	p-th norm in $\mathbb{R}^n$, namely $\left\vert x \right\vert_p := \left(\sum_{i=1}^n \lvert x_i \rvert^p\right)^{1/p}$ for $p \in [1, \infty)$ and $\left\vert x \right\vert_p := \max_{i=1,\dots,n} \lvert x_i \rvert$ for $p = \infty$, where $x \in \mathbb{R}^n$.

In addition, we do not want to bother with separate notation for norms of $\mathbb{R}^n$-valued functions, hence we denote the standard norm of $(L^p(\mathbb{E}))^n$ simply as $\lVert . \rVert_{p,\mathbb{E}}$. Similarly, we denote the norms of $(L^p(\Omega))^n$ and $L^q(0, T; (L^p(\Omega))^n)$ by $\lVert . \rVert_p$ or $\lVert . \rVert_{p,q}$, respectively. The standard scalar product in $\left(L^2(\mathbb{E})\right)^n$ will be denoted as $(. , .)_{L^2(\mathbb{E})}$.

Moreover, for $p \in [1, \infty)$, the space $L^p(0, T; L^p(\Omega))$ can be identified with the space $L^p(Q_T)$. The inclusion $L^p(Q_T) \subseteq L^p(0, T; L^p(\Omega))$ follows by arguments as in the proof of Example 23.4 in Chap. 23.2 in [51], the other inclusion follows by approximation with step functions. Thus, in the present work, we will use these two spaces interchangeably. In particular, we assume that for an arbitrary $F \in L^p(0, T; L^p(\Omega))$ it is legal to evaluate the norm $\lVert F \rVert_{L^p(Q_T)}$ and *vice versa.*

The definitions Lebesgue and Sobolev spaces are contained e.g. in [1, Chap. 2 & Chap. 3], [45, Chap. 2.2] or [21, App.A.3 & Chap. 5.2]. The Bochner spaces are introduced e.g. in [1, Par.

7.4], [21, Chap. 5.9.2], [45, Chap. 3.4.1] or [51, Chap. 23.2]. Space $C([0,T];X)$ is defined e.g. in [45, Chap. 3.4.1], [21, Chap. 5.9.2] or [51, Def. 23.1, Chap. 23.2]. The norms of Lebesgue, Sobolev, Bochner and $C([0,T];X)$ spaces are also defined in the given references.

Notation for differentiation

Let $\mathbb{D} \subseteq \mathbb{R}^n$ be a domain, for certain $n \in \mathbb{N} \setminus \{0\}$. In the present work, for a given function $F\colon \mathbb{D} \to \mathbb{R}$, partial derivative sign $\partial_i F$, for $i = 1, \dots, n$, can refer both to the classical partial derivative and the weak partial derivative. Similarly, $\nabla F(x)$, for $x \in \mathbb{D}$, can denote the vector of classical partial derivatives or weak partial derivatives in x. Analogous remarks hold if $F\colon \mathbb{D} \to \mathbb{R}^m$, for certain $m \in \mathbb{N} \setminus \{0\}$.

The „prim" operator for functions of one variable also can have various meanings. Let $\mathbb{I} \subseteq \mathbb{R}$ be an open interval (finite or infinite) and let F be an X-valued function on $\mathbb{I}$, where X is a given Banach space. Then, depending on the context, F' can refer both to the classical derivative of F or to the vector-valued distributional derivative of F.

To sum up, the „∂_i" and „∇" operators, if not understood in classical sense, refer to weak partial derivatives. The „prim" operator, if not understood in classical sense, refer to the vector-valued distributional derivative of a vector-valued function of one variable. In particular places of the text, the meaning of the subject differential operators should be clear by the context. Otherwise, we will explicitly stress which meaning of the differential operators is involved.

In addition to the above, in the present work, for a given function $F\colon Q_T \to \mathbb{R}$, symbol ∇F always refers to the gradient with respect to the spatial variables. In other words, ∇F does not include the partial derivative with respect to the time variable, associated with interval $(0,T)$, regardless of the meaning of the partial derivatives (classical or weak).

We understand the concept of the weak derivative as in [51, Def. 21.2, Chap. 21.1], [1, Par. 1.62] or [21, Chap. 5.2.1]. The vector-valued distributional derivative concept that we use is described e.g. in [51, Def. 23.15, Chap. 23.5] or [45, Chap. 3.4.3].

Chapter 1

Thermostat control mechanism — properties

The fundamental results for the reaction-diffusion model with an additive control term not involving the automatic correction mechanism (see model (0.A)), as the existence and uniqueness of solutions or stability results, are known. However, introducing the automatic correction mechanism to the control term can turn the original reaction-diffusion model into a model of different algebraic type. This is the case for the model (0.1) - (0.3), which can be understood as the model of reaction-diffusion process with control by a particular automatic correction mechanism. It is straightforward that the results concerning a single reaction-diffusion equation do not apply to the system (0.1) - (0.3). Hence, the analysis of the properties of (0.1) - (0.3) is necessary.

Therefore, in the present chapter, we focus on fundamental analysis of the system (0.1) - (0.3). By fundamental analysis, we understand in particular the results on existence and uniqueness of solutions for (0.1) - (0.3). We present also the results on stability of (0.1) - (0.3) under perturbations of the control and of the initial condition.

The plan of the present chapter is as follows. In Section 1.1, we begin with analysis of the system (0.1) - (0.3) in the case where the switching functions w_k, $k = 1, \ldots, K$, are upper semicontinuous multivalued mappings. This approach has the following advantages:

1. It is possible to prove existence for w_k being upper semicontinuous multivalued mappings,

2. For a discontinuous function, it is possible to find an upper semicontinuous multivalued mapping related to this function in certain sense (see Proposition A.5.5).

Thus, the above approach is an attempt to indirectly handle the case of discontinuous switching functions w_k, including the $-sgn$ function.

A drawback of the proposed approach is that, to our knowledge, no method for proving uniqueness of solutions is known for models with control by thermostats with switching functions being upper semicontinuous multivalued mappings. In the beginning of Section 1.1, we indicate some reference works where the subject approach was exploited. In none of the indicated works, uniqueness was obtained for switching functions being upper semicontinuous multivalued mappings.

In Section 1.2 we investigate the case of stronger restrictions for the switching functions w_k. This restriction consists in assuming that w_k are single-valued, Lipschitz continuous mappings, for $k = 1, \ldots, K$. With the latter assumption, we obtain not only existence but also uniqueness results for the system (0.1) - (0.3). In addition, in Section 1.2 we provide the analysis of stability, with respect to both the control and the initial condition, of the system (0.1) - (0.3) with single-valued Lipschitz w_k. Nevertheless, imposing the latter assumption excludes the possibility of the above proposed approach for dealing with the case of discontinuous switching functions, including $w_k(s) = -sgn(s)$, in the system (0.1) - (0.3). Thus, one may say that in Section 1.2 we trade a method of indirect handling the situation of $w_k = -sgn$ in the system (0.1) - (0.3) for fundamental results for the latter system. On the other hand, a method of indirect handling the case of $w_k = -sgn$ is available also with the assumption of Lipschitz switching functions — with the latter assumption, the function $-sgn$ can be approximated with Lipschitz functions of a very steep slope near point zero.

The purpose of the announced stability analysis is twofold. First, the mentioned uniqueness result for the system (0.1) - (0.3) is in fact proven by using the stability with respect to the initial condition. Second, the results concerning stability w.r.t. the control are useful from the point of view of the optimal control theory, for proving differentiability of so-called state operators. Our results concerning stability w.r.t. the control will be used in Chapter 3 of the present work, exactly for the latter purpose.

Notation remarks

In Chapter 1, we use the following definitions of spaces:

$$\begin{aligned}
X^0 &= L^2(\Omega) \times \mathbb{R}^J \\
X^1 &= L^2(Q_T) \times \left(L^2(0,T)\right)^J \\
X^2 &= \Big\{(y, \kappa_1, \ldots, \kappa_J) \in L^\infty(0,T; L^2(\Omega)) \times \left(L^\infty(0,T)\right)^J : \\
&\qquad y' \in L^2(0,T; H^1(\Omega)^*),\ \nabla y \in \left(L^2(Q_T)\right)^{\mathbf{d}}, \\
&\qquad \kappa_j' \in L^2(0,T) \text{ for } j = 1, \ldots, J\Big\}
\end{aligned}$$

and

$$\begin{aligned}
X^y &= \Big\{y \in L^\infty(0,T; L^2(\Omega)) \colon \nabla y \in \left(L^2(Q_T)\right)^{\mathbf{d}},\ y' \in L^2(0,T; H^1(\Omega)^*)\Big\} \\
X^\kappa &= \Big\{(\kappa_1, \ldots, \kappa_J) \in \left(L^2(0,T)\right)^J : \kappa_j' \in L^2(0,T),\ j = 1, \ldots, J\Big\}
\end{aligned}$$

where natural number J is the same as J appearing in the system (0.1) - (0.3). In the above definitions of spaces: 1) the derivatives y' and κ_j' are assumed to exist in the sense of vector-valued distributional derivatives (see *Notation conventions*) and 2) ∇y is assumed to exist as the vector of the weak partial derivatives of y w.r.t. the spatial variables (see *Notation conventions*).

The topologies of X^0, X^1, X^2, X^y and X^κ are given by the following norms:

$$\begin{aligned}
\left\|(y,\kappa_1,\dots,\kappa_J)\right\|_{X^0} &= \left\|y\right\|_2 + \sum_{j=1}^{J}|\kappa_j| \\
\left\|(y,\kappa_1,\dots,\kappa_J)\right\|_{X^1} &= \left\|y\right\|_{2,2} + \sum_{j=1}^{J}\left\|\kappa_j\right\|_{L^2(0,T)} \\
\left\|(y,\kappa_1,\dots,\kappa_J)\right\|_{X^2} &= \left\|y\right\|_{2,\infty} + \left\|\nabla y\right\|_{2,2} + \left\|y'\right\|_{H^1(\Omega)^*,2} + \\
&\quad + \sum_{j=1}^{J}\left\|\kappa_j\right\|_{L^\infty(0,T)} + \sum_{j=1}^{J}\left\|\kappa_j'\right\|_{L^2(0,T)} \\
\left\|y\right\|_{X^y} &= \left\|y\right\|_{2,\infty} + \left\|\nabla y\right\|_{2,2} + \left\|y'\right\|_{H^1(\Omega)^*,2} \\
\left\|(\kappa_1,\dots,\kappa_J)\right\|_{X^\kappa} &= \sum_{j=1}^{J}\left\|\kappa_j\right\|_{L^2(0,T)} + \sum_{j=1}^{J}\left\|\kappa_j'\right\|_{L^2(0,T)}
\end{aligned}$$

It is known that $L^2(0,T;L^2(\Omega))$ can be identified with $L^2(Q_T)$ and that $\left\|F\right\|_{2,2} = \left\|F\right\|_{2,Q_T}$ for $F \in L^2(Q_T)$ (see Example 23.4 in Chap. 23.2 in [51]). An analogous fact holds for spaces $L^2(0,T;\left(L^2(\Omega)\right)^{\mathbf{d}})$ and $\left(L^2(Q_T)\right)^{\mathbf{d}}$. Therefore, the above definitions of norms are meaningful.

Moreover, we define the following spaces:

$$U = U_g \times U_h \times U_\alpha, \quad U_g = \left(L^2(\Omega)\right)^J, \; U_h = \left(L^2(\Omega)\right)^K, \; U_\alpha = \mathbb{R}^{KJ}$$

where natural numbers J, K are the same as J, K appearing in the system (0.1) - (0.3). U will be called *the control space*. We equip it with standard product topology and scalar product. For a given element $\hat{u} \in U$ we denote the coordinates of $\hat{u}$ in the following way:

$$\hat{u} = (\hat{u}_{g_j}, \hat{u}_{h_k}, \hat{u}_{\alpha_{jk}})_{j=1,\dots,J}^{k=1,\dots,K}$$

$$\text{where} \quad (\hat{u}_{g_1}\dots,\hat{u}_{g_J}) \in U_g, \quad (\hat{u}_{h_1},\dots,\hat{u}_{h_k}) \in U_h, \quad (\hat{u}_{\alpha_{j,k}})_{j=1,\dots,J}^{k=1,\dots,K} \in U_\alpha$$

An arbitrary sufficiently integrable control $(g_j, h_k, \alpha_{jk})_{j=1,\dots,J}^{k=1,\dots,K}$ in the system (0.1) - (0.3) can be interpreted as an element of U and *vice versa* — an arbitrary element $\hat{u} \in U$ gives a control for the system (0.1) - (0.3) by putting $g_j := \hat{u}_{g_j}$, $h_k := \hat{u}_{h_k}$ and $\alpha_{j,k} := \hat{u}_{\alpha_{j,k}}$.

For technical reason, we define also the following space:

$$\widetilde{U} = \left(L^2(\Omega)\right)^{2J}$$

We equip $\widetilde{U}$ with standard product topology and scalar product. For a given $\hat{u} \in \widetilde{U}$, we denote the coordinates of $\hat{u}$ as follows:

$$\hat{u} = (\hat{u}_{g_1},\dots,\hat{u}_{g_J},\hat{u}_{h_1},\dots,\hat{u}_{h_J}) = (\hat{u}_{g_j},\hat{u}_{h_j})_{j=1}^{J}$$

REMARK. Concerning the weights $\alpha_{j,k}$ in (0.3), one can expect an assumption that $\alpha_{j,k}$ are nonnegative and summable to unity over $k = 1,\dots,K$, for all $j = 1,\dots,J$. But this assumption does not play any role in our considerations, hence we do not impose it and allow $\alpha_{j,k}$ to be arbitrary real numbers. This is reflected in the structure of the control space U, whose component space U_α can be understand as a space of admissible $(\alpha_{j,k})_{j=1,\dots,J}^{k=1,\dots,K}$. ▲

1.1 Multivalued switching function — existence results

This section is devoted to investigate the existence of solutions for the model of reaction-diffusion process with control by thermostats, described by the system (0.1) - (0.3). Consider an abstract operator defined as the operator assigning the solution y of (0.1) to a given $(\kappa_1, \ldots, \kappa_J)$, and than solution of (0.2) to y, denote it $(\bar{\kappa}_1, \ldots, \bar{\kappa}_J)$. The problem is to show that there exists $(\kappa_1, \ldots, \kappa_J)$ such that $(\bar{\kappa}_1, \ldots, \bar{\kappa}_J) = (\kappa_1, \ldots, \kappa_J)$. In other words, we wish to employ the fixed-point method for proving the existence of solutions.

Nevertheless, for the sake of limitations of the mathematical techniques utilized below, we need to modify (0.1) - (0.3) slightly before we proceed further.

Let us explain the latter comment in more detail. The natural candidate for the switching function w_k in (0.3) is the discontinuous function $w_k(s) = -sgn(s)$. The lack of continuity of the switching function is an obstacle for proving the existence in models with the variant of thermostat control mechanism without hysteresis in the work of the switching mechanism, which is our variant. This obstacle was the case in works [33], [15] and [19], which took into account models with the non-hysteresis variant of the thermostat control mechanism (more precisely, [19] focused only on a non-hysteresis thermostat control mechanism while [33] and [15] accounted, in addition to non-hysteresis controls, controls involving hysteresis in the work of the switching mechanism). In none of these works, for the variant of switching mechanism without hysteresis, the existence of solutions was proven under assumptions covering the case of discontinuous switching functions being equal $-sgn$. Works [33], [19] required considering a switching function being an upper semicontinuous multivalued mapping in order to obtain the existence result. In [15], a switching function being a maximal monotone mapping whose graph contained the graph of $-sgn$ was considered. The maximal monotonicity of the switching function was essential in the existence proof in [15].

Within this setting, $-sgn$ cannot be viewed directly as an admissible switching function, because it is not upper semicontinuous in the sense of multivalued mappings, nor it is maximal monotone. However, it is possible to take a switching function being a maximal monotone multivalued mapping whose graph contains the graph of $-sgn$ into consideration. Thus in some sense, it is allowed to consider switching functions „somehow related" to $-sgn$ within this setting. But, this abstract approach has only technical reasons and makes the model less realistic.

Nevertheless, we will adapt this approach here and allow the switching functions to be multivalued mappings, obeying certain additional conditions. From the mathematical point of view allowing a multivalued w_k makes the model (0.1) - (0.3) more general, thus results shown with this approach will apply also for a certain class of the single-valued switching functions (which, as we will see, unfortunately occurs to exclude the $-sgn$ switching function).

Assuming that w_k are multivalued mappings forces us to understand the ordinary differential equations (0.2) as an ordinary differential inclusions. Hence, in this section we will consider the following modification of the system (0.1) - (0.3) instead of (0.1) - (0.3) itself:

$$\begin{cases} y_t(x,t) - D\Delta y(x,t) = f(y(x,t)) + \sum_{j=1}^{J} g_j(x)\kappa_j(t) & \text{on } Q_T \\ \dfrac{\partial y}{\partial n} = 0 & \text{on } \partial\Omega \times (0,T) \\ y(x,0) = y_0(x) & \text{for } x \in \Omega \end{cases} \tag{1.1}$$

together with

$$\begin{cases} \beta_1\kappa_1'(t) + \kappa_1(t) \in W_1\big(y(\,.\,,t), y^*(x,t)\big) & \text{on } [0,T] \\ \vdots & \vdots \\ \beta_J\kappa_J'(t) + \kappa_J(t) \in W_J\big(y(\,.\,,t), y^*(x,t)\big) & \text{on } [0,T] \\ \kappa_j(0) = \kappa_{j0} \in \mathbb{R} & \text{for } j = 1,\dots,J \end{cases} \tag{1.2}$$

where the notation is as in the system (0.1) - (0.3) with the exception that W_j are multivalued functions now, defined by:

$$W_j(y(\,.\,,t), y^*(\,.\,,t)) = \sum_{k=1}^{K} \alpha_{jk} w_k\Big(\int_\Omega h_k(x)\big(y(x,t) - y^*(x,t)dx\big)\Big) \tag{1.3}$$

where $\alpha_{jk} \in \mathbb{R}$, $h_k\colon \Omega \to \mathbb{R}$ are functions and $w_k\colon \mathbb{R} \to 2^{\mathbb{R}}$ are multivalued mappings, for $k = 1,\dots,K$.

The present section utilizes the theory of multivalued mappings, in the scope of Appendix A.5.

We will follow the methods exploiting upper semicontinuity of w_k in the sense of multivalued mappings (see Definition A.5.2 in Appendix A.5), as it was the case in [33] or [19]. This is reflected in the following assumptions for the system (1.1) - (1.3):

(A-1) $\Omega \subset \mathbb{R}^{\mathbf{d}}$ is a bounded domain, such that the embedding $W^{1,2}(\Omega) \hookrightarrow L^2(\Omega)$ is compact (e.g. a bounded domain satisfying the cone condition is sufficient, see the Rellich-Kondrachov theorem presented e.g. in [1, Th. 6.3.]; for definition of the cone condition, see [1, par. 4.6.]),

(A-2) K, J are given positive natural numbers, $T > 0$, $D > 0$ and $\beta_j > 0$ for all $j = 1,\dots,J$,

(A-3) f is globally Lipschitz continuous; we denote its Lipschitz constant by L,

(A-4) w_k is a multivalued function, $w_k\colon \mathbb{R} \to 2^{\mathbb{R}}$, satisfying the following conditions, for $k = 1,\dots,K$:

   a) w_k has nonempty, closed and convex values,

   b) w_k is upper semicontinuous in the sense of upper semicontinuity of multivalued mappings,

   c) w_k is bounded; we define constant $C_{w_k} > 0$ as constant such that $w_k(t,s) \subseteq [-C_{w_k}, C_{w_k}]$ for all $s \in \mathbb{R}$,

(A-5) $y_0 \in L^2(\Omega)$, $\kappa_{j0} \in \mathbb{R}$ for $j = 1,\dots,J$,

(A-6) $y^* \in C([0,T]; L^2(\Omega)_w)$.

In the present section, we will use the following definition of solutions for the system (1.1) - (1.3):

Definition 1.1.1 *An element $(y, \kappa_1,\dots,\kappa_J)$ of the space X^2 is a weak solution of the system (1.1) - (1.3) if there exists $(\mathbf{W}_1,\dots,\mathbf{W}_J) \in \big(L^2(0,T)\big)^J$ such that:*

(a) $y(\,.\,,0) = y_0$ in $L^2(\Omega)$ and $\kappa_j(0) = \kappa_{j0}$ for $j = 1,\dots,J$,

(b) for all $\phi \in L^2(0,T; H^1(\Omega))$, there holds

$$\int_0^T \langle y', \phi\rangle + D\big(\nabla y, \nabla\phi\big)_{L^2(\Omega)} + \big(-f(y) - \kappa_1 g_1 - \ldots - \kappa_J g_J\,,\, \phi\big)_{L^2(\Omega)}\, dt \;=\; 0$$

(c) for all $\xi \in L^2(0,T)$, for $j = 1,\dots,J$, there holds

$$\int_0^T \big(\beta_j \kappa_j' + \kappa_j - \mathbf{W}_j\big)\,\xi\,dt \;=\; 0$$

(d) $\mathbf{W}_j(t) \in W_j\big(y(\,.\,,t), y^*(\,.\,,t)\big)$ for a.e. $t \in (0,T)$.

The point (a) in Definition 1.1.1 is meaningful because if $(y, \kappa_1, \dots, \kappa_J) \in X^2$ then $y \in C([0,T]; L^2(\Omega))$ and $(\kappa_1,\dots,\kappa_J) \in C([0,T])$. For justification, note that the spaces $H^1(\Omega)$, $L^2(\Omega)$ and $H^1(\Omega)^*$ form so-called evolution triple (defined e.g. in [51]) with embeddings $H^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow H^1(\Omega)^*$. Having this, see [51, Prop. 23.23] to conclude that $y \in C([0,T]; L^2(\Omega))$. Then, use the Sobolev embedding theorem, see [1, Th. 4.12, p. 85], or apply [51, Prop. 23.23] again to get $(\kappa_1,\dots,\kappa_J) \in C([0,T])$.

The main Theorem of Section 1.1 is the following existence result:

Theorem 1.1.2 *Let assumptions (A-1) - (A-6) be fulfilled. Assume also that $(y_0, \kappa_{10}, \dots, \kappa_{J0}) \in X^0$ and $(g_j, h_k, \alpha_{j,k})_{j=1,\dots,J}^{k=1,\dots,K} \in U$. Then, there exists a weak solution of the system (1.1) - (1.3).*

We present the proof of Theorem 1.1.2 in Section 1.1.1. Earlier, in Section 1.1.2, we give some technical lemmas necessary for the proof.

1.1.1 Auxiliary lemmas

This section presents some auxiliary facts that will be necessary for the proof of Theorem 1.1.2.

We will need to consider the following auxiliary systems of equations:

$$\begin{cases} y_t(x,t) - D\Delta y(x,t) = f(y(x,t)) + \sum_{j=1}^{J} g_j(x)k_j(t) & \text{on } Q_T \\ \dfrac{\partial y}{\partial n} = 0 & \text{on } \partial\Omega \times (0,T) \\ y(0) = y_0 & \text{on } \Omega \end{cases} \tag{1.4}$$

$$\begin{cases} \beta_j \kappa_j'(t) + \kappa_j(t) = \mathbf{V}_j(t) & \text{on } [0,T] \\ \kappa_j(0) = \kappa_{j0} \end{cases} \qquad \text{for } j = 1,\dots,J \tag{1.5}$$

where $k_j \in L^2(0,T)$, $\mathbf{V}_j \in L^2(0,T)$ for $j = 1,\dots,J$ are given and the rest of the notation is as in the system (0.1) - (0.3).

Definition 1.1.3 *A weak solution of (1.4) is a function $y \in X^y$ that satisfies $y(0) = y_0$ and*

$$\int_0^T \langle y', \phi\rangle + D\big(\nabla y, \nabla\phi\big)_{L^2(\Omega)} + \Big(-f(y) - \sum_{j=1}^{J} g_j k_j\,,\,\phi\Big)_{L^2(\Omega)}\,dt \;=\; 0 \tag{1.6}$$

for all $\phi \in L^2(0,T; H^1(\Omega))$.

Definition 1.1.4 *A weak solution of (1.5) is a function $\kappa = (\kappa_1,\dots,\kappa_J) \in X^\kappa$ that satisfies $\kappa_j(0) = \kappa_{j0}$ and*

$$\int_0^T \big(\beta_j \kappa_j' + \kappa_j - \mathbf{V}_j\big)\,\xi\,dt \;=\; 0 \tag{1.7}$$

for all $\xi \in L^2(0,T)$, for $j = 1,\dots,J$.

For weak solutions of both (1.4) and (1.5), initial conditions are well defined, by the same arguments as the ones on page 6, concerning Definition 1.1.1.

Now, we give some lemmas describing properties of the weak solutions to (1.4) and (1.5):

Lemma 1.1.5 *Let Ω, T, D, J, f, y_0 be as in assumptions (A-1), (A-2), (A-3), (A-5), respectively, and let $g_j \in L^2(\Omega)$ for $j = 1, \dots, J$. In addition:*

1. *Let $k_j \in L^2(0,T)$ for $j = 1, \dots, J$. Then the weak solution of (1.4) exists and is unique.*

2. *Let y^1 and y^2 be two weak solutions of (1.4) corresponding to $k_j = k_j^1$ and $k_j = k_j^2$ respectively, for $j = 1, \dots, J$,where $k_j^1 \in L^2(0,T)$ and $k_j^2 \in L^2(0,T)$. Then*

$$\|y^1 - y^2\|_{X^y} \;\le\; C_1 \Big\| \sum_{j=1}^{J} g_j(x)(k_j^1(t) - k_j^2(t)) \Big\|_{2,2} \;\le\; C_2 \sum_{j=1}^{J} \|k_j^1 - k_j^2\|_{L^2(0,T)} \tag{1.8}$$

where $C_1 = C_1(T, D, L)$ and $C_2 = C_2\big(T, D, L, \|g_1\|_2, \dots, \|g_J\|_2\big)$.

Proof. It is a known result that under the imposed assumptions the weak solution of the equation (1.4) exists and is unique. Thus we do not prove it here but only give some comments on the addressed matter.

The existence of solutions of (1.4) can be shown by Galerkin method. See [40, Chap. 8] for example realization of this method for a semilinear reaction-diffusion equation. A case of homogeneous Dirichlet boundary data and a growth condition for f other than ours is considered there, also the solutions are defined in other spaces. Nevertheless, the method presented there can be adapted to our case, after adequate modifications.

One may conduct the proof of the existence with the above mentioned method to find that our assumptions concerning Ω, f, y_0, g_j, k_j and D are essential for the assertion. The assumptions concerning T and J are necessary just to make the problem well defined.

The stability of the system (1.4), expressed by the first inequality in (1.8), also is a known result for the case of the Lipschitz nonlinearity f, but we present its proof here for the sake of completeness of the presented content. The first inequality in (1.8) can be shown as follows. For estimates for $\|y^1 - y^2\|_{2,\infty}$ we subtract the identity (1.6) corresponding to $k_j = k_j^1$, $j = 1, \dots, J$ and the same identity corresponding to $k_j = k_j^2$, $j = 1, \dots, J$. We test the resulting identity by $\phi = \mathbf{1}_{[0,t]}(y^1 - y^2)$ for a given $t \in [0,T]$. This results in:

$$\begin{aligned} &\int_0^t \langle {y^1}' - {y^2}', y^1 - y^2 \rangle \, ds \;+\; D \int_0^t \|\nabla(y^1 - y^2)\|_2^2 \, ds \;= \\ &\quad = \int_0^t \big(f(y^1) - f(y^2), y^1 - y^2\big)_{L^2(\Omega)} \, ds \;+\; \int_0^t \int_\Omega \sum_{j=1}^{J} g_j \big(k_j^1 - k_j^2\big)\big(y^1 - y^2\big) \, dx \, ds \end{aligned} \tag{1.9}$$

Next, the following identity holds:

$$\int_0^t \langle {y^1}' - {y^2}', y^1 - y^2 \rangle \, dt \;=\; \frac{1}{2}\|y^1(\,.\,,t) - y^2(\,.\,,t)\|_2^2 \;-\; \frac{1}{2}\|y^1(\,.\,,0) - y^2(\,.\,,0)\|_2^2 \tag{1.10}$$

(see Prop. 23.23 in [51] and note that spaces $H^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow H^1(\Omega)^*$ form an evolution triple, defined as in Chap. 23.4 in [51]). Using the above in (1.9) and recalling that $y^1(\,.\,,0) = y^2(\,.\,,0)$, we obtain:

$$\begin{aligned} &\frac{1}{2}\big\|y^1(\,.\,,t)-y^2(\,.\,,t)\big\|_2^2 \;+\; D\int_0^t \big\|\nabla(y^1-y^2)\big\|_2^2\,ds \;= \\ &\leq\; \big(L+\tfrac{1}{2}\big)\int_0^t \big\|y^1(t)-y^2(t)\big\|_2^2\,dt + \frac{1}{2}\Big\|\sum_{j=1}^{J} g_j\,\big(k^1-k^2\big)\Big\|_{2,2}^2 \end{aligned} \tag{1.11}$$

where the Lipschitz continuity of f and the Young inequality were used to estimate the right hand side of (1.9). Now, we neglect the gradient term (which is nonnegative) and by the Grönwall inequality we conclude that

$$\big\|y^1-y^2\big\|_{2,\infty} \leq C_{10}\Big\|\sum_{j=1}^{J} g_j\,\big(k_j^1-k_j^2\big)\Big\|_{2,2} \tag{1.12}$$

for some constant $C_{10}>0$, $C_{10}=C_{10}(T,L)$.

To get the estimates for $\big\|\nabla\,\big(y^1-y^2\big)\big\|_{2,2}$, we again use (1.11). Neglecting the term $\big\|y^1(\,.\,,t)-y^2(\,.\,,t)\big\|_2$ and taking $t=T$, it follows that:

$$D\int_0^T \big\|\nabla(y^1-y^2)\big\|_2^2\,dt \;\leq\; (L+\tfrac{1}{2})T\big\|y^1-y^2\big\|_{2,\infty}^2 \;+\; \frac{1}{2}\Big\|\sum_{j=1}^{J} g_j\,\big(k_j^1-k_j^2\big)\Big\|_{2,2}^2$$

where we have used the estimate $\big\|y^1-y^2\big\|_{2,2}\leq T^{1/2}\big\|y^1-y^2\big\|_{2,\infty}$. Now, we can use the above inequality and (1.12) to get that

$$\big\|\nabla\,\big(y^1-y^2\big)\big\|_{2,2} \leq C_{11}\Big\|\sum_{j=1}^{J} g_j\,\big(k_j^1-k_j^2\big)\Big\|_{2,2} \tag{1.13}$$

where $C_{11}=C_{11}(T,D,L)$.

To obtain estimates for $\big(y^1-y^2\big)'$ in $L^2(0,T;H^1(\Omega)^*)$, we again subtract two copies of (1.6) and treat the resulting integral identity as a condition for a functional on the space $L^2(0,T;H^1(\Omega))$. We conclude that the below holds:

$$(y^1-y^2)' + D\mathbf{A}(y^1-y^2) - \big(\mathbf{F}y^1-\mathbf{F}y^2\big) - \mathbf{G} = 0 \quad \text{in } L^2(0,T;H^1(\Omega)^*) \tag{1.14}$$

where $\mathbf{A}\colon L^2(0,T;H^1(\Omega)) \to L^2(0,T;H^1(\Omega)^*)$, $\mathbf{F}\colon L^2(0,T;H^1(\Omega)) \to L^2(0,T;H^1(\Omega)^*)$ and $\mathbf{G}\in L^2(0,T;H^1(\Omega)^*)$ are defined by

$$\begin{aligned} \int_0^T \langle \mathbf{A}\widetilde{y},\phi\rangle\,dt \;&=\; \int_0^T \Big(\nabla\widetilde{y},\nabla\phi\Big)_{L^2(\Omega)}\,dt \\ \int_0^T \langle \mathbf{F}\widetilde{y},\phi\rangle\,dt \;&=\; \int_0^T \Big(f(\widetilde{y}),\phi\Big)_{L^2(\Omega)}\,dt \\ \int_0^T \langle \mathbf{G},\phi\rangle\,dt \;&=\; \int_0^T \Big(\sum_{j=1}^{J} g_j\,\big(k_j^1-k_j^2\big)\,,\phi\Big)_{L^2(\Omega)}\,dt \end{aligned} \tag{1.15}$$

for a given $\widetilde{y}\in L^2(0,T;H^1(\Omega))$ and all $\phi\in L^2(0,T;H^1(\Omega))$.

It follows by definition of $\mathbf{A}$, $\mathbf{F}$ and $\mathbf{G}$ that

$$\begin{aligned}\|\mathbf{A}\widetilde{y}^1\|_{H^1(\Omega)^*,2} &\leq \|\nabla\widetilde{y}\|_{2,2}\\ \|\mathbf{F}\widetilde{y}^1 - \mathbf{F}\widetilde{y}^2\|_{H^1(\Omega)^*,2} &\leq \|f(\widetilde{y}^1) - f(\widetilde{y}^2)\|_{2,2}\\ \|\mathbf{G}\|_{H^1(\Omega)^*,2} &\leq \Big\|\sum_{j=1}^{J} g_j\left(k_j^1 - k_j^2\right)\Big\|_{2,2}\end{aligned} \tag{1.16}$$

for given $\widetilde{y}^1, \widetilde{y}^2 \in L^2(0,T;H^1(\Omega))$. This, together with (1.14), yields:

$$\begin{aligned}\|(y^1 - y^2)'\|_{H^1(\Omega)^*,2} &\leq \|\nabla y^1 - \nabla y^2\|_{2,2} + \|f(y^1) - f(y^2)\|_{2,2} + \Big\|\sum_{j=1}^{J} g_j\left(k_j^1 - k_j^2\right)\Big\|_{2,2}\\ &\leq \|\nabla y^1 - \nabla y^2\|_{2,2} + L\|y^1 - y^2\|_{2,2} + \Big\|\sum_{j=1}^{J} g_j\left(k_j^1 - k_j^2\right)\Big\|_{2,2}\end{aligned}$$

Now, recalling that $\|y^1 - y^2\|_{2,2}$ can be estimated by $\|y^1 - y^2\|_{2,\infty}$, we use (1.12) and (1.13) to conclude that

$$\|(y^1 - y^2)'\|_{H^1(\Omega)^*,2} \leq C_{12}\Big\|\sum_{j=1}^{J} g_j\left(k_j^1 - k_j^2\right)\Big\|_{2,2} \tag{1.17}$$

where $C_{12} = C_{12}(T,D,L)$.

To sum up, by (1.12), (1.13) and (1.17), the first inequality in (1.8) follows. The second inequality in (1.8) follows straight by the Fubini theorem.

The proof of uniqueness can be conducted by application of the Grönwall inequality, analogously to the above proof of (1.12). Take $y_0^1, y_0^2 \in L^2(\Omega)$ and denote by y^1, y^2 given weak solutions of (1.1) corresponding to y_0^1, y_0^2 respectively. Then, subtract two copies of identity (1.6) corresponding to y_0^1 and y_0^2, respectively, and test the resulting identity by $\phi = \mathbf{1}_{[0,t]}(y^1 - y^2)$, for a given $t \in [0,T]$. This gives:

$$\int_0^t \langle y^{1\prime} - y^{2\prime}, y^1 - y^2\rangle\, ds \ + \ D\int_0^t \|\nabla(y^1 - y^2)\|_2^2\, ds = \int_0^t \ \left(f(y^1) - f(y^2), y^1 - y^2\right)_{L^2(\Omega)}\, dt$$

In the above, use identity (1.10), recall the Lipschitz continuity of f with constant L and neglect the gradient term (which is nonnegative):

$$\frac{1}{2}\|y^1(\,.\,,t) - y^2(\,.\,,t)\|_2^2 \ \leq L\int_0^t \ \|y^1(t) - y^2(t)\|_2^2\, dt + \frac{1}{2}\|y_0^1 - y_0^2\|_2^2$$

Now, the Grönwall inequality yields $\|y^1 - y^2\|_{2,\infty} \leq C_{13}\|y_0^1 - y_0^2\|_2$, for certain $C_{13} = C_{13}(T,L)$. Thus, for $y_0^1 = y_0^2$ in $L^2(\Omega)$ we have $y^1(t) = y^2(t)$ in $L^2(\Omega)$ for a.e. $t \in [0,T]$, what concludes the proof of the uniqueness. ■

Lemma 1.1.6 *Let T, J, K and β_j for $j = 1,\ldots,J$ be as in the assumption (A-2). Then, the following statements are true:*

1. *Let $\mathbf{V}_j \in L^2(0,T)$ for $j = 1,\ldots,J$. Then, the weak solution of (1.5) exists and is unique.*

2. *Moreover, if* $\kappa = (\kappa_1, \dots, \kappa_J) \in X^\kappa$ *is the weak solution of (1.5) corresponding to a given* $(\mathbf{V}_1, \dots, \mathbf{V}_J) \in (L^\infty(0,T))^J$, *then*

$$\|\kappa\|_{X^\kappa} \;\le\; C_3\Big(\sum_{j=1}^{J}|\kappa_{j0}| \;+\; \sum_{j=1}^{J}\|\mathbf{V}_j\|_{L^2(0,T)}\Big) \tag{1.18}$$

where $C_3 = C_3(\beta_1, \dots, \beta_J, T)$.

3. *Moreover, assume that* $\widetilde{\mathbf{V}}^n \in \big(L^2(0,T)\big)^J$ *for* $n \in \mathbb{N}$ *and that* $\widetilde{\kappa}^n \in X^\kappa$ *are the weak solutions of (1.5) corresponding to* $\widetilde{\mathbf{V}}^n$, *by putting* $\mathbf{V}_j := \widetilde{\mathbf{V}}_j^n$ *in (1.5). In addition, assume that* $\widetilde{\mathbf{V}}^n \rightharpoonup \widetilde{\mathbf{V}}$ *in* $\big(L^2(0,T)\big)^J$ *for certain* $\widetilde{\mathbf{V}} \in \big(L^2(0,T)\big)^J$ *and that* $\widetilde{\kappa}^n \rightharpoonup \widetilde{\kappa}$ *in* X^κ *for certain* $\widetilde{\kappa} \in X^\kappa$. *Then,* $\widetilde{\kappa}$ *is the weak solution of (1.5) corresponding to* $\widetilde{\mathbf{V}}$, *by putting* $\mathbf{V}_j := \widetilde{\mathbf{V}}_j$ *in (1.5).*

Proof. For the existence and uniqueness of solutions, first observe that the above introduced notion of the weak solution of (1.5) is actually a Carathéodory solution. The Carathéodory solution of (1.5) is an absolutely continuous function from $[0,T]$ to $\mathbb{R}^J$ satisfying the ODE in (1.5) a.e. on $[0,T]$ and satisfying the initial condition in (1.5). The Carathéodory solutions, also for ordinary differential equations more general than (1.5), were investigated e.g. in handbooks [14] or [22].

Let us briefly justify the above observation. An arbitrary weak solution κ of (1.5) belongs to X^κ and hence is Hölder continuous by the Sobolev embedding theorem (see [1, Th. 4.12]). In particular, κ is absolutely continuous. Moreover, it satisfies the identity $\beta_j\kappa_j' + \kappa_j - \mathbf{V}_j = 0$ a.e. on $[0,T]$ for $j = 1, \dots, J$, because by the definition of the weak solution of (1.5), $\beta_j\kappa_j' + \kappa_j - \mathbf{V}_j$ is the zero element of $L^2(0,T)$. Hence, κ being a weak solution of (1.5) is a Carathéodory solution of (1.5) as well.

Conversely, let κ be a Carathéodory solution of (1.5). Since it fulfills $\beta_j\kappa_j' + \kappa_j - \mathbf{V}_j = 0$ a.e. on $[0,T]$ for $j = 1, \dots, J$, it fulfills also the integral identity in (1.5). Moreover, as a continuous function on a closed interval, κ_j is square integrable, for $j = 1, \dots, J$. κ_j' also is square integrable because $\kappa_j' = \beta_j^{-1}(-\kappa_j + \mathbf{V}_j)$ and κ_j, $\mathbf{V}_j$ are square integrable. Hence, $\kappa \in X^\kappa$. In total, κ occurs to be a weak solution of (1.5) as well.

Thus the question on existence and uniqueness of weak solutions of (1.5) can be replaced by the question on existence and uniqueness of the Carathéodory solutions of (1.5). The existence of Carathéodory solutions can be concluded by Theorem 1.1 in Chapter 2 in [14] or by Theorem 1 in Chapter 1 in [22], concerning the existence of Carathéodory solutions for ODEs more general than ours (the formulation of Theorem 1.1, Chap. 2 in [14] does not specify precisely the interval of existence, but analysis of the proof of this theorem indicates that in our case the existence on $[0,T]$ can be obtained; the formulation of Theorem 1, Chap. 1 in [22] is more precise and does not cause this kind problems). The uniqueness of Carathéodory solutions of (1.5) follows by Theorem 2 in Chapter 1 in [22].

Alternatively, instead of referring to the general theory presented in [14] and [22], one can prove the demanded existence and uniqueness assertion as follows. Simply note that the function κ is a Carathéodory solution of (1.5) if and only if

$$\kappa_j(t) = \exp\Big(-\frac{1}{\beta_j}t\Big)\kappa_{j0} + \frac{1}{\beta_j}\int_0^t \exp\Big(-\frac{1}{\beta_j}(t-s)\Big)\mathbf{V}_j(s)\,ds \qquad \text{for } j = 1, \dots, J$$

Since the integral in the right hand side of the latter identity is well defined for a given $\mathbf{V}_j \in L^2(0,T)$, the Carathéodory solution of (1.5) exists and is unique.

Now, let $\kappa = (\kappa_1, \ldots, \kappa_J) \in X^\kappa$ be the weak solution of (1.5) corresponding to $(\mathbf{V}_1, \ldots, \mathbf{V}_J) \in (L^\infty(0,T))^J$. By testing the weak form (1.7) of the equation (1.5) by $\xi = \kappa_j \mathbf{1}_{[0,t]}$ we have

$$\beta_j \int_0^t {\kappa_j}' \kappa_j \, ds \; + \; \int_0^t \left|\kappa_j\right|^2 \; = \; \int_0^t \mathbf{V}_j \kappa_j \, ds \tag{1.19}$$

for $t \in [0,T]$, for $j = 1, \ldots, J$. By integrability of ${\kappa_j}'$, we have the absolute continuity of κ_j. Thus, by the integration by parts, the relation $\int_0^t \kappa_j' \kappa_j \; = \frac{1}{2}\left|\kappa_j(t)\right|^2 - \frac{1}{2}\left|\kappa_j(0)\right|^2$ is valid. Applying the latter in (1.19), neglecting the $\left|\kappa_j\right|^2$ term (which is nonnegative) and applying the Young inequality yields:

$$\left|\kappa_j(t)\right|^2 \; \leq \; \left|\kappa_{j0}\right|^2 \; + \beta_j^{-1} \|\mathbf{V}_j\|^2_{L^2(0,T)} + \beta_j^{-1} \int_0^t \left|\kappa_j(s)\right|^2 ds$$

By applying the integral Grönwall inequality to the above:

$$\|\kappa_j\|^2_{L^\infty(0,T)} \leq C_{30,j} \left(\left|\kappa_{j0}\right|^2 + \|\mathbf{V}_j\|^2_{L^2(0,T)} \right) \tag{1.20}$$

for $j = 1, \ldots, J$, where $C_{30,j} = C_{30,j}(\beta_j, T)$.

Next, the weak form (1.7) implies that

$$\beta_j {\kappa_j}' + \kappa_j = \mathbf{V}_j \qquad \text{in } L^2(0,T)$$

for $j = 1, \ldots, J$ and therefore

$$\|\kappa_j'\|_{L^2(0,T)} \leq \beta_j^{-1} \|\kappa_j\|_{L^2(0,T)} + \beta_j^{-1} \|\mathbf{V}_j\|_{L^2(0,T)} \tag{1.21}$$

Inequalities (1.20) and (1.21) together imply the estimate (1.18).

Proving the remaining part of the assertions of the present lemma is straightforward. Let $\widetilde{\mathbf{V}}^n$, $\widetilde{\mathbf{V}}$, $\widetilde{\kappa}^n$ and $\widetilde{\kappa}$ be as in the assumptions of the lemma. Then

$$\beta_j \left(\widetilde{\kappa}_j^n\right)' + \widetilde{\kappa}_j^n - \widetilde{\mathbf{V}}_j^n \;\rightharpoonup\; \beta_j \widetilde{\kappa}_j' + \widetilde{\kappa}_j - \widetilde{\mathbf{V}}_j \qquad \text{in } L^2(0,T)$$

for $j = 1, \ldots, J$. The above convergence suffices to pass to the limit in the weak form (1.7) of the equation (1.5) and infer the desired assertion. ■

Remark. It can be verified that the proof of Lemma 1.1.6, after minor modifications, would be valid also for $\beta_j < 0$. ▲

The following two lemmas also will be required in the proof of Theorem 1.1.2:

Lemma 1.1.7 *Let $\widetilde{\mathbb{W}} \colon \mathbb{R} \to 2^{\mathbb{R}}$ be a bounded upper semicontinuous multivalued mapping (see definitions in Appendix A.5) with nonempty and closed values. Let $\widetilde{\mathbf{v}} \in C([0,T])$. Then $\widetilde{\mathbb{W}} \circ \widetilde{\mathbf{v}}$ has a measurable selection, i.e. there exists at least one function $\widetilde{\mathbf{V}} : [0,T] \to \mathbb{R}$ which is measurable and $\widetilde{\mathbf{V}}(t) \in \widetilde{\mathbb{W}} \circ \widetilde{\mathbf{v}}(t)$ for a.e. $t \in [0,T]$.*

Proof. The proof of Lemma 1.1.7 is analogous to that of [26, Lemma 3.4], but we include it here for completeness of the presented content.

By Corollary 1.1 on p. 237 in [20], if

1. the image of $\widetilde{\mathbb{W}} \circ \widetilde{\mathbf{v}}$ is contained in some compact $\mathbb{K} \subset \mathbb{R}$,
2. $G(\widetilde{\mathbb{W}} \circ \widetilde{\mathbf{v}})$ is a Borel set of $\mathbb{R} \times \mathbb{K}$ and
3. $\widetilde{\mathbb{W}} \circ \widetilde{\mathbf{v}}$ has closed and nonempty values a.e. on $[0, T]$

then $\widetilde{\mathbb{W}} \circ \widetilde{\mathbf{v}}$ has a measurable selection, as demanded in the assertion of the present lemma.

A compact $\mathbb{K}$ as above exists by the assumption on boundedness of $\widetilde{\mathbb{W}}$.

Next, $\widetilde{\mathbb{W}} \circ \widetilde{\mathbf{v}}$ has closed and nonempty values because the same applies to $\widetilde{\mathbb{W}}$.

Moreover, $\widetilde{\mathbb{W}} \circ \widetilde{\mathbf{v}}$ is upper semicontinuous in sense of multivalued functions because $\widetilde{\mathbb{W}}$ and $\widetilde{\mathbf{v}}$ are so (see Prop. 6, Sec. 1, Chap. 3 in [4]). An upper semicontinuous multivalued mapping with closed values has closed graph (see Prop. 7, Sec. 1, Chap. 3 in [4]), hence $G(\widetilde{\mathbb{W}} \circ \widetilde{\mathbf{v}})$ is closed. Hence, $G(\widetilde{\mathbb{W}} \circ \widetilde{\mathbf{v}})$ is Borel as well.

This concludes the proof. ■

Lemma 1.1.8 *Let $\widetilde{\mathbb{W}}\colon \mathbb{R} \to 2^{\mathbb{R}}$ be a bounded upper semicontinuous multivalued mapping with nonempty, closed and convex values. Assume that $\widetilde{\mathbf{v}}_n \to \widetilde{\mathbf{v}}$ in $C([0,T])$, $\widetilde{\mathbf{V}}_n \overset{*}{\rightharpoonup} \widetilde{\mathbf{V}}$ in $L^\infty(0,T)$ and that $\widetilde{\mathbf{V}}_n(t) \in \widetilde{\mathbb{W}} \circ \widetilde{\mathbf{v}}_n(t)$ for a.e. $t \in [0,T]$, for $n \in \mathbb{N}$. Then $\widetilde{\mathbf{V}}(t) \in \widetilde{\mathbb{W}} \circ \widetilde{\mathbf{v}}(t)$ for a.e. $t \in [0,T]$.*

Lemma 1.1.8 can be viewed as a particular case of Lemma 3.6 in [26]. ▲

1.1.2 The proof of the existence theorem (Theorem 1.1.2)

In this section, we prove Theorem 1.1.2 with the use of auxiliary facts from Section 1.1.1. The proof will base on the following fixed-point theorem for multivalued mappings:

Theorem 1.1.9 (generalized Kakutani theorem) *Let X be a real Banach space and let $M \subset X$ be its convex, compact and nonempty subset. Let $T\colon M \to 2^M$ be a multivalued mapping having the following properties:*

a) the values $T(x)$ are nonempty and convex for all $x \in M$,

b) $G(T)$ is closed in $X \times X$.

Then T has a fixed point in M, i.e. there exists $\bar{x} \in M$ such that $\bar{x} \in T(\bar{x})$.

For the proof of Theorem 1.1.9, see [9, Th. 4] or [27]. The proof in [27] covers the more general case of convex Hausdorff linear topological spaces. Alternatively, Theorem 1.1.9 can be viewed as a direct consequence of Corollary 9, Chap. 3, Sec. 1 in [4] and Theorem 13, Chap. 6, Sec. 4 in [4], for the general case of Hausdorff locally convex spaces.

Remark. The formulation of Theorem 4 in [9] lacks the assumption that the sets $T(x)$ are convex but the proof presented there shows that this assumption is necessary and perhaps was accidentally missed in the theorem statement. ▲

Proof of Theorem 1.1.2. Define the following operators:

- $P_1\colon \left(L^2(0,T)\right)^J \to C([0,T]; L^2(\Omega))$ is assigns the solution of (1.4) to a given $(k_1, \ldots, k_J) \in \left(L^2(0,T)\right)^J$.

- $P_2\colon C([0,T];L^2(\Omega)) \to (C([0,T]))^K$ assigns $(\mathbf{v}_1,\dots,\mathbf{v}_K) \in (C([0,T]))^K$ determined by the formula
$$\mathbf{v}_k(t) = \int_\Omega h_k(x)\,(Y(x,t) - y^*(x,t))\;dx \quad \text{on } [0,T], \text{ for } k = 1,\dots,K \tag{1.22}$$
to a given $Y \in C([0,T];L^2(\Omega))$.

- $P_3\colon\ (C([0,T]))^K \to 2^{(L^\infty(0,T))^J}$ is a multivalued mapping assigning to a given $(\mathbf{v}_1,\dots,\mathbf{v}_K) \in (C([0,T]))^K$ the set $(\mathbb{W}_1,\dots,\mathbb{W}_J) \subseteq (L^\infty(0,T))^J$ determined by the following condition: for $j = 1,\dots,J$, $\mathbf{V}_j \in \mathbb{W}_j$ if and only if
$$\mathbf{V}_j(t) \in \sum_{k=1}^{K} \alpha_{j,k}\,(w_k \circ \mathbf{v}_k(t)) \quad \text{a.e. on } [0,T] \tag{1.23}$$

- $P_4\colon\ (L^\infty(0,T))^J \to \left(L^2(0,T)\right)^J$ assigns the solution of (1.5) to a given $(\mathbf{V}_1,\dots,\mathbf{V}_J) \in (L^\infty(0,T))^J$.

- $P := P_4 \circ P_3 \circ P_2 \circ P_1\colon\ \left(L^2(0,T)\right)^J \to \left(L^2(0,T)\right)^J$.

The meaning of the above operators in the context of the system (1.1) - (1.3), involving the thermostat control mechanism, is explained in Figure 1.1.

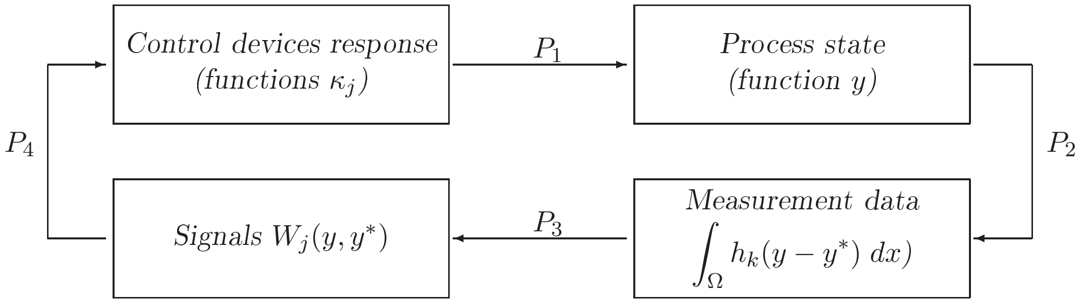

Control devices response (functions κ_j)
P_1
Process state (function y)
P_4
P_2
Signals $W_j(y,y^*)$
P_3
Measurement data $\int_\Omega h_k(y-y^*)\,dx)$

Figure 1.1: A schematic representation of the role of the operators P_1, P_2, P_3 and P_4, considered in the proof of Theorem 1.1.2, in the context of the thermostat control mechanism, present in the system (1.1) - (1.3). The notation in the picture is as in the subject system.

The existence of a weak solutions of (1.1) - (1.3) is equivalent to the existence of a fixed point of P, i.e. of $\bar{k} \in \left(L^2(0,T)\right)^J$ with $\bar{k} \in P(\bar{k})$. Indeed, by the definition of the operator P_4, such $\bar{k}$ belongs to the space X^κ (defined in Section 1.1.1), and $P_1(\bar{k})$ belongs to the space X^y (also defined there), hence the element $\left(P_1(\bar{k}), \bar{k}_1,\dots,\bar{k}_J\right)$ belongs to X^2. Moreover, by definitions of operators P_1, P_2, P_3 and P_4, the latter element fulfills Definition 1.1.1 with $y = P_1(\bar{k})$, $\kappa_j = \bar{k}_j$ for $j = 1,\dots,J$ and with $(\mathbf{W}_1,\dots,\mathbf{W}_J) \in \left(L^2(0,T)\right)^J$ given by $\mathbf{W}_j = \beta_j \bar{k}_j' + \bar{k}_j = \left({P_4}^{-1}(\bar{k})\right)_j$.

Now, we shall verify that the assumptions of Theorem 1.1.9 are satisfied for the operator P restricted to a suitable subset (which we will indicate in the sequel). This will justify that P has a fixed point and allow us to conclude the proof.

Nonempty values. By Lemma 1.1.5, P_1 is well defined. By the assumption (A-6) and by the structure of (1.22), P_2 is well defined. By Lemma 1.1.6, P_4 is well defined. Moreover, P_3 has nonempty values, because, by Lemma 1.1.7, each of multivalued mappings $\mathbf{v}_k \mapsto w_k \circ \mathbf{v}_k$, $k =$

$1, \ldots, K$, entering the definition of P_3, has nonempty values. More precisely, by the continuity of $\mathbf{v}_k$ and properties of w_k, Lemma 1.1.7 yields the existence of a measurable selection for the multivalued mapping $s \mapsto w_k \circ \mathbf{v}_k(s)$. By the boundedness of w_k, this measurable selection must be bounded and hence must be an element of $L^\infty(0,T)$. Thus, the set $w_k \circ \mathbf{v}_k \subset L^\infty(0,T)$ is nonempty for a given $\mathbf{v}_k \in C([0,T])$.

Therefore, the superposition $P_4 \circ P_3 \circ P_2 \circ P_1$ has nonempty values.

Convex values. By point (a) in the assumption (A-4), the values of P_3 are convex. Indeed, for a given $t \in [0,T]$, $\mathbf{w}_k(t) := w_k \circ \mathbf{v}_k(t)$ is a convex set and hence the collection $\widetilde{\mathbb{W}_k}$ of all $\widetilde{\mathbf{w}}_k \in L^\infty(0,T)$ such that $\widetilde{\mathbf{w}}_k(t) \in \mathbf{w}_k(t)$ a.e. on $[0,T]$ is convex. Next, $\mathbb{W}_j = \sum_{j=1}^{J} \alpha_{j,k} \widetilde{\mathbb{W}}_k$, i.e. $\mathbb{W}_j$ is a linear combination of convex sets, and as such is convex. It follows straight that the product over $j = 1, \ldots, J$ of $\mathbb{W}_j$ is convex in $(L^\infty(0,T))^J$. Thus the convexity of values of P_3 is justified.

Next, the operator P_4 is affine thus it maps convex sets to convex sets, i.e. $P_4 \circ P_3(\mathbf{v})$ is convex for an arbitrary $\mathbf{v} \in (C([0,T]))^K$. But the latter means that $P_4 \circ P_3 \circ P_2 \circ P_1(k)$ is convex for an arbitrary $k \in \left(L^2(0,T)\right)^J$.

Convex and compact image. Theorem 1.1.9, to hold, requires a multivalued mapping to act from a compact, convex and nonempty set into itself. Now we shall determine a set that is suitable for Theorem 1.1.9 in our case. Define auxiliary sets $\mathbb{A}$ and $\mathbb{B}$ as follows:

$$\mathbb{A} := \left\{ (\mathbf{V}_1, \ldots, \mathbf{V}_J) \in (L^\infty(0,T))^J : \|\mathbf{V}_j\|_{L^\infty(0,T)} \le C_{W_j} \ \forall_{j=1,\ldots,J} \right\}$$

where $C_{W_j} := \sum_{k=1}^{K} \alpha_{j,k} C_{w_k}$, for $j = 1 \ldots, J$ and for C_{w_k} being the constants from point (c) in the assumption (A-4),

$$\mathbb{B} := \left\{ k \in \left(L^2(0,T)\right)^J : \|k\|_{X^\kappa} \le C_3 \sum_{j=1}^{J} \left(|\kappa_{j0}| + T C_{W_j} \right) \ \forall_{j=1,\ldots,J} \right\}$$

where κ_{j0} are the initial conditions assumed for (1.2) in the assumption (A-5) and C_3 is the constant appearing in the estimate (1.18) in Lemma 1.1.6.

It follows from the definition of P_3 and from point (c) in the assumption (A-4) that $P_3(\mathbf{v}) \subseteq \mathbb{A}$ for an arbitrary $\mathbf{v} \in (C([0,T]))^K$. Next, the estimate (1.18) in Lemma 1.1.6 allows to infer that P_4 maps the set $\mathbb{A}$ into the set $\mathbb{B}$. Hence, $P_4 \circ P_3 \circ P_2 \circ P_1(k) \subseteq \mathbb{B}$ for an arbitrary $k \in \left(L^2(0,T)\right)^J$.

Denote by $\mathbf{B}$ the closure of $\mathbb{B}$ in $\left(L^2(0,T)\right)^J$. $\mathbb{B}$ is nonempty and convex, and hence the same holds for its closure. By the Rellich-Kondrachov Theorem (see [1, Th. 6.3]), $\mathbb{B}$ is precompact in $\left(L^2(0,T)\right)^J$. Moreover, $P(k) \in \mathbf{B}$ for $k \in \left(L^2(0,T)\right)^J$. Thus in total, $\mathbf{B}$ is nonempty, convex and compact and $P|_{\mathbf{B}} : \mathbf{B} \to 2^{\mathbf{B}}$.

Closed graph. Now, we will verify that $G(P|_{\mathbf{B}})$ is closed in $\left(L^2(0,T)\right)^J \times \left(L^2(0,T)\right)^J$. Since we are in a metric space, it is sufficient to check that $G(P|_{\mathbf{B}})$ is sequentially closed. Thus let $k^n, \xi^n \in \mathbf{B}$, $\xi^n \in P(k^n)$ for $n \in \mathbb{N}$ and assume that $k^n \to k$ and $\xi^n \to \xi$ in $\left(L^2(0,T)\right)^J$, for certain $k, \xi \in \left(L^2(0,T)\right)^J$. Since $\mathbb{B}$ is closed, $k, \xi \in \mathbf{B}$. We are left to show that $\xi \in P|_{\mathbf{B}}(k) = P_4 \circ P_3 \circ P_2 \circ P_1(k)$.

For $n \in \mathbb{N}$, there exist $\mathbf{V}^n \in (L^\infty(0,T))^J$ such that $\xi^n = P_4(\mathbf{V}^n)$ and $\mathbf{V}^n = P_3 \circ P_2 \circ P_1(k)$. Since $\mathbf{V}^n$ are in the image of P_3, $\mathbf{V}^n$ are bounded w.r.t. n in $(L^\infty(0,T))^J$. Hence, a weakly-$*$ convergent subsequence $\mathbf{V}^n \overset{*}{\rightharpoonup} \mathbf{V}$ can be extracted, for some $\mathbf{V} \in (L^\infty(0,T))^J$ (for brevity of notation, we denote this subsequence with the original indexes). It remains to verify that $\xi = P_4(\mathbf{V})$ and $\mathbf{V} \in P_3 \circ P_2 \circ P_1(k)$.

Weak-$*$ convergence of $\mathbf{V}^n$ to $\mathbf{V}$ in $(L^\infty(0,T))^J$ implies weak convergence in $\big(L^2(0,T)\big)^J$, therefore, by Lemma 1.1.6, $\xi = P_4(\mathbf{V})$. To conclude the inclusion $\mathbf{V} \in P_3 \circ P_2 \circ P_1(k)$, note that P_1 and P_2 are continuous. The continuity of P_1 follows by Lemma 1.1.5. The continuity of P_2 follows from the Hölder inequality. Having this and denoting $\mathbf{v}^n := P_2 \circ P_1(k^n)$ and $\mathbf{v} := P_2 \circ P_1(k)$, we infer that convergence $k^n \to k$ in$\big(L^2(0,T)\big)^J$ implies convergence $\mathbf{v}^n \to \mathbf{v}$ in $(C([0,T]))^K$. By definition of $\mathbf{V}^n$ and $\mathbf{v}^n$, we have $\mathbf{V}^n \in P_3(\mathbf{v}^n)$. To obtain the inclusion $\mathbf{V} \in P_3 \circ P_2 \circ P_1(k)$, it suffices to show that $\mathbf{V} \in P_3(\mathbf{v})$.

To show the latter, we will use Lemma 1.1.8, proceeding as follows. We have $\mathbf{v} = (\mathbf{v}_1, \ldots, \mathbf{v}_K)$ and $\mathbf{v}^n = (\mathbf{v}_1^n, \ldots, \mathbf{v}_K^n)$, where $\mathbf{v}_k^n \to \mathbf{v}_k$ in $C([0,T])$. Moreover, we have $\mathbf{V}^n = (\mathbf{V}_1^n, \ldots, \mathbf{V}_J^n)$, where, by the definition of the operator P_3, elements $\mathbf{V}_j^n$, for $j = 1, \ldots, J$, can be represented as

$$\mathbf{V}_j^n = \sum_{k=1}^{K} \mathbf{V}_{(j,k)}^n$$

where, for all $j = 1, \ldots, J$ and $k = 1, \ldots, K$,

$$\mathbf{V}_{(j,k)}^n \in \alpha_{j,k}(w_k \circ \mathbf{v}_k^n) \qquad \text{in } L^\infty(0,T) \tag{1.24}$$

By the assumption that w_k are bounded (see the part c) of the assumption (A-4)), $\mathbf{V}_{(j,k)}^n$ are bounded in $L^\infty(0,T)$ w.r.t. n, for all $j = 1, \ldots, J$, $k = 1, \ldots, K$. Thus, we can extract weakly-$*$ convergent subsequences $\mathbf{V}_{(j,k)}^n \overset{*}{\rightharpoonup} \widetilde{\mathbf{V}}_{(j,k)}$, for certain $\widetilde{\mathbf{V}}_{(j,k)} \in L^\infty(0,T)$. In consequence, on the subsequences we have $\mathbf{V}^n \overset{*}{\rightharpoonup} \widetilde{\mathbf{V}}$, where $\widetilde{\mathbf{V}} = (\widetilde{\mathbf{V}}_1, \ldots, \widetilde{\mathbf{V}}_J)$ and $\widetilde{\mathbf{V}}_j = \sum_{k=1}^K \widetilde{\mathbf{V}}_{(j,k)}$.

Now, by (1.24), by convergences $\mathbf{v}_k^n \to \mathbf{v}_k$ and $\mathbf{V}_{(j,k)}^n \overset{*}{\rightharpoonup} \widetilde{\mathbf{V}}_{(j,k)}$ and by an application of Lemma 1.1.8 to functions $\alpha_{j,k} w_k$, we obtain $\widetilde{\mathbf{V}}_{(j,k)} \in \alpha_{j,k}(w_k \circ \mathbf{v}_k)$. Thus, by definitions of P_3 and $\widetilde{\mathbf{V}}$, we can write $\widetilde{\mathbf{V}} \in P_3(\mathbf{v})$. Note also that $\widetilde{\mathbf{V}} = \mathbf{V}$, otherwise the convergence $\mathbf{V}^n \overset{*}{\rightharpoonup} \widetilde{\mathbf{V}}$ would be a contradiction to the convergence $\mathbf{V}^n \overset{*}{\rightharpoonup} \mathbf{V}$. Therefore, $\mathbf{V} \in P_3(\mathbf{v})$, as required. The proof of the closedness of $G(P|_{\mathbf{B}})$ is complete.

Now, apply Theorem 1.1.9 with $X = \big(L^2(0,T)\big)^J$, $M = \mathbf{B}$ and $T = P|_{\mathbf{B}}$ to get the existence of a fixed point of $P|_{\mathbf{B}}$ and hence of P as well. The proof of Theorem 1.1.2 is complete. ■

Remark. By definition, in the case of a single-valued function, the upper semicontinuity in the multivalued sense reduces to the usual continuity. Thus, any result holding for (1.1) - (1.3) under the assumption (A-4) from beginning of Section 1.1, holds in particular for bounded, continuous single-valued switching functions. ▲

Remark. One can say that Theorem 1.1.2 offers a method of indirect handling of the case of discontinuous switching functions in the thermostat control mechanism. Assume that a discontinuous single-valued function $\widetilde{w}_k \colon \mathbb{R} \to \mathbb{R}$ is given. In the case where the switching function w_k in the system (1.1) - (1.3) is defined by $w_k := \widetilde{w}_k$, it is not possible to apply Theorem 1.1.2. However, assuming that right and left limits of $\widetilde{w}_k$ exist in an arbitrary point $s \in \mathbb{R}$, it is possible to take into account a switching function $\widetilde{\widetilde{w}}_k$ associated with $\widetilde{w}_k$ by the formula (A.5.5) in the statement of Proposition A.5.5 in Appendix A.5. The assertion of Proposition A.5.5 together with the formula (A.5.5) guarantee that $\widetilde{\widetilde{w}}_k$ fulfills the assumption (A-4). In consequence, Theorem 1.1.2 apply for $w_k := \widetilde{\widetilde{w}}_k$ in the system (1.1) - (1.3). Thus, Theorem 1.1.2, however does not allow discontinuous switching functions directly, allows to consider, instead of a given discontinuous switching function $\widetilde{w}_k$, a multivalued switching function $\widetilde{\widetilde{w}}_k$ related to $\widetilde{w}_k$ (related — in the sense of the formula (A.5.5)).

Note, that the above comment is valid in particular for $\widetilde{w}_k(s) = -sgn(s)$, which is a natural candidate for the switching function in the thermostat control mechanism (see §1 of *Introduction*). In this case, $\widetilde{\widetilde{w}}_k$ generated by the formula (A.5.5) is

$$\widetilde{\widetilde{w}}_k = \begin{cases} +1 & \text{for } s < 0 \\ [-1, +1] & \text{for } s = 0 \\ -1 & \text{for } s > 0 \end{cases} \tag{1.25}$$

▲

Remark. An alternative approach could be employed to justify the closedness of the operator P_3 in the proof of Theorem 1.1.2. The subject approach refers to the theory of maximal monotone multivalued mappings. However, such approach would be less general to the one present in the proof of Theorem 1.1.2. Let us explain this matter in more detail.

In the proof of Theorem 1.1.2, the assumption (A-4) from beginning of Section 1.1, concerning switching functions w_k in the system (1.1) - (1.3), was crucial. It was the property which allowed us to conclude that the multivalued operator P_3, utilized in the proof, was closed in suitable topology. More precisely, P_3 can be interpreted as $P_3 = \big((P_3)_1, \ldots, (P_3)_j\big)$, where $(P_3(\mathbf{v}))_j = \sum_{k=1}^{K} \alpha_{j,k}(w_k \circ \mathbf{v}_k)$, for $j = 1, \ldots, J$ (compare with (1.23)). A given operator $(P_3)_j$ is thus a weighted sum of multivalued superposition operators $w_k \circ \mathbf{v}_k$, induced by the multivalued mappings w_k. In the proof of Theorem 1.1.2, each of these superposition operators occurred to be closed in suitable topology due to Lemma 1.1.8, basing strongly on the properties of multivalued mappings indicated in the assumption (A-4).

However, it is possible to prove the closedness of the superposition operator associated with a given multivalued mapping also with other means, e.g. assuming that the multivalued mapping is maximal monotone. If this is the case, then the associated superposition operator also is a maximal monotone mapping, in suitable spaces. At the same time, in certain function spaces, maximal monotonicity of multivalued mappings suffices to imply their closedness — results of this kind are given e.g. in Proposition 3, Ch. 6, Sec. 7 in [4] or Lemma 1.3, Chap. 2, Sec. 1.2, p. 42 in [6].

This argument was exploited in [15], also investigating a model with a control by thermostats, to prove closedness of the superposition operator associated with a multivalued switching function, denote it w, such that $-w$ was maximal monotone. In addition to the maximal monotonicity of the negative of the switching function, boundedness of the switching function was necessary in [15], as in our case (see the part c) of the assumption (A-4)).

In our situation, after suitable modification of the employed function spaces, applying the subject method for proving closedness of P_3 would be possible for the case of bounded and maximal monotone $-w_k$ (maximal monotonicity of w_k itself also would work but then the case of w_k as in (1.25) would be excluded, because the latter, in opposite to its negative, is not a monotone multivalued mapping). We skip the details because do not intend to develop this approach here.

Nevertheless, the method employed in the proof of Theorem 1.1.2, involving the assumption (A-4), is more general than the method basing on boundedness and maximal monotonicity of $-w_k$. The reason for this is that the assumption of boundedness and maximal monotonicity is stronger than the assumption (A-4). Indeed, it is straightforward that there exist w_k fulfilling the assumption (A-4) from beginning of Section 1.1 but such that w_k, nor $-w_k$, is not maximal monotone. On the other hand, an arbitrary bounded maximal monotone $-w_k$ obeys the assumption (A-4), and so w_k does. The latter is true because a maximal monotone multivalued mapping has closed and convex values (see Proposition A.5.8) and, if it additionally has the image contained in a compact set, it is upper semicontinuous (Proposition A.5.7) and has nonempty values (Proposition A.5.9). Thus, from the condition of boundedness and maximal monotonicity of a multivalued mapping, one can recover the properties indicated in the assumption (A-4). ▲

1.2 Single-valued switching function — existence, uniqueness, stability

The modification of the system (0.1) - (0.3) considered in Section 1.1 allowed to prove an existence result for the case where discontinuous switching functions are replaced with a multivalued mappings satisfying sufficiently strong assumptions (assumption (A-4)). However, these assumptions, being strong enough for the existence, still are not sufficient for obtaining the uniqueness result.

This was the case e.g. in works [33], [15] or [19]. These works, similarly to Section 1.1 of the present work, concern models with the variant of the thermostat control mechanism without hysteresis in the work of the switching mechanism and with multivalued switching functions (work [19] concern only this variant, works [33] and [15] concern also variants where the work of the switching mechanism involves hysteresis). Works [33] and [19] take into account the case of multivalued switching functions fulfilling assumptions analogous to the assumption (A-4). Work [15] exploited even stronger properties of the there considered multivalued switching function, namely the boundedness and the maximal monotonicity. At the same time, in none of the works [33], [19], [15] the uniqueness for the models with there considered variants of the thermostat control mechanism was proven.

Hence, in the present section we aim in strengthening the assumptions concerning the switching functions in the system (1.1) - (1.3) in order to be able to prove the uniqueness result. For this end, we shall assume that the switching functions are single-valued Lipschitz continuous functions.

Note, that the latter assumption implies that the inclusion (1.2) becomes equality again. Thus, we return to analysis of primary the system (0.1) - (0.3) instead of its modification (1.1) - (1.3) from Section 1.1.

Moreover, the assumption of the Lipschitz continuity of the switching function excludes the possibility of taking the switching function w_k equal the $-sgn$ function. It also excludes the approach from Section 1.1, providing a method for indirect handling of the case of $w_k = -sgn$ by replacing the original w_k by an upper semicontinuous multivalued mapping in some sense related to w_k (see Section 1.1 for details). Nevertheless, a sort of indirect method of handling the situation of $w_k = -sgn$ is available also under the presently considered assumption. Namely, the assumption of the Lipschitz continuity of w_k allows to approximate the function $-sgn$ by Lipschitz functions of a very steep slope near point zero.

Example. For instance, for $\widetilde{w}_k = -sgn$, one can define functions $\widetilde{w}_k^n$ by $\widetilde{w}_k^n(s) := -\max(\min(ns, 1), -1)$, for $s \in \mathbb{R}$, $n \in \mathbb{N}$. It follows straight that $\widetilde{w}_k^n$ are Lipschitz continuous functions. Moreover, for all $k = 1, \ldots, K$, $\widetilde{w}_k^n \to \widetilde{w}_k$, both pointwise and in the Lebesgue norm $\|\,.\,\|_{L^p(\mathbb{R})}$, for arbitrary $p \in [1, \infty)$ (cf Figure 1.2). Instead switching functions $w_k := \widetilde{w}_k$ in the system (0.1) - (0.3), which are not Lipschitz continuous, one may consider switching functions $w_k := \widetilde{w}_k^n$, which are Lipschitz continuous and approximate $\widetilde{w}_k$ in the latter sense. ▲

Thus, in certain sense, the assumption of Lipschitz continuity of the switching functions is no

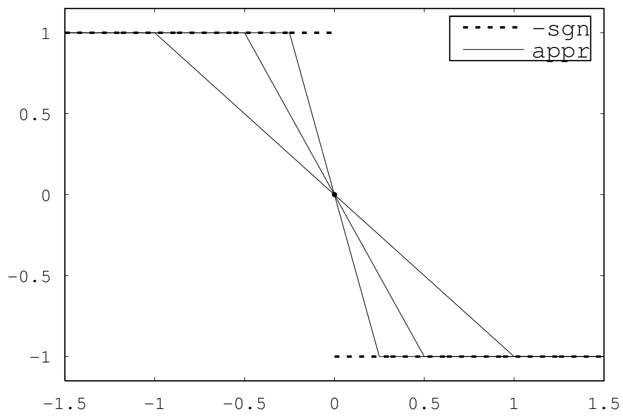


Figure 1.2: An example of a sequence of Lipschitz continuous functions approximating the function $-sgn$, both pointwise and in the $L^p(\mathbb{R})$-norm, for $p \in [1, \infty)$. The lines denoted as `appr` correspond to approximating functions given by $s \mapsto -\max(\min(ns, 1), -1)$, for $n = 1, 2, 4$.

waste in comparison to the situation considered in Section 1.1, because 1) in both cases, direct treatment of $w_k = -sgn$ is not possible, 2) in both cases, an indirect way to deal with $w_k = -sgn$ is available. The above proposed approach for dealing with discontinuous w_k was exploited in the numerical simulations described in Chapter 2.

Also, the assumption that the switching functions are Lipschitz continuous will be sufficient for proving the stability of the system (0.1) - (0.3) with respect to perturbations of the control. Results concerning this kind of stability will be crucial in Chapter 3, concerning the mathematical analysis of the optimal targeting problem. This gives a motivation to consider the above announced assumption that the switching functions are Lipschitz continuous.

We proceed in the following order. Section 1.2.1 focuses on existence of solutions of the system (0.1) - (0.3). The existence is shown for the case of Lipschitz switching functions w_k in the system (0.1) - (0.3) being additionally bounded. Section 1.2.1 contains two existence theorems. The first of them is just a consequence of Theorem 1.1.2 in Section 1.1. The second of these theorems generalizes the first in sense of weakening the assumptions for the reference trajectory y^*. It is the main theorem of Section 1.2.1.

In Section 1.2.2, existence, uniqueness and stability results are presented and justified, for Lipschitz w_k without the restriction of boundedness. Dismissing the restriction of boundedness of w_k in existence results in Section 1.2.2 involves slightly stronger assumptions for the reference trajectory y^* in (0.1) - (0.3) than in the main theorem in Section 1.2.1. The uniqueness and stability results in Section 1.2.2 are proven for Lipschitz w_k. The latter results do not require the restriction of boundedness and do not require the assumptions for the reference trajectory to be stronger than in the main theorem in Section 1.2.1.

In Section 1.2.3, estimates as well as existence and uniqueness for weak solutions of the system (0.1) - (0.3) are proven under the assumption that f is locally Lipschitz, fulfills the growth condition $f(s)s \leq 0$ for big $|s|$ and that $y_0 \in L^\infty(\Omega)$. These assumptions are different that the assumptions utilized in Section 1.2.2, where f is assumed to be Lipschitz and y_0 is assumed to belong to $L^2(\Omega)$. The assumptions that f is locally Lipschitz and y_0 is bounded were used in the numerical simulations for the system (0.1) - (0.3) which are described in further parts of the present work. Thus, Section 1.2.3 provides theoretical results which cover the data utilized in the subject simulations. Moreover, the results of Section 1.2.3 will be used also in some places of Chapter 3 of the present work, providing analytical background for the optimal targeting problem.

Section 1.2.4 concerns a modification of the system (0.1) - (0.3), assuming modified structure of the equations. We state the results concerning existence, uniqueness and estimates for the solutions of the modified system. For technical reasons, the subject results for the modified system will be necessary in Chapter 3. The modification of the system (0.1) - (0.3) considered in Section 1.2.4 and the original the system (0.1) - (0.3) are similar enough to apply the same methods for the analysis of the modified system. For this reason, in Section 1.2.4, we do not contain the proofs of the results described there, but we only give some remarks concerning the proofs. The results described in Section 1.2.4 will play an auxiliary role in Chapter 3, concerning the analytical aspects of the optimal targeting problem.

Remark. As mentioned above, Lipschitz continuous switching functions in the system (0.1) - (0.3) can be utilized to approximate discontinuous switching functions, as $-sgn$. We stress that switching functions equal $-sgn$ are not allowed in our results, however, instead, certain multivalued switching functions containing $-sgn$ were allowed in the results in Section 1.1, concerning the modified system (1.1) - (1.3). Assuming notation as in the example given above, results concerning the convergence of solutions of (0.1) - (0.3) corresponding to switching functions $\widetilde{w}_k^n$ to a solution of (1.1) - (1.3) corresponding to suitable multivalued switching functions containing $\widetilde{w}_k$ would be interesting. This matter was not covered in the present work and can be a field for further research. ▲

Let us proceed to the mathematical details. The below assumptions for the system (0.1) - (0.3) will be necessary in the present section:

(B-1) $\Omega \subset \mathbb{R}^{\mathbf{d}}$ is a domain that:

a) is bounded,

b) satisfies the cone condition (definition of the cone condition can be found e.g. in [1, par. 4.6.]),

(B-2) K, J are given positive natural numbers, $T > 0$, $D > 0$ and $\beta_j > 0$ for all $j = 1, \ldots, J$,

(B-3) f is globally Lipschitz continuous; we denote its Lipschitz constant by L and put $f_0 := f(0)$,

(B-4) w_k is globally Lipschitz continuous, where we denote the Lipschitz constant of w_k by L_k and put $w_{k0} := w_k(0)$, for all $k = 1, \ldots, K$,

(B-5) $y_0 \in L^2(\Omega)$, $\kappa_{j0} \in \mathbb{R}$ for $j = 1, \ldots, J$.

The necessary regularity of the reference trajectory y^* in (0.1) - (0.3) will differ in particular theorems of this section. The following two variants of the assumption concerning y^* will be in use:

(C-1) $y^* \in L^2(0, T; L^2(\Omega))$,

(C-2) $y^* \in L^\infty(0, T; L^2(\Omega))$,

The following definition of solutions for the system (0.1) - (0.3) will be utilized in the present section:

Definition 1.2.1 *An element $(y, \kappa_1, \ldots, \kappa_J) \in X^2$ is a weak solution of the system (0.1) - (0.3) if:*

(a) $y(\,.\,,0) = y_0$ *in* $L^2(\Omega)$ *and* $\kappa_j(0) = \kappa_{j0}$ *for* $j = 1, \ldots, J$,

(b) for all $\phi \in L^2(0,T;H^1(\Omega))$, *there holds*

$$\int_0^T \langle y', \phi\rangle + D\big(\nabla y, \nabla\phi\big)_{L^2(\Omega)} + \big(-f(y) - \kappa_1 g_1 - \ldots - \kappa_J g_J\,,\,\phi\big)_{L^2(\Omega)}\,dt \;=\; 0 \qquad (1.26)$$

(c) for all $\xi \in L^2(0,T)$, *for* $j = 1, \ldots, J$, *there holds*

$$\int_0^T \big(\beta_j \kappa_j' + \kappa_j - W_j(y, y^*)\big)\,\xi\,dt \;=\; 0 \qquad (1.27)$$

The point (a) in Definition 1.2.1 is meaningful, because, by arguments similar as in the case of Definition 1.1.1 (see page 6), if $(y, \kappa_1, \ldots, \kappa_J) \in X^2$ then $y \in C([0,T];L^2(\Omega))$ and $(\kappa_1, \ldots, \kappa_J) \in C([0,T])$.

1.2.1 Existence for bounded switching functions

Below, we prove existence of weak solutions for the system (0.1) - (0.3). Nevertheless, we make an assumption that the switching functions w_k, for $k = 1, \ldots, K$, not only fulfill the assumption (B-4) but moreover are bounded. If the reference trajectory y^* fulfills the assumption (A-6) in Section 1.1, then the existence result can be obtained as a consequence of results of Section 1.1. But, with the above restrictions for w_k, it is possible to prove the existence for y^* satisfying the assumption (C-1) only. It will be done below.

The restriction of boundedness of w_k is temporary — in Section 1.2.2, we will show how to dismiss it in the existence results for price of strengthening the assumptions for the reference trajectory y^* from (C-1) to (C-2).

Let us begin with short justification that the results of Section 1.1 can be applied here, under suitable assumptions. Compare Definition 1.2.1 of weak solutions for the system (0.1) - (0.3) with Definition 1.1.1 of weak solutions for the system (1.1) - (1.3), given in Section 1.1. Assume that w_k in the system (1.1) - (1.3) are single-valued functions. Then, the only possible choice of $\mathbf{W}_j$ in Definition 1.1.1 is $\mathbf{W}_j(t) := W_j(y(\,.\,,t), y^*(\,.\,,t))$ for a.e. $t \in [0,T]$. Consequently, conditions in points (c) and (d) in Definition 1.1.1 reduce to the point (c) in Definition 1.2.1. Hence, Definition 1.1.1 is equivalent to Definition 1.2.1 if w_k in the system (1.1) - (1.3) are single-valued functions.

Hence, under suitable assumptions, results concerning weak solutions of the system (1.1) - (1.3) can be transmitted to weak solutions of the system (0.1) - (0.3). Thus we conclude the below:

Theorem 1.2.2 *Let assumptions (B-1) - (B-5) be fulfilled and* $(g_j, h_k, \alpha_{jk})_{j=1,\ldots,J}^{k=1,\ldots,K} \in U$, $y^* \in C([0,T];L^2(\Omega)_w)$. *Assume additionally that* w_k *are bounded for* $k = 1, \ldots, K$. *Then, there exists a weak solution of the system (0.1) - (0.3).*

This is true, because under imposed assumptions, switching functions w_k fulfill the assumption (A-4) and the reference trajectory y^* fulfills the assumption (A-6). Thus, Theorem 1.1.2 can be applied. This theorem, together with the above remark on the equivalence of definitions of solutions, yields the assertion.

One can follow the lines of the proof of Theorem 1.1.2 to find out that the assumption $y^* \in C([0,T];L^2(\Omega)_w)$ was essential there. It was used to ensure that the operator P_2 (given by formula (1.22)) is well defined as an operator into $(C([0,T]))^K$. Enforcing the image space of P_2 to be $(C([0,T]))^K$ was required because, in the proof of Theorem 1.1.2, it was necessary to make the image space of P_2 be not larger than the domain space of the operator P_3, which was actually $(C([0,T]))^K$ (see (1.23) for the definition of P_3 in the subject proof). Next, it was needed to take $(C([0,T]))^K$ as the domain space of P_3 because it allowed to apply Lemma 1.1.7 and Lemma 1.1.8 to P_3, what was an essential step of the proof of Theorem 1.1.2 (more precisely, the subject lemmas were applied not to P_3 directly, but to certain operators entering its definition; nevertheless, one can verify that the latter does not change the conclusion concerning the requirement on the domain space of P_3). To sum up, assumption $y^* \in C([0,T]; L^2(\Omega)_w)$ was essential for Theorem 1.1.2 and hence cannot be relaxed in Theorem 1.2.2, as long as we derive the latter as a corollary of the former.

On the other hand, it is not necessary to derive the theorem on the existence of weak solutions of (0.1) - (0.3) as a corollary of Theorem 1.1.2. One can prove it separately and, due to the strengthened assumption concerning the switching functions w_k, obtain a result allowing a weakened assumption for the reference trajectory y^*. The below theorem realizes the latter postulate:

Theorem 1.2.3 *Assume that general assumptions (B-1) - (B-5) together with (C-1) hold and* $(g_j, h_k, \alpha_{jk})_{j=1,\dots,J}^{k=1,\dots,K} \in U$. *Assume moreover that functions* w_k *are bounded for* $k = 1, \dots, K$. *Then the system (0.1) - (0.3) has a weak solution.*

The proof bases on the Schauder fixed theorem, formulated below for convenience. The Schauder theorem is less general that the generalized Kakutani theorem (Theorem 1.1.9), utilized for the proof of Theorem 1.1.2, but sufficient for the proof of Theorem 1.2.3.

Theorem 1.2.4 (Schauder theorem) *Let* X *be a Banach space. Let* M *be a convex, compact and nonempty subset of* X. *Let* $T\colon M \to M$ *be continuous. Then* T *has a fixed point, i.e. there exists* $\bar{x} \in M$ *such that* $\bar{x} = T(\bar{x})$.

The above version of the Schauder fixed point theorem is given in Corollary 2.13 in Chap. 2.6 in [50].

Proof of Theorem 1.2.3. We define following operators:

- $P_1\colon \left(L^2(0,T)\right)^J \to C([0,T]; L^2(\Omega))$ is the operator assigning the solution of (1.4) to a given $(k_1, \dots, k_j) \in \left(L^2(0,T)\right)^J$. It is well defined since, by Lemma 1.1.5, for $(k_1, \dots, k_j)$ as declared, the solution of (1.4) exists in X^y, is unique and $X^y \hookrightarrow C([0,T]; L^2(\Omega))$ (by [51, Prop. 23.23]).
- $P_2\colon C([0,T]; L^2(\Omega)) \to \left(L^2(0,T)\right)^J$ assigns $(\mathbf{V}_1, \dots, \mathbf{V}_J)$ given by formula
$$\mathbf{V}_j(t) = \sum_{k=1}^{K} \alpha_{j,k} w_k \Big(\int_\Omega h_k(x) \left(Y(x,t) - y^*(x,t)\right) dx \Big) \quad \text{a.e. on } [0,T] \tag{1.28}$$
to a given $Y \in C([0,T]; L^2(\Omega))$. We can verify that P_2 is well defined. More precisely, Hölder inequality allows to infer that $\mathbf{v}_k$ defined for $k = 1, \dots, K$ by
$$\mathbf{v}_k := \int_\Omega h_k(x)(Y(x,t) - y^*(x,t))\, dx$$
belong to $L^2(0,T)$, for Y as declared and y^* as in the assumption (C-1). If $\mathbf{V} = P_2(Y)$, then $\mathbf{V}_j = \sum_{k=1}^K \alpha_{j,k} w_k \circ \mathbf{v}_k$. Hence, $\mathbf{V}_j$ are measurable as sums of superpositions of continuous w_k with measurable $\mathbf{v}_k$. In addition, $\mathbf{V}_j$ are also bounded because w_k are bounded. Thus, $\mathbf{V}_j$ belongs not only to $L^2(0,T)$ but even to $L^\infty(0,T)$, for $j = 1, \dots, J$.

- $P_3\colon \big(L^2(0,T)\big)^J \to \big(L^2(0,T)\big)^J$ assigns the solution of (1.5) for a given $(\mathbf{V}_1,\dots,\mathbf{V}_J) \in \big(L^2(0,T)\big)^J$. It is well defined since, by Lemma 1.1.6, for $(\mathbf{V}_1,\dots,\mathbf{V}_J)$ as declared, the solution of (1.5) exists in X^κ and is unique, and $X^\kappa \hookrightarrow \big(L^2(0,T)\big)^J$.

The role of the above operators in the context of the system (0.1) - (0.3) is illustrated in Figure 1.3.

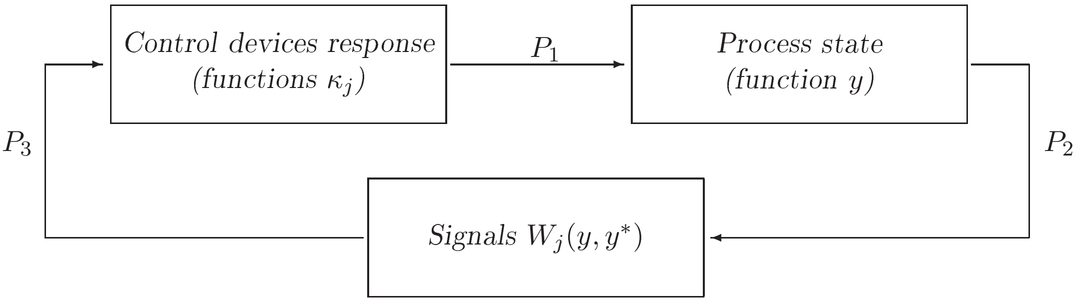


Figure 1.3: A schematic representation of the role of the operators P_1, P_2 and P_3, considered in the proof of Theorem 1.2.3, in the context of the thermostat control mechanism, present in the system (0.1) - (0.3). The notation in the picture is as in the subject system. Comparing to the proof of Theorem 1.1.2, the state-to-measurement and measurement-to-signal operators considered there (see Figure 1.1) are „merged" in the present proof into the state-to-signal operator. The latter simplification is made because in the present situation the necessary properties of the state-to-signal operator are easy enough to obtain „in one turn", without splitting the subject mapping into two separate operator.

Proving that $P := P_3 \circ P_2 \circ P_1$ has a fixed point in $L^2(0,T)$ is equivalent to proving the assertion of the theorem. In other words, we need to prove that there exists $\bar{k} \in L^2(0,T)$ such that $\bar{k} = P_3(\mathbf{V})$, $\mathbf{V} = P_2(Y)$, $Y = P_1(\bar{k})$.

By Lemma 1.1.5, the operator P_1 is continuous.

By the assumption that w_k are Lipschitz continuous for $k = 1,\dots,K$, we also verify the continuity of P_2. Let $\mathbf{V}^1 = P_2(Y^1)$ and $\mathbf{V}^2 = P_2(Y^2)$ for given $Y^1, Y^2 \in C([0,T];L^2(\Omega))$. Then:

$$\begin{aligned}\big\|\mathbf{V}_j^1 - \mathbf{V}_j^2\big\|_{L^2(0,T)} &\le T^{1/2}\big\|\mathbf{V}_j^1 - \mathbf{V}_j^2\big\|_{L^\infty(0,T)}\\ &\le T^{1/2}\ \operatorname{ess\,sup}_{t\in[0,T]}\sum_{k=1}^{K}\alpha_{j,k}L_k\Big|\int_\Omega h_k(x)(Y^1(x,t) - Y^2(x,t))\,dx\Big|\\ &\le T^{1/2}\Big(\sum_{k=1}^{K}\alpha_{j,k}L_k\big\|h_k\big\|_2\Big)\big\|Y^1 - Y^2\big\|_{2,\infty}\end{aligned}$$

for $j = 1,\dots,J$, where L_k are the Lipschitz constants of w_k, as in the assumption (B-4).

Moreover, by the linear structure of (1.5), the operator P_3 is affine. By the estimate (1.18) in Lemma 1.1.6, the operator P_3 is also bounded. Therefore, as a bounded affine operator, P_3 is continuous from $\big(L^2(0,T)\big)^J$ to X^κ. Since X^κ can be embedded continuously into $\big(L^2(0,T)\big)^J$, P_3 is also continuous with values in $\big(L^2(0,T)\big)^J$.

Summing up the above considerations, $P_3 \circ P_2 \circ P_1$ is continuous from $\big(L^2(0,T)\big)^J$ to itself.

Next, recall the assumption that w_k are bounded. We denote $C_{w_k} := \big\|w_k\big\|_{L^\infty(\mathbb{R})}$ for $k =$

$1, \ldots, K$. It is straightforward, that $P_2 \colon C([0,T]; L^2(\Omega)) \to \mathbb{A}$ for

$$\mathbb{A} := \left\{ (\mathbf{V}_1, \ldots, \mathbf{V}_J) \in \left(L^2(0,T)\right)^J : \|\mathbf{V}_j\|_{L^2(0,T)} \leq T\|\mathbf{V}_j\|_{L^\infty(0,T)} \leq TC_{W_j} \; \forall_{j=1,\ldots,J} \right\}$$

where $C_{W_j} := \sum_{k=1}^K \alpha_{j,k} C_{w_k}$, for $j = 1 \ldots, J$. By estimate (1.18) in Lemma 1.1.6, we also get that $P_3|_{\mathbb{A}} \colon \mathbb{A} \to \mathbb{B}$ for

$$\mathbb{B} := \left\{ k \in \left(L^2(0,T)\right)^J : \|k\|_{X^\kappa} \leq C_3 \sum_{j=1}^J \left(|\kappa_{j0}| + TC_{W_j}\right) \; \forall_{j=1,\ldots,J} \right\}$$

where κ_{j0} are the initial conditions assumed for (1.2) in the assumption (A-5) and C_3 is the constant appearing in the estimate (1.18) in Lemma 1.1.6. Thus superposition $P_3 \circ P_2 \circ P_1$ takes values in $\mathbb{B}$ as well.

The set $\mathbb{B}$ is nonempty and convex. The closure of $\mathbb{B}$ in $\left(L^2(0,T)\right)^J$, denote it $\mathbf{B}$, is in addition compact (by Rellich-Kondrachov theorem, see [1, Th. 6.3]).

To sum up, we have shown that $P = P_3 \circ P_2 \circ P_1 \colon \left(L^2(0,T)\right)^J \to \mathbf{B}$, where $\mathbf{B}$ is nonempty, convex and compact in $\left(L^2(0,T)\right)^J$ and P is continuous from $\left(L^2(0,T)\right)^J$ to itself, and thus from $\mathbf{B}$ to itself. Hence, P has a fixed point in $\mathbf{B}$ by the Schauder theorem (Theorem 1.2.4). ■

Remark. The only step in the proof of Theorem 1.2.3 where the condition $\beta_j > 0$, being a part of the assumption (B-2), was used was the application of Lemma 1.1.6, which also assumes $\beta_j > 0$. However, it is possible to prove a version of Lemma 1.1.6 allowing $\beta_j < 0$ (what was pointed out in the remark on page 11). Hence, a version of Theorem 1.2.3 allowing $\beta_j < 0$ also would be valid.

An analogous remark hold for Theorem 1.1.2, and hence for Theorem 1.2.2, being a corollary of the former result, as well. ▲

The result given in Theorem 1.2.3 detaches us from the requirement of the weak continuity of the reference trajectory, present in Theorem 1.2.2. This can be essential in certain situations. For example, it seems natural to allow the user of the thermostat control mechanism to change the reference state that he would like to keep. Thus, there can be some switching moment during the experiment. E.g., for time from 0 up to a given $t_1 < T$, the user may want to keep the state of the process close to some state $y_1^* \colon \Omega \to \mathbb{R}$ and then, for times grater than t_1, he may decide to change the state that he want to be close to from y_1^* to some $y_2^* \colon \Omega \to \mathbb{R}$. It would be inconvenient for the user to force him to focus on how he should change his target from y_1^* to y_2^* in order not to break the requirement of the weak continuity. In this sense, it would be better if the thermostat control mechanism allowed the user to just switch the state that he wants to keep. Here, Theorem 1.2.3 have the advantage over Theorem 1.2.2.

For concrete example of situation of the above kind, consider two square integrable functions y_1^* and y_2^*, $y_1^*, y_2^* \colon \Omega \to \mathbb{R}$, such that $\int_\Omega y_1^*(x)\,dx \neq \int_\Omega y_2^*(x)\,dx$. Let the reference trajectory y^* in the system (0.1) - (0.3) be given by

$$y^*(x,t) = \begin{cases} y_1^*(x) & \text{for } t \leq t_1 \\ y_2^*(x) & \text{for } t > t_1 \end{cases}$$

where $t_1 \in (0,T)$ is known. Then, y^* is an element of $L^2(0,T;L^2(\Omega))$ but is not an element of $C([0,T];L^2(\Omega)_w)$. To justify the latter, note that, by assumptions on y_1^* and y_2^*, integral $\int_\Omega y^*(x,t)\phi(x)\,dx$ can be discontinuous in time, what is the case e.g. for $\phi \equiv 1$ on Ω. Therefore, for the reference trajectory y^* as above, it is possible to apply Theorem 1.2.3 but not Theorem 1.2.2.

1.2.2 Existence, uniqueness and stability for general case

In Section 1.2.1, we have proven the existence of weak solutions of (0.1) - (0.3) for the case of switching functions fulfilling the assumption (B-4), being additionally bounded. Here, we are going to extend this results and prove not only existence but also uniqueness and stability for arbitrary switching functions fulfilling the assumption (B-4). Nevertheless, the existence results from Section 1.2.1 form a base, necessary for some of arguments utilized in the present section.

The stability of (0.1) - (0.3) will be investigated w.r.t. both the control and the initial condition. We will also prove the weak subsequential stability of (0.1) - (0.3) when the control space is considered with its weak topology.

The price for obtaining the above mentioned existence results for arbitrary switching functions w_k obeying the assumption (B-4) will be a slightly stronger assumption for y^*, in comparison to Theorem 1.2.3 in Section 1.2.1. More precisely, the new existence result will require the assumption (C-2) instead of the assumption (C-1). Fortunately, the strengthened assumption for y^* is still weaker than that indicated in Theorem 1.2.2 in Section 1.2.2.

The above announced existence result will involve some additional estimates for weak solutions of the system (0.1) - (0.3). Moreover, the uniqueness result will rely on the stability of the system (0.1) - (0.3) with respect to perturbations of the initial condition. Hence, we start this section with proving the necessary estimates and the stability results. Next, we proceed to existence and uniqueness results. In the final part of the present section, we focus on the results concerning the weak subsequential stability of (0.1) - (0.3).

Theorem 1.2.5 *Let the part a) in the assumption (B-1) and assumptions (B-2) - (B-4) together with (C-1) be fulfilled, let $\hat{u} \in U$ and $(y_0, \kappa_{10}, \dots, \kappa_{J0}) \in X^0$. Assume also that $\|\hat{u}\|_U \le R^U$ for some $R^U > 0$ and that $\|(y_0, \kappa_{10}, \dots, \kappa_{J0})\|_{X^0} \le R^0$ for some $R^0 > 0$. Let $(y, \kappa_1, \dots, \kappa_J) \in X^2$ be a weak solution of the system (0.1) - (0.3) corresponding to $g_j := \hat{u}_{g_j}$, $h_k := \hat{u}_{h_k}$, $\alpha_{j,k} := \hat{u}_{\alpha_{j,k}}$ and the initial condition $(y_0, \kappa_{10}, \dots, \kappa_{J0})$. Then the following estimate holds:*

$$\|(y, \kappa_1, \dots, \kappa_J)\|_{X^2} \;\le\; C$$

where

$$C = C(T, |\Omega|, K, J, L, f_0, L_1, \dots, L_K, w_{10}, \dots, w_{K0}, R^U, R^0, \|y^*\|_{2,2}, D, \beta_1, \dots, \beta_J)$$

and where the appearing quantities are the same as those in the general assumptions referred to above.

Proof. We test the weak form (1.26) of the equation for y by $\phi(x,s) := y(x,s)\mathbf{1}_{(0,t)}(s)$, for certain $t \in [0,T]$, and obtain:

$$\int_0^t \langle y', y\rangle + D\|\nabla y\|_2^2 \, ds \;=\; \int_0^t (f(y), y)_{L^2(\Omega)} + \sum_{j=1}^J (\kappa_j \hat{u}_{g_j}, y)_{L^2(\Omega)} \, ds \tag{1.29}$$

Next, we estimate term $(f(y), y)_{L^2(\Omega)}$ in (1.29) by using $|f(s)| \le |f_0| + L|s|$ (what is true by the assumption (B-3)), by the Hölder inequality and by the Young inequality and our structural assumptions:

$$\begin{aligned}\int_\Omega f(y)y \, dx \;&\le\; \int_\Omega L|y|^2 \, dx + f_0 \int_\Omega |y| \, dx \;\le L\|y\|_2^2 + f_0\|y\|_2\|\mathbf{1}_\Omega\|_2 \\ &\le L\|y\|_2^2 + \frac{f_0}{2}\|y\|_2^2 + \frac{f_0}{2}\|\mathbf{1}_\Omega\|_2^2\end{aligned} \tag{1.30}$$

By the Hölder and Young inequalities and the definition of constant R^U, term $(\kappa_j \hat{u}_{g_j}, y)_{L^2(\Omega)}$ in (1.29) can be estimated, for each $j = 1, \ldots, J$, by:

$$(\kappa_j \hat{u}_{g_j}, y)_{L^2(\Omega)} \;=\; |\kappa_j| \|\hat{u}_{g_j}\|_2 \|y\|_2 \;\leq\; \frac{1}{2}\|y\|_2^2 + \frac{1}{2}\big(R^U\big)^2 |\kappa_j|^2 \tag{1.31}$$

Spaces $H^1(\Omega)$, $L^2(\Omega)$ and $H^1(\Omega)^*$ form an evolution triple with embeddings $H^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow H^1(\Omega)^*$, hence the identity $\int_0^t \langle y', y\rangle = \frac{1}{2}\|y(\,.\,,t)\|_2^2 - \frac{1}{2}\|y(\,.\,,0)\|_2^2$ holds (see Prop. 23.23 in [51]). By the latter, by the relation $y(.,0) = y_0$ and by (1.29), (1.30) and (1.31), we obtain:

$$\begin{aligned} \frac{1}{2}\|y(.,t)\|_2^2 \;+\; D\int_0^t \|\nabla y\|_2^2\, ds \;\leq \frac{1}{2}\int_0^t C_1\|y\|_2^2 \;+\; \big(R^U\big)^2 \sum_{j=1}^{J} |\kappa_j|^2\, ds \quad + \\ +\; \frac{1}{2}C_2 \;+\; \frac{1}{2}\|y_0\|_2^2 \end{aligned} \tag{1.32}$$

where

$$C_1 = \Big(2L + |f_0| + J\Big), \qquad C_2 = Tf_0|\Omega|$$

Above, the assumption that Ω is bounded was necessary to ensure that $\|\mathbf{1}_\Omega\|_2$ is finite.

At the same time, testing the weak form (1.27) of the equation for κ_j by $\xi(s) := \kappa_j(s)\mathbf{1}_{(0,t)}(s)$, neglecting the appearing $|\kappa_j|^2$ term (which is nonnegative), expanding the definition of W_j (given in (0.3)) and using the Young inequality yields:

$$\begin{aligned} \beta_j \int_0^t \kappa_j' \kappa_j\, ds \;&\leq\; \int_0^t \sum_{k=1}^{K} \hat{u}_{\alpha_{jk}} w_k\Big(\int_\Omega \hat{u}_{h_k}(y-y^*)\, dx\Big)\, \kappa_j\, ds \\ \leq \frac{1}{2}\int_0^t \sum_{k=1}^{K} \hat{u}_{\alpha_{jk}}^2 w_k &\Big(\int_\Omega \hat{u}_{h_k}(y-y^*)\, dx\Big)^2 ds \;+\; \frac{1}{2}\int_0^t K|\kappa_j|^2\, ds \end{aligned} \tag{1.33}$$

By the assumption (B-4), the Hölder inequality and the definition of R^U, the first term appearing in the sum obeys:

$$\begin{aligned} &\Big|\hat{u}_{\alpha_{jk}} w_k\Big(\int_\Omega \hat{u}_{h_k}(y-y^*)\, dx\Big)\Big|^2 \;\leq\; |\hat{u}_{\alpha_{jk}}|^2 \Big(|w_{k0}| + L_k\|h_k\|_2\|y-y^*\|_2\Big)^2 \\ &\leq \big(R^U\big)^2\Big(|w_{k0}| + L_k R^U\big(\|y\|_2 + \|y^*\|_2\big)\Big)^2 \\ &\leq 2\big(R^U\big)^2 w_{k0}^2 \;+\; 2\big(R^U\big)^4 L_k^2\big(\|y\|_2 + \|y^*\|_2\big)^2 \\ &\leq 2\big(R^U\big)^2 w_{k0}^2 \;+\; 4\big(R^U\big)^4 L_k^2\|y^*\|_2^2 \;+\; 4\big(R^U\big)^4 L_k^2\|y\|_2^2 \end{aligned}$$

From the above, we derive the following:

$$\int_0^t \sum_{k=1}^{K} \Big|\hat{u}_{\alpha_{jk}}\, w_k\Big(\int_\Omega \hat{u}_{h_k}(y-y^*)\, dx\Big)\Big|^2 ds \;\leq\; C_{3,j}\int_0^t \|y\|_2^2\, ds \;+\; C_{4,j} \;+\; C_{5,j} \tag{1.34}$$

where

$$
\begin{aligned}
C_{3,j} &= 4\big(R^U\big)^4 \sum_{k=1}^{K} L_k^2 \\
C_{4,j} &= 4\big(R^U\big)^4 \sum_{k=1}^{K} L_k^2 \, \|y^*\|_{2,2}^2 \\
C_{5,j} &= 2T\big(R^U\big)^2 \sum_{k=1}^{K} w_{k0}^2
\end{aligned}
$$

As $\kappa_j{}'$ is integrable, κ_j is absolutely continuous. Thus, by integration by parts, identity $\int_0^t \kappa_j' \kappa_j = \frac{1}{2}\big|\kappa_j(t)\big|^2 - \frac{1}{2}\big|\kappa_j(0)\big|^2$ holds. Combining the latter with the relation $\kappa_j(0) = \kappa_{j0}$ and with estimates (1.33) and (1.34) yields, for $j = 1, \ldots, J$:

$$
\frac{1}{2}\big|\kappa_j(t)\big|^2 \;\le\; \frac{1}{2\beta_j}\int_0^t C_{3,j}\big\|y\big\|_2^2 \;+\; K\big|\kappa_j\big|^2\, ds \quad+\; \frac{1}{2\beta_j}\big(C_{4,j} + C_{5,j}\big) \;+\; \frac{1}{2}\big|\kappa_{j0}\big|^2 \tag{1.35}
$$

After summation of (1.32) and (1.35) for every j and neglecting the gradient term (which is nonnegative), we obtain:

$$
\begin{aligned}
\big\|y(.,t)\big\|_2^2 + \sum_{j=1}^{J}\big|\kappa_j(t)\big|^2 \;\le\; & \int_0^t C_6\big\|y\big\|_2^2 \;+\; C_7\sum_{j=1}^{J}\big|\kappa_j\big|^2\, ds \quad + \\
& + \; C_8 \;+\; \big\|y_0\big\|_2^2 \;+\; \sum_{j=1}^{J}\big|\kappa_{j0}\big|^2
\end{aligned}
\tag{1.36}
$$

where

$$
\begin{aligned}
C_6 &= C_1 + \sum_{j=1}^{J} \beta_j^{-1} C_{3,j} \\
C_7 &= \big(R^U\big)^2 + K\sum_{j=1}^{J} \beta_j^{-1} \\
C_8 &= C_2 \;+\; \sum_{j=1}^{J} \beta_j^{-1}\big(C_{4,j} + C_{5,j}\big)
\end{aligned}
$$

Now, by the definition of R^0, one can verify that

$$
\big\|y_0\big\|_2^2 + \sum_{j=1}^{J}\big|\kappa_{j0}\big|^2 \;\le\; (J+1)\big\|(y_0, \kappa_{10}, \ldots, \kappa_{J0})\big\|_{X^0}^2 \;\le\; (J+1)\big(R^0\big)^2
$$

Using the above in (1.36) and applying the integral Grönwall inequality allows to find that

$$
\begin{aligned}
& \big\|y\big\|_{2,\infty}^2 \;+\; \sum_{j=1}^{J}\big\|\kappa_j\big\|_{L^\infty(0,T)}^2 \;\le \\
& \le \Big(C_8 \;+\; (J+1)\big(R^0\big)^2\Big) \cdot \Big(1 \;+\; T\max\{C_6, C_7\} e^{T\max\{C_6,C_7\}}\Big)
\end{aligned}
\tag{1.37}
$$

The structure of the constants C_6, C_7, C_8 guarantees that the right hand side of the above depends only on the quantities stated in the assertion of the theorem.

Still, to complete the proof we need to estimate norms $\|\nabla y\|_{2,2}$, $\|y'\|_{H^1(\Omega)^*,2}$ and $\|\kappa_j'\|_{L^2(0,T)}$, since they enter the definition of the norm of the space X^2. For estimating the gradient term, we again use the inequality (1.32) with $t = T$, neglecting $\|y(.,t)\|_2^2$ term:

$$\begin{aligned} D\|\nabla y\|_{2,2}^2 &\leq \frac{1}{2}\int_0^T C_1\|y\|_2^2 \;+\; \big(R^U\big)\sum_{j=1}^J |\kappa_j(s)|^2\, ds \;+\; \frac{1}{2}C_2 \;+\; \frac{1}{2}\|y_0\|_2^2 \\ &\leq \frac{T}{2}C_1\|y\|_{2,\infty}^2 \;+\; \frac{T}{2}\big(R^U\big)\sum_{j=1}^J \|\kappa_j\|_{L^\infty(0,T)}^2 \;+\; \frac{1}{2}C_2 \;+\; \frac{1}{2}\|y_0\|_2^2 \end{aligned} \tag{1.38}$$

Next, use the relation $\|y_0\|_2 \leq R^0$ and apply (1.37) to estimate the right hand side of the above inequality in terms of C_1, C_2, C_6, C_7, C_8, T, J, R^U and R^0, which depend at most on the quantities stated in the theorem.

To obtain estimates for the time derivative of y, we treat the weak form (1.26) of (0.1) as an equality of functionals on the space $L^2(0,T;H^1(\Omega))$. We rewrite it in the below form:

$$y' + D\mathbf{A}y - \mathbf{F}y - \mathbf{G} = 0 \quad \text{in } L^2(0,T;H^1(\Omega)^*) \tag{1.39}$$

where $\mathbf{A}y$, $\mathbf{F}y$ and G are defined by

$$\begin{aligned} \int_0^T \langle \mathbf{A}y, \phi\rangle\, dt &= \int_0^T \Big(\nabla y, \nabla\phi\Big)_{L^2(\Omega)}^2 dt \\ \int_0^T \langle \mathbf{F}y, \phi\rangle\, dt &= \int_0^T \Big(f(y), \phi\Big)_{L^2(\Omega)}^2 dt \\ \int_0^T \langle \mathbf{G}, \phi\rangle\, dt &= \int_0^T \Big(\sum_{j=1}^J \kappa_j g_j, \phi\Big)_{L^2(\Omega)} dt \end{aligned}$$

for $\phi \in L^2(0,T;H^1(\Omega))$.

It follows by the definition of the above functionals that

$$\|\mathbf{A}y\|_{H^1(\Omega)^*,2} \leq \|\nabla y\|_{2,2}, \quad \|\mathbf{F}y\|_{H^1(\Omega)^*,2} \leq \|f(y)\|_{2,2}, \quad \|\mathbf{G}\|_{H^1(\Omega)^*,2} \leq \sum_{j=1}^J \|\kappa_j g_j\|_{2,2} \tag{1.40}$$

This, along with (1.39), yields:

$$\begin{aligned} \|y'\|_{H^1(\Omega)^*,2} &\leq D\|\nabla y\|_{2,2} \;+\; \|f(y)\|_{2,2} \;+\; \sum_{j=1}^J \|\kappa_j \hat{u}_{g_j}\|_{2,2} \\ &\leq D\|\nabla y\|_{2,2} \;+\; \||f_0| + L|y|\|_{2,2} \;+\; \sum_{j=1}^J \|\hat{u}_{g_j}\|_2 \|\kappa_j\|_{L^2(0,T)} \\ &\leq D\|\nabla y\|_{2,2} \;+\; T^{1/2}L\|y\|_{2,\infty} \;+\; TR^U\sum_{j=1}^J \|\kappa_j\|_{L^\infty(0,T)} \;+\; (T|\Omega|)^{1/2}|f_0| \end{aligned} \tag{1.41}$$

where we have used the Lipschitz continuity of f, the Hölder inequality and the definition of R^U. Now, (1.37) and (1.38) can be applied to estimate the right hand side of (1.41) in terms of C_1, C_2, C_6, C_7, C_8, D, $|\Omega|$, T, L, f_0, J, R^U and R^0.

Moreover, by (1.27), one can infer that

$$\beta_j \kappa_j' \;+\; \kappa_j \;=\; W_j(y, y^*) \qquad \text{in } L^2(0,T)$$

for $j = 1, \ldots, J$. By the above, expanding the definition of W_j given in (0.3), we have

$$\beta_j^2 \big\|\kappa_j'\big\|^2_{L^2(0,T)} \;\leq\; 2\big\|\kappa\big\|^2_{L^2(0,T)} \;+\; 2\Big\|\sum_{k=1}^{K} \hat{u}_{\alpha_{jk}} w_k \Big(\int_\Omega \hat{u}_{h_k}(y - y^*)\, dx\Big)\Big\|^2_{L^2(0,T)} \tag{1.42}$$

Dividing (1.42) by β_j^2 and using (1.34) to estimate the second term in the right hand side, we obtain the following:

$$\begin{aligned} \big\|\kappa_j'\big\|^2_{L^2(0,T)} &\leq 2\beta_j^{-2}\Big(\big\|\kappa\big\|^2_{L^2(0,T)} \;+\; KC_{3,j}\big\|y\big\|^2_{2,2} \;+\; KC_{4,j} \;+\; KC_{5,j}\Big) \\ &\leq 2\beta_j^{-2}\Big(T\big\|\kappa\big\|^2_{L^\infty(0,T)} \;+\; KC_{3,j}T\big\|y\big\|^2_{2,\infty} \;+\; KC_{4,j} \;+\; KC_{5,j}\Big) \end{aligned} \tag{1.43}$$

Constant K above appears due to moving the square power to the terms under the sum $\sum_{k=1}^K$, according to general inequality $\big|\sum_k a_k\big|^2 \leq K\sum_k |a_k|^2$. Now, (1.37) can be applied to estimate terms $\big\|\kappa_j\big\|_{L^\infty(0,T)}$ and $\big\|y\big\|_{2,\infty}$. This gives a bound for the right hand side of (1.43)in terms of β_j, $C_{3,j}$, $C_{4,j}$, $C_{5,j}$, C_6, C_7, C_8, T, K, J and R^0, which depend at most on the quantities stated in the theorem.

Altogether, (1.37), (1.38), (1.41) and (1.43) guarantee that all the investigated norms can be estimated in terms of the constants which depend at most on the quantities stated in the assertion of the theorem. ■

We now proceed to the stability of the system (0.1) - (0.3). During the lecture of the proof of the below stability theorem, one can note that the proof utilizes the above proven Theorem 1.2.5, concerning the estimates of the weak solutions of the system (0.1) - (0.3).

Theorem 1.2.6 *Let the part a) in the assumption (B-1) and assumptions (B-2) - (B-4) together with (C-1) be fulfilled, let $\hat{u}^1, \hat{u}^2 \in U$ and*

$$(y_0^1, \kappa_{10}^1, \ldots, \kappa_{J0}^1),\; (y_0^2, \kappa_{10}^2, \ldots, \kappa_{J0}^2) \;\in\; X^0$$

Assume also that $\big\|\hat{u}^i\big\|_U \leq R^U$ for some $R^U > 0$ and that $\big\|(y_0^i, \kappa_{10}^i, \ldots, \kappa_{J0}^i)\big\|_{X^0} \leq R^0$ for some $R^0 > 0$, for $i = 1, 2$. Let $(y^i, \kappa_1^i, \ldots, \kappa_J^i) \in X^2$ be a weak solution of the system (0.1) - (0.3) corresponding to $g_j := \hat{u}^i_{g_j}$, $h_k := \hat{u}^i_{h_k}$, $\alpha_{j,k} := \hat{u}^i_{\alpha_{j,k}}$ and the initial condition $(y_0^i, \kappa_{10}^i, \ldots, \kappa_{J0}^i)$, for $i = 1, 2$. Denote $y = y^1 - y^2$, $\kappa_j = \kappa_j^1 - \kappa_j^2$, $\hat{u} = \hat{u}^1 - \hat{u}^2$, $y_0 = y_0^1 - y_0^2$ and $\kappa_{j0} = \kappa_{j0}^1 - \kappa_{j0}^2$. Then:

$$\big\|(y, \kappa_1, \ldots, \kappa_J)\big\|_{X^2} \;\leq\; C\left(\big\|\hat{u}\big\|_U^2 + \big\|(y_0, \kappa_{10} \ldots, \kappa_{J0})\big\|^2_{X^0}\right)^{1/2}$$

where

$$C = C(T, |\Omega|, K, J, L, f_0, L_1, \ldots, L_K, w_{10}, \ldots, w_{K0}, R^U, R^0, \big\|y^*\big\|_{2,2}, D, \beta_1, \ldots, \beta_J)$$

and where the appearing quantities are the same as those in the general assumptions referred to above.

Proof. For $i = 1, 2$, the function y^i satisfies the identity (1.26) with $\kappa_j := \kappa^i_j$ and $g_j := g^i_j$, for $j = 1, \dots, J$. For $i = 1, 2$ and for $j = 1, \dots, J$, the function κ^i_j satisfies the identity (1.27), with $y = y^i$ and with $W_j := W^i_j$, where

$$W^i_j(y(\,.\,,t), y^*(\,.\,,t)) := \sum_{k=1}^{K} \alpha^i_{jk} w_k\Big(\int_\Omega h^i_k(x)\big(y(x,t) - y^*(x,t)dx\big)\Big) \qquad \text{for } i = 1,2$$

Subtracting by sides the identities corresponding to y^1 and y^2 and subtracting by sides the identities corresponding to κ^1_j and κ^2_j, for $j = 1, \dots, J$, we obtain:

$$\begin{aligned}\int_0^T \langle y', \phi\rangle + D\big(\nabla y, \nabla\phi\big)_{L^2(\Omega)}\, ds \;&=\; \int_0^T \big(f(y^1) - f(y^2)\,,\, \phi\big)_{L^2(\Omega)}\, ds \;+ \\ &\quad+ \int_0^T \big(\sum_{j=1}^{J} \kappa^1_j g^1_j - \sum_{j=1}^{J} \kappa^2_j g^2_j\,,\, \phi\big)_{L^2(\Omega)}\, ds\end{aligned} \tag{1.44}$$

for all $\phi \in L^2(0,T; H^1(\Omega))$ and

$$\int_0^T \big(\beta_j\kappa'_j + \kappa_j\big)\,\xi\, dt \;=\; \int_0^T \big(W^1_j(y^1, y^*) - W^2_j(y^2, y^*)\big)\,\xi\, dt \tag{1.45}$$

for all $\xi \in L^2(0,T)$, for $j = 1, \dots, J$.

Now, we proceed as in the proof of Theorem 1.2.5. The present proof is very similar however requires longer calculations, which involves multiple use of the triangle inequality.

Testing the identity (1.44) by $\phi(x,s) := y(x,s)\mathbf{1}_{(0,t)}(s)$ yields:

$$\begin{aligned}\int_0^t \langle y', y\rangle + D\big\|\nabla y\big\|_2^2\, ds \;&=\; \int_0^t (f(y^1) - f(y^2), y^1 - y^2)_{L^2(\Omega)} \;+ \\ &\quad+ \sum_{j=1}^{J} (\hat{u}^1_{g_j}\kappa^1_j - \hat{u}^2_{g_j}\kappa^2_j, y^1 - y^2)_{L^2(\Omega)}\, ds\end{aligned} \tag{1.46}$$

By the Lipschitz continuity of f we have:

$$(f(y^1) - f(y^2), y^1 - y^2)_{L^2(\Omega)} \;\le\; L\big\|y^1 - y^2\big\|_2 \tag{1.47}$$

while for the second term on the right hand side of (1.46) we can write

$$\begin{aligned}&(\hat{u}^1_{g_j}\kappa^1_j - \hat{u}^2_{g_j}\kappa^2_j, y^1 - y^2)_{L^2(\Omega)} \;=\; (\hat{u}^1_{g_j}\kappa^1_j - \hat{u}^2_{g_j}\kappa^1_j, y)_{L^2(\Omega)} \;+\; (\hat{u}^2_{g_j}\kappa^1_j - \hat{u}^2_{g_j}\kappa^2_j, y)_{L^2(\Omega)} \\ &\le \big|\kappa^1_j\big|\big\|\hat{u}_{g_j}\big\|_2\big\|y\big\|_2 \;+\; \big|\kappa_j\big|\big\|\hat{u}^2_{g_j}\big\|_2\big\|y\big\|_2 \\ &\le \frac{1}{2}\big|\kappa^1_j\big|^2\big\|\hat{u}_{g_j}\big\|_2^2 \;+\; \frac{1}{2}\big\|y\big\|_2^2 \;+\; \frac{1}{2}\big\|\hat{u}^2_{g_j}\big\|_2^2\big|\kappa_j\big|^2 \;+\; \frac{1}{2}\big\|y\big\|_2^2 \\ &\le \frac{1}{2}C_1\big\|\hat{u}_{g_j}\big\|_2^2 \;+\; \frac{1}{2}(R^U)^2\big|\kappa_j\big|^2 \;+\; \big\|y\big\|_2^2\end{aligned} \tag{1.48}$$

where C_1 denotes the constant from the assertion of Theorem 1.2.5 — it states that the square of the supremum of each κ_j is bounded by this constant. Note, that the imposed assumptions cover the assumptions of Theorem 1.2.5, hence the latter can be applied.

Now, the relation $\int_0^t \langle y', y\rangle = \frac{1}{2}\big\|y(\,.\,,t)\big\|_2^2 - \frac{1}{2}\big\|y(\,.\,,0)\big\|_2^2$ (see the comments preceding (1.32) in the proof of Theorem 1.2.5) and relations $y(\,.\,,0) = y_0$, (1.46), (1.47), (1.48) together imply

$$\begin{aligned} &\frac{1}{2}\big\|y(\,.\,,t)\big\|_2^2 \;-\; \frac{1}{2}\big\|y_0\big\|_2^2 \;+\; D\big\|\nabla y\big\|_2^2\,ds \;\le \\ &\le \int_0^t (L+J)\big\|y\big\|_2^2 \;+\; \frac{1}{2}(R^U)^2\sum_{j=1}^{J}\big|\kappa_j\big|^2\,ds \quad+\; \frac{1}{2}TC_1\sum_{j=1}^{J}\big\|\hat{u}_{g_j}\big\|_2^2 \end{aligned} \tag{1.49}$$

A similar procedure can be performed for the equation for κ_j — for $j = 1, \ldots, J$, we test the identity (1.45) by $\xi(s) := \kappa_j(s)\mathbf{1}_{(0,t)}(s)$, neglect the $\big|\kappa_j\big|^2$ term (being nonnegative) and expand the definition of W_j^1 and W_j^2 what gives:

$$\begin{aligned} &\beta_j\int_0^t \kappa_j'\kappa_j\,ds \;\le \\ &\le \int_0^t \sum_{k=1}^{K}\Big|\hat{u}^1_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^1_{h_k}(y^1-y^*)\,dx\Big) - \hat{u}^2_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^2_{h_k}(y^2-y^*)\,dx\Big)\Big|\big|\kappa_j\big|\,ds \\ &\le \int_0^t \frac{K}{2}\big|\kappa_j\big|^2 + \frac{1}{2}\sum_{k=1}^{K}\Big|\hat{u}^1_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^1_{h_k}(y^1-y^*)\,dx\Big) - \hat{u}^2_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^2_{h_k}(y^2-y^*)\,dx\Big)\Big|^2\,ds \end{aligned} \tag{1.50}$$

where the second inequality follows by the Young inequality. The right hand side term containing w_k is the term requiring the most calculations in the present proof. The subject term fulfills the below inequality:

$$\begin{aligned} &\Big|\hat{u}^1_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^1_{h_k}(y^1-y^*)\,dx\Big) - \hat{u}^2_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^2_{h_k}(y^2-y^*)\,dx\Big)\Big|^2 \;\le \\ &\le 3\Big|\hat{u}^1_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^1_{h_k}(y^1-y^*)\,dx\Big) - \hat{u}^1_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^1_{h_k}(y^2-y^*)\,dx\Big)\Big|^2 + \\ &+ 3\Big|\hat{u}^1_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^1_{h_k}(y^2-y^*)\,dx\Big) - \hat{u}^1_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^2_{h_k}(y^2-y^*)\,dx\Big)\Big|^2 + \\ &+ 3\Big|\hat{u}^1_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^2_{h_k}(y^2-y^*)\,dx\Big) - \hat{u}^2_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^2_{h_k}(y^2-y^*)\,dx\Big)\Big|^2 \end{aligned} \tag{1.51}$$

We estimate separately the three terms appearing in the right hand side of (1.51). In the first term, by the Lipschitz continuity of w_k we get:

$$\begin{aligned} &\Big|\hat{u}^1_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^1_{h_k}(y^1-y^*)\,dx\Big) - \hat{u}^1_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^1_{h_k}(y^2-y^*)\,dx\Big)\Big|^2 \;\le \\ &\le L_k^2\,\big|\hat{u}^1_{\alpha_{jk}}\big|^2\,\big\|\hat{u}^1_{h_k}\big\|_2^2\,\big\|y^1-y^2\big\|_2^2 \\ &\le L_k^2\,\big(R^U\big)^4\,\big\|y^1-y^2\big\|_2^2 \end{aligned} \tag{1.52}$$

The second term in the right hand side of (1.51) is estimated as follows:

$$\begin{aligned} &\Big|\hat{u}^1_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^1_{h_k}(y^2-y^*)\,dx\Big) - \hat{u}^1_{\alpha_{jk}}w_k\Big(\int_\Omega \hat{u}^2_{h_k}(y^2-y^*)\,dx\Big)\Big|^2 \;\le \\ &\le L_k^2\,\big|\hat{u}^1_{\alpha_{jk}}\big|^2\,\big\|y^2-y^*\big\|_2^2\,\big\|\hat{u}^1_{h_k}-\hat{u}^2_{h_k}\big\|_2^2 \\ &\le L_k^2\,\big(R^U\big)^2\,\big(C_1+\big\|y^*\big\|_2\big)^2\,\big\|\hat{u}^1_{h_k}-\hat{u}^2_{h_k}\big\|_2^2 \\ &\le 2\,L_k^2\,\big(R^U\big)^2\,\big(C_1^2+\big\|y^*\big\|_2^2\big)\,\big\|\hat{u}^1_{h_k}-\hat{u}^2_{h_k}\big\|_2^2 \end{aligned} \tag{1.53}$$

because $\|y^2(\,.\,,t)\|_2 \le C_1$ for $t \in [0,T]$. The latter is true since $\|y^2\|_{2,\infty} \le C_1$ (by Theorem 1.2.5) and $y^2 \in C([0,T];X)$ (see the comments after Definition 1.2.1). The third term in the right hand side of (1.51) obeys:

$$\begin{aligned}
&\Big|\hat{u}^1_{\alpha_{jk}} w_k\Big(\int_\Omega \hat{u}^2_{h_k}(y^2-y^*)\,dx\Big) - \hat{u}^2_{\alpha_{jk}} w_k\Big(\int_\Omega \hat{u}^2_{h_k}(y^2-y^*)\,dx\Big)\Big|^2 \le \\
&\le |\hat{u}^1_{\alpha_{jk}} - \hat{u}^2_{\alpha_{jk}}|^2 \Big|w_k\Big(\int_\Omega \hat{u}^2_{h_k}(y^2-y^*)\,dx\Big)\Big|^2 \\
&\le |\hat{u}^1_{\alpha_{jk}} - \hat{u}^2_{\alpha_{jk}}|^2 \Big(w_{k0} + L_k\|\hat{u}^2_{h_k}\|_2\|y^2-y^*\|_2\Big)^2 \\
&\le |\hat{u}^1_{\alpha_{jk}} - \hat{u}^2_{\alpha_{jk}}|^2 \Big(w_{k0} + L_k R^U\big(C_1 + \|y^*\|_2\big)\Big)^2 \\
&\le |\hat{u}^1_{\alpha_{jk}} - \hat{u}^2_{\alpha_{jk}}|^2 \Big(2w_{k0}^2 + 4L_k^2\,(R^U)^2\,(C_1^2 + \|y^*\|_2^2)\Big)
\end{aligned} \tag{1.54}$$

where we have again used the fact that $\|y^2(\,.\,,t)\|_2 \le C_1$ for $t \in [0,T]$. In total, by inequalities (1.51), (1.52), (1.53) and (1.54) we infer that:

$$\begin{aligned}
&\int_0^t \sum_{k=1}^K \Big|\hat{u}^1_{\alpha_{jk}} w_k\Big(\int_\Omega \hat{u}^1_{h_k}(y^1-y^*)\,dx\Big) - \hat{u}^2_{\alpha_{jk}} w_k\Big(\int_\Omega \hat{u}^2_{h_k}(y^2-y^*)\,dx\Big)\Big|^2 ds \le \\
&\le C_{2,j}\int_0^t \|y\|_2^2\,ds \;+\; C_{3,j}\sum_{k=1}^K \|\hat{u}_{h_k}\|_2^2 \;+\; C_{4,j}\sum_{k=1}^K |\hat{u}_{\alpha_{jk}}|^2
\end{aligned} \tag{1.55}$$

where, for $j = 1,\ldots,J$,

$$\begin{aligned}
C_{2,j} &= 3\sum_{k=1}^K L_k^2 (R^U)^4 \\
C_{3,j} &= 3\max_{k=1,\ldots,K}\Big\{2L_k^2(R^U)^2(TC_1^2 + \|y^*\|_{2,2}^2)\Big\} \\
C_{4,j} &= 3\max_{k=1,\ldots,K}\Big\{2Tw_{k0}^2 + 4L_k^2(R^U)^2(TC_1^2 + \|y^*\|_{2,2}^2)\Big\}
\end{aligned}$$

From the relation $\int_0^t \kappa_j'\kappa_j = \frac{1}{2}|\kappa_j(t)|^2 - \frac{1}{2}|\kappa_j(0)|^2$ (see the comments preceding (1.35) in the proof of Theorem 1.2.5) and from relations $\kappa_j(0) = \kappa_{j0}$, (1.50), (1.55) we infer that, for $j = 1,\ldots,J$:

$$\begin{aligned}
\frac{1}{2}|\kappa_j(t)|^2 - \frac{1}{2}|\kappa_{j0}|^2 \;\le\; &\frac{K}{2\beta_j}\int_0^t |\kappa_j|^2\,ds \;+\; \frac{1}{2\beta_j}C_{2,j}\int_0^t \|y\|_2^2\,ds \;+ \\
&+\; \frac{1}{2\beta_j}C_{3,j}\sum_{k=1}^K \|\hat{u}_{h_k}\|_2^2 \;+\; \frac{1}{2\beta_j}C_{4,j}\sum_{k=1}^K |\hat{u}_{\alpha_{jk}}|^2
\end{aligned} \tag{1.56}$$

We sum (1.49) and (1.56) for every $j = 1,\ldots,J$ and neglect the gradient term, which is nonnegative. As the result, we get:

$$
\begin{aligned}
\big\|y(\,.\,,t)\big\|_2^2 \;+\; \sum_{j=1}^{J}\big|\kappa_j(t)\big|^2 \;\le\; &\big\|y_0\big\|_2^2 \;+\; \sum_{j=1}^{J}\big|\kappa_{j0}\big|^2 \;+ \\
&+\; C_5\int_0^t\big\|y\big\|_2^2\,ds \;+\; C_6\sum_{j=1}^{J}\int_0^t\big|\kappa_j\big|^2\,ds \;+ \\
&+\; TC_1\sum_{j=1}^{J}\big\|\hat u_{g_j}\big\|_2^2 \;+\; C_7\sum_{k=1}^{K}\big\|\hat u_{h_k}\big\|_2^2 \;+\; C_8\sum_{j=1}^{J}\sum_{k=1}^{K}\big|\hat u_{\alpha_{jk}}\big|^2
\end{aligned}
\tag{1.57}
$$

where

$$
\begin{aligned}
C_5 &= 2L + 2J + \sum_{j=1}^{J}\beta_j^{-1}C_{2,j} \qquad & C_7 &= \sum_{j=1}^{J}\beta_j^{-1}C_{3,j} \\
C_6 &= \big(R^U\big)^2 + \max_j\{K\beta_j^{-1}\} \qquad & C_8 &= \max_{j=1,\ldots,J}\beta_j^{-1}C_{4,j}
\end{aligned}
$$

By the integral Grönwall inequality we infer from (1.57) that

$$
\begin{aligned}
\big\|y\big\|_{2,\infty}^2 \;+\; \sum_{j=1}^{J}\big\|\kappa_j\big\|_{L^\infty(0,T)}^2 \;\le\; &\Big(1 + T\max\{C_5,C_6\}e^{T\max\{C_5,C_6\}}\Big)\cdot \\
&\cdot\Big(\big\|y_0\big\|_2^2 \;+\; \sum_{j=1}^{J}\big|\kappa_{j0}\big|^2 \;+\; \max\{TC_1,C_7,C_8\}\big\|\hat u\big\|_U^2\Big)
\end{aligned}
\tag{1.58}
$$

where constants C_1, C_5, C_6, C_7, C_8 depend only on the quantities stated in the assertion of the theorem.

To close the proof, it suffices to show that

$$
\begin{aligned}
&\big\|\nabla y\big\|_{2,2} \;+\; \big\|y'\big\|_{H^1(\Omega)^*,2} \;+\; \big\|\kappa_j'\big\|_{L^2(0,T)} \;\le \\
&\le C_9\Big(\big\|y\big\|_{2,\infty} \;+\; \sum_{j=1}^{J}\big\|\kappa_j\big\|_{L^\infty(0,T)} \;+\; \big\|y_0\big\|_2^2 \;+\; \sum_{j=1}^{J}\big|\kappa_{j0}\big|^2 \;+\; \big\|\hat u\big\|_U^2\Big)
\end{aligned}
\tag{1.59}
$$

for certain positive C_9 depending only on the quantities stated in the assertion of the theorem. If (1.59) holds, then (1.58) can be applied to complete our reasoning. The necessary estimates for particular norms in the left hand side of (1.59) can be obtained with methods similar as in the proof of Theorem 1.2.5, but, for completeness, we derive the subject estimates.

We start with term $\big\|\nabla y\big\|_{2,2}$. By (1.49), neglecting $\big\|y(\,.\,,t)\big\|_2$ term (which is nonnegative), setting $t = T$ and taking into account $\sum_{j=1}^{J}\big\|\hat u_{g_j}\big\|_2^2 \le \big\|\hat u\big\|_U^2$, we derive

$$
\begin{aligned}
D\big\|\nabla y\big\|_{2,2}^2\,ds \;\le T(L+J)\big\|y\big\|_{2,\infty}^2 \;+\; &\frac{1}{2}T(R^U)^2\sum_{j=1}^{J}\big\|\kappa_j\big\|_{L^\infty(0,T)}^2 \;+ \\
&+\; \frac{1}{2}\big\|y_0\big\|_2^2 \;+\; \frac{1}{2}TC_1\big\|\hat u\big\|_U^2
\end{aligned}
\tag{1.60}
$$

To estimate term $\big\|y'\big\|_{H^1(\Omega)^*,2}$, we treat (1.44) as an equality in $L^2(0,T;H^1(\Omega)^*)$, which can be rewritten as:

$$
(y^1 - y^2)' + D\mathbf{A}(y^1 - y^2) - \big(\mathbf{F}y^1 - \mathbf{F}y^2\big) - \mathbf{K} = 0 \quad \text{in } L^2(0,T;H^1(\Omega)^*)
\tag{1.61}
$$

where we define **A** and **F** as in (1.15) while **K** is defined by

$$\int_0^T \langle \mathbf{K}, \phi \rangle \, dt \;=\; \int_0^T \Big(\sum_{j=1}^{J} \hat{u}^1_{g_j} \kappa^1_j - \hat{u}^2_{g_j} k^2_j , \phi \Big)_{L^2(\Omega)} dt \qquad \text{for } \phi \in L^2(0,T;H^1(\Omega))$$

The below follow straight from the definition of **K** and basic inequalities:

$$\begin{aligned} \big\|\mathbf{K}\big\|_{H^1(\Omega)^*,2} \;&\le\; \Big\| \sum_{j=1}^{J} \hat{u}^1_{g_j} \kappa^1_j - \hat{u}^2_{g_j} \kappa^2_j \Big\|_{2,2} \;\le\; \Big\| \sum_{j=1}^{J} \hat{u}^1_{g_j} (\kappa^1_j - \kappa^2_j) \Big\|_{2,2} \;+\; \Big\| \sum_{j=1}^{J} (\hat{u}^1_{g_j} - \hat{u}^2_{g_j}) \kappa^2_j \Big\|_{2,2} \\ &\le\; \sum_{j=1}^{J} \big\|\hat{u}^1_{g_j}\big\|_2 \big\|\kappa_j\big\|_{L^2(0,T)} + \sum_{j=1}^{J} \big\|\hat{u}_{g_j}\big\|_2 \big\|\kappa^2_j\big\|_{L^2(0,T)} \\ &\le R^U \sum_{j=1}^{J} \big\|\kappa_j\big\|_{L^2(0,T)} + T^{1/2} C_1 \sum_{j=1}^{J} \big\|\hat{u}_{g_j}\big\|_2 \end{aligned}$$

where we have used Theorem 1.2.5 to estimate $\big\|\kappa^2_j\big\|_{L^\infty(0,T)} \le C_1$, for $j = 1, \dots, J$ and for C_1 as above in the present proof. From the above estimate for **K**, from the estimates for **A** and **F** given in (1.16) and from (1.61), we derive the following:

$$\begin{aligned} \big\|y'\big\|_{H^1(\Omega)^*,2} \;&\le\; \big\|\nabla y\big\|_{2,2} + \big\|f(y^1) - f(y^2)\big\|_{2,2} + R^U \sum_{j=1}^{J} \big\|\kappa_j\big\|_{L^2(0,T)} + T^{1/2} C_1 \sum_{j=1}^{J} \big\|\hat{u}_{g_j}\big\|_2 \\ &\le\; \big\|\nabla y\big\|_{2,2} + T^{1/2} L \big\|y\big\|_{2,\infty} + T^{1/2} R^U \sum_{j=1}^{J} \big\|\kappa_j\big\|_{L^\infty(0,T)} + T^{1/2} C_1 J^{1/2} \big\|\hat{u}\big\|_2 \end{aligned} \tag{1.62}$$

where we have used the Lipschitz continuity of f with constant L and inequality $\big(\sum_j \big\|\hat{u}_{g_j}\big\|_2\big)^2 \le J \sum_j \big\|\hat{u}_{g_j}\big\|_2^2 \le J \big\|\hat{u}\big\|_U^2$.

To estimate $\big\|\kappa'_j\big\|_{L^2(0,T)}$, we proceed as follows. From (1.45) we conclude that

$$\beta_j {\kappa_j}' \;+\; \kappa_j \;=\; W^1_j(y^1, y^*) - W^2_j(y^2, y^*) \qquad \text{in } L^2(0,T)$$

for $j = 1, \dots, J$. By the above, expanding the definition of W^1_j and W^2_j, one obtain:

$$\begin{aligned} \beta_j \big\|\kappa'_j\big\|^2_{L^2(0,T)} \;&\le\; 2 \big\|\kappa_j\big\|^2_{L^2(0,T)} \;+ \\ +\; 2 \Big\| &\sum_{k=1}^{K} \hat{u}^1_{\alpha_{jk}} w_k \Big(\int_\Omega \hat{u}^1_{h_k} (y^1 - y^*) \, dx \Big) - \hat{u}^2_{\alpha_{jk}} w_k \Big(\int_\Omega \hat{u}^2_{h_k} (y^2 - y^*) \, dx \Big) \Big\|^2_{L^2(0,T)} \end{aligned} \tag{1.63}$$

The second term in the right hand side of (1.63) can be estimated with the use of (1.55), what yields:

$$\begin{aligned} \beta_j \big\|\kappa'_j\big\|^2_{L^2(0,T)} \;&\le\; 2T \big\|\kappa_j\big\|^2_{L^\infty(0,T)} \;+ \\ &+\; 2K \Big(T C_{2,j} \big\|y\big\|^2_{2,\infty} \;+\; C_{3,j} \sum_{k=1}^{K} \big\|\hat{u}_{h_k}\big\|_2^2 \;+\; C_{4,j} \sum_{k=1}^{K} \big|\hat{u}_{\alpha_{jk}}\big|^2 \Big) \\ &\le\; 2T \big\|\kappa_j\big\|^2_{L^\infty(0,T)} \;+\; 2KT C_{2,j} \big\|y\big\|^2_{2,\infty} \;+\; 2K \max\{C_{3,j}, C_{4,j}\} \big\|\hat{u}\big\|_U^2 \end{aligned} \tag{1.64}$$

Above, constant K appears as a result of moving the square to the terms under the sum sign $\sum_{k=1}^{K}$, as in general inequality $\left|\sum_k a_k\right|^2 \le K \sum_k |a_k|^2$.

Altogether, by (1.60), (1.62) and (1.64), the estimate (1.59) holds with constant C_9 depending only on the quantities appearing in (1.60), (1.62) and (1.64), i.e. on C_1, $C_{2,j}$, $C_{3,j}$, $C_{4,j}$, T, D, β_j, K, J, L, R^U. This closes the proof. ■

As the next result shows, in consequence of the existence result provided by Theorem 1.2.3 and the estimates given in Theorem 1.2.5, it is possible to prove the existence of solutions for unbounded switching functions w_k in the system (0.1) - (0.3). The latter is the case not covered by Theorem 1.2.3. However, note that the below result requires a stronger assumption concerning the reference trajectory y^* in the system (0.1) - (0.3), in comparison to Theorem 1.2.3.

Theorem 1.2.7 *Assume that assumptions (B-1) - (B-5) and (C-2) hold and* $(g_j, h_k, \alpha_{jk})_{j=1,\dots,J}^{k=1,\dots,K} \in U$*. Then the system (0.1) - (0.3) has a weak solution.*

Proof. Theorem 1.2.3 assumes that w_k functions are bounded, i.e. $\left\|w_k\right\|_{L^\infty(\mathbb{R})} < \infty$. But Theorem 1.2.5 gives a bound for solutions of (0.1) - (0.3) that is independent of $\left\|w_k\right\|_{L^\infty(\mathbb{R})}$. Thus the standard truncation technique can be utilized to dismiss the assumption that w_k are bounded.

More precisely, for a given w_k as in the assumption (B-4), consider its truncation w_k^n given by

$$w_k^n(s) \;:=\; \begin{cases} w_k(-n) & \text{for } s < -n \\ w_k(s) & \text{for } s \in [-n, n] \\ w_k(n) & \text{for } s > n \end{cases}$$

Let $(y^n, \kappa_1^n, \dots, \kappa_J^n) \in X^2$ denote the weak solution of the system (0.1) - (0.3) with w_k^n in place of w_k. By Theorem 1.2.5, $\left\|y^n\right\|_{2,\infty} = C_1 < \infty$ where C_1 does not depend on $\left\|w_k\right\|_{L^\infty(\mathbb{R})}$.

Let $C_2 := \left\|y^*\right\|_{2,\infty}$ and choose $\widetilde{n} > \left\|h_k\right\|_2 (C_1 + C_2)$. The switching function w_k^n in (0.3) can be replaced by $w_k^{\widetilde{n}}$ with no side effect to the weak solution $(y^{\widetilde{n}}, \kappa_1^{\widetilde{n}}, \dots, \kappa_J^{\widetilde{n}})$. Indeed, for the above choice of $\widetilde{n}$ we have

$$\begin{aligned} \mathbf{v}_k(t) \;:=\; & \int_\Omega h_k(y^{\widetilde{n}}(\,.\,,t) - y^*(\,.\,,t))\,dx \quad \le \\ \le\; & \left\|h_k\right\|_2 (C_1 + C_2) \qquad\quad < \widetilde{n} \end{aligned} \qquad \text{for a.e. } t \in [0,T] \tag{1.65}$$

where we have used the Hölder inequality. Therefore

$$w_k(\mathbf{v}_k(t)) = w_k^{\widetilde{n}}(\mathbf{v}_k(t)) \quad \text{for a.e. } t \in [0,T] \tag{1.66}$$

Thus, from the above and from the definition of the weak solution we conclude what follows — for $\widetilde{n}$ as indicated, an arbitrary weak solution of (0.1) - (0.3) with switching functions $w_k^{\widetilde{n}}$ is also a weak solution of (0.1) - (0.3) with switching functions w_k. Now, Theorem 1.2.3 can be applied to obtain existence of the weak solution for (0.1) - (0.3) with switching functions $w_k^{\widetilde{n}}$. Hence the assertion follows. ■

Remark. In the above proof the assumption that $y^* \in L^\infty(0,T;L^2(\Omega))$ was essential to obtain the estimate (1.65) for a.e. $t \in [0,T]$. The assumption $y^* \in L^2(0,T;L^2(\Omega))$, imposed in Theorem 1.2.3, would not allow to obtain this estimate a.e. on $[0,T]$ and hence the identity (1.66) could fail on some subset of $[0,T]$ of positive measure. This would make impossible to identify the weak solutions of the system (0.1) - (0.3) with an unbounded switching function w_k and the weak solutions of (0.1) - (0.3) with the switching function $w_k^{\widetilde{n}}$ defined as in the above proof. Hence, the final argument of the proof would be not valid.

Thus, comparing Theorem 1.2.3 with Theorem 1.2.7, we have traded the unboundedness of $\left\|y^*(\,.\,,t)\right\|_2$ for unboundedness of w_k. ▲

The below corollaries are straightforward due to existence Theorems 1.2.3, 1.2.7 and stability Theorem 1.2.6.

Corollary 1.2.8 *Let assumptions (B-1) - (B-5) and (C-1) be satisfied and $(g_j, h_k, \alpha_{jk})_{j=1,\dots,J}^{k=1,\dots,K} \in U$. Assume moreover that functions w_k entering the system (0.1) - (0.3) are bounded. Then the system (0.1) - (0.3) has a unique weak solution.*

Corollary 1.2.9 *Let assumptions (B-1) - (B-5) and (C-2) be fulfilled and $(g_j, h_k, \alpha_{jk})_{j=1,\dots,J}^{k=1,\dots,K} \in U$. Then the system (0.1) - (0.3) has a unique weak solution.*

This closes the part concerning the uniqueness and existence of the weak solutions of (0.1) - (0.3). However, Theorem 1.2.5 and Theorem 1.2.6 are necessary not only for the uniqueness and existence results in Corollaries 1.2.8 and 1.2.9. The stability result in Theorem 1.2.6 will be crucial in Chapter 3, concerning theoretical aspects of the optimal targeting problem, announced in §2 of *Introduction*.

But there are also other properties concerning the behavior of the system (0.1) - (0.3) under the perturbations of the control which we would like to present. Assume that there is a sequence of controls $\hat{u}^n \in U$ given and one have only the knowledge on the weak convergence of these controls. This does not allow to utilize the former theorems of the present section to infer about anything more than boundedness of $(y^n, \kappa_1^n, \dots, \kappa_J^n)$ in X^2, where $(y^n, \kappa_1^n, \dots, \kappa_J^n)$ denotes the solution of (0.1) - (0.3) corresponding to $\hat{u}^n$. Here, the following result may be useful:

Theorem 1.2.10 *Let assumptions (B-1) - (B-5) and (C-1) be fulfilled. Let the sequence $\hat{u}^n$ converge weakly to $\hat{u}$ in U. Denote by $(y^n, \kappa_1^n, \dots, \kappa_J^n)$ the weak solution of (0.1) - (0.3) corresponding to $\hat{u}^n$ and by $(\widetilde{y}, \widetilde{\kappa}_1, \dots, \widetilde{\kappa}_J)$ the weak solution of (0.1) - (0.3) corresponding to $\hat{u}$. Then there exists a sequence of natural indexes $n_1 < n_2 < \dots$ such that subsequence $(y^{n_k}, \kappa_1^{n_k}, \dots, \kappa_J^{n_k})$ converges weakly-$*$ to $(\widetilde{y}, \widetilde{\kappa}_1, \dots, \widetilde{\kappa}_J)$ in X^2 when $k \to \infty$.*

Proof. Let $\hat{u}^n \rightharpoonup \hat{u}$ in U, as in the assumptions. A weakly convergent sequence is bounded, thus by Theorem 1.2.5 $(y^n, \kappa_1^n, \dots, \kappa_J^n)$ is bounded in X^2. This allows us to extract a weakly-$*$ convergent subsequence (for simplicity, we relabel it and keep the original indexes): $(y^n, \kappa_1^n, \dots, \kappa_J^n) \overset{*}{\rightharpoonup} (\bar{y}, \bar{\kappa}_1, \dots, \bar{\kappa}_J)$ in X^2 for certain $(\bar{y}, \bar{\kappa}_1, \dots, \bar{\kappa}_J) \in X^2$. In particular:

$$\begin{aligned}
&y^n \overset{*}{\rightharpoonup} \bar{y} && \text{in } L^\infty(0,T;L^2(\Omega)) \\
&y^{n\prime} \rightharpoonup \bar{y}' && \text{in } L^2(0,T;H^1(\Omega)^*) \\
&\nabla y^n \rightharpoonup \nabla \bar{y} && \text{in } \left(L^2(Q_T)\right)^{\mathbf{d}} \\
&\kappa_j^n \overset{*}{\rightharpoonup} \bar{\kappa}_j && \text{in } L^\infty(0,T) \\
&\kappa_j^{n\prime} \rightharpoonup \bar{\kappa}_j' && \text{in } L^2(0,T)
\end{aligned} \tag{1.67}$$

It suffices to show that $(\bar{y}, \bar{\kappa}_1, \dots, \bar{\kappa}_J) = (\widetilde{y}, \widetilde{\kappa}_1, \dots, \widetilde{\kappa}_J)$. For this reason we need to prove that we can pass with n to infinity in all terms appearing in the weak formulation given in Definition

1.2.1 The passage in linear terms follows straight due to (1.67). We are left to deal with the terms

$$\int_0^T (\kappa_j^n \hat{u}_{g_j}^n, \phi)_{L^2(\Omega)}\, dt, \qquad \int_0^T (f(y^n), \phi)_{L^2(\Omega)}\, dt, \qquad \int_0^T W_j(y^n, y^*)\, \xi\, dt$$

for $\phi \in L^2(0,T;H^1(\Omega))$, $\xi \in L^2(0,T)$.

Let us begin with the term corresponding to $\kappa_j^n \hat{u}_{g_j}^n$. By the assumption and by (1.67), $\hat{u}_{g_j} \rightharpoonup \hat{u}$ in $L^2(\Omega)$ and $\kappa_j^n \rightharpoonup \bar{\kappa_j}$ in $L^2(0,T)$. But this means that for an arbitrary $\phi^\Omega \in C(\bar{\Omega})$ and $\phi^T \in C([0,T])$ we have

$$\begin{aligned} \int_0^T (\kappa_j^n \hat{u}_{g_j}^n, \phi^\Omega \phi^T)_{L^2(\Omega)}\, dt \quad &= \\ = \int_0^T \kappa_j^n \phi^T\, dt \int_\Omega \hat{u}_{g_j}^n \phi^\Omega\, dx \quad &\longrightarrow \int_0^T \bar{\kappa}_j \phi^T\, dt \int_\Omega \hat{u}_{g_j} \phi^\Omega\, dx \; = \\ &= \int_0^T (\bar{\kappa}_j \hat{u}_{g_j}, \phi^\Omega \phi^T)_{L^2(\Omega)}\, dt \end{aligned}$$

To conclude that the weak convergence of $\kappa_j^n \hat{u}_{g_j}^n$ to $\bar{\kappa}_j \hat{u}_{g_j}$ in $L^2(Q_T)$ holds it suffices to justify that $\kappa_j^n \hat{u}_{g_j}^n$ is bounded in $L^2(Q_T)$ and the set of functions ϕ of form $\phi(x,t) = \phi^\Omega(x)\phi^T(t)$, where ϕ^Ω and ϕ^T are as above, is linearly dense in $L^2(Q_T)$. The former is straightforward by the weak convergence properties of κ_j^n and $\hat{u}_{g_j}^n$. Concerning the latter, by the Stone-Weierstrass theorem (see [49, Chap. 0.2, p.9]), the set of all possible ϕ is dense in $C(\bar{Q}_T)$ and the latter set is linearly dense in $L^2(Q_T)$. Altogether, the following can be stated:

$$\kappa_j^n \hat{u}_{g_j}^n \rightharpoonup \bar{\kappa}_j \hat{u}_{g_j} \quad \text{in } L^2(Q_T) \tag{1.68}$$

Guaranteeing the convergence of the remaining two terms will involve the knowledge on the strong convergence of y^n in $L^2(Q_T)$. But this can be concluded by the Aubin-Lions lemma (see [43, Chap III.1. Prop. 1.3] for the probably most common formulation of the lemma or [44, Sec. 8 Cor. 4] for a more general statement). More precisely, spaces $H^1(\Omega)$, $L^2(\Omega)$ and $H^1(\Omega)^*$ form an evolution triple with continuous embeddings $H^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow H^1(\Omega)^*$ (see [51, Chap. 23.4]), where the first embedding is in addition compact, by the Rellich-Kondrachov theorem (see [1, par. 4.6.]). Moreover, the bounds for y^n and $y^{n\prime}$ in (1.67) hold. Thus the conditions of the Aubin-Lions lemma are fulfilled and it can be applied to conclude that there exists a subsequence such that

$$y^n \to \bar{y} \quad \text{in } L^2(Q_T) \tag{1.69}$$

The limit in (1.69) is exactly $\bar{y}$ since otherwise it would be a contradiction to (1.67). This is the point where the assumption (B-1) was necessary since the above referred Rellich-Kondrachov theorem version requires that Ω is bounded and satisfies the cone condition.

By the Lipschitz continuity of f and (1.69) the convergence

$$f(y^n) \to f(\bar{y}) \quad \text{in } L^2(Q_T) \tag{1.70}$$

is a straightforward conclusion.

We are left to investigate the convergence of the term corresponding to $W_j(y^n, y^*)$. Note that by the definition (see (0.3)), W_j has an implicit dependence on $\hat{u}_{h_k}^n$ and $\hat{u}_{\alpha_{jk}}^n$. Thus in the present context we should interpret W_j as $W_j(\hat{u}_{h_k}^n, \hat{u}_{\alpha_{jk}}^n, y^n, y^*)$. By (0.3) and the Lipschitz continuity of w_k we can write, using the triangle inequality:

$$
\begin{aligned}
&\int_0^T \big|W_j\big(\hat{u}^n_{h_k}, \hat{u}^n_{\alpha_{jk}}, y^n(t), y^*(t)\big) - W_j\big(\hat{u}_{h_k}, \hat{u}_{\alpha_{jk}}, \bar{y}(t), y^*(t)\big)\big|^2\, dt \;\le \\
&\le 2\sum_{k=1}^K L_k \Big\{ \;\big|\hat{u}^n_{\alpha_{jk}}\big|^2 \big\|\hat{u}^n_{h_k}\big\|_2^2 \int_0^T \big\|y^n - \bar{y}\big\|_2^2\, dt \;+ \\
&\qquad + \;\big|\hat{u}^n_{\alpha_{jk}}\big|^2 \int_0^T \Big|\int_\Omega (\hat{u}^n_{h_k} - \hat{u}_{h_k})(\bar{y} - y^*)\, dx\Big|^2\, dt \;+ \\
&\qquad + \;\big|\hat{u}^n_{\alpha_{jk}} - \hat{u}_{\alpha_{jk}}\big|^2 \big\|\hat{u}_{h_k}\big\|_2^2 \int_0^T \big\|\bar{y} - y^*\big\|_2^2\, dt \quad \Big\}
\end{aligned}
\tag{1.71}
$$

Let us consider each of the three terms appearing in the right hand side of the above.

The first term in the right hand side of (1.71) converges to zero since the sequence of controls $\hat{u}^n$ is bounded and (1.69) holds.

The third term in the right hand side of (1.71) is convergent to zero since by $\hat{u}^n \rightharpoonup \hat{u}$ in U we have $\hat{u}^n_{\alpha_{jk}} \to \hat{u}_{\alpha_{jk}}$.

To treat the second term, consider a function

$$F^n(t) \;=\; \int_\Omega (\hat{u}^n_{h_k} - \hat{u}_{h_k})(\bar{y}(t) - y^*(t))\, dx$$

As the sequence of numbers $\big|\hat{u}^n_{\alpha_{j,k}}\big|^2$ in the considered term is bounded, it is enough to show the convergence of F^n to zero in $L^2(0,T)$. We have $\bar{y}(t), y^*(t) \in L^2(\Omega)$ a.e. on $[0,T]$. Thus, by the weak convergence $\hat{u}^n_{h_k} \rightharpoonup \hat{u}_{h_k}$ in $L^2(\Omega)$ for every $k = 1,\dots,K$ we infer that $F^n(t)$ converges to zero a.e. on $[0,T]$, as $n \to \infty$. Moreover, a.e. on $[0,T]$

$$\big|F^n(t)\big| \;\le\; \big\|\hat{u}^n_{h_k} - \hat{u}_{h_k}\big\|_2 \big\|\bar{y}(t) - y^*(t)\big\|_2 \;\le\; C_U \big\|\bar{y}(t) - y^*(t)\big\|_2$$

where $C_U = \sup_n \big\|\hat{u}^n\big\|_U$ is finite and the term $\big\|\bar{y}(t) - y^*(t)\big\|_2$ is square integrable due to $\bar{y}, y^* \in L^2(Q_T)$. These observations concerning $F^n(t)$ allow us to apply the Lebesgue dominated convergence theorem (see [41, Chap. 1] or [21, App. E.3, Th. 5]) and get the convergence

$$F^n \;\to\; 0 \quad \text{in } L^2(0,T)$$

Altogether, we conclude that the right hand side of (1.71) converges to zero thus:

$$W_j(\hat{u}^n_{h_k}, \hat{u}^n_{\alpha_{jk}}, y^n, y^*) \;\to\; W_j(\hat{u}_{h_k}, \hat{u}_{\alpha_{jk}}, \bar{y}, y^*) \quad \text{in } L^2(0,T) \tag{1.72}$$

To sum up, the convergence results (1.67), (1.68), (1.70), (1.72) allow us to infer that $(\bar{y}, \bar{\kappa}_1, \dots, \bar{\kappa}_J)$ is the weak solution of the system (0.1) - (0.3) in sense of the Definition 1.2.1, corresponding to $\hat{u}$, i.e. $(\bar{y}, \bar{\kappa}_1, \dots, \bar{\kappa}_J) = (\widetilde{y}, \widetilde{\kappa}_1, \dots, \widetilde{\kappa}_J)$ what concludes the proof. ■

Remark. Note that, in the proof of Theorem 1.2.10, we did not require *a priori* knowledge on validity of theorems concerning existence of weak solutions. We simply assumed that $(y^n, \kappa^n_1, \dots, \kappa^n_J)$ and $(\widetilde{y}, \widetilde{\kappa}_1, \dots, \widetilde{\kappa_J})$ are weak solutions of the system (0.1) - (0.3). Thus, the assumptions of Theorem 1.2.10 did not need to cover the assumptions of the existence results provided by Theorem 1.2.3 or Theorem 1.2.7. Analogous remark holds for Theorem 1.2.5 and Theorem 1.2.6, which also did not base on the existence results and hence did not require to cover the assumptions of the latter results. ▲

Remark. In the content of the present section, the condition $\beta_j > 0$, being a part of the assumption (B-2), was utilized directly only in the proofs of Theorem 1.2.5 and Theorem 1.2.6, e.g. to preserve the direction of inequalities when dividing by β_j. In the rest of the statements of Section 1.2.2, the condition β_j was necessary only because they refer to Theorem 1.2.5 and Theorem 1.2.6 (or to Theorem 1.2.3, but the latter actually could be proven also for $\beta_j < 0$, see the remark on page 23).

However, we expect that, after suitable modifications, versions of Theorem 1.2.5 and Theorem 1.2.6 allowing $\beta_j < 0$ also could be proven. In consequence, the rest of the results of Section 1.2.2 also would be valid for $\beta_j < 0$.

We also expect that the results presented in Section 1.2.3 and Section 1.2.4, which also assume $\beta_j > 0$, would be valid for $\beta_j < 0$ as well.

The above, if true, have consequences also for analytical results in Chapter 3 of the present work, which rely on the theorems given here, in Section 1.2.2, as well as in Section 1.2.3 and Section 1.2.4. Perhaps, all of the analytical results of Chapter 3, as well as the rest of the present work, would be valid if we allowed $\beta_j < 0$. However, a careful verifications of the proofs would be necessary to guarantee the above hypotheses. ▲

1.2.3 Generalizations for locally Lipschitz reactive term

In Section 1.2.3, we focus on the system (0.1) - (0.3) with assumptions concerning nonlinear term f different than in Section 1.2.2. More precisely, we assume below that f is locally Lipschitz continuous only. However, to compensate this loose of strength of assumptions, we assume that f obeys certain growth condition, which will be precisely formulated below. In addition, we impose assumptions for the initial condition component y_0 that are stronger in comparison to the assumptions imposed in Section 1.2.2, namely $y_0 \in L^\infty(\Omega)$. Also, we put more restrictive assumptions for the integrability of the functions describing the control devices actions, denoted in the system (0.1) - (0.3) by g_j, $j = 1, \dots, J$.

Te reasons of considering the system (0.1) - (0.3) with the above mentioned modified assumptions are twofold. First, numerical experiments described in further chapters of the present work involved data with locally Lipschitz f and bounded initial condition. Hence, our intention is to give analytical results that cover the data utilized in the mentioned experiment. Second, the results presented in Section 1.2.3 will be used also in the chapter concerning mathematical analysis of the optimal targeting problem.

The results of the present subsection rely strongly on a theorem for boundedness of the weak solutions of (1.4). The subject theorem requires the nonlinear term to satisfy certain growth condition, the initial condition to be bounded and the free term to be integrable with sufficiently high power. In the result, the assumptions concerning the growth of f, the boundedness of y_0 and for the integrability of functions g_j in (0.1) - (0.3) are inherited by most of the results of the present subsection.

In Section 1.2.3, we prove estimates analogous to those given in Theorem 1.2.5, but for the system (0.1) - (0.3) with the modified assumptions, mentioned above. Next, using the boundedness of the weak solutions of (1.4) and the derived estimates, we prove that the weak solutions of the system (0.1) - (0.3) with the modified assumptions also are bounded. Having the latter boundedness result, we prove the existence and uniqueness result for the system (0.1) - (0.3) with the modified assumptions. For this end, we base on a truncation argument, reducing the problem with the modified assumptions to the problem with the assumptions originally considered in the results of Section 1.2.2.

Let us proceed to the mathematical details. The above mentioned growth condition for $f\colon \mathbb{R} \to \mathbb{R}$ is as follows.

$$sf(s) \leq 0 \qquad \text{if } |s| > C_f \tag{1.73}$$

for certain $C_f > 0$.

In the sequel, we will need also the following conditions. Recall that $\mathbf{d}$ denotes the space dimension of domain Ω, entering the system (0.1) - (0.3). The following conditions constituting a relation between two numbers $s_1, s_2 \in [1, \infty]$ will be utilized in Section 1.2.3:

$$\frac{1}{2s_2'} + \frac{\mathbf{d}}{4s_1'} = \frac{\mathbf{d}}{4} \tag{1.74}$$

$$\begin{cases} s_1 \in [1, \infty], \quad s_2 \in [1, 2] & \text{for } \mathbf{d} = 1 \\ s_1 \in (1, \infty], \quad s_2 \in [1, \infty) & \text{for } \mathbf{d} = 2 \\ s_1 \in [\frac{\mathbf{d}}{2}, \infty], \quad s_2 \in [1, \infty] & \text{for } \mathbf{d} \geq 3 \end{cases} \tag{1.75}$$

where s_1' and s_2' denote the Hölder conjugate of s_1 and s_2, respectively. Notation „$\frac{1}{\infty} = 0$” is utilized in the above conditions.

The below theorem concerning the boundedness of the weak solutions of parabolic differential equations will be crucial:

Theorem 1.2.11 *Let Ω, T, D, J, f be as in assumptions (B-1), (B-2), (B-3). Let $y_0 \in L^\infty(\Omega)$. Let also $g_j \in L^{s_1}(\Omega)$, $k_j \in L^{s_2}(0, T)$ for $j = 1, \ldots, J$, where numbers s_1 and s_2 obey conditions (1.74) and (1.75). Let C_∞, C_F be nonnegative numbers such that*

$$\|y_0\|_\infty \;\leq\; C_\infty, \qquad \Big\|\sum_{j=1}^{J} g_j k_j\Big\|_{s_1,s_2} \;\leq\; C_F$$

Let f fulfill the condition (1.73) with a constant C_f. Assume that y is a weak solution of the system (1.4), corresponding to the above data. Then y belongs to $L^\infty(Q_T)$ and

$$\|y\|_{L^\infty(Q_T)} \;\leq\; C$$

where $C = C(\mathbf{d}, \Omega, T, D, s_1, s_2, C_\infty, C_F, C_f)$.

Theorem 1.2.11 can be proved with the same methods as Theorem 7.1 in Chapter III of [37]. The case treated there is in some details different than ours. In the referred theorem it is *a priori* assumed that the values of the solution on $\partial\Omega \times (0, T)$ are bounded what is an information that we do not assume to have (instead, we assume to control the values of the derivative of the solution on $\partial\Omega \times (0, T)$, in the direction normal to $\partial\Omega$). Besides, the referred theorem treats the case of a linear parabolic equation while the state equation in (1.4) is semilinear. In spite of that, we have verified that the methods utilized in the proof of Theorem 7.1 in Chapter 3 of [37] can be applied in our situation. The above listed differences do not change the main steps of the proof.

Now, we proceed to the estimates for the weak solutions of the system (0.1) - (0.3). The following result is a variant of Theorem 1.2.5, assuming a modified assumption for the reactive term f in the system (0.1) - (0.3):

Theorem 1.2.12 *In the system (0.1) - (0.3), let the part a) of the assumption (B-1) and assumptions (B-2), (B-4), (C-1) hold. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a locally Lipschitz continuous function, satisfying the condition (1.73) for a given constant $C_f > 0$. Denote $f_0 := f(0)$ and let L_{C_f} be* *the Lipschitz constant of f on interval $[-C_f, C_f]$. Let also $\hat{u} \in U$ and $(y_0, \kappa_{10}, \ldots, \kappa_{J0}) \in X^0$. Assume that R^U and R^0 are positive numbers such that*

$$\|\hat{u}\|_U \;\le\; R^U, \qquad \|(y_0, \kappa_{10}, \ldots, \kappa_{J0})\|_{X^0} \;\le\; R^0$$

Assume that $(y, \kappa_1, \ldots, \kappa_J) \in X^2$ is a weak solution of the system (0.1) - (0.3) with the above data and with $g_j := \hat{u}_{g_j}$, $h_k := \hat{u}_{h_k}$, $\alpha_{j,k} := \hat{u}_{\alpha_{j,k}}$. Then

$$\|y\|_{2,\infty} \;+\; \|\nabla y\|_{2,2} \;+\; \sum_{j=1}^{J} \|\kappa_j\|_{L^\infty(0,T)} \;+\; \sum_{j=1}^{J} \|\kappa_j'\|_{L^2(0,T)} \;\le\; C_1 \tag{1.76}$$

where

$$C_1 = C_1(T, |\Omega|, K, J, D, \beta_1, \ldots, \beta_J, L_{C_f}, f_0, L_1, \ldots, L_K, w_{10}, \ldots, w_{K0}, R^U, R^0, \|y^*\|_{2,2})$$

where the quantities on which constant C_1 depends are as in the above assumptions.

If, in addition, $\|y\|_{L^\infty(Q_T)} \le C_0$, then

$$\|y'\|_{H^1(\Omega)^*,2} \;\le\; C_2 \tag{1.77}$$

where

$$C_2 = C_2(f(C_0), C_1, T, |\Omega|, D, R^U)$$

Proof. We start with the proof of the estimate (1.76). The proof is analogous to a part of the proof of Theorem 1.2.5. The differences are minor. Therefore, we do not present the full proof but only discuss the subject differences.

The only difference occurs in the estimate (1.30). Estimating term $(f(y), y)_{L^2(\Omega)}$ needs to be done slightly different in the present situation than in the proof of Theorem 1.2.5. More precisely, denote

$$\mathbb{A}_{C_f} := \Big\{(x,t) \in \Omega \times (0,T) : \; |y(x,t)| \le C_f\Big\}$$

Now we use property (1.73), Lipschitz continuity of f on $[-C_f, C_f]$, the Hölder inequality and the Young inequality to find that:

$$\begin{aligned}
(f(y), y)_{L^2(\Omega)} \;&=\; \int_\Omega f(y) y \, dx \;\le\; \int_{\mathbb{A}_{C_f}} f(y) y \, dx \\
&\le L_{C_f} \int_{\mathbb{A}_{C_f}} |y|^2 \, dx \;+\; f_0 \int_{\mathbb{A}_{C_f}} |y| \, dx \\
&\le L_{C_f} \|y\|_2^2 \;+\; f_0 \|y\|_2 \|\mathbf{1}_\Omega\|_2 \;\le\; L_{C_f} \|y\|_2^2 \;+\; \frac{f_0}{2} \|y\|_2^2 \;+\; \frac{1}{2} \|\mathbf{1}_\Omega\|_2^2
\end{aligned}$$

In the proof of Theorem 1.2.5, we insert the above estimate instead of the estimate (1.30). The further part of the proof, until the estimate (1.38), remains valid, with the side effect that constant L, whenever appears in the subject part of the proof, should be replaced by L_{C_f}. In particular, estimates (1.37) and (1.38) hold (for L replaced by L_{C_f}), what gives the demanded estimates for $\|y\|_{2,\infty}$, $\|\nabla y\|_{2,2}$ and $\|\kappa_j\|_{L^\infty(0,T)}$, for $j = 1, \ldots, J$.

Similarly, one can verify that estimates (1.42) and (1.43) remain valid, assuming that constant L is replaced by L_{C_f}. Thus, by the estimate (1.43) (for L replaced by L_{C_f}), we have the estimate for $\|\kappa_j'\|_{L^2(0,T)}$, for $j = 1, \ldots, J$. This gives the estimate (1.76).

To obtain the estimate (1.77), we cannot proceed exactly as in the proof of Theorem 1.2.5. The reason for this is that in the estimate (1.41), crucial for estimating $\|y'\|_{H^1(\Omega)^*,2}$, term $\|f(y)\|_{2,2}$ appears. Under the present assumptions for f, the subject term can be ill defined if y belongs to $L^\infty(0,T;L^2(\Omega))$ only. This makes the estimates for $\|y'\|_{H^1(\Omega)^*,2}$ derived in the proof of Theorem 1.2.5 invalid. To overcome the subject obstacle, we use the assumption $\|y\|_{L^\infty(Q_T)} \leq C_0$.

More precisely, (1.39) and (1.40) in the proof of Theorem 1.2.5 still hold, with the same arguments as given there. Thus, from (1.39) and (1.40) we infer that:

$$\|y'\|_{H^1(\Omega)^*,2} \leq D\|\nabla y\|_{2,2} + \|f(y)\|_{2,2} + \sum_{j=1}^{J}\|\hat{u}_{g_j}\|_2\|\kappa_j\|_{L^2(0,T)}$$

By the assumption $\|y\|_{L^\infty(Q_T)} \leq C_0$, by the Hölder inequality and by the definition of constant R^U, we can estimate the right hand side of the above and obtain:

$$\|y'\|_{H^1(\Omega)^*,2} \leq D\|\nabla y\|_{2,2} + f(C_0)\big(T|\Omega|\big)^{1/2} + TR^U\sum_{j=1}^{J}\|\kappa_j\|_{L^\infty(0,T)}$$

Now, (1.76) can be used to estimate norms $\|\nabla y\|_{2,2}$ and $\|\kappa_j\|_{L^\infty(0,T)}$ for $j = 1, \ldots, J$ appearing above by C_1. In total, the right hand side of the above can be estimated in terms of $f(C_0)$, C_1, D, T, $|\Omega|$ and R^U. Hence (1.77) follows. ■

The below theorem requires both Theorem 1.2.11 and Theorem 1.2.12 for the proof. It will be a crucial technical result in our method of proving the uniqueness and existence results given in the further part of Section 1.2.3.

Theorem 1.2.13 *In the system (0.1) - (0.3), let the part a) of the assumption (B-1) and assumptions (B-2), (B-4), (C-1) hold. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a locally Lipschitz continuous function, satisfying the condition (1.73) for a given constant $C_f > 0$. Denote $f_0 := f(0)$ and let L_{C_f} be the Lipschitz constant of f on interval $[-C_f, C_f]$. Let also $\hat{u} \in U$ and $(y_0, \kappa_{10}, \ldots, \kappa_{J0}) \in X^0$. Assume that R^U and R^0 are positive numbers such that*

$$\|\hat{u}\|_U \leq R^U, \qquad \|(y_0, \kappa_{10}, \ldots, \kappa_{J0})\|_{X^0} \leq R^0$$

In addition, assume that $y_0 \in L^\infty(\Omega)$ and that $\hat{u}_{g_j} \in L^{s_1}(\Omega)$ for certain $s_1 \geq \max\{2, \frac{\mathbf{d}}{2}\}$, for $j = 1, \ldots, J$. Let C_∞ and R^g be nonnegative number such that

$$\|y_0\|_{L^\infty(\Omega)} \leq C_\infty, \qquad \max_{j=1,\ldots,J}\|\hat{u}_{g_j}\|_{s_1} \leq C_g$$

Assume that $(y, \kappa_1, \ldots, \kappa_J) \in X^2$ is a weak solution of the system (0.1) - (0.3) with the above data and with $g_j := \hat{u}_{g_j}$, $h_k := \hat{u}_{h_k}$, $\alpha_{j,k} := \hat{u}_{\alpha_{j,k}}$. Then

$$\|y\|_{L^\infty(Q_T)} \leq C \tag{1.78}$$

where

$$C = C(\mathbf{d}, T, \Omega, K, J, D, \beta_1, \ldots, \beta_J, L_{C_f}, f_0, L_1, \ldots, L_K, w_{10}, \ldots, w_{K0},$$
$$R^U, R^0, \|y^*\|_{2,2}, C_\infty, C_f, s_1, C_g)$$

where the quantities on which C depends are as in the assumptions of the theorem.

Proof. Let s_1 be as in the assumption of the theorem and let $s_2 \in [1, \infty]$. We will need to have estimates for norm $\big\|\sum_{j=1}^{J} \hat{u}_{g_j} \kappa_j\big\|_{s_1,s_2}$. We derive them as follows. By independence of variables being arguments for $\hat{u}_{g_j}$ and κ_j and by the definition of C_g:

$$\Big\|\sum_{j=1}^{J} \hat{u}_{g_j} \kappa_j\Big\|_{s_1,s_2} \;\leq\; \sum_{j=1}^{J} \big\|\hat{u}_{g_j}\big\|_{s_1} \big\|\kappa_j\big\|_{L^{s_2}(0,T)} \;\leq\; \sum_{j=1}^{J} C_g \big\|\kappa_j\big\|_{L^{s_2}(0,T)} \tag{1.79}$$

The assumptions concerning the estimate (1.76) in Theorem 1.2.12 are fulfilled. Thus, by the Hölder inequality and by Theorem 1.2.12, term $\big\|\kappa_j\big\|_{L^{s_2}(0,T)}$ can be estimated by:

$$\big\|\kappa_j\big\|_{L^{s_2}(0,T)} \;\leq\; C_0 \big\|\kappa_j\big\|_{L^{\infty}(0,T)} \;\leq\; C_0 C_1 \tag{1.80}$$

where $C_0 = T^{1/s_2}$ for $s_2 < \infty$, $C_0 = 1$ for $s_2 = \infty$ and where C_1 stands for the constant from (1.76). Note that the assumption $s_1 \geq 2$ is necessary here due to the fact that Theorem 1.2.12 assumes $\hat{u}_{g_j} \in L^2(\Omega)$, $j = 1, \ldots, J$.

Combining (1.79) and (1.80) together, we have

$$\Big\|\sum_{j=1}^{J} \hat{u}_{g_j} \kappa_j\Big\|_{s_1,s_2} \;\leq\; C_F \tag{1.81}$$

where

$$C_F \;:=\; J C_g C_0 C_1$$

The estimate (1.81) is true for an arbitrary $s_2 \in [1, \infty]$. In particular, we can choose

$$\begin{cases} s_2 = \dfrac{2s_1}{2s_1 - \mathbf{d}} & \text{for } s_1 > \mathbf{d}/2 \\ s_2 = \infty & \text{for } s_1 = \mathbf{d}/2 \end{cases} \tag{1.82}$$

One can verify that for s_1 as in the assumptions of the theorem and for s_2 given in (1.82), pair of numbers s_1, s_2 obeys conditions (1.74) and (1.75). This is the point of the proof where the assumption $s_1 \geq \max\{2, \frac{\mathbf{d}}{2}\}$ is necessary because it guarantees that s_1 obeys the restrictions given in (1.75).

Let s_2 be as in (1.82), so as conditions (1.74) and (1.75) were valid. This, along with (1.81) and with the assumptions of the present theorem, implies that the assumptions of Theorem 1.2.11 are fulfilled for the system (1.4) with $k_j := \kappa_j$ and with $g_j := \hat{u}_{g_j}$, $j = 1, \ldots, J$. Observe that y is a weak solution of the system (1.4), with the mentioned assignments (see Definition 1.1.3). Thus, by Theorem 1.2.11 we find that

$$\big\|y\big\|_{L^{\infty}(Q_T)} \leq C_3$$

where C_3 is the constant from the assertion of Theorem 1.2.11. Taking into account the list of quantities on which constant C_3 depends, the construction of constant C_F above and the meaning of C_1, the assertion follows. ■

Basing on Theorem 1.2.13, we will show the following modifications of the existence and uniqueness results given in Corollary 1.2.8 and Corollary 1.2.9:

Theorem 1.2.14 *Let the assumptions of Corollary 1.2.8 be fulfilled, with the following modifications:*

- *we assume that* $f\colon \mathbb{R} \to \mathbb{R}$ *is locally Lipschitz continuous and obeys (1.73) with constant* $C_f > 0$, *instead of the condition for* f *given in the assumption (B-3),*
- *we assume that* $y_0 \in L^\infty(\Omega)$, *instead of the condition for* y_0 *given in the assumption (B-5),*
- *we assume that* $g_j \in L^{s_1}(\Omega)$, *for* $s_1 \geq \max\{2, \frac{\mathbf{d}}{2}\}$, *for* $j = 1, \ldots, J$, *instead of assuming that* g_j *belongs to* $L^2(\Omega)$ *only.*

Then, there exists a unique weak solution of the system (0.1) - (0.3).

Theorem 1.2.15 *Let the assumptions of Corollary 1.2.9 be fulfilled, with the modifications as in Theorem 1.2.14. Then, there exists a unique weak solution of the system (0.1) - (0.3).*

Remark. Note that the condition $g_j \in L^2(\Omega)$ for $j = 1, \ldots, J$ allows the assumptions of Theorem 1.2.14 and Theorem 1.2.15 be fulfilled only for domain dimension $\mathbf{d} \in \{1, 2, 3, 4\}$. One can verify that for higher dimension of the domain, higher integrability of functions g_j would be required. ▲

For conciseness, we present only the proof of Theorem 1.2.14. The proof of Theorem 1.2.15 follows the same lines.

Proof of Theorem 1.2.14. The proof relies on the concept of truncations. For a given $n > 0$, we define truncation $f^n\colon \mathbb{R} \to \mathbb{R}$ as follows:

$$f^n(s) := \begin{cases} f(n) & \text{for } s > n \\ f(s) & \text{for } s \in [-n, n] \\ f(-n) & \text{for } s < -n \end{cases}$$

Note that the function f^n is Lipschitz continuous for an arbitrary $n > 0$ (by local Lipschitz continuity of f) and, for $n \geq C_f$, obeys (1.73) with the same constant C_f as the original function f.

Denote by $\big((0.1)\text{ - }(0.3)\big)^n$ the modification of the system (0.1) - (0.3) consisting in putting f^n instead of f in the main equation of (0.1). The system $\big((0.1)\text{ - }(0.3)\big)^n$ certainly is a particular case of (0.1) - (0.3), hence all definitions and theorems concerning (0.1) - (0.3) apply to $\big((0.1)\text{ - }(0.3)\big)^n$ as well.

In particular, a weak solution of the system $\big((0.1)\text{ - }(0.3)\big)^n$ (see Definition 1.2.1) exists and is unique, for an arbitrary $n > 0$ — see Corollary 1.2.8 and recall the Lipschitz continuity of f^n. The assumption that $s_1 \geq 2$ also is necessary to apply Corollary 1.2.8.

Assume that $(y^n, \kappa_1^n, \ldots, \kappa_J^n) \in X^2$ is the weak solution of the system $\big((0.1)\text{ - }(0.3)\big)^n$ for certain $n > 0$. Now, we will justify that y^n is bounded on Q_T by a constant independent of n, for n big enough.

Functions $\hat{u}_{g_j} := g_j$ obey the requirements of Theorem 1.2.13, for s_1 as presently assumed. As mentioned, f^n is Lipschitz and, for $n \geq C_f$, f^n fulfills (1.73) with constant C_f independent of n. By the latter, and under other assumptions of the present theorem, the system $\big((0.1)\text{ - }(0.3)\big)^n$ obeys the assumptions of Theorem 1.2.13, for $n > C_f$. Thus, by Theorem 1.2.13, we find that

$$\big\| y^n \big\|_{L^\infty(Q_T)} \leq C_0 \qquad \text{for } n > C_f \tag{1.83}$$

where C_0 is the constant from the assertion of Theorem 1.2.13. C_0 is independent of n because none of the quantities on which C_0 depends (Theorem 1.2.13) is dependent on n (what in particular concerns constant C_f, which is the constant for the condition (1.73) for the function f^n with $n > C_f$).

Let us choose number $\widetilde{n}$ greater than $\max\{C_0, C_f\}$. Taking into account the estimate (1.83) and the definition of f^n we obtain:

$$f^{\widetilde{n}}(y^{\widetilde{n}}) = f(y^{\widetilde{n}}) \qquad \text{for a.e. } (x,t) \in Q_T$$

Therefore we conclude that $(y^{\widetilde{n}}, \kappa_1^{\widetilde{n}}, \ldots, \kappa_J^{\widetilde{n}})$ is also a weak solution of the system (0.1) - (0.3).

Above, we have proven that an arbitrary weak solution of $\big((0.1) \text{ - } (0.3)\big)^{\widetilde{n}}$ is a weak solution of (0.1) - (0.3). Thus, by existence of weak solutions for $\big((0.1) \text{ - } (0.3)\big)^{\widetilde{n}}$ (Corollary 1.2.8) we conclude the existence of weak solutions of (0.1) - (0.3). To infer the uniqueness, we need justify that an arbitrary weak solution of (0.1) - (0.3) is a weak solution of $\big((0.1) \text{ - } (0.3)\big)^n$, for certain $n > 0$, and recall the uniqueness result for $\big((0.1) \text{ - } (0.3)\big)^n$ (Corollary 1.2.8). This will close the proof.

But the fact that a weak solution of (0.1) - (0.3) is also a weak solution of $\big((0.1) \text{ - } (0.3)\big)^n$, for certain $n > 0$, follows by arguments analogous to the above ones. Assume that $(y, \kappa_1, \ldots, \kappa_J) \in X^2$ is a weak solution of (0.1) - (0.3). Under the assumptions of the present theorem, the system (0.1) - (0.3) obeys the requirements of Theorem 1.2.13. Thus, we can apply Theorem 1.2.13 again to infer that

$$\big\|y\big\|_{L^\infty(Q_T)} \leq C_0$$

where constant C_0 is the same as in (1.83). Having this, by arguments analogous as above, we see that

$$f^{\widetilde{n}}(y) = f(y) \qquad \text{for a.e. } (x,t) \in Q_T$$

for $\widetilde{n}$ greater than C_0. Therefore, $(y, \kappa_1, \ldots, \kappa_J)$ is a weak solution of $\big((0.1) \text{ - } (0.3)\big)^{\widetilde{n}}$. The uniqueness of the weak solutions for $\big((0.1) \text{ - } (0.3)\big)^{\widetilde{n}}$ follows by Corollary 1.2.8. ■

Remark. The proof of Theorem 1.2.15 is exactly the same as the above proof, with the sole difference that every reference to Corollary 1.2.8 appearing in the proof should be replaced with a reference to Corollary 1.2.9. ▲

Remark. The estimate (1.77) in Theorem 1.2.12 assumes *a priori* knowledge that $y \in L^\infty(Q_T)$, what can be impractical. Theorem 1.2.13 allows to specify more concrete assumptions under which the estimate (1.77) is valid. Namely,

- let the assumptions necessary for the estimate (1.76) in Theorem 1.2.12 hold,
- and in addition, assume that $\big\|y_0\big\|_\infty \leq C_\infty$ and $\hat{u}_{g_j} \in L^{s_1}(\Omega)$, for certain $s_1 \geq \max\{2, \frac{\mathbf{d}}{2}\}$, for $j = 1, \ldots, J$.

Then, the assumptions of Theorem 1.2.13 are fulfilled. Now, Theorem 1.2.13 can be applied to conclude that $y \in L^\infty(Q_T)$. In consequence of the latter and the fact that we impose the assumptions required for (1.76), the assumptions necessary for (1.77) in Theorem 1.2.12 hold.

Provided the above reasoning, constant C_0 entering the structure of C_2 in the estimate (1.77) becomes the constant from the assertion of Theorem 1.2.13 and depends on the quantities indicated therein. ▲

1.2.4 Other generalizations

For technical reasons, in further parts of the present work it will be necessary to deal also with systems of structure slightly different than the structure of (0.1) - (0.3). These are the system

(3.9) - (3.10) (called *linearized system*) and the system (3.30) - (3.31) (called *adjoint system*), introduced in Chapter 3. It will be necessary to have existence and uniqueness results for the mentioned systems, moreover we will need to have estimates for the solutions of the linearized system. Hence, below we introduce a system of structure sufficiently general to let the linearized system and the adjoint system be particular cases of the subject system, and, next, provide uniqueness and existence results along with the necessary estimates for the subject system.

The announced system, which covers the case of both the linearized system and the adjoint system, is the following one:

$$\begin{cases} y_t(x,t) - D\Delta y(x,t) = \widetilde{f}(x,t,y(x,t)) + & \\ \qquad + \sum_{j=1}^{J} \Xi_j(x,t)\kappa_j(t) + \sum_{j=1}^{J} \widetilde{g}_j(x)\Theta_j(x,t) & \text{on } Q_T \\ \dfrac{\partial y}{\partial n} = 0 & \text{on } \partial\Omega \times (0,T) \\ y(0,x) = \widetilde{y}_0(x) & \text{for } x \in \Omega \end{cases} \tag{1.84}$$

$$\begin{cases} \beta_1 \kappa_1'(t) + \kappa_1(t) = \widetilde{W}_1\big(y(\,.\,,t), \mathbf{Y}(\,.\,,t)\big) & \text{on } [0,T] \\ \vdots & \vdots \\ \beta_J \kappa_J'(t) + \kappa_J(t) = \widetilde{W}_J\big(y(\,.\,,t), \mathbf{Y}(\,.\,,t)\big) & \text{on } [0,T] \\ \kappa_j(0) = \widetilde{\kappa}_{j0} \in \mathbb{R} & \text{for } j = 1,\dots,J \end{cases} \tag{1.85}$$

$$\widetilde{W}_j(y(\,.\,,t), \mathbf{Y}(\,.\,,t)) = \mathbf{Z}_j(t)\Big(\int_\Omega \widetilde{h}_j(x)\mathbf{Y}(x,t)\,dx + \widetilde{w}_j\Big(\int_\Omega \mathbf{h}_j(x)y(x,t)dx\Big)\Big) \tag{1.86}$$

where unknown are $\kappa_j\colon (0,T) \to \mathbb{R}$ for $j = 1,\dots,J$ and $y\colon Q_T \to \mathbb{R}$. In the system (1.84) - (1.86), as in previous sections, $T > 0$ and Ω, being a domain in $\mathbb{R}^{\mathbf{d}}$, are given, and $Q_T := \Omega \times (0,T)$. Moreover, $D, \beta_1, \dots, \beta_J > 0$, $\widetilde{f}\colon \Omega \times (0,T) \times \mathbb{R} \to \mathbb{R}$, $\widetilde{w}_j\colon \mathbb{R} \to \mathbb{R}$, $\Xi_j, \Theta_j, \mathbf{Y}\colon Q_T \to \mathbb{R}$, $\widetilde{y}_0, \widetilde{g}_j, \widetilde{h}_j, \mathbf{h}_j\colon \Omega \to \mathbb{R}$, $\mathbf{Z}_j\colon (0,T) \to \mathbb{R}$ and $\widetilde{\kappa}_{j0} \in \mathbb{R}$ are given, for $j = 1,\dots,J$.

In the present section, we provide existence and uniqueness results for the system (1.84) - (1.86), together with estimates for its solutions. The system (1.84) - (1.86) cannot be viewed as a particular case of the system (0.1) - (0.3), thus the results concerning (0.1) - (0.3) are not transmittable to the system (1.84) - (1.86). Nevertheless, the proofs of the existence, uniqueness and stability theorems for (1.84) - (1.86), which will be formulated below, utilize the same methods as the proofs of the analogous theorems concerning (0.1) - (0.3). For this reason, we do not present the proofs in the present section.

The following assumptions will be necessary in this section:

(D-1) $\Omega \subset \mathbb{R}^{\mathbf{d}}$ is as in the assumption (B-1), i.e. Ω:

a) is bounded,

b) satisfies the cone condition,

(D-2) J, T, D and β_j, for all $j = 1,\dots,J$, are as in the assumption (B-2),

(D-3) $\widetilde{f}\colon (\hat{x}, \hat{t}, \hat{y}) \mapsto \hat{f} \in \mathbb{R}$, acting on $\Omega \times (0,T) \times \mathbb{R}$, is:

a) globally Lipschitz continuous w.r.t. $\hat{y}$ for a.e. $(\hat{x}, \hat{t}) \in Q_T$, with a Lipschitz constant independent of $(\hat{x}, \hat{t}) \in Q_T$; we denote this Lipschitz constant by $\widetilde{L}$ and put $\widetilde{f}_0 := f(0)$

b) measurable w.r.t. $(\hat{x},\hat{t})$ for all $\hat{y}\in\mathbb{R}$,

c) $\widetilde{f}^0$, defined by $\widetilde{f}^0(\hat{x},\hat{t}):=\widetilde{f}(\hat{x},\hat{t},0)$ for $(\hat{x},\hat{t})\in Q_T$, belongs to $L^2(Q_T)$,

(D-4) $\widetilde{w}_j$ is globally Lipschitz continuous; we denote the Lipschitz constant of $\widetilde{w}_j$ by $\widetilde{L}_j$ and put $\widetilde{w}_{j0}:=\widetilde{w}_j(0)$, for all $j=1,\dots,J$,

(D-5) $\widetilde{y}_0\in L^2(\Omega)$ and $\widetilde{\kappa}_{j0}\in\mathbb{R}$, for $j=1,\dots,J$,

(D-6) $\mathbf{Y}\in L^2(0,T;L^2(\Omega))$, $\Xi_j\in L^\infty(0,T;L^2(\Omega))$, $\Theta_j\in L^\infty(Q_T)$, $\mathbf{Z}_j\in L^\infty(0,T)$ and $\mathbf{h}_j\in L^2(\Omega)$, for $j=1,\dots,J$.

The solutions of the system (1.84) - (1.86) are understood in the sense analogous to that given in Definition 1.2.1:

Definition 1.2.16 *We say that $(y,\kappa_1,\dots,\kappa_J)\in X^2$ is a weak solution to the system (1.84) - (1.86) if:*

(a) $y(\,.\,,0)=\widetilde{y}_0$ in $L^2(\Omega)$ and $\kappa_j(0)=\widetilde{\kappa}_{j0}$ for $j=1,\dots,J$,

(b) for all $\phi\in L^2(0,T;H^1(\Omega))$, there holds

$$\int_0^T \langle y',\phi\rangle + D\big(\nabla y,\nabla\phi\big)_{L^2(\Omega)} + \big(-\widetilde{f}(\,.\,,t,y)-\sum_{j=1}^J \Xi_j\kappa_j-\sum_{j=1}^J\Theta_j\widetilde{g}_j\,,\,\phi\big)_{L^2(\Omega)}\,dt \;=\; 0$$

(c) for all $\xi\in L^2(0,T)$, for $j=1,\dots,J$, there holds

$$\int_0^T \big(\beta_j\kappa_j'+\kappa_j-\widetilde{W}_j(y,\mathbf{Y})\big)\,\xi\,dt \;=\; 0$$

The point (a) in the above definition makes sense, because, by arguments as in the case of Definition 1.1.1 (see page 6), the condition $(y,\kappa_1,\dots,\kappa_J)\in X^2$ implies $y\in C([0,T];L^2(\Omega))$ and $(\kappa_1,\dots,\kappa_J)\in C([0,T])$.

The below analogues of results presented in Theorem 1.2.5 (estimates in X^2 norm) and Corollary 1.2.9 (existence and uniqueness) are valid:

Theorem 1.2.17 *Let the part a) of the assumption (D-1) and assumptions (D-2) - (D-4), (D-6) be fulfilled. Let $\hat{u}\in\widetilde{U}$ and $(y_0,\kappa_{10},\dots,\kappa_{J0})\in X^0$. Assume also that $\|\hat{u}\|_{\widetilde{U}}\le R^U$ for some $R^U>0$ and that $\|(\widetilde{y}_0,\widetilde{\kappa}_{10},\dots,\widetilde{\kappa}_{J0})\|_{X^0}\le R^0$ for some $R^0>0$. Let $(y,\kappa_1,\dots,\kappa_J)\in X^2$ be a weak solution of the system (1.84) - (1.86) corresponding to $\widetilde{g}_j:=\hat{u}_{g_j}$, $\widetilde{h}_j:=\hat{u}_{h_j}$, for $j=1,\dots,J$, and the initial condition $(\widetilde{y}_0,\widetilde{\kappa}_{10},\dots,\widetilde{\kappa}_{J0})$. Then the following estimate holds:*

$$\|(y,\kappa_1,\dots,\kappa_J)\|_{X^2}\;\le\;C$$

where C depends only on

$$T,J,\widetilde{L},\|\widetilde{f}^0\|_{2,2},\widetilde{L}_1,\dots,\widetilde{L}_J,\widetilde{w}_{10},\dots,\widetilde{w}_{J0},D,\beta_1,\dots,\beta_J,$$

$$R^U,R^0,\|\mathbf{Y}\|_{2,2},\|\Xi_j\|_{2,\infty},\|\Theta_j\|_{L^\infty(Q_T)},\|\mathbf{Z}_j\|_{L^\infty(0,T)},\|\mathbf{h}\|_{L^2(\Omega)}.$$

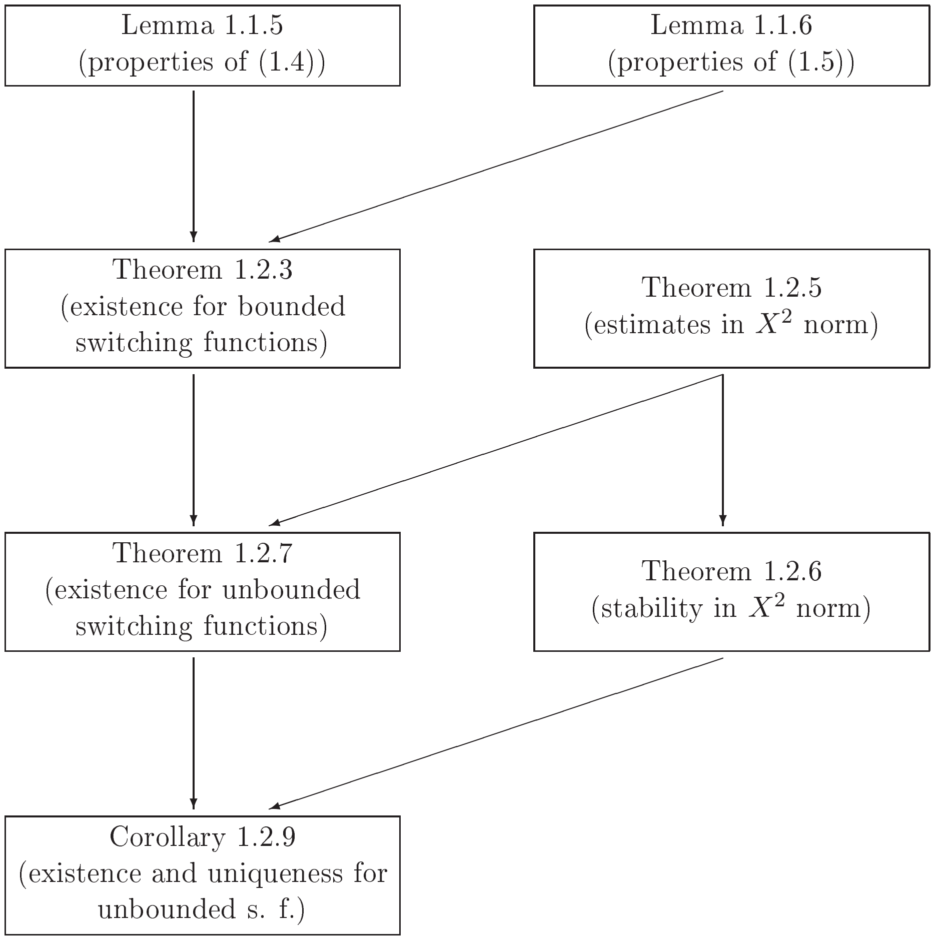

Lemma 1.1.5
(properties of (1.4))
Lemma 1.1.6
(properties of (1.5))
Theorem 1.2.3
(existence for bounded switching functions)
Theorem 1.2.5
(estimates in X^2 norm)
Theorem 1.2.7
(existence for unbounded switching functions)
Theorem 1.2.6
(stability in X^2 norm)
Corollary 1.2.9
(existence and uniqueness for unbounded s. f.)

Figure 1.4: Dependencies between some of theorems in Chapter 1, concerning the system (0.1) - (0.3). Lemmas 1.1.5 and 1.1.6 concern auxiliary equations, while the rest of the results indicated in the above graph concern the system (0.1) - (0.3) directly. In the graph, an arrow leading from A to B means that A was utilized in the proof of B.

Theorem 1.2.18 *Let assumptions (D-1) - (D-6) be fulfilled. Let* $\left(\widetilde{g}_j, \widetilde{h}_j\right)_{j=1}^{J} \in \widetilde{U}$*. Then, the system (1.84) - (1.86) has a unique weak solution.*

REMARK. We have verified that Theorem 1.2.17 and Theorem 1.2.18, as analogues of Theorem 1.2.5 and Corollary 1.2.9, respectively, can be proven with the same methods as the latter statements. Corollary 1.2.9 depend also on other results proven in Chapter 1, see Figure 1.4. Fortunately, analogues of these results also can be proven for the system (1.84) - (1.86) with the same methods.

We give one necessary comment concerning the above matter. One of the necessary results is an analogue of Lemma 1.1.5. We remark that the appropriate analogue of Lemma 1.1.5, necessary here, should be proven (and can be proven), not for auxiliary the system (1.4) (which was considered in Lemma 1.1.5), but for the following modification of (1.4):

$$\begin{cases} y_t(x,t) - D\Delta y(x,t) = \widetilde{f}(x,t,y(x,t)) + & \\ \qquad + \sum_{j=1}^{J} \Xi_j(x,t)k_j(t) + \sum_{j=1}^{J} \Theta_j(x,t)\widetilde{g}_j(x) & \text{on } Q_T \\ \dfrac{\partial y}{\partial n} = 0 & \text{on } \partial\Omega \times (0,T) \\ y(0) = \widetilde{y}_0 & \text{on } \Omega \end{cases}$$

▲

Since, according to the above remark, the proofs of Theorems 1.2.17 and 1.2.18 can be conducted with the methods as the other proofs of Chapter 1, we skip them.

Chapter 2

Thermostat control mechanism — numerical prototypes

The present chapter is devoted to numerical simulations concerning the thermostat control mechanism, utilized in (0.1) - (0.3).

The aim of the simulations is twofold. First, we intended to investigate the efficiency of the thermostat control mechanism, understood as the ability of the latter to bring the state of the process close to some neighborhood of the reference state y^*. In our simulations, we observe how the efficiency changes with changes of the reference state, of the initial state and of the number of the control and measurement devices. Note that the results described in Chapter 1 do not say anything about the efficiency of the thermostat control mechanism, in the mentioned sense. Thus, the observations concerning the efficiency, made within the scope of the numerical simulations, complement the qualitative results given in Chapter 1.

Second, we were interested in the question whether the state of the process controlled by thermostats, for large time, becomes independent of the initial state of the process or not. This kind of independence is essential for the optimal targeting problem, announced in §2 of *Introduction*, because the independence on the initial state gives additional practical advantage to the cost functional (0.8).

Being more precise, assume that the process, controlled by thermostats, stabilizes close to a certain state, independent of the initial state. Then, the cost functional (0.8) with T_0 close to T, also becomes independent of the initial state of the controlled process. In consequence, still assuming T_0 close to T, the optimal targeting problem, which bases on the latter cost functional, has solutions independent of the initial state. Nevertheless, we mention the above only to signalize certain issues concerning the optimal targeting problem. We postpone the analysis of the latter problem until Chapter 3 and Chapter 4.

As mentioned above, the efficiency of the thermostat control mechanism will be understood as the ability to bring the state of the process to a neighborhood of the reference state. To work with this approach, it is necessary to observe whether the state of the process indeed stays, for large time, in some neighborhood of the reference state or not. Assuming that this is the case, we can introduce an intuitive criterion to compare the efficiency of the thermostat control mechanism in two distinct situations. For example, let situations A and B differ in the initial state of the process. We will say that the thermostat control mechanism is more efficient in situation A than in situation B if in situation A the controlled process stays in a neighborhood of the reference state of a diameter smaller than in situation B. In particular, assume that, after some time, the process evolution stabilizes near to some time-invariant state. Then, the efficiency of the thermostat control mechanism can be measured in terms of the gap between the process state, at time moment large enough to observe the stabilization, and the reference state. In the present chapter, we refer to the latter understanding of efficiency. For this purpose, we measure the gap between the process state and the reference state in terms of $W^{1,2}(\Omega_N)$ norm, where Ω_N denotes the triangulated domain utilized in the simulations.

Mathematically, in the present chapter, by the initial state of the process controlled by thermostats we mean y_0 component of the initial condition $(y_0, \kappa_{10}, \dots, \kappa_{J0})$ in the system (0.1) - (0.3).

In the simulations described in the present chapter, the main equation of the system (0.1) - (0.3) was discretized in space with the use of the finite element method. A square domain, triangulated with triangular elements, was considered. The finite element space was the space of continuous functions, linear on each element. The time discretization was performed by employing the implicit Euler scheme. The nonlinear terms entering the system (0.1) - (0.3) were treated by means of the Picard iterations method.

Three experiments were performed. The first concerns the properties of the thermostat control mechanism when it is focused on a task of preserving an unstable state. The second one concerns an attempt of comparison of efficiency of the thermostat control mechanism for various initial states. The third one compares the properties of the thermostat control mechanism when two different numbers of the control and measurement devices are considered.

In the results of the simulations, we observe that the efficiency of the thermostat control mechanism, understood in the above mentioned sense, changes with the changes of the number of the control and measurement devices. The efficiency varies also with changes of the size of the supports of functions g_j and h_k, describing the control and measurement devices actions.

Concerning the independence of the behavior of the controlled process on the initial state for large time, varying results were observed. In some of the performed simulations, the results suggest that the alleged independence is possible. However, there were also simulations suggesting the opposite, namely that a change of the initial state possibly could result, even for long time horizon, in an essentially different state.

The order of the present chapter is as follows. In Section 2.1, we describe the structural assumptions imposed in the system (0.1) - (0.3) in our simulations, i.e. we specify the domain, the nonlinear terms etc. Next, in Section 2.2, we describe the utilized numerical scheme in more detail. Eventually, we proceed to Section 2.3, which is devoted to presentation and discussion of the results of the simulations.

2.1 Structural assumptions

In the experiments described in Section 2.3, the below assumptions were made.

We assumed that every control device in the thermostat control mechanism distributes energy uniformly in a disc centered at given $x_j \in \Omega$. We treated the measurement devices analogously, assuming that every measurement device observes a disc-shaped area. Moreover, we assumed that the numbers of the control and measurement devices are equal. More precisely, in the system (0.1) - (0.3), functions g_j and h_k, characterizing the devices actions, were determined by

$$K = J \tag{2.1}$$

$$g_j := \hat{u}_{g_j} := \sigma_g(\,.\, - x_j)|_\Omega, \qquad h_j(x) := \hat{u}_{h_j} := \sigma_h(\,.\, - x_j)|_\Omega \tag{2.2}$$

for $j = 1, \dots, J$, where $x_j \in \mathbb{R}^{\mathbf{d}}$ and $\sigma_g, \sigma_h \colon \mathbb{R}^{\mathbf{d}} \to \mathbb{R}$, and where σ_g and σ_h are given by:

$$\sigma_g(x) = C_g \mathbf{1}_{B(0,r_\sigma)}(x), \qquad \sigma_h(x) = C_h \mathbf{1}_{B(0,r_\sigma)}(x) \tag{2.3}$$

for certain $r_\sigma, C_g, C_h > 0$. In other words, the area of actions of every control device coincided with area of actions of exactly one measurement device.

We imposed the following assumption for the weights α_{jk}:

$$\alpha_{jk} := \hat{u}_{\alpha_{jk}} := \delta_{j,k} \tag{2.4}$$

for $j, k = 1, \ldots, J$, where $\delta_{j,k}$ denotes the Kronecker delta function of j and k (see *Notation conventions*). The assumption (2.4) is natural in the context of assumptions (2.1), (2.2), (2.3).

Having (2.1), (2.2), (2.3) and (2.4), the control $(g_j, h_j, \alpha_{jk})_{j=1,\ldots,J} \in U$, applied in the system (0.1) - (0.3), is determined once a selection of the points $x_1, \ldots, x_J$ and the parameters $r_\sigma, C_g, C_h > 0$ is made.

The above assumptions result in a simplified version of the model (0.1) - (0.3), which is a focus of our interest in the present chapter, concerning the numerical results:

$$\begin{cases} y_t(x,t) - D\Delta y(x,t) = f(y(x,t)) + \sum_{j=1}^{J} g_j(x)\kappa_j(t) & \text{on } Q_T \\ \dfrac{\partial y}{\partial n} = 0 & \text{on } \partial\Omega \times (0,T) \\ y(x) = y_0(x,0) & \text{for } x \in \Omega \end{cases} \tag{2.5}$$

together with

$$\begin{cases} \beta_1 \kappa_1'(t) + \kappa_1(t) = w_1\Big(\displaystyle\int_\Omega h_1(x)(y - y^*)dx\Big) & \text{on } [0,T] \\ \vdots & \vdots \\ \beta_J \kappa_J'(t) + \kappa_J(t) = w_J\Big(\displaystyle\int_\Omega h_J(x)(y - y^*)dx\Big) & \text{on } [0,T] \\ \kappa_j(0) = \kappa_{j0} \in \mathbb{R} & \text{for } j = 1, \ldots, J \end{cases} \tag{2.6}$$

for functions g_j and h_j defined by (2.2) and (2.3).

The experiments were performed for a two-dimensional rectangular domain:

$$\Omega = (-1,1) \times (-1,1) \subset \mathbb{R}^2 \tag{2.7}$$

It was assumed that y^* was time independent: $y^* = y^*(x)$.

The reactive term f treated in the experiments was:

$$f(s) = -s^3 + s \tag{2.8}$$

together with w_j given by

$$w_j(s) = H_w \max(\min(L_w s, 1), -1) \tag{2.9}$$

for certain L_w, H_w, for $j = 1, \ldots, J$.

Remark. In fact, our intention was to use w_j defined by $w_j(s) = -H_w sgn(s)$ for a certain H_w, because, according to remarks in §1 of *Introduction*, $-sgn$ is a natural example of a switching function in thermostat control mechanism. Nevertheless, we wanted the data for the simulations to be covered by the analytical results presented in Section 1.2, concerning in particular existence and uniqueness of solutions for the system (0.1) - (0.3). The results of Section 1.2 are proven under assumption that the switching functions are Lipschitz continuous, what excludes the choice of $-sgn$ or $-H_w sgn$. Therefore, for the simulations, we have decided to choose Lipschitz functions of a steep slope in point $s = 0$, approximating in a certain sense the ideal function $-H_w sgn$. Basing on the reasoning as in the example on page 17, we have chosen the switching function as in (2.9). ▲

For a given r_σ, we considered the value of C_h to be determined by the following relation:

$$C_{switch} \int_{\mathbb{R}^{\mathbf{d}}} \sigma_h = 1/\left|L_w\right| \tag{2.10}$$

for certain $C_{switch} > 0$. In the above, C_h is present in the definition of σ_h. The identity (2.10) along with definition of σ_h in (2.3) allows to infer that

$$C_h = \left(\pi\left|L_w\right| C_{switch}\, r_\sigma^2\right)^{-1} \tag{2.11}$$

Remark. For better explanation of the meaning of the constant $C_{switch} > 0$, we make the following remark. Due to assumptions (2.1) and (2.4), the term $w_j\left(\int_\Omega h_j(y - y^*)\right)$ in the right hand side of (2.6) is the signal generated by the signal generator associated with j-th control device (see the nomenclature introduced in §1 of *Introduction*). The concept is that C_{switch} defines a threshold gap between the solution y and the reference state y^* after exceeding which the extremal value of signal is returned by the signal generators. Being more precise, for a given measurement device, (which actions are characterized by the function h_j) we want the signal to achieve its maximal value when $y - y^* \approx C_{switch}$ or $y - y^* \approx -C_{switch}$ in the area observed by the measurement device (i.e. in the support of h_j). Taking the formula for w_j into account, the extremal signal value is achieved for $\int_\Omega h_j(y - y^*) = \pm 1/\left|L_w\right|$ (or for higher values of the latter integral; nevertheless, in our idea, we are interested in the smallest gap between y and y^* for which the extremal signal value is achieved; hence the latter condition with sign „=", not „≥", expressing that we want the value of the integral to coincide with the closest to zero extremal points of w_j). Processing the above conditions yields

$$1/\left|L_w\right| = \int_\Omega h_j\left|y - y^*\right| \approx C_{switch}\int_\Omega h_j$$

This gives the relation (2.10), after assuming that „≈" sign can be replaced by the equality sign and after assuming that $\int_\Omega h_j = \int_{\mathbb{R}^d} \sigma_h$. The latter is correct if $\mathrm{supp}(\sigma_h(\,.\, - x_j)) \subseteq \Omega$. For simplicity of the above reasoning, referring rather to general concepts than to precise calculations, we assumed it to be true. However, it can be not the case in general. ▲

Altogether, for Ω given by (2.7), the reactive term as in (2.8), the switching function w_j as in (2.9), g_j, h_j, $\alpha_{j,k}$ defined by conditions (2.1), (2.2), (2.3), (2.4) and C_h as in the formula (2.11), the system (2.5) - (2.6) is uniquely determined by the choice of the following quantities:

$$\begin{array}{llll} y_0, & \kappa_{10}, \ldots, \kappa_{J0}, \quad y^* & J, & x_1, \ldots, x_J \\ T, & D, \beta_1, \ldots, \beta_J, & & r_\sigma, C_g, C_{switch}, L_w, H_w \end{array}$$

The values of the above quantities utilized in the particular experiments will be specified in Section 2.3.

Remark. One may verify that the above Ω, f, w_j, g_j, h_j for $j = 1, \ldots, J$ fits the assumptions of the existence, uniqueness and stability results from Section 1.2.3. Moreover, for particular experiments described in Section 2.3, we will choose y_0 and y^* which also fulfill the assumptions of the subject existence, uniqueness and stability results. ▲

2.2 Numerical methods

The below numerical methods were utilized in the experiments described in Section 2.3.

For numerical treatment of the system (2.5) - (2.6) we utilized the finite element method to solve the component y corresponding to the parabolic equation.

The triangulation of Ω, see (2.7), was of the type presented on Figure 2.1. The finite element space chosen for the simulations was the space of continuous functions, linear on every element of the triangulation. The time interval was discretized by selecting a uniformly distributed in the set $[0, T]$ of time points. The implicit Euler scheme was used to solve the model w.r.t. the time variable.

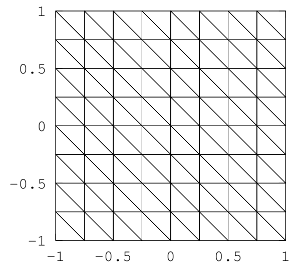


Figure 2.1: The type of triangulation of Ω utilized in the experiment. The triangulation is such that the mesh associated with the triangulation has the same number of nodes along each spatial direction.

The nonlinear terms f and w were treated with the use of the Picard iterations technique. A constant number of the Picard iterations for every time step was utilized. We preferred a constant number of Picard iterations instead of applying the error-based stop criterion in order to control the computational time.

In the further part of our work, we will use the following notation concerning the above described numerical scheme:

$N+1$ — the number of nodes along each spatial direction, for the mesh associated with the triangulation,
τ_N — the length of the mesh step along each spatial direction,
$M+1$ — the number of time points in the time discretization,
τ_M — the length of the time step,
N_{Picard} — the number of Picard iterations in every time step.

According to the above notation, the total number of nodes in the triangulation equals $(N+1)^2$. Moreover, relations $\tau_N = N^{-1}$ and $\tau_M = M^{-1}$ hold.

Let us sketch in more detail the numerical scheme applied for the system (2.5) - (2.6). Denote the triangulation of type presented in Figure 2.1, corresponding to $N+1$ nodes along each spatial direction, as Ω_N. Denote the finite element space of functions on Ω_N being continuous on Ω_N and linear on every element of Ω_N as $P_1(\Omega_N)$.

Moreover, for a given function $F \colon \Omega \to \mathbb{R}$, denote the continuous linear interpolation of F, taking exact values in the nodes of the mesh associated with Ω_N, by $[F]_N$. In addition, denote by $\vec{F}$ the vertical vector of the values of F in the nodes of the mesh associated with Ω_N. It follows by the definitions that $\vec{F}=\overrightarrow{[F]_N}$.

Remark. Note that, Ω_N, understood as a subset of $\mathbb{R}^2$, equals Ω. As a consequence, it is legal to write $P_1(\Omega_N) \subseteq L^2(\Omega)$ or $L^2(\Omega_N) = L^2(\Omega)$. ▲

We begin with discretization in space, proceeding as follows. In the system (2.5) - (2.6), we take $[g_j]_N$, $[h_j]_N$, $[y_0]_N$ and $[y^*]_N$ instead of g_j, h_j, y_0 and y^*, respectively. Next, we transform this modification of (2.5) - (2.6) to the following variational problem, using the $P_1(\Omega_N)$ space:

$$\begin{cases} \frac{d}{dt}\big(y_N, \phi\big)_{L^2(\Omega_N)} \;+\; D\big(\nabla y_N, \nabla\phi\big)_{L^2(\Omega_N)} = \\ \qquad = \big([f(y_N)]_N, \phi\big)_{L^2(\Omega_N)} + \sum_{j=1}^{J}\big([g_j]_N, \phi\big)_{L^2(\Omega_N)}\kappa_{j,N} \quad \text{on } [0,T],\ \forall_{\phi\in P_1(\Omega_N)} \\ y_N(0) = [y_0]_N \end{cases} \tag{2.12}$$

and

$$\begin{cases} \beta_j \frac{d}{dt}\kappa_{j,N} + \kappa_{j,N} = w_j\Big(\big([h_j]_N, (y_N - [y^*]_N)\big)_{L^2(\Omega_N)}\Big) \quad \text{on } [0,T] \\ \kappa_{j,N}(0) = \kappa_{j0} \end{cases} \tag{2.13}$$

for $j = 1, \ldots, J$, where $(y_N, \kappa_{1,N}, \ldots, \kappa_{J,N})$, with $y_N(t) \in P_1(\Omega_N)$ and $\kappa_{j,N}(t) \in \mathbb{R}$ for $t \in [0,T]$, is the desired solution. Note, that the term $f(y_N)$ is not in $P_1(\Omega_N)$. This is the reason for which, defining the above variational problem, we use $[f(y_N(t))]_N$ in (2.12) instead of $f(y_N(t))$ (for the sake of readability, the time dependence in (2.12) is hidden). Note also that term $(\nabla y_N, \nabla\phi_N)_{L^2(\Omega_N)}$ above is well defined, since $P_1(\Omega_N) \subseteq H^1(\Omega_N)$ (see Theorem 2.1.1. in [13]).

Remark. Since, as a subset of $\mathbb{R}^2$, Ω_N equals Ω, using notation „Ω_N" instead of „Ω" in (2.12) - (2.13) is not necessary. Nevertheless, in (2.12) - (2.13) we use notation „Ω_N" in order to stress that we are working with a space discretization of original the system (2.5) - (2.6). ▲

Define the following matrices:

$$\mathbb{M}_N = \Big((\phi_m, \phi_n)_{L^2(\Omega_N)}\Big)_{n,m=1}^{(N+1)^2}, \qquad \mathbb{A}_N = \Big((\nabla\phi_m, \nabla\phi_n)_{L^2(\Omega_N)}\Big)_{n,m=1}^{(N+1)^2}$$

where ϕ_n, for $n = 1, \ldots, (N+1)^2$, denotes the standard „hat" basis of the finite element space $P_1(\Omega_N)$.

Note that, given $F, G \in P_1(\Omega)$, we can represent them as $F = \sum_{n=1}^{(N+1)^2} \vec{F}_n\ \phi_n$ and $G = \sum_{n=1}^{(N+1)^2} \vec{G}_n\ \phi_n$, respectively. Hence:

$$(F,G)_{L^2(\Omega)} = (\vec{F})^T \mathbb{M}_N\ \vec{G}, \qquad (\nabla F, \nabla G)_{L^2(\Omega)} = (\vec{F})^T \mathbb{A}_N\ \vec{G} \tag{2.14}$$

Now, note that $\overrightarrow{[f(y_N)]}_N = f\big(\vec{y_N}\big)$. Using this and the above observation concerning products of $P_1(\Omega_N)$ functions, we transform the system (2.12) - (2.13) further, to the matrix form:

$$\begin{cases} \frac{d}{dt}\mathbb{M}_N\ \vec{y_N}\ + D\mathbb{A}_N\ \vec{y_N} = \mathbb{M}_N f\big(\vec{y_N}\big) + \sum_{j=1}^{J} \mathbb{M}_N\ \overrightarrow{[g_j]}_N\ \kappa_{j,N} \quad \text{on } [0,T] \\ \vec{y_N}\ (0) = \overrightarrow{[y_0]}_N \end{cases} \tag{2.15}$$

with

$$\begin{cases} \beta_j \frac{d}{dt}\kappa_{j,N} + \kappa_{j,N} = w_j\Big(\overrightarrow{[h_j]}_N^{\ T}\ \mathbb{M}_N\big(\vec{y_N} - \overrightarrow{[y^*]_N}\big)\Big) \quad \text{on } [0,T] \\ \kappa_{j,N}(0) = \kappa_{j0} \end{cases} \tag{2.16}$$

for $j = 1, \dots, J$. The unknown solution of (2.15) - (2.16) is $\big(\overrightarrow{y_N}, \kappa_{1,N}, \dots, \kappa_{J,N}\big)$.

We approximate the solution of (2.15) - (2.16), as mentioned, by using the implicit Euler scheme with $M + 1$ time points, uniformly distributed in interval $[0, T]$, and by using the method of Picard iterations with N_{Picard} iterations to treat the nonlinear terms in each time step. Denote the approximation of solution of (2.15) - (2.16) obtained with these methods by $(\overrightarrow{Y_N}, \hat{k}_{1,N}, \dots, \hat{k}_{J,N})$. The latter approximation is a function defined in the time discretization points, $t = m\tau_M$, $m = 0, 1, \dots, M$, with values in $\mathbb{R}^{(N+1)^2} \times \mathbb{R}^J$.

Having this, we construct the following function $(Y_N, k_{1,N}, \dots, k_{J,N})$, defined in time discretization points, i.e. in $t = m\tau_M$, $m = 0, \dots, M$, and taking values in $P_1(\Omega_N) \times \mathbb{R}^J$. For $t = m\tau_M$, $m = 0, \dots, M$, we put $Y_N(t) = \sum_{n=1}^{(N+1)^2} (\overrightarrow{Y_N}\,(t))_n\, \phi_n$ and $k_{j,N} = \hat{k}_{j,N}$ for $j = 1, \dots, J$.

The function $(Y_N, k_{1,N}, \dots, k_{J,N})$ is the output of the above numerical scheme for the system (2.5) - (2.6). In other words, we treat $(Y_N, k_{1,N}, \dots, k_{J,N})$ as an approximation of the weak solution of (2.5) - (2.6) (since (2.5) - (2.6) is a particular case of (0.1) - (0.3), we understand the weak solution of (2.5) - (2.6) in sense of Definition 1.2.1).

All simulations which results are presented in Section 2.3 were performed with the use of the above described scheme.

For the purpose of our experiments, the matrices $\mathbb{M}_N$ and $\mathbb{A}_N$ were computed explicitly, with no use of numerical integration methods.

Note, that the above described numerical scheme is fully determined by the choice of the parameters determining the finite element space, the time discretization scheme and the nonlinear term treatment method, i.e. by the following parameters:

$$N, \qquad M, \qquad N_{Picard}$$

The values of the above parameters utilized in the particular experiments will be specified in Section 2.3.

2.3 Results of simulations

Now we proceed to presentation of the results announced in the introduction to Chapter 2. The experiments described below were performed with the use of the numerical scheme from Section 2.2 and under the structural assumptions from Section 2.1.

In the below discussion of the results, we put stress on the efficiency of the thermostat control mechanism, understood in terms of the gap between the process state and the reference state for large time. To realize the subject objective, we proceed with the following strategy. We observe whether stabilization of the process occurred at the terminal time, $t = T$, of our simulations and scrutinize the gap at $t = T$.

We are also interested in observing whether the behavior of the process controlled by thermostats exhibits independence on the initial state for large time. The idea to investigate this matter is to wait until the process, considered with distinct initial states, stabilizes, an then to compare the observed process states.

Our approach to the both of the above questions (efficiency and independence on the initial state) assume that the behavior of the process stabilizes after some initial period, in which oscillations possibly occur. Hence, throughout the results discussion in the present section, we will stress whether we observed stabilization in the behavior of the controlled process or not. Above, as everywhere else in the further part of the present chapter, by stabilization we mean that the process remains close to certain time-invariant state. By oscillations we mean rapid changes of the process state.

Moreover, to realize the above ideas concerning efficiency, it is necessary to have some measure of the distance between the reference state the process state in a given time $t \in [0, T]$. For this end, we measure the distance between two given states in terms of $W^{1,2}(\Omega_N)$ norm, where Ω_N is as in Section 2.2 (this is implemented by means of functions E_{Y_N} and $E^{grad}_{Y_N}$, defined below).

In Section 2.3.1 and Section 2.3.3, we describe experiments illustrating the behavior of the thermostat control mechanism for varying numbers of the control and measurement devices. In Section 2.3.2, we take a look at behavior of the subject system in a situation where the initial state of the process varies.

Section 2.3.1 concerns the case where y^* is an unstable equilibrium of the process and the supports of functions g_j and h_j cover the domain tightly. The cases of various sizes of the supports of g_j and h_j are compared. It is observed that the efficiency of the thermostat control mechanism improves as the size of the supports of g_j and h_j decreases. In Section 2.3.2, we assume that the number of the control and measurement devices, as well as the targeting of their actions, are fixed and we do not assume that y^* is an unstable equilibrium (y^* is chosen as a state representing some free boundary). We observe that the efficiency of the thermostat control mechanism is similar for two distinct variants of the initial state. In Section 2.3.3, we consider y^* as in Section 2.3.2. We also assume that the initial state and the sizes of the supports of g_j and h_j are fixed. We compare the behavior of the thermostat control mechanism for varying numbers of the control and measurement devices. It is observed that the efficiency of the thermostat control mechanism decreases as the number of the devices decreases.

In all cases considered in Section 2.3.1, Section 2.3.2 and Section 2.3.3 some stabilization of the behavior of the process was observed, after an initial period of oscillations. In other words, the thermostat control mechanism seemed to bring the process near to some time-invariant state. Nevertheless, in some cases the achieved approximate time-invariant state seems to be dependent on the initial state of the process. We comment on this matter more broadly in Section 2.3.4.

Below, by *numerical solution* of the system (2.5) - (2.6) we mean the approximation of a solution of (2.5) - (2.6), denoted in Section 2.2 as $(Y_N, k_{1,N}, \dots, k_{J,N})$. For convenience, here we also keep notation $(Y_N, k_{1,N}, \dots, k_{J,N})$ for denoting the numerical solution of (2.5) - (2.6). In addition, by *numerical process* we mean „numerical approximation of the process controlled by thermostats". Mathematically, the notion of numerical process below coincide with Y_N.

In the presentation of the results, some plots appear and thus we give a short clarification of the utilized plot convention here. The plots can be grouped into certain classes: 1) plots of functions from $P_1(\Omega_N)$, 2) plots concerning configuration of the control devices utilized in the experiments and 3) error plots.

By *configuration of the control and measurement devices* we mean the choice of the supports of functions g_j and h_j, which characterize the control and measurement devices actions.

The error plots are self-describing. The rest of the plots need to be commented.

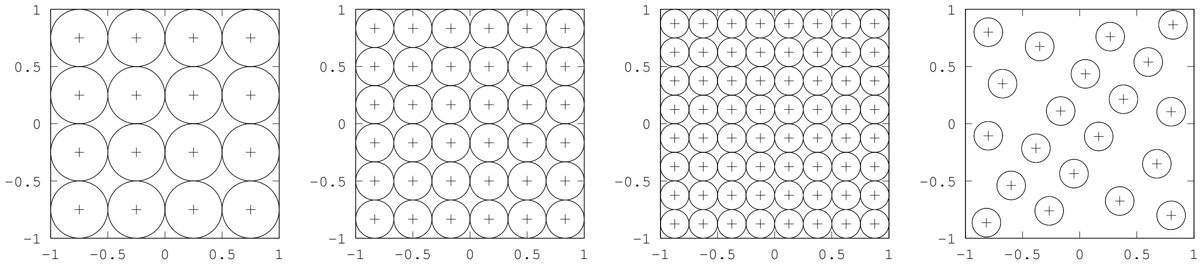
The plots of functions from $P_1(\Omega_N)$ are plots:

- of the main component Y_N of the numerical solution of the system (2.5) - (2.6), in a given moment of time,
- of the initial state y_0 of the process or of the reference state y^*, utilized in the experiments.

In the plots of functions from $P_1(\Omega_N)$, the color map extends from black to white. The values below a down threshold value of the color map are plotted in black and the values exceeding an upper threshold value are plotted in white. The threshold values of the color map are indicated in the plots. The maximal and minimal values of the plotted data also are indicated there.

The plots concerning the configuration of the control and measurement devices are visualizations of supports of functions g_j and h_j. An essential remark is that, due to the structural assumptions in the Section 2.1, the supports of the functions g_j and h_j are pairwise equal. Thus, one disc in a plot concerning the configuration of the devices represents a pair of supports — the support of g_j and the support of h_j, for certain $j \in \{1, \ldots, J\}$.

The mentioned visualizations of supports, if sufficiently precise, give a unique characterization of the parameter r_σ and of the utilized sequence of the central points, $x_1, \ldots, x_J$, appearing in (2.3) (up to permutation). The latter information, along with information concerning parameters C_g and C_h (which will be provided explicitly in the description of the experiments), gives full information about the functions g_j and h_j.



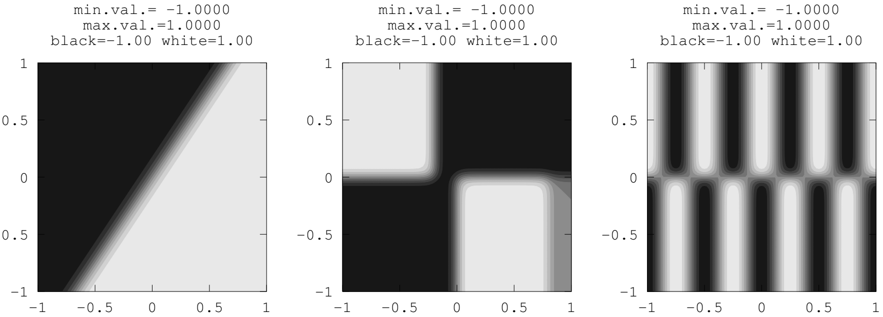
(a) 16 devices

(b) 36 devices

(c) 64 devices

(d) 20 devices

Figure 2.2: Control and measurement devices configurations for Section 2.3.



(a) A reference state.

(b) Init. cond., 1st variant

(c) Init. cond., 2nd variant

Figure 2.3: A part of data employed for simulations in Section 2.3. The plotted functions are given by formulas (2.17) for Fig. 2.3a, (2.18) for Fig. 2.3b and (2.19) for Fig. 2.3c.

Figures 2.2 and 2.3 present data which shall be utilized in the experiments below. The data employed in particular experiments will be specified in their description by reference to these figures. The functions plotted in Figure 2.3 are given by the following formulas:

$$\hat{y}(x_1, x_2) = 1 - 2\big(1 + e^{-15\frac{3\sqrt{13}}{13}(x_2 - 1.5\,x_1)}\big) \tag{2.17}$$

$$\begin{aligned} \hat{y}(x_1, x_2) &= -1 + \Big(2\big(1 + e^{-30\,x_1}\big)^{-1} - \big(1 + e^{-30(x_1 - 0.8)}\big)^{-1}\Big) \cdot \big(1 + e^{30\,x_2}\big)^{-1} + \\ &\quad + 2\big(1 + e^{30(x_1 + 0.2)}\big)^{-1} \cdot \big(1 + e^{-30\,x_2}\big)^{-1} \end{aligned} \tag{2.18}$$

$$\hat{y}(x_1, x_2) = \cos\big(4\pi x_1\big) \cdot \Big(1 - 2\big(1 + e^{30\,x_2}\big)^{-1}\Big) \tag{2.19}$$

Moreover, assume that $y^* \in H^1(\Omega)$ and that Y_N is the main component of numerical solution of (2.12) - (2.13), obtained with the methods described in Section 2.2. For the time discretization points $t = m\tau_M$, $m = 0, \dots, M$ we denote by $E_{Y_N}(t)$ the L^2 error between Y_N and $[y^*]_N$:

$$E_{Y_N}(t) = \left\| Y_N(t) - [y^*]_N \right\|_{L^2(\Omega)}$$

and by $E^{grad}_{Y_N}(t)$ the gradient error between y_N and $[y^*]_N$, or more precisely:

$$E^{grad}_{Y_N}(t) = \left\| \nabla \left(Y_N(t) - [y^*]_N \right) \right\|_{L^2(\Omega)}$$

where $[y^*]_N$ is defined as in Section 2.2.

For brevity, below, values $E_{Y_N}(t)$ and $E^{grad}_{Y_N}(t)$ will be called *error values*.

The below described simulations have been performed with the use of the GNU Octave software.

2.3.1 Experiment 1 — unstable equilibrium

The present experiment is intended to illustrate properties of the thermostat control mechanism in a situation where the reference state is unstable.

The following data were exploited for the present experiment:

$$\begin{array}{llll} T = 24 & C_g = 16/\pi & L_w = -10 & \kappa_{j0} = 0 \ \forall_{j=1,\dots,J} \\ D = 0.01 & C_{switch} = 0.2 & H_w = 10 & \end{array}$$

together with the numerical scheme specification given by:

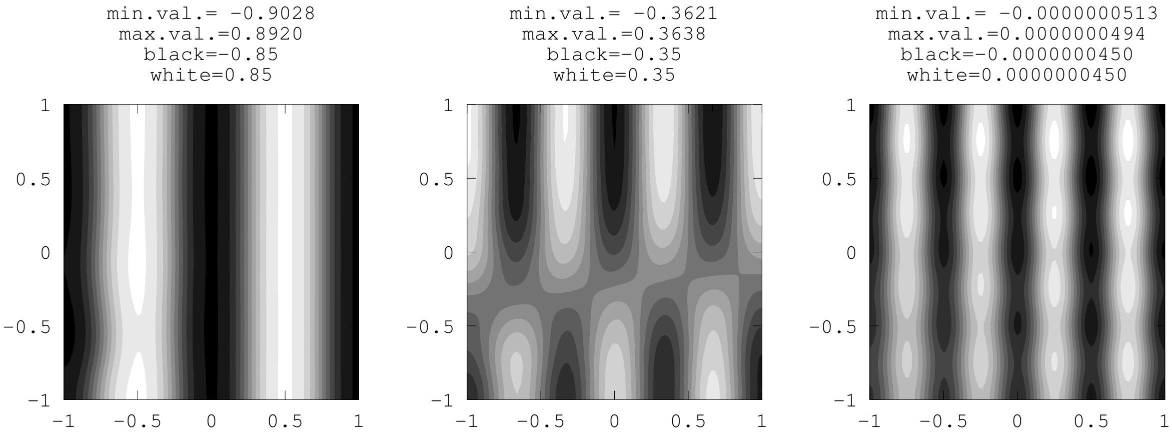
$$N = 100, \qquad M = 2400, \qquad N_{Picard} = 3$$

We considered the initial state y_0 as on Figure 2.3b and the reference state $y^* \equiv 0$. Note that the y^* taken into account indeed is an unstable state for the assumed reactive term f.

We have performed three simulations, basing on various configurations of the control and measurement devices. The cases of $J = 16, 36, 64$, with the devices tightly covering the domain with their effects, but varying in the size of the areas affected by a single device, have been considered. The utilized devices configurations are presented on Figures 2.2a, 2.2b and 2.2c. One can say that these configurations differ with resolution of measurement abilities and with resolution of control abilities.

In each of the three simulations, oscillations in the process behavior faded after certain initial period. It could be observed that, after this initial period, there emerged certain patterns which did not underwent further rapid changes. However still, some slow evolution of the numerical process could be observed in longer time horizon. Nevertheless, by the evolution of the process which we observed, the process states achieved for the time $t = T$ seemed to be close to certain time-invariant states of the considered model (however, the latter require further work for better verification).

Now, let us comment on the efficiency of thermostat control mechanisms associated with the addressed devices configurations. Probably, for many users the result on Figure 2.4a (corresponding to only 16 devices) cannot be considered to be precise solution in the context of the problem of leading the state of the process to the state $y^* \equiv 0$. Nevertheless, the situation was changing as we were increasing the number of the devices, keeping uniform distribution of their actions through the domain. Comparing Figures 2.4a, 2.4b and 2.4c suggests that the greater

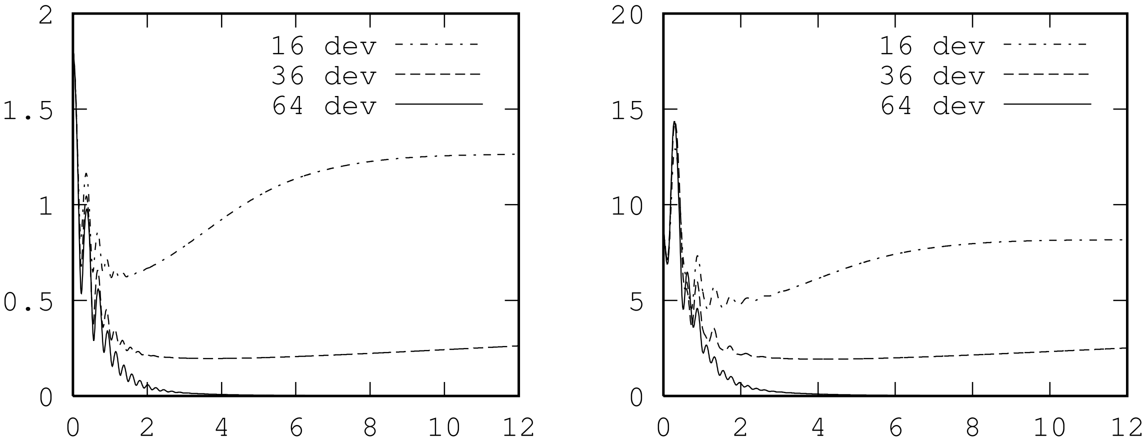


(a) 16 dev., $t = T$

(b) 36 dev., $t = T$

(c) 64 dev., $t = T$

Figure 2.4: Numerical process at time $t = T$, for the devices configurations considered in Section 2.3.1. Fig. 2.4a corresponds to the dev. conf. in Fig. 2.2a; Fig. 2.4b — to Fig. 2.2b; Fig. 2.4c — to Fig. 2.2c.



(a) $E_{Y_N}(t)$ values (vert. axis) in time

(b) $E^{grad}_{Y_N}(t)$ values (vert. axis) in time

Figure 2.5: $E_{Y_N}(t)$ and $E^{grad}_{Y_N}(t)$ for time points $t = m\tau_M$, $m = 0, \ldots, M/2$ for simulations corresponding to the devices configurations considered in Section 2.3.1. For the sake of readability, the time horizon of the error plots is limited to $[0, 12]$. After time $t = 12$ the error values still evolves, however slowly, without rapid changes.

the number of the control and measurement devices is, the more precise response of the control devices can be expected. This stays consistent with the natural intuition.

The drastic difference between the efficiency of the thermostat control mechanism for 16 devices and the efficiency for the cases of 36 and 64 devices is well visible on the error plots in Figures 2.5a and 2.5a. The Reader may also compare the obtained error values at time $t = T$ in Table 2.1.

Remark. The above described results suggest that, in the situation of the present experi-

y part for:	16 dev.	36 dev.	64 dev.
$E_y(T)$	1.3006	0.3568	5.5550e-08
$E_y^{grad}(T)$	8.2791	3.4143	6.9999e-07

Table 2.1: The values of error at the terminal time ($t = T$) for the devices configurations considered in Section 2.3.1. The presented values are rounded.

ment, the main question concerning the efficiency of the control by thermostats can be reduced to the question on the number of the devices which would be sufficient to achieve demanded precision. This is much simpler adjustment procedure than procedures that often can be necessary in the case of systems with an open-loop control. Suppose that we consider a system with an open-loop control in which the user is responsible for the choice of right number of the control devices as well as for the choice of the power functions, κ_j. In other words, equations (2.6) are not taken into account. Such open-loop control is more difficult to handle than our closed-loop control, utilized in the model (2.5) - (2.6), because the user has to control more variables. Necessary is the choice of the devices together with the power functions in the introduced open-loop case, versus the choice of the devices only in the case of our closed-loop control. Moreover, in the open-loop situation a proper choice of the power functions κ_j is hard to be done by intuition. Probably, proper power functions would be searched by some optimization procedure, what additionally increases the complexity of efforts necessary to deal with the open-loop case. In addition, it is reasonable to expect that the choice of the power functions depend on the initial state of the process. Thus, it would be necessary to repeat the optimization procedure concerning the power functions after every change of the initial state.

To sum up, the observed simplicity of adjustment of the thermostat control mechanism stays in accordance with the expected advantages of the models with automatic correction mechanisms, expressed in *Introduction*. ▲

2.3.2 Experiment 2 — various initial conditions

Below, we present numerical results which illustrate behavior occurring in the investigated model with control by thermostats when perturbations of the initial state are induced.

In the present experiment, the following data were used :

$$\begin{array}{llll} T = 4 & r_\sigma = 1/8 & L_w = -10 & C_{switch} = 0.2 \\ D = 0.02 & C_g = 16/\pi & H_w = 10 & \kappa_{j0} = 0 \ \forall_{j=1,\ldots,J} \end{array}$$

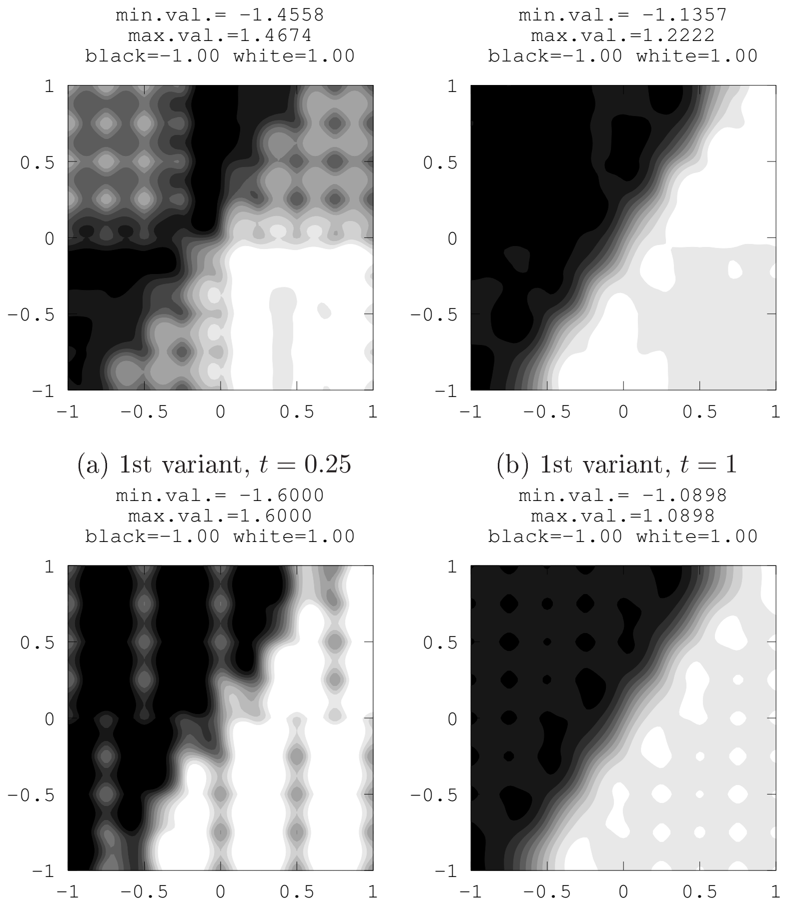
together with the numerical scheme specification given by:

$$N = 100, \qquad M = 400, \qquad N_{Picard} = 3$$

The configuration of the control and measurement devices was assumed to be as the devices configuration with $J = 64$ utilized in the experiment from the Section 2.3.1, i.e. as on Figure 2.2c. The reference state was as in Figure 2.3a.

Two simulations has been performed, with two variants of the initial state y_0. The first of them was as in Figure 2.3b, the second initial state was as in Figure 2.3c.

For the both simulations, stabilization of the numerical process occurred after initial period of oscillations, i.e. certain states which did not underwent further visible changes emerged.



(a) 1st variant, $t = 0.25$

(b) 1st variant, $t = 1$

(c) 2nd variant, $t = 0.25$

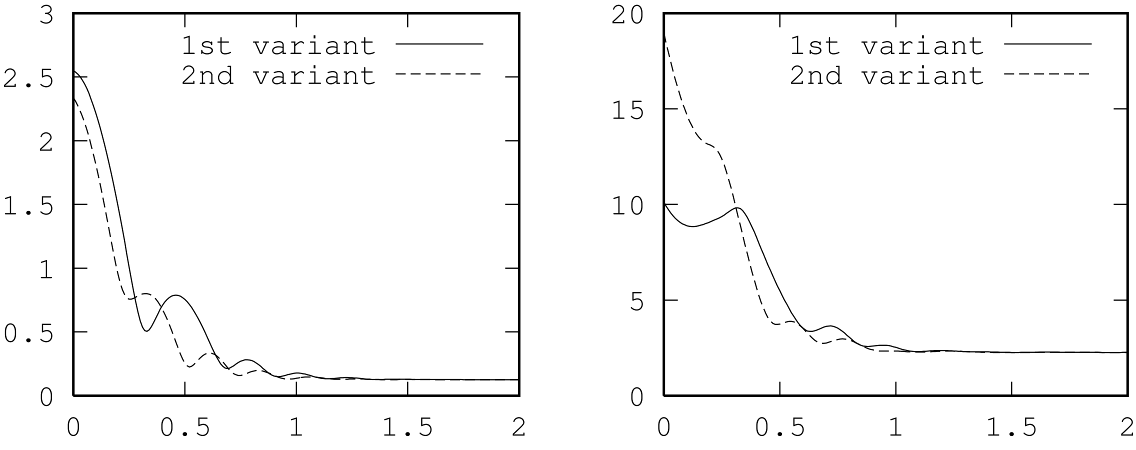
(d) 2nd variant, $t = 1$

Figure 2.6: Numerical process at time $t = 0.25$ and $t = 1$, for two initial state variants considered in Section 2.3.2. Fig. 2.6a, 2.6b correspond to the i. cond. in Fig. 2.3b; Fig. 2.6c, 2.6d — to Fig. 2.3c.

The subject stable states seemed to match the reference state at some rate of accuracy, at least visually. Moreover, the numerical process generated in both simulations occurred to achieve a high level of likeness in a short time. This is visible on Figures 2.6a - 2.6d — in particular, the figures corresponding to the time $t = 1$ (Figures 2.6b and 2.6d) represent process states which can be considered to be visually similar. It suggests that the efficiency of the thermostat control mechanism is similar for the two subject simulations.

The error plots in Figures 2.7a and 2.7b confirm that the components Y_N of the both numerical solutions fall into the same neighborhood of the reference state, in the sense of the error metric considered in the present chapter. Moreover, the ratio of the error at the terminal time of the experiment is close to 1 (see Table 2.2). Thus, indeed, the efficiency of the thermostat control mechanism, observed in the above numerical simulations, can be considered to be similar for the two initial state cases.

As an outcome of the above observations, we propose the following hypothesis: the thermostat control mechanism has the very useful property of preserving the efficiency under perturbations of the initial state.



(a) $E_{Y_N}(t)$ values (vert. axis) in time

(b) $E^{grad}_{Y_N}(t)$ values (vert. axis) in time

Figure 2.7: $E_{Y_N}(t)$ and $E^{grad}_{Y_N}(t)$ for time points $t = m\tau_M$, $m = 0, \ldots, M/2$, for simulations corresponding to the two initial state variants considered in Section 2.3.2. The time interval for the plots is limited to $[0, 2]$ for the sake of readability. No significant fluctuations of the error values were observed after time $t = 2$.

y part for:	1st variant	2nd variant	ratio
$E_y(T)$	0.12569814	0.12569916	1.00000812
$E^{grad}_y(T)$	2.26541586	2.26541453	0.99999941

Table 2.2: The values of error at the terminal time ($t = T$) for the initial state y_0 considered in Section 2.3.2 (with rounding to 8 significant digits).

Remark. In Figures 2.7a and 2.7b, it can be observed that the initial error was leveled within a similar time, approximately equal $t \approx 1$, in both cases. However, the reason of the latter can be e.g. the comparable rank of values of the considered initial states. It is reasonable to expect that if we had considered two initial states where one of them was defined as ten thousand times the other then the time of leveling the initial error would differ. Nevertheless, the above observation suggests the following hypothesis concerning the properties of the investigated thermostat control mechanism: if the family of initial states satisfy certain common bound, then the time of convergence of the controlled process to a given neighborhood of the stable state is similar for all initial states in the subject family. ▲

Remark. An interesting observation can be made by comparing the results discussed in Section 2.3.2 with the result concerning the case of 64 devices, discussed in Section 2.3.1. The simulations which generated the subject results share the same configuration of the control and measurement devices. As we already have noted, in all the subject simulations the numerical process behavior eventually stabilize. The error values (see Figures 2.5a, 2.5a, 2.7a, 2.7b) also seem to stabilize at some stable value. Compare the error values in Table 2.2 and Table 2.1 (for 64 devices). An observation can be made that the stable error value is much lower in the case of the reference state $y^* \equiv 0$ than in the case of y^* as in Figure 2.3a. This is interesting since one could expect the opposite, as the behavior of the process near $y^* \equiv 0$ is perhaps, roughly speaking, more unstable than near y^* as in Figure 2.3a. ▲

2.3.3 Experiment 3 — various numbers of thermostats

This experiment is devoted to compare behavior of the thermostat control mechanism for two different configurations of the control and measurement devices, where the size of the areas affected by particular devices equals in both cases but the number of the devices differs. This is a situation different than in Section 2.3.1, where the considered devices configurations differed not only with number of the devices but also with the sizes of the areas affected by the devices.

The following data was exploited for the present experiment:

$$\begin{array}{llll} T = 4 & r_\sigma = 1/8 & L_w = -10 & C_{switch} = 0.2 \\ D = 0.02 & C_g = 16/\pi & H_w = 10 & \kappa_{j0} = 0 \ \forall_{j=1,\dots,J} \end{array}$$

together with the numerical scheme specification given by:

$$N = 100, \qquad M = 400, \qquad N_{Picard} = 3$$

The initial state chosen for the present experiment was as in Figure 2.3b and the reference state was as in Figure 2.3a.

Two simulations, corresponding to two configurations of the control and measurement devices, were performed. The considered configurations of the devices one with $J = 64$ and the other with $J = 20$, are presented in Figures 2.2c and 2.2d.

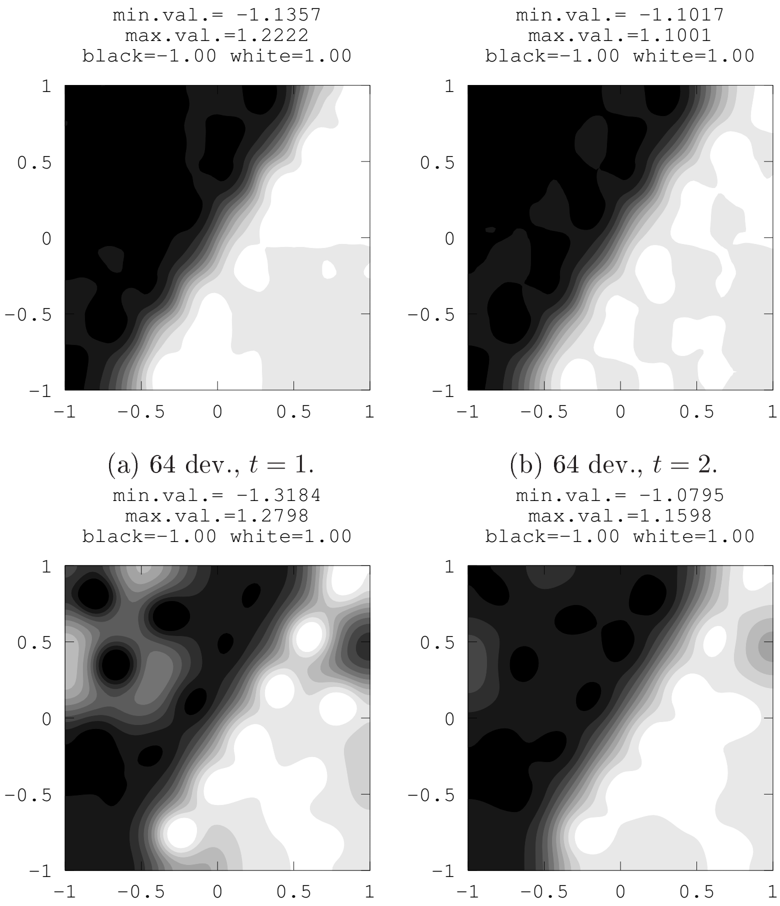
In both simulations, stabilization of the numerical process took place after some initial period of time. In other words, certain states which did not underwent further visible changes emerged. In the case of 64 devices, the numerical process occurred to stabilize quickly at some state similar to the reference state, see Figures 2.8a and 2.8b. We can say that the process falls to some relatively small neighborhood of the reference state in this case. For the case of 20 devices, as we see on Figures 2.8c and 2.8d, the process also seems to fall into some neighborhood of the reference state. However, the difference between Figures 2.8c and 2.8d seems to be bigger than between Figures 2.8a and 2.8b, at least visually. Therefore, it is possible that for 20 devices, the evolution toward the reference state is slower than in case of the simulation with 64 devices.

In the error plots in Figures 2.9a and 2.9b we observe that the error values for both considered simulations stabilize at some level. The subject error plots also suggest that the efficiency of the thermostat control mechanism, understood as the error at time $t = T$, differs for the two considered devices configurations. The latter is also confirmed by the error values at time $t = T$, presented in Table 2.3.

y part for:	64 dev.	20 dev.	ratio
$E_y(T)$	0.2609	0.1257	0.4817
$E_y^{grad}(T)$	3.0757	2.2654	0.7366

Table 2.3: The values of error at the terminal time ($t = T$) for the devices configurations considered in Section 2.3.3 (with rounding to 4 significant digits).

As a conclusion, the above observations stays consistent with the intuitive hypothesis that the efficiency of the thermostat control mechanism looses its efficiency as the number of the control and measurement devices is decreased.



(a) 64 dev., $t = 1$.

(b) 64 dev., $t = 2$.

(c) 20 dev., $t = 1$.

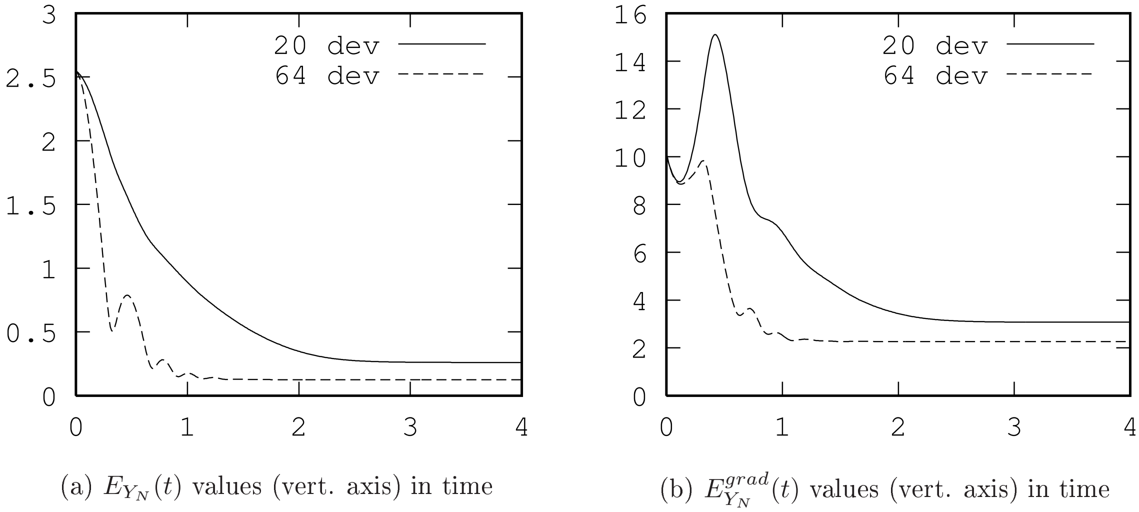
(d) 20 dev., $t = 2$.

Figure 2.8: Numerical process at time $t = 1$ and $t = 2$ for the devices configurations considered in Section 2.3.3. Fig. 2.8a, 2.8b correspond to the dev. conf. in Fig. 2.2c; Fig. 2.8c, 2.8d — to Fig. 2.2d.

Remark. We already remarked above that in the case of 20 devices the thermostat control mechanism may drive the process state toward some stable state slower than in the case of 64 devices. This is visible also in Figures 2.9a and 2.9b. For both plots, the error line concerning the case of 20 devices tends to the terminal value slower, in comparison to the error line concerning 64 control and measurement devices.

Hence, by the above observations, we propose the following hypothesis: when the number of the devices is decreased, the thermostat control mechanism loose not only its efficiency, understood in terms of the gap between the process state and the reference state for large time, but also looses the speed of stabilizing the process. Note that this stays in opposite to the situation considered in Section 2.3.2. There, we concluded with a hypothesis that, for a given configuration of the devices, the speed of stabilization is approximately the same for varying initial data. ▲

Remark. Summing up the observations made in Section 2.3.3, one can say that the 20 devices thermostat control mechanism seems to loose in the contest with the 64 devices thermostat control mechanism. However, a situation where we have not enough control devices



(a) $E_{Y_N}(t)$ values (vert. axis) in time

(b) $E^{grad}_{Y_N}(t)$ values (vert. axis) in time

Figure 2.9: $E_{Y_N}(t)$ and $E^{grad}_{Y_N}(t)$ for time points $t = m\tau_M$, $m = 0, \ldots, M$, for simulations corresponding to the devices configurations considered in Section 2.3.3.

to cover the domain tightly with their effects, i.e. the situation of 20 devices considered above, seems to be more natural than the situation of 64 devices.

This leads to further questions. The configuration of the 20 control and measurement devices presented on Figure 2.2d has been chosen for our experiments by intuition. Hence it is natural to ask whether the actions of these devices could be localized in the domain Ω better. Or, whether we could remove more control devices and still obtain a result which would be called satisfactory with respect to a given criterion. Here, the realm of optimization begins. ▲

2.3.4 Remarks on large time behavior

In the above described experiments, observations concerning stabilization of the numerical process near to some time-invariant state were made. This allows to pose hypotheses on the dependence of these time-invariant states on the initial state. It will be convenient to express the hypotheses in question in the language of hypotheses concerning the asymptotic behavior of the system (0.1) - (0.3), understood in terms of existence and characterization of attracting sets. For example, to say that the time-invariant state is probably independent of the initial state means to say that the attracting set is probably a singleton (if exists).

It is not straightforward what should be the precise form of the hypotheses in question. The numerical prototypes in Section 2.3.1, Section 2.3.2 and Section 2.3.3 suggest that the behavior of the model (0.1) - (0.3) for large times varies depending its configuration. By the configuration of the model (0.1) - (0.3) we understand the choice of particular parameters, as the initial state y_0, the reference state y^* and functions g_j and h_k, characterizing the control and measurement devices actions.

In the situations taken into account in the simulations in Section 2.3.2 and Section 2.3.3, intuition suggests that the process stabilizes at certain state which is relatively close to the reference state. Thus, for these configurations of the model, existence of a one-point or a very small attracting set can be expected.

The situation in the simulations concerning the reference state being an unstable equilibrium, what was the case in Section 2.3.1, is different. If the numerical process states in terminal time, presented on the Figures 2.4a, 2.4b and 2.4c, are close to certain time-invariant state of the real process then, by symmetry, the transposed states are close to a time-invariant state as well. The transposed state should be obtained at time $t = T$ in the simulation with the transposed initial state. By a transposed state we mean a state with swapped role of axis of the coordinate system in $\mathbb{R}^2$. In consequence, in the case of $J = 16$ control and measurement devices, the hypothetic attracting set, if exists, cannot be expected to be small in the sense of diameter. The reason for this is that in the subject case the process state obtained at the terminal time (Figure 2.4a) is quite distant from its transposed state. The attracting set, if exists, should contain states which are close to both the original and transposed state.

To sum up the above, the numerical results presented in this chapter suggest that the attracting set for the dynamical system associated with the model (0.1) - (0.3), if exists, has the structure varying significantly with changes of the configuration of the model. There are configurations for which the results suggest a small, or even one-point attracting set, as well as there are configurations for which a rather big attracting set can be expected.

Besides the above question on the structure of the attracting set, one can also be interested in the question on time necessary to bring the process near to the time-invariant state. The subject information also is essential, if one wants to rank the thermostat control mechanism with respect to the gap between the state obtained for large times and the reference state.

In this field, the differences also occurred between particular simulations. For simulations described in Section 2.3.2, time interval $[0, 4]$ was enough for the numerical process to achieve some state that seemed time invariant. This is also reflected on the error plots on Figures 2.7a, 2.7b, 2.9a, 2.9b. In contrary, for experiment described in Section 2.3.1, for the cases of $J = 16$ and $J = 36$ devices, the evolution of the numerical process toward states which seemed time-invariant was very slow. This is the main reason for which we have chosen the time interval for this experiment equal to $[0, 24]$, what is six times longer than the time intervals in other experiments. At time $t = 4$, the numerical process still evolved, for the cases of $J = 16$ and $J = 36$ devices described described in Section 2.3.1. This is visible in the error plots in Figures 2.5a and 2.5b.

Thus, the numerical results described in the present chapter suggest that the time necessary to bring the state of the controlled process near a time-invariant state varies wit changes of the configuration of the model (0.1) - (0.3).

Nevertheless, the above hypotheses concerning the structure of the alleged attracting set and the speed of evolution of the process base on the error graphs and on visual inspection of the numerical solution plots. Therefore, these hypotheses require further verification. It will be not the subject of the present work.

Chapter 3

Optimal targeting problem — properties

In the simulations described in Chapter 2, we have observed that the efficiency of the thermostat control mechanism, understood as the gap between the state of the controlled process and the reference state at the terminal time T, may differ for different choice of parameters in the thermostat control mechanism (e.g. for different reference states or different numbers of the control and measurement devices). Hence the natural question concerning improving the efficiency of the thermostat control mechanism.

The problem of improving efficiency of the thermostat control mechanism can be understood as the problem of optimizing the feedback law in this system, with respect to a cost functional which reflects the above understanding of efficiency (where the feedback law is the algorithm for computing the response functions κ_j in the system (0.1) - (0.3)). However, the problem of optimizing the feedback law require a parametrization of the feedback law.

In many situations, it can be a natural assumption that the user of the thermostat control mechanism cannot freely manipulate the patterns of energy distributed in the domain by a given control device but only can decide on the location of the pattern. Analogous remark concerns the actions of the measurement devices. We will thus parametrize the feedback law by assuming that the patterns associated with the actions of both control and measurement devices are given and that the control parameter is the set of locations of the subject patterns. Moreover, to exclude the problems associated with the choice of weights $\alpha_{j,k}$, we will assume that $\alpha_{j,k}$ are given.

The above assumptions lead us to the optimal targeting problem, announced in §2 of *Introduction.* The latter problem will be the subject of the present chapter.

To recall, the optimal targeting problem bases on the system (0.1) - (0.3) with additional conditions (0.4) - (0.7). The latter conditions allow to transform the system (0.1) - (0.3) to the following system:

$$\begin{cases} y_t(x,t) - D\Delta y(x,t) = & \\ \qquad = f(y(x,t)) + \sum_{j=1}^{J} \big(\mathcal{P}^{R,\Omega}\mathcal{T}_{\sigma_g}(x_j)\big)(x)\kappa_j(t) & \text{on } Q_T \\ \dfrac{\partial y}{\partial n} = 0 & \text{on } \partial\Omega \times (0,T) \\ y(x,0) = y_0(x) & \text{for } x \in \Omega \end{cases} \tag{3.1}$$

together with

$$\begin{cases} \beta_1\kappa_1'(t)+\kappa_1(t)= & \\ \qquad = w_1\Big(\displaystyle\int_\Omega \big(\mathcal{P}^{R,\Omega}\mathcal{T}_{\sigma_h}(x_1)\big)(x)\big(y(x,t)-y^*(x,t)\big)\,dx\Big) & \text{on } [0,T] \\ \vdots & \vdots \\ \beta_J\kappa_J'(t)+\kappa_J(t)= & \\ \qquad = w_J\Big(\displaystyle\int_\Omega \big(\mathcal{P}^{R,\Omega}\mathcal{T}_{\sigma_h}(x_J)\big)(x)\big(y(x,t)-y^*(x,t)\big)\,dx\Big) & \text{on } [0,T] \\ \kappa_j(0)=\kappa_{j0}\in\mathbb{R} & \text{for } j=1,\dots,J \end{cases} \tag{3.2}$$

where $(y,\kappa_1,\dots,\kappa_J)$ is the unknown and: $\sigma_g,\sigma_h\colon \mathbb{R}^{\mathbf{d}}\to\mathbb{R}$; $x_j\in\mathbb{R}^{\mathbf{d}}$; Ω is a domain in $\mathbb{R}^{\mathbf{d}}$; $T,D,\beta_j>0$, $y^*\colon Q_T\to\mathbb{R}$; $y_0\colon\Omega\to\mathbb{R}$; $\kappa_{j0}\in\mathbb{R}$; $f,w_j\colon\mathbb{R}\to\mathbb{R}$; where $j=1,\dots,J$. Operators $\mathcal{T}_{\sigma_g}$ and $\mathcal{T}_{\sigma_h}$ are defined as in Appendix A.4. The operator $\mathcal{P}^{R,\Omega}$ is the operator of restriction to Ω of a function from $\mathbb{R}^{\mathbf{d}}$ to $\mathbb{R}$.

For convenience, in the present chapter, we will refer to the system (3.1) - (3.2) rather than to the system (0.1) - (0.3) with conditions (0.4) - (0.7). Note that conditions (2.1), (2.2) and (2.4), utilized in Chapter 2, are equivalent to conditions (0.4) - (0.7), constituting the optimal targeting problem. The difference is that in Chapter 2 we considered a particular choice of the pattern functions, given by the additional condition (2.3), while in the present chapter we dismiss the latter condition, taking aim at allowing a more general choice the pattern functions.

Recall the nomenclature introduced in §2 of *Introduction.* In (3.1) - (3.2), functions σ_g and σ_h are called *the pattern functions*. The sequence $(x_1,\dots,x_J)$ is called *the control parameter*, because it determines the control uniquely.

The cost functional which we will investigate is the following:

$$(x_1,\dots,x_J)\ \mapsto\ \widetilde{\lambda}\int_{T_0}^{T}\int_\Omega \big|y(x,t)-y^*(x,t)\big|^2\,dx\,dt \tag{3.3}$$

where $\widetilde{\lambda}>0$, $T_0\in(0,T)$ and $y\colon Q_T\to\mathbb{R}$ is as in (3.1) - (3.2) — in particular, y depends on the control parameter $(x_1,\dots,x_J)$. *The optimal targeting problem* is to minimize the cost functional (3.3).

Recall that, for T_0 close to T, the cost functional (3.3) can be understood as an approximate measure of the gap between the process state and the reference state at the terminal time T (see the remarks in §2 of *Introduction*), i.e. as an approximate measure of efficiency of thermostat control mechanism. Recall also that, since we do not consider the functions g_j and h_j to represent material objects (see §1 of *Introduction*), intersection of their supports with each other and with the exterior of Ω are allowed. In consequence, we do not put any constraints in the optimal targeting problem (see §2 of *Introduction*).

In this chapter, we intend to perform mathematical analysis of the optimal targeting problem. The main results of this analysis concern existence of minimizers and characterization of the gradient of the cost functional defined by (3.3), in a form of an explicit formula. The formula for the gradient of the cost functional is a result of a great practical meaning. An explicit formula for the gradient of (3.3) is necessary for performing many optimization procedures which approximate the local minimizers of (3.3). In Chapter 4, we describe results of numerical optimization experiments in which the formula for gradient of (3.3), derived in the present chapter, was utilized. Moreover, an explicit formula for the gradient of (3.3) has also a meaning for formulating explicit necessary optimality conditions for the considered optimization problem.

The more detailed order of the present chapter is as follows. In Section 3.1, the main goal is to investigate the properties of the operator assigning solutions of (3.1) - (3.2) to a given control parameter $(x_1, \ldots, x_J)$, let us call it the state operator. Knowledge on this properties is necessary for further analysis, concerning the cost functional (3.3), because the subject cost functional can be viewed as a superposition of the squared second Lebesgue norm, of translation by $-y^*$ and of the mentioned state operator. In Section 3.1, the main results rely strongly on the properties of the system (0.1) - (0.3) which were investigated in Section 1.2. Consequently, the main results of Section 3.1 are shown under structural assumptions concerning the system (3.1) - (3.2) similar to the assumptions imposed in Section 1.2 for the system (0.1) - (0.3), with some modifications and supplements, if necessary. To describe briefly the mentioned results, we show that, depending on pattern functions σ_g and σ_h, the mentioned state operator is continuous (for $\sigma_g, \sigma_h \in L^2(\mathbb{R}^{\mathbf{d}})$), or even Lipschitz continuous and weakly Gâteaux differentiable (for $\sigma_g, \sigma_h \in W^{1,2}(\mathbb{R}^{\mathbf{d}})$).

In Section 3.2, we focus directly on analysis of the cost functional (3.3). The analysis involves also the results for the state operator obtained in Section 3.1. We derive a simple criterion for existence of minimizers for the cost functional (3.3). This criterion is shown under conditions sufficient for continuity of the state operator (in particular, $\sigma_g, \sigma_h \in L^2(\mathbb{R}^{\mathbf{d}})$) and additionally assumes that the supports of the pattern functions σ_g and σ_h are compact. The latter assumption is strong but sufficient for our purposes because, in the numerical optimization experiments described in Chapter 4, we operate with the pattern functions with compact support. Next, we proceed to analysis of differentiability of the cost functional (3.3). In brief, the cost functional (3.3) is Gâteaux differentiable if the above mentioned state operator is weakly Gâteaux differentiable. Therefore, the Gâteux differentiability of the cost functional (3.3) is shown under the assumption $\sigma_g, \sigma_h \in W^{1,2}(\mathbb{R}^{\mathbf{d}})$ in particular, as it is one of conditions necessary for weak Gâteaux differentiability of the state operator in Section 3.1. Under the same assumption, we also derive a formula characterizing the gradient of the cost functional, what is a main result of Section 3.2.

Before we proceed to realization of the above objectives, let us introduce the definition of weak solutions of the system (3.1) - (3.2). PDE-ODE the system (3.1) - (3.2) is a particular case of (0.1) - (0.3). Thus, we assume the definition of weak solutions for (3.1) - (3.2) to be exactly the same as for (0.1) - (0.3) — see Definition 1.2.1. To be clear:

Definition 3.0.1 *An element $(y, \kappa_1, \ldots, \kappa_J)$ belonging to X^2 is a weak solution of the system (3.1) - (3.2) if it is a weak solution for the system (0.1) - (0.3) corresponding to:*

$$g_j := \mathcal{P}^{R,\Omega}\mathcal{T}_{\sigma_g}(x_j), \quad h_j := \mathcal{P}^{R,\Omega}\mathcal{T}_{\sigma_h}(x_j), \quad \alpha_{j,k} = \delta_{j,k}$$

for $j, k = 1, \ldots, J$.

Above, the space X^2 is as in Chapter 1. Uniqueness and existence of weak solutions of (3.1) - (3.2) will be one of results of Section 3.1.2, thus we do not touch this matter now.

In many results of the present chapter, assumptions concerning the system (3.1) - (3.2) will cover, in particular, assumptions utilized in previous chapters for the system (0.1) - (0.3). More precisely, assumptions (B-1) - (B-5) and (C-1) - (C-2) from Section 1.2 will be in use in this chapter as well. Nevertheless, some of the results in the present chapter will require additional assumptions. These assumptions are:

(E-1) $f'(s)$ exists for all $s \in \mathbb{R}$, in classical sense,

(E-2) $w_j'(s)$ exists for all $s \in \mathbb{R}$ and all $j = 1, \ldots, J$, in classical sense,

(E-3) a) $p_2 \in (2, 4 - \frac{4}{p_1}]$, where p_1 is a given number satisfying $p_1 > 2$ (in case $\mathbf{d} = 1, 2$) or $2\mathbf{d}/(\mathbf{d} - 2) \geq p_1 > 2$ (in case $\mathbf{d} > 2$),

b) $y^* \in L^{p_2}(0,T;L^2(\Omega))$, for p_2 as in a).

Moreover, assumptions concerning pattern functions σ_g and σ_h are necessary. Depending on situation, a subset of the following set of assumptions will be utilized:

(F-1) $\sigma_g, \sigma_h \in L^2(\mathbb{R}^{\mathbf{d}})$,

(F-2) $\sigma_g, \sigma_h \in W^{1,2}(\mathbb{R}^{\mathbf{d}})$,

(F-3) σ_g and σ_h have compact supports in $\mathbb{R}^{\mathbf{d}}$.

Notation remarks

In the present chapter, spaces X^1, X^2, U and $\widetilde{U}$ are as in Chapter 1. In addition, we define the following space:

$$X^{3,p} = L^p(0,T;L^p(\Omega)) \times \left(L^2(0,T)\right)^J$$

where $p \in [1,\infty]$ is given and natural number J is the same as J appearing in the system (0.1) - (0.3). We endow $X^{3,p}$ with the standard product topology, hence we consider the following norm for $X^{3,p}$:

$$\big\|(y,\kappa_1,\ldots,\kappa_J)\big\|_{X^{3,p}} = \big\|y\big\|_{p,p} + \sum_{j=1}^{J} \big\|\kappa_j\big\|_{L^2(0,T)}$$

We also define

$$V = \left(\mathbb{R}^{\mathbf{d}}\right)^J$$

where natural number J is the same as J appearing in the system (0.1) - (0.3). V will be called *the control parameter space.* For a given element $\hat{v} \in V$ we denote its components as follows:

$$\hat{v} = (\hat{v}_1,\ldots,\hat{v}_J)$$

Note, that an arbitrary control parameter $(x_1,\ldots,x_J)$ in the system (3.1) - (3.2) can be understood as an element of V and *vice versa* — an element $\hat{v} \in V$ determines a control parameter for the system (3.1) - (3.2), by relations $x_j := \hat{v}_j$, $j = 1,\ldots,J$.

For a given $T_0 \in (0,T)$, we use the following notation:

$$Q_T^{T_0} := \Omega \times (T_0,T)$$

In addition, for given functions $F_1\colon \mathbb{R}^{\mathbf{d}} \to \mathbb{R}$, $F_2\colon \Omega \to \mathbb{R}$, $F_3\colon Q_T \to \mathbb{R}$ and $F_4\colon (0,T) \to \mathbb{R}$ and a given index $j \in \{1,\ldots,J\}$, the following definitions of operators will be valid in the present chapter:

$\mathcal{P}^{R,\Omega}$ — restriction operator defined by $\mathcal{P}^{R,\Omega}(F_1) = F_1|_{\Omega}$ (already used in the system (3.1) - (3.2)),

$\mathcal{P}^{E,\Omega}$ — extension by zero operator defined by $\mathcal{P}^{E,\Omega}(F_2) = F_2$ on Ω and $\mathcal{P}^{E,\Omega}(F_2) = 0$ on Ω^c,

$\mathcal{P}^{R,T_0}$ — restriction operator defined by $\mathcal{P}^{R,T_0}(F_3) = F_3|_{Q_T^{T_0}}$,

$\mathcal{P}^{i}_{Q_T}$ — inverse time operator defined by $\mathcal{P}^{i}_{Q_T}(F_3)(x,t) := F_3(x,T-t)$, for all $(x,t) \in Q_T$,

$\mathcal{P}^{i}_{T}$ — inverse time operator defined by $\mathcal{P}^{i}_{Q_T}(F_4)(t) := F_4(T-t)$, for $t \in (0,T)$,

$\mathcal{P}_j^{R,V}$ — operator for extraction of j-th component of $\hat{v} \in V$, i.e. $\mathcal{P}_j^{R,V}(\hat{v}) = \hat{v}_j$ for $\hat{v} \in V$,

$\mathcal{P}_j^{E,V}$ — operator for extension of a vector in $\mathbb{R}^{\mathbf{d}}$ by zero to a vector in V, i.e. $\mathcal{P}_j^{E,V}(a) = \hat{v}$ for $a \in \mathbb{R}^{\mathbf{d}}$, where $\hat{v} \in V$ is such that $\hat{v}_j = a$ and $\hat{v}_k = \mathbf{0}$ for $k \neq j$ and where $\mathbf{0}$ is the zero vector in $\mathbb{R}^{\mathbf{d}}$.

By definition, $\mathcal{P}_j^{R,V} : V \to \mathbb{R}^{\mathbf{d}}$ and $\mathcal{P}_j^{E,V} : \mathbb{R}^{\mathbf{d}} \to V$. Concerning the rest of the above operators, in general, their domain and range spaces can be chosen in various ways. In the present chapter, we understand operators $\mathcal{P}^{R,\Omega}$ and $\mathcal{P}^{E,\Omega}$ as $\mathcal{P}^{R,\Omega} \colon L^2(\mathbb{R}^{\mathbf{d}}) \to L^2(\Omega)$ and $\mathcal{P}^{E,\Omega} \colon L^2(\Omega) \to L^2(\mathbb{R}^{\mathbf{d}})$, the operator $\mathcal{P}^{R,T_0}$ as $\mathcal{P}^{R,T_0} \colon L^2(Q_T) \to L^2(Q_T^{T_0})$, the operator $\mathcal{P}^i_{Q_T}$ as $\mathcal{P}^i_{Q_T} \colon L^2(Q_T) \to L^2(Q_T)$ and the operator $\mathcal{P}^i_T$ as $\mathcal{P}^i_T \colon L^2(0,T) \to L^2(0,T)$. This requires understanding the above definitions in the „almost everywhere" sense which involves acting on the equivalence classes of functions in the relation of being equal almost everywhere instead of acting on functions themselves.

Besides the above preliminaries, the present chapter utilizes theory concerning differentiability in Banach spaces, properties of the Nemytskii operators and properties of translation operators. The required material is contained in Appendix A.1, Appendix A.3 and Appendix A.4, respectively. In particular, Appendix A.1 introduces the notion of the weak sequential directional derivative, which will be necessary in the present chapter and which is probably not common in the literature.

In the present chapter, for a given $F \colon \mathbb{R}^n \to \mathbb{R}$, $n \in \mathbb{N} \setminus \{0\}$, the associated translation operator $\mathcal{T}_F$ is always understood as $\mathcal{T}_F \colon \mathbb{R}^n \to L^2(\mathbb{R}^n)$.

3.1 State operators

Below, we will precisely define and formulate properties of two operators: 1) the operator $\mathcal{S}$, assigning the weak solution of (0.1) - (0.3) to a given control $(g_j, h_k, \alpha_{jk})_{j=1,\dots,J}^{k=1,\dots,K}$ and 2) the operator $\mathcal{Z}$, assigning the weak solution of (3.1) - (3.2) to a given control parameter $x_1, \dots, x_J \in \mathbb{R}^{\mathbf{d}}$. Since the idea of both $\mathcal{S}$ and $\mathcal{Z}$ is to assign a realization of the process to given data, both of these operators will be called *state operators.*

The state operator $\mathcal{Z}$ will be utilized in the analysis of the optimal targeting problem, in Section 3.2. For this reason, we need to have some information about the properties of $\mathcal{Z}$. The properties which will be necessary in Section 3.2, are continuity and differentiability properties of $\mathcal{Z}$. Both of them will be investigated below.

Nevertheless, the operator $\mathcal{S}$ also is helpful because, as we will see, it can be used to conclude certain informations about $\mathcal{Z}$. Thus, we start with precise definition and Lipschitz continuity of $\mathcal{S}$. This is done in Section 3.1.1. There included material is brief — the Lipschitz continuity of $\mathcal{S}$ is a simple conclusion of theorems presented in Section 1.2.2, concerning the stability result in the space X^2. However, we show that the Lipschitz continuity of $\mathcal{S}$ with values in X^2 implies also the Lipschitz continuity of $\mathcal{S}$ with values in the space X^{3,p_2}, with suitably chosen $p_2 > 2$.

In Section 3.1.2 and Section 3.1.3, we will focus on the operator $\mathcal{Z}$. In Section 3.1.2, we present precise definition of $\mathcal{Z}$. Moreover, we briefly indicate conditions under which $\mathcal{Z}$ inherits the Lipschitz continuity property of $\mathcal{S}$. Next, in Section 3.1.3, we proceed to investigating the differentiability of $\mathcal{Z}$. This differentiability will be shown to hold in sense of weak Gâteaux differentiability. Proving this will rely on the Lipschitz continuity of $\mathcal{Z}$, thus the conditions required in Section 3.1.2 for the Lipschitz continuity are required also in Section 3.1.3 for the weak Gâteaux differentiability.

3.1.1 Control-to-state operator — definition and continuity

We define *the state operator*

$$\mathcal{S} = (\mathcal{S}_y, \mathcal{S}_{\kappa_1}, \ldots, \mathcal{S}_{\kappa_J}) \colon\ U \ \longrightarrow\ X^2$$

as the operator assigning to a given control $\hat{u} \in U$ the weak solution of the system (0.1) - (0.3) corresponding to $g_j := \hat{u}_{g_j}$, $h_k := \hat{u}_{h_k}$ and $\alpha_{jk} := \hat{u}_{\alpha_{jk}}$ in the subject system.

Below, we justify briefly that $\mathcal{S}$ is well posed and Lipschitz continuous, in suitable spaces. These properties of $\mathcal{S}$ will be required in Section 3.1.2.

It follows straight that under assumptions of Corollary 1.2.8 or Corollary 1.2.9, $\mathcal{S}(\hat{u})$ is well defined, for an arbitrary $\hat{u} \in U$. In addition, Theorem 1.2.6 allows to conclude the Lipschitz continuity of $\mathcal{S}$, under suitable assumptions. We summarize these observations in the following theorem:

Theorem 3.1.1 *In the system (0.1) - (0.3), let assumptions (B-1) - (B-5) and at least one of the following:*

- *y^* fulfills the assumption (C-1) and functions w_k are bounded for $k = 1, \ldots, K$,*
- *y^* fulfills the assumption (C-2)*

be fulfilled. Then, the operator $\mathcal{S} \colon U \to X^2$ is well defined and Lipschitz continuous on bounded subsets of U, with respect to the norms of the considered spaces.

In the sequel, we will need to have the Lipschitz continuity of $\mathcal{S}$ in a space different than X^2, what is the subject of the next theorem.

Theorem 3.1.2 *Let the assumptions of Theorem 3.1.1 be fulfilled. Assume also that p_2 is as in the part a) of the assumption (E-3). Then the operator $\mathcal{S}$ understood as*

$$\mathcal{S} \colon\ U \ \longrightarrow\ X^{3,p_2}$$

is well defined and is Lipschitz continuous on bounded subsets of U, with respect to the norms of the considered spaces.

Theorem 3.1.2 is a direct consequence of Theorem 3.1.1 and the below lemma:

Lemma 3.1.3 *Assume that p_2 is as in the part a) of the assumption (E-3). Then, $X^2 \subseteq X^{3,p_2}$ and $X^2 \hookrightarrow X^{3,p_2}$.*

Proof. By definition of X^{3,p_2}, to justify the demanded inclusion and continuous embedding, it is enough to verify that

$$\begin{aligned} &L^2(0,T;H^1(\Omega)) \cap L^\infty(0,T;L^2(\Omega)) \subseteq L^{p_2}(Q_T) \\ &L^2(0,T;H^1(\Omega)) \cap L^\infty(0,T;L^2(\Omega)) \hookrightarrow L^{p_2}(Q_T) \end{aligned} \tag{3.4}$$

Take p_1 and p_2 as in the part a) of the assumption (E-3). Then

$$\begin{aligned} L^\infty(0,T;L^2(\Omega)) \cap L^2(0,T;L^{p_1}(\Omega)) \ &\subseteq\ L^{p_2}(0,T;L^{p_2}(\Omega)) \\ \|y\|_{p_2,p_2} \ &\leq\ C_1 \max\left\{\tfrac{1}{q}, \tfrac{q-1}{q}\right\} \left(\|y\|_{2,\infty} + \|y\|_{p_1,2}\right) \end{aligned} \tag{3.5}$$

for certain constant $C_1 = C(p_1, p_2, \Omega)$. Indeed, by the Hölder inequality:

$$\begin{aligned}
\|y\|_{p_2,p_2}^{p_2} &= \int_0^T \int_\Omega |y|^{p_2-2}|y|^2\,dx\,dt \\
\leq & \int_0^T \Big(\int_\Omega |y|^{p_2-2\frac{p_1}{p_1-2}}\,dx\Big)^{\frac{p_1-2}{p_1}} \Big(\int_\Omega |y|^{2\frac{p_1}{2}}\,dx\Big)^{\frac{2}{p_1}}\,dt \\
\leq & \sup_{[0,T]} \|y\|^{p_2-2}_{\frac{p_1(p_2-2)}{p_1-2}} \int_0^T \|y\|^2_{p_1}\,dt \\
\leq & \, C_1 \|y\|^{p_2-2}_{2,\infty} \|y\|^2_{p_1,2}
\end{aligned}$$

where we have used the fact that the Hölder conjugate of $\frac{p_1}{2}$ is $\frac{p_1}{p_1-2}$ and that $L^{\frac{p_1(p_2-2)}{p_1-2}}(\Omega) \subseteq L^2(\Omega)$ since by the assumptions it can be verified that $\frac{p_1(p_2-2)}{p_1-2} \leq 2$. The constant C_1 is the constant appearing in estimation of the $L^{\frac{p_1(p_2-2)}{p_1-2}}(\Omega)$ norm by the $L^2(\Omega)$ norm, hence $C_1 = C_1(p_1, p_2, \Omega)$. This justifies the inclusion in (3.5).

Now, still having the assumptions for p_1 and p_2 in mind, we can estimate the right hand side by the Young inequality, taking an arbitrary exponent $1 < q < \infty$:

$$\|y\|^{p_2}_{p_2,p_2} \leq C_1 \|y\|^{(p_2-2)/p_2}_{2,\infty} \|y\|^{2/p_2}_{p_1,2} \leq C_1 \left(\frac{1}{q}\|y\|^{\frac{p_2-2}{p_2}q}_{2,\infty} + \frac{q-1}{q}\|y\|^{\frac{2}{p_2}\frac{q}{q-1}}_{p_1,2} \right)$$

since the Hölder conjugate of q is $\frac{q}{q-1}$. Let us set $q = \frac{p_2}{p_2-2}$ or, equivalently, $p_2 = \frac{2q}{q-1}$. Then both exponents appearing in the right hand side of the above reduce: $\frac{p_2-2}{p_2}q = 1$ and $\frac{2}{p_2}\frac{q}{q-1} = 1$. Hence the inequality in (3.5).

Moreover, for p_1 as in the part a) of the assumption (E-3), we have

$$\begin{aligned}
L^2(0,T;H^1(\Omega)) &\subseteq L^2(0,T;L^{p_1}(\Omega)) \\
\|\cdot\|_{p_1,2} &\leq C_2\|\cdot\|_{H^1(\Omega),2}
\end{aligned} \tag{3.6}$$

where $C_2 = C_2(p_1, \mathbf{d}, \Omega)$. This is straightforward by the Sobolev embedding theorem (see [1, Theorem 4.12]).

(3.5) and (3.6) together yield the inclusion and continuous embedding (3.4) for $p_2 \in (2, 4 - (4/p_1)]$, what concludes the proof. ■

3.1.2 Targeting-to-state operator — definition and continuity

We define *the state operator*

$$\mathcal{Z} = (\mathcal{Z}_y, \mathcal{Z}_{\kappa_1}, \ldots, \mathcal{Z}_{\kappa_J}) \colon V \longrightarrow X^2$$

as the operator assigning to a given control parameter $\hat{v} \in V$ the weak solution of the system (3.1) - (3.2) corresponding to $x_j := \hat{v}_j$ for $j = 1, \ldots, J$ in the subject system.

We are interested in Lipschitz continuity and weak Gâteaux differentiability of $\mathcal{Z}$. The differentiability of $\mathcal{Z}$ is the subject of Section 3.1.3. Here, we focus on the continuity matter. To deal with it, we will represent $\mathcal{Z}$ as the superposition of $\mathcal{S}$ with certain other operator. This kind of representation immediately allows to see that continuity properties of $\mathcal{Z}$ depend strongly on continuity properties of $\mathcal{S}$.

Assuming that (2.1) holds and that pattern functions $\sigma_g, \sigma_h : \mathbb{R}^{\mathbf{d}} \to \mathbb{R}$ in (2.2) are given, we define the operator

$$\Upsilon = \left(\Upsilon_{g_j}, \Upsilon_{h_j}, \Upsilon_{\alpha_{j,k}}\right)_{j,k=1,\ldots,J} : V \longrightarrow U$$

by the following relations:

$$\begin{array}{lclcl} (\Upsilon(\hat{v}))_{g_j} & := & \Upsilon_{g_j}(\hat{v}) & := & \mathcal{P}^{R,\Omega}\mathcal{T}_{\sigma_g}(x_j) \\ (\Upsilon(\hat{v}))_{h_j} & := & \Upsilon_{h_j}(\hat{v}) & := & \mathcal{P}^{R,\Omega}\mathcal{T}_{\sigma_h}(x_j) \\ (\Upsilon(\hat{v}))_{\alpha_{j,k}} & := & \Upsilon_{\alpha_{j,k}}(\hat{v}) & := & \delta_{j,k} \end{array} \tag{3.7}$$

for $j, k = 1, \ldots, J$, where $\delta_{j,k}$ is defined as in *Notation conventions*. We recall that, in the present chapter, the particular operators above are understood as $\mathcal{T}_{\sigma_g}, \mathcal{T}_{\sigma_h} : \mathbb{R}^{\mathbf{d}} \to L^2(\mathbb{R}^{\mathbf{d}})$ and $\mathcal{P}^{R,\Omega} : L^2(\mathbb{R}^{\mathbf{d}}) \to L^2(\Omega)$. Due to (3.7), the operator Υ is fully determined by the choice of σ_g and σ_h. The operator Υ can be understood as an operator assigning a control to a given control parameter.

To conclude properties of the operator $\mathcal{Z}$, it first will be useful to know how properties of σ_g and σ_h are related with properties of the operator Υ, which definition depends on σ_g and σ_h. Informations concerning these relations are summarized in the below lemma:

Lemma 3.1.4 *The following implications are true:*

a) if $\sigma_g, \sigma_h \in L^2(\mathbb{R}^{\mathbf{d}})$, then operators $\Upsilon_{g_j} : V \to L^2(\Omega)$, $\Upsilon_{h_j} : V \to L^2(\Omega)$ and $\Upsilon_{\alpha_{j,k}} : V \to \mathbb{R}$, for $j, k = 1, \ldots, J$, are well-defined and continuous and hence so $\Upsilon : V \to U$ is,

b) if $\sigma_g, \sigma_h \in W^{1,2}(\mathbb{R}^{\mathbf{d}})$, then operators $\Upsilon_{g_j} : V \to L^2(\Omega)$, $\Upsilon_{h_j} : V \to L^2(\Omega)$ and $\Upsilon_{\alpha_{j,k}} : V \to \mathbb{R}$, for $j, k = 1, \ldots, J$, are Lipschitz continuous (globally) and hence so $\Upsilon : V \to U$ is,

c) if $\sigma_g, \sigma_h \in W^{1,2}(\mathbb{R}^{\mathbf{d}})$, then operators $\Upsilon_{g_j} : V \to L^2(\Omega)$, $\Upsilon_{h_j} : V \to L^2(\Omega)$ and $\Upsilon_{\alpha_{j,k}} : V \to \mathbb{R}$, for $j, k = 1, \ldots, J$, are weakly Gâteaux differentiable and hence so $\Upsilon : V \to U$ is.

Proof. It is straightforward that operators $\Upsilon_{\alpha_{j,k}}$ are well-defined, Lipschitz continuous, weak Gâteaux differentiable. We are left to deal with the remaining operators Υ_{g_j} and Υ_{g_j}.

The operators Υ_{g_j} and Υ_{g_j}, for $j = 1, \ldots, J$, can be expressed as

$$\Upsilon_{g_j} = \mathcal{P}^{R,\Omega} \circ \mathcal{T}_{\sigma_g} \circ \mathcal{P}_j^{R,V}, \qquad \Upsilon_{g_j} = \mathcal{P}^{R,\Omega} \circ \mathcal{T}_{\sigma_h} \circ \mathcal{P}_j^{R,V} \tag{3.8}$$

Operators $\mathcal{P}^{R,\Omega} : L^2(\mathbb{R}^{\mathbf{d}}) \to L^2(\Omega)$ and $\mathcal{P}_j^{R,V} : V \to \mathbb{R}^{\mathbf{d}}$ are linear and continuous. Thus, the question on the properties of Υ_{g_j} and Υ_{h_j}, for $j = 1, \ldots, J$, reduces in its most essential part to the question on the properties of $\mathcal{T}_{\sigma_g}$.

Operators $\mathcal{P}^{R,\Omega}$ and $\mathcal{P}_j^{R,V}$ are well defined in respective spaces, for $j = 1, \ldots, J$. Moreover, for an arbitrary $\sigma_g \in L^2(\mathbb{R}^{\mathbf{d}})$, translation operators $\mathcal{T}_{\sigma_g}$ and $\mathcal{T}_{\sigma_h}$ are well defined from $\mathbb{R}^{\mathbf{d}}$ to $L^2(\mathbb{R}^{\mathbf{d}})$. Thus, by (3.8), Υ_{g_j} and Υ_{h_j}, for $j = 1, \ldots, J$, are well defined.

For an arbitrary $\sigma_g \in L^2(\mathbb{R}^{\mathbf{d}})$, the translation operator $\mathcal{T}_{\sigma_g} : \mathbb{R}^{\mathbf{d}} \to L^2(\mathbb{R}^{\mathbf{d}})$ is continuous (see Theorem A.4.2). This, together with (3.8) and the continuity of $\mathcal{P}^{R,\Omega}$ and $\mathcal{P}_j^{R,V}$, gives the continuity of Υ_{g_j} for $j = 1, \ldots, J$. Analogous argument holds for operators Υ_{h_j}, for $j = 1, \ldots, J$.

For $\sigma_g \in W^{1,2}(\mathbb{R}^{\mathbf{d}})$, the translation operator $\mathcal{T}_{\sigma_g} : \mathbb{R}^{\mathbf{d}} \to L^2(\mathbb{R}^{\mathbf{d}})$ is Lipschitz continuous (see Theorem A.4.4). Moreover, operators $\mathcal{P}^{R,\Omega}$ and $\mathcal{P}_j^{R,V}$, as linear and continuous operators, are Lipschitz continuous for $j = 1, \ldots, J$. Hence, by (3.8), Υ_{g_j} is so, for $j = 1, \ldots, J$. Similarly, $\sigma_h \in W^{1,2}(\mathbb{R}^{\mathbf{d}})$ implies Lipschitz continuity of Υ_{h_j}, for $j = 1, \ldots, J$.

Also, for $\sigma_g \in W^{1,2}(\mathbb{R}^{\mathbf{d}})$, Theorem A.4.5 gives weak Gâteaux differentiability of $\mathcal{T}_{\sigma_g}\colon \mathbb{R}^{\mathbf{d}} \to L^2(\mathbb{R}^{\mathbf{d}})$. Therefore, by (3.8) and by the rules for differential calculus in Banach spaces (see Theorem A.1.4, Observation A.1.7 and Observation A.1.11 in Appendix A.1), Υ_{g_j} is weakly Gâteaux differentiable for $\sigma_g \in W^{1,2}(\mathbb{R}^{\mathbf{d}})$. Analogously, Υ_{h_j} is weakly Gâteaux differentiable for $\sigma_h \in W^{1,2}(\mathbb{R}^{\mathbf{d}})$. ■

Now, we proceed to investigating properties of the state operator $\mathcal{Z}$. Note that, under the assumption that (2.1) holds, the weak solution of (3.1) - (3.2) is exactly the weak solution of (0.1) - (0.3) associated with $g_j := \Upsilon_{g_j}(\hat{v})$, $h_j := \Upsilon_{h_k}(\hat{v})$ and $\alpha_{j,k} := \Upsilon_{\alpha_{j,k}}(\hat{v})$. Hence,

$$\mathcal{Z} = \mathcal{S} \circ \Upsilon$$

In particular, the properties of $\mathcal{Z}$ are determined by properties of $\mathcal{S}$ and Υ.

Having made the above observation, Lemma 3.1.4 together with Theorems 3.1.1 and 3.1.2 allow to justify the below:

Theorem 3.1.5 *In the system (3.1) - (3.2), let assumptions (B-1) - (B-5) be fulfilled, with additional restriction $K = J$. Assume also that at least one of the following is true:*

- *y^* fulfills the assumption (C-1) and functions w_j are bounded, for $j = 1, \ldots, J$,*
- *y^* fulfills the assumption (C-2).*

Then, the following statements are true:

a) if σ_g, σ_h fulfill the assumption (F-1), then $\mathcal{Z}\colon V \to X^2$ is well defined and continuous,

b) if σ_g, σ_h fulfill the assumption (F-3), then $\mathcal{Z}\colon V \to X^2$ is in addition Lipschitz continuous (globally).

Moreover, let p_2 be as in the part a) of the assumption (E-3). Then, the above statements hold also with X^2 replaced by X^{3,p_2}.

Remark. Note, that Theorem 3.1.5 in particular asserts that the weak solution of (3.1) - (3.2) exists and is unique. ▲

Remark. Note, that in contrary to the Lipschitz continuity on the bounded sets stated for $\mathcal{S}$ in theorems of Section 3.1.1, the Lipschitz continuity of $\mathcal{Z}$ in Theorem 3.1.5 is global. The reason for the latter is the following. $\mathcal{Z} = \mathcal{S} \circ \Upsilon$, hence, for an arbitrary subset $\mathbb{A}$ of V, the Lipschitz constant of $\mathcal{Z}$ on $\mathbb{A}$ is lesser on equal to product of Lipschitz constant of Υ on $\mathbb{A}$ and Lipschitz constant of $\mathcal{S}$ on $\Upsilon(\mathbb{A})$. The Lipschitz constant of Υ is global (see Lemma 3.1.4). Moreover, for all $\hat{v} \in V$, the corresponding control $\hat{u} = \Upsilon(\hat{v})$ belongs to a ball $B_U(0, r_\sigma)$ in U, with radius r_σ depending only on $\left\|\sigma_g\right\|_{2,\mathbb{R}^{\mathbf{d}}}$ and $\left\|\sigma_h\right\|_{2,\mathbb{R}^{\mathbf{d}}}$. By Theorems 3.1.1 and 3.1.2, $\mathcal{S}$ is Lipschitz continuous on $B_U(0, r_\sigma)$. Thus, we can take $\mathbb{A} = B(0, r_\sigma)$ to justify the global Lipschitz continuity of $\mathcal{Z}$. ▲

3.1.3 Targeting-to-state operator — differentiability

Now, we will focus on the matter of weak Gâteaux differentiability of the operator $\mathcal{Z}$ understood as an operator from V to X^1. $\mathcal{Z}$ is certainly well defined in this sense, because $X^2 \subseteq X^1$. Nevertheless, investigating differentiability of $\mathcal{Z}\colon V \to X^1$ involves longer justification.

We will begin with presenting an auxiliary system of equations, which we call *the linearized system* and justifying some basic properties of the subject system. Next, we will formulate the main theorem of the present section, i.e. theorem concerning weak Gâteaux differentiability of $\mathcal{Z}$. This theorem, as well as its proof, involves strongly the linearized system, therefore the linearized system is essential for the present section.

Let us start. The below system, which we call *the linearized system*, will be utilized later for characterizing the weak Gâteaux differential of $\mathcal{Z}$:

$$\begin{cases} y_t - D\Delta y - f'(\hat{y})y = \sum_{j=1}^{J} \Upsilon_{g_j}(\hat{v})\kappa_j + \sum_{j=1}^{J} D_{G,w}\Upsilon_{g_j}(\hat{v})(\hat{\eta})\hat{\kappa}_j & \text{on } Q_T \\ \dfrac{\partial y}{\partial n} = 0 & \text{on } \partial\Omega \times (0,T) \\ y(x,0) \equiv 0 & \text{for } x \in \Omega \end{cases} \tag{3.9}$$

together with

$$\begin{cases} \beta_1\kappa_1' + \kappa_1 = w_1'\big(\big(\Upsilon_{h_1}(\hat{v}), \hat{y} - y^*\big)_{L^2(\Omega)}\big)\cdot \\ \qquad \cdot\Big(\big(D_{G,w}\Upsilon_{h_1}(\hat{v})(\hat{\eta}), \hat{y} - y^*\big)_{L^2(\Omega)} + \big(\Upsilon_{h_1}(\hat{v}), y\big)_{L^2(\Omega)}\Big) & \text{on } [0,T] \\ \vdots & \vdots \\ \beta_J\kappa_J' + \kappa_J = w_J'\big(\big(\Upsilon_{h_J}(\hat{v}), \hat{y} - y^*\big)_{L^2(\Omega)}\big)\cdot \\ \qquad \cdot\Big(\big(D_{G,w}\Upsilon_{h_J}(\hat{v})(\hat{\eta}), \hat{y} - y^*\big)_{L^2(\Omega)} + \big(\Upsilon_{h_J}(\hat{v}), y\big)_{L^2(\Omega)}\Big) & \text{on } [0,T] \\ \kappa_j(0) = 0 & \text{for } j = 1, \dots, J \end{cases} \tag{3.10}$$

where: Ω is a domain, $T > 0$, $Q_T := \Omega \times (0,T)$; $D, \beta_j > 0$; $f, w_j \colon \mathbb{R} \to \mathbb{R}$; $\hat{\kappa}_j \colon (0,T) \to \mathbb{R}$; $\hat{y}, y^* \colon Q_T \to \mathbb{R}$; $\hat{v}, \hat{\eta} \in V$; Υ_{g_j} and Υ_{h_j} correspond to given $\sigma_g, \sigma_h \colon \mathbb{R}^{\mathbf{d}} \to \mathbb{R}$ (see (3.7) for the explanation of the latter correspondence); where $j = 1, \dots, J$. In the system (3.9) - (3.10), the unknown is the function $(y, \kappa_1, \dots, \kappa_J) \colon Q_T \to \mathbb{R}^{J+1}$.

The system (3.9) - (3.10) is a particular case of the system (1.84) - (1.86) in Section 1.2.4, with

$$\begin{array}{rclrcl} \widetilde{g}_j & := & D_{G,w}\Upsilon_{g_j}(\hat{v})(\hat{\eta}), & \mathbf{h}_j & := & \Upsilon_{h_j}(\hat{v}), \\ \widetilde{h}_j & := & D_{G,w}\Upsilon_{h_j}(\hat{v})(\hat{\eta}), & \mathbf{Z}_j & := & w_j'\big(\big(\Upsilon_{h_j}(\hat{v}), \hat{y} - y^*\big)_{L^2(\Omega)}\big) \\ \mathbf{Y} & := & \hat{y} - y^*, & \widetilde{y}_0(x) & := & 0, \\ \Theta_j(x,t) & := & \hat{\kappa}_j(t), & \widetilde{\kappa}_{j0} & := & 0, \\ \Xi_j(x,t) & := & \Upsilon_{g_j}(\hat{v}), & \widetilde{f}(x,t,s) & := & f'(\hat{y}(x,t))s, \\ & & & \widetilde{w}_j(s) & := & s, \end{array} \tag{3.11}$$

for $j = 1, \dots, J$, $x \in \Omega$, $t \in (0,T)$, $s \in \mathbb{R}$. Hence the below definition:

Definition 3.1.6 $(y, \kappa_1, \dots, \kappa_J) \in X^2$ *is a weak solution of (3.9) - (3.10) if it is a weak solution of (1.84) - (1.86) with conditions (3.11) (see Definition 1.2.16).*

The following lemma summarizes those properties of (3.9) - (3.10) which will be necessary for us in the sequel.

Lemma 3.1.7 *Let assumptions (B-1) - (B-4) be fulfilled, with additional restriction $K = J$. Let also assumptions (E-1) - (E-2) and (F-2) hold. Moreover, assume that $\hat{y}, y^* \in L^2(0,T; L^2(\Omega))$ and $\kappa_j \in L^\infty(0,T)$, for $j = 1, \dots, J$.*

Then, the weak solution of the system (3.9) - (3.10) exists, is unique and moreover belongs to X^{3,p_2}, for p_2 as in the assumption (E-3), for arbitrary $\hat{v}, \hat{\eta} \in V$. In addition, for a given $\hat{v} \in V$, the operator assigning the weak solution of (3.9) - (3.10) to $\hat{\eta} \in V$ belongs to $L(V, X^1)$, to $L(V, X^2)$ and to $L(V, X^{3,p_2})$ with p_2 as in the assumption (E-3).

Proof. We will verify that the functions defined by relations (3.11) fulfill assumptions (D-1) - (D-6) from Section 1.2.4.

First, f' is a Borel measurable function, as the classical derivative of a continuous function (see assumptions (B-4) and (E-1)). Thus composition of f' with the measurable function $\hat{y}$ is measurable. Hence, $\widetilde{f}$ in (3.11) is measurable in $(x,t) \in Q_T$ for an arbitrary $s \in \mathbb{R}$. Moreover, f' is bounded (by the assumption (B-3)), hence $\widetilde{f}$ is Lipschitz continuous in s, with the same constant for every $(x,t) \in Q_T$. Also, function $\widetilde{f}(\,.\,,\,.\,,0)$ belongs to $L^2(Q_T)$. Therefore, $\widetilde{f}$ defined in (3.11) fulfills the assumption (D-3) in Section 1.2.4.

Next, $(\Upsilon_{h_j}(\hat{v}), \hat{y} - y^*)_{L^2(\Omega)}$, understood as a function of variable t, is measurable. To see this, note that this function can be understood as a composition of a strongly measurable function $\hat{y} - y^*$, from $[0,T]$ to $L^2(\Omega)$, with a continuous linear functional on $L^2(\Omega)$ given by $\Upsilon_{h_j}(\hat{v})$ and apply the Pettis theorem (see [3, Th. 1.1.1], [21, App. E.5], [49, Chap. V.4] or [52, p. 1012]; [21] and [52] do not contain the proof of the theorem). The function w_j' is Borel measurable, as a classical derivative of a continuous function (see assumptions (B-5) and (E-2)). Thus, the composition of w_j' with a measurable function is measurable, for $j = 1, \ldots, J$. Moreover, w_j' is bounded (by the assumption (B-4)), for $j = 1, \ldots, J$. Hence, for $j = 1, \ldots, J$, $\mathbf{Z}_j$ defined in (3.11) is an element of $L^\infty(0,T)$ and as such, obeys the assumption (D-6) in Section 1.2.4.

The observation that, for $j = 1, \ldots, J$, $\mathbf{Y}$, Θ_j, Ξ_j, $\mathbf{h}_j$, $\widetilde{y}_0$, $\widetilde{\kappa}_{j0}$ and $\widetilde{w}_j$ defined in (3.11) obey assumptions (D-4), (D-5), (D-6) and (C-2), respectively, follows straight. Moreover, by the assumption (F-2) and Lemma 3.1.4, $\widetilde{g}_j$ and $\widetilde{h}_j$ in (3.11) belong to $L^2(\Omega)$, for $j = 1, \ldots, J$. Hence, $(\widetilde{g}_j, \widetilde{h}_j)_{j=1,\ldots,J} \in \widetilde{U}$.

Therefore, the system (3.9) - (3.10) fulfills the assumptions of Theorems 1.2.17 and 1.2.18 in Section 1.2.4. By Theorem 1.2.18, we conclude that the weak solution of (3.9) - (3.10) exists in X^2 and is unique. In addition, for p_2 as assumed, $X^2 \subseteq X^{3,p_2}$, $X^2 \hookrightarrow X^{3,p_2}$ (see Lemma 3.1.3). Hence, the weak solution of (3.9) - (3.10) belongs also to X^{3,p_2}. Moreover:

- by the definition of weak Gâteaux differential, the operator

$$\hat{\eta} \mapsto (D_{G,w}\Upsilon_{g_1}(\hat{v})(\hat{\eta}), \ldots, D_{G,w}\Upsilon_{g_J}(\hat{v})(\hat{\eta}), D_{G,w}\Upsilon_{h_1}(\hat{v})(\hat{\eta}), \ldots, D_{G,w}\Upsilon_{h_J}(\hat{v})(\hat{\eta})) =: \hat{u}^{\hat{v},\hat{\eta}}$$

  is linear and bounded from V to $\widetilde{U}$,

- by the structure of (3.9) - (3.10) and by Theorem 1.2.17, the operator assigning the weak solution of (3.9) - (3.10) to a given element $\hat{u}^{\hat{v},\hat{\eta}}$ is linear and bounded from $\widetilde{U}$ to X^2.

Hence, the operator assigning the weak solution of (3.9) - (3.10) to a given $\hat{\eta} \in V$, as the superposition of the above operators, is linear and bounded from V to X^2. Since $X^1 \subseteq X^2$, $X^1 \hookrightarrow X^2$, the subject operator is also linear and bounded from V to X^1. Moreover, since $X^2 \subseteq X^{3,p_2}$ and $X^2 \hookrightarrow X^{3,p_2}$, the subject operator is linear and bounded from V to X^{3,p_2}. ■

Now, we formulate the main theorem of Section 3.1.3:

Theorem 3.1.8 *In the system (3.1) - (3.2), let assumptions (B-1) - (B-5) be fulfilled, with additional restriction $K = J$. Assume also that at least one of the following is true:*

- *y^* fulfills the assumption (C-1) and functions w_j are bounded, for $j = 1, \ldots, J$,*

- *y^* fulfills the assumption (C-2).*

Moreover, let assumptions (E-1) - (E-3) and (F-2) be fulfilled.

Then, the operator $\mathcal{Z}$ understood as

$$\mathcal{Z} : V \longrightarrow X^1$$

is well defined and weakly Gâteaux differentiable. Moreover, the value of the weak Gâteaux differential of $\mathcal{Z}$ in a point $\hat{v} \in V$ applied to a direction $\hat{\eta} \in V$, i.e. the value $D_{G,w}\mathcal{Z}(\hat{v})(\hat{\eta})$, can be identified with the element $(\widetilde{y}, \widetilde{\kappa}_1, \ldots, \widetilde{\kappa}_J) \in X^1$ which is the weak solution to the system (3.9) - (3.10) with conditions $\hat{y} = \mathcal{Z}_y(\hat{v})$ and $\hat{\kappa}_j = \mathcal{Z}_{\kappa_j}(\hat{v})$.

Remark. In the assumptions of the above theorem, the assumption (E-3) is not necessary if y^* fulfills the assumption (C-2). But if y^* fulfills the assumption (C-1) only, then the assumption (E-3) is essential. ▲

Remark. Note that, under assumptions of Theorem 3.1.8, Lemma 3.1.7 can be applied. Hence, the element $(\widetilde{y}, \widetilde{\kappa}_1, \ldots, \widetilde{\kappa}_J) \in X^1$ in Theorem 3.1.8 is well defined. Moreover, as Lemma 3.1.7 states, for a given $\hat{v}$, the operator assigning $(\widetilde{y}, \widetilde{\kappa}_1, \ldots, \widetilde{\kappa}_J) \in X^1$ to $\hat{\eta} \in V$, denote it $\widehat{\mathcal{Z}}^{\hat{v}} \colon V \to X^1$, is linear and bounded. Hence indeed, the operator $\widehat{\mathcal{Z}}^{\hat{v}}$ is meaningful as the weak Gâteaux differential of $\mathcal{Z}$ in point $\hat{v}$. Therefore, Theorem 3.1.8, asserting in fact that $D_{G,w}\mathcal{Z}(\hat{v})(\hat{\eta}) = \widehat{\mathcal{Z}}^{\hat{v}}(\hat{\eta})$ for all $\hat{\eta} \in V$, makes sense. ▲

Remark. Note also, that equality $D_{G,w}\mathcal{Z}(\hat{v})(\hat{\eta}) = \widehat{\mathcal{Z}}^{\hat{v}}(\hat{\eta})$ for all $\hat{\eta} \in V$, where $\widehat{\mathcal{Z}}^{\hat{v}}$ is as above, explains why we call the system (3.9) - (3.10) the linearized system. ▲

The following observation will be useful in the proof of Theorem 3.1.8:

Lemma 3.1.9 *Let Banach spaces X, Y and an operator $T : X \to Y$, point $\hat{u} \in X$ and direction $\hat{v} \in X$ be given. Assume that T is Lipschitz continuous and Y is reflexive. Consider collection $\mathbb{E}$ of all sequences $\{\varepsilon_n\}_{n=1}^{\infty}$ such that $\varepsilon_n \neq 0$, $\varepsilon_n \to 0$ for $n \to \infty$ and the difference quotients ${\varepsilon_n}^{-1}(T(\hat{u} + \varepsilon_n \hat{v}) - T(\hat{u}))$ are weakly convergent to some limit in Y for $n \to \infty$. Assume that this limit is independent of the choice of $\{\varepsilon_n\}_{n=1}^{\infty} \in \mathbb{E}$, or more precisely, that there exists $\mathbb{L} \in Y$ such that*

$$\{\varepsilon_n\}_{n=1}^{\infty} \in \mathbb{E} \quad \Longrightarrow \quad \frac{T(\hat{u} + \varepsilon_n \hat{v}) - T(\hat{u})}{\varepsilon_n} \overset{n\to\infty}{\rightharpoonup} \mathbb{L}$$

Then $\delta_w T(\hat{u}; \hat{v})$ exists and equals $\mathbb{L}$.

To our knowledge, results of the above type are rarely formulated in the literature on PDEs. We have derived the below simple proof by our own considerations. The proof is not technically complex, thus the result is probably not new. However, we do not known a literature reference for citing here.

Proof. For brevity, for a given $\varepsilon \neq 0$ denote $T^{\varepsilon}(\hat{u}; \hat{v}) := \varepsilon^{-1}(T(\hat{u} + \varepsilon \hat{v}) - T(\hat{u}))$. Let $\widetilde{\mathbb{E}}$ denote the collection of all real sequences $\{\varepsilon_n\}_{n=1}^{\infty}$ such that $\varepsilon_n \neq 0$, $\varepsilon_n \to 0$ for $n \to \infty$. Establishing equality $\widetilde{\mathbb{E}} = \mathbb{E}$ will conclude the proof, since, under the axiom of choice, Cauchy and Heine limit of a function definitions are equivalent in metric spaces. The inclusion $\widetilde{\mathbb{E}} \supseteq \mathbb{E}$ follows straight. The inclusion $\widetilde{\mathbb{E}} \subseteq \mathbb{E}$ can be justified as follows.

First, $\bigcup \mathbb{E}$ contains a set $(-\bar{\varepsilon}, \bar{\varepsilon}) \setminus \{0\}$, for some $\bar{\varepsilon} > 0$. It comes by contradiction: if not, then, by the axiom of choice, there exists $\widetilde{\epsilon} \in \widetilde{\mathbb{E}}$, $\widetilde{\epsilon} = \{\widetilde{\varepsilon}_n\}_{n=1}^{\infty}$ such that $\widetilde{\epsilon} \cap (\bigcup \mathbb{E}) = \emptyset$. But, by the Lipschitz continuity of T, the corresponding difference quotients, $T^{\widetilde{\varepsilon}_n}(\hat{u}; \hat{v})$, are bounded in

Y w.r.t. n. Thus, by reflexivity of Y, sequence $\widetilde{\epsilon}$ contains a subsequence $\widetilde{\widetilde{\epsilon}} = \{\widetilde{\widetilde{\varepsilon}}_n\}_{n=1}^{\infty}$ such that $T^{\widetilde{\widetilde{\varepsilon}}_n}(\hat{u};\hat{v})$ converges weakly in Y as $n \to \infty$. Hence, $\widetilde{\widetilde{\epsilon}} \in \mathbb{E}$, what contradicts $\widetilde{\epsilon} \cap (\bigcup \mathbb{E}) = \emptyset$.

Having this, an arbitrary sequence belonging $\widetilde{\mathbb{E}}$ consists of elements of sequences belonging to $\mathbb{E}$. The inclusion $\widetilde{\mathbb{E}} \subseteq \mathbb{E}$ will be shown once we justify that an arbitrary sequence consisting of elements of sequences belonging to $\mathbb{E}$ is still in $\mathbb{E}$. For this end, it is now enough to verify that all sequences from $\mathbb{E}$ have the same modulus of convergence, i.e. for all $\phi \in Y^*$ for all $\lambda > 0$ there exists $\gamma > 0$ such that for all $\epsilon = \{\varepsilon_n\}_{n=1}^{\infty} \in \mathbb{E}$ for all elements satisfying $\varepsilon_n < \gamma$ there holds $\left|\langle \phi, T^{\varepsilon_n}(\hat{u};\hat{v}) - \mathbb{L}\rangle_{Y^*,Y}\right| < \lambda$.

But this also comes by contradiction. If this is not true, then, by the axiom of choice, we would be able to construct a sequence $\bar{\epsilon} = \{\bar{\varepsilon}_n\}_{n=1}^{\infty}$ consisting of elements of sequences from $\mathbb{E}$ such that $\left|\langle \phi, T^{\bar{\varepsilon}_n}(\hat{u};\hat{v}) - \mathbb{L}\rangle_{Y^*,Y}\right| \geq \lambda$ for certain $\lambda > 0$ and $\phi \in Y^*$. Hence, $\bar{\epsilon}$ cannot have any weakly convergent to $\mathbb{L}$ subsequence. But this is not possible: by the Lipschitz continuity of T, the difference quotients $T^{\bar{\varepsilon}_n}(\hat{u};\hat{v})$ are bounded in Y w.r.t. n, and therefore, by reflexivity of Y, $\bar{\epsilon}$ has a subsequence $\bar{\bar{\epsilon}} = \{\bar{\bar{\varepsilon}}_n\}_{n=1}^{\infty}$ such that $T^{\bar{\bar{\varepsilon}}_n}(\hat{u};\hat{v})$ converges weakly in Y as $n \to \infty$. By assumption, the weak limit of $\bar{\bar{\epsilon}}$ equals $\mathbb{L}$, what is a contradiction. ■

Now, we are ready to proceed to the proof of the main theorem of the present section.

Proof of Theorem 3.1.8. The fact that $\mathcal{Z}$ is well defined from V to X^1 is clear by Theorem 3.1.5 and by $X^2 \subseteq X^1$, $X^2 \hookrightarrow X^1$. Concerning the differentiability matter, we will prove that, in fact, the operator $\mathcal{Z}$ is weakly Gâteaux differentiable from V to X^{3,p_2}, with p_2 as assumed. This yields the asserted differentiability from V to X^1, since $p_2 > 2$ and thus $X^{3,p_2} \subseteq X^1$, $X^{3,p_2} \hookrightarrow X^1$.

For $\varepsilon \neq 0$, denote difference quotients of $\mathcal{Z}$ in $\hat{v}$ in direction $\hat{\eta}$ as

$$\mathcal{Z}^{\varepsilon}(\hat{v};\hat{\eta}) \;:=\; \varepsilon^{-1}\left(\mathcal{Z}(\hat{v}+\varepsilon\hat{\eta}) - \mathcal{Z}(\hat{v})\right)$$

Assume that $\epsilon = \{\varepsilon_n\}_{n=1}^{\infty}$ is a sequence such that $\varepsilon_n \neq 0$, $\varepsilon_n \to 0$ for $n \to \infty$ and that the corresponding difference quotients are weakly convergent to certain $\widetilde{Z}_{\epsilon}(\hat{v};\hat{\eta}) \in X^{3,p_2}$:

$$\mathcal{Z}^{\varepsilon_n}(\hat{v};\hat{\eta}) \rightharpoonup \widetilde{Z}_{\epsilon}(\hat{v};\hat{\eta}) \quad \text{in } X^{3,p_2}, \text{ as } n \to \infty \tag{3.12}$$

Let $\mathbb{E}$ denote the collection of all sequences $\epsilon = \{\varepsilon_n\}_{n=1}^{\infty}$ satisfying the above conditions. To justify that $\mathcal{Z}$ is weakly Gâteaux differentiable from V to X^1, we need to establish that the following hypotheses hold:

(Hyp-1) $\widetilde{Z}_{\epsilon}(\hat{v};\hat{\eta}) \in X^{3,p_2}$ is independent of sequence $\epsilon \in \mathbb{E}$, i.e. there exists $\widetilde{Z}(\hat{v};\hat{\eta}) \in X^{3,p_2}$ such that $\widetilde{Z}_{\epsilon}(\hat{v};\hat{\eta}) = \widetilde{Z}(\hat{v};\hat{\eta})$ for all $\epsilon \in \mathbb{E}$.

(Hyp-2) $\widetilde{Z}(\hat{v};\,.\,)$ is a bounded linear operator from V to X^{3,p_2}.

The above two hypotheses together, if proven, imply that $\mathcal{Z}$ is weakly Gâteaux differentiable from V to X^{3,p_2}. To justify it, assume temporarily that hypotheses (Hyp-1) and (Hyp-2) hold. Having this, note that, by (Hyp-1), Lemma 3.1.9 can be applied. Indeed, X^{3,p_2} is reflexive and Banach and, by Theorem 3.1.5, $\mathcal{Z}$ is Lipschitz continuous with values in X^{3,p_2}. Therefore, since (Hyp-1) holds, all assumptions of Lemma 3.1.9 are satisfied. Thus it can be used to conclude that $\delta_w\mathcal{Z}(\hat{v};\hat{\eta})$ exists in X^{3,p_2} and equals $\widetilde{Z}(\hat{v};\hat{\eta})$. Now, if $\delta_w\mathcal{Z}(\hat{v};\,.\,)$ is linear and bounded from V to X^{3,p_2}, then it can be identified with the weak Gâteaux differential of $\mathcal{Z}\colon V \to X^{3,p_2}$, in point $\hat{v} \in V$. But the linearity and boundedness follows by the relation $\delta_w\mathcal{Z}(\hat{v};\hat{\eta}) = \widetilde{Z}(\hat{v};\hat{\eta})$ and by (Hyp-2).

Therefore, we are left to justify hypotheses (Hyp-1) and (Hyp-2). But the considered hypotheses will be straightforward once we prove that, for an arbitrary sequence $\epsilon \in \mathbb{E}$, $\widetilde{Z}_\epsilon(\hat{v};\hat{\eta})$ is the element of the space X^{3,p_2} which is the weak solution of the system (3.9) - (3.10). Indeed, by Lemma 3.1.7, the weak solution of (3.9) - (3.10) exists in X^{3,p_2} and is unique. Hence, if $\widetilde{Z}_\epsilon(\hat{v};\hat{\eta})$ is the weak solution of (3.9) - (3.10) for an arbitrary $\epsilon \in \mathbb{E}$, then (Hyp-1) holds — we can write that $\widetilde{Z}_\epsilon(\hat{v};\hat{\eta}) = \widetilde{Z}(\hat{v};\hat{\eta})$, for $\widetilde{Z}(\hat{v};\hat{\eta})$ being the weak solution of (3.9) - (3.10). Moreover, if $\widetilde{Z}(\hat{v};\hat{\eta})$ is the weak solution of (3.9) - (3.10), then Lemma 3.1.7 states that the operator $\widetilde{Z}(\hat{v};\,.\,)$ is linear and bounded from V to X^{3,p_2}. Thus, (Hyp-2) also holds. Altogether, it remains to show that $\widetilde{Z}_\epsilon(\hat{v};\hat{\eta})$ solves the system (3.9) - (3.10) for an arbitrary $\epsilon \in \mathbb{E}$ to complete the proof.

Thus fix $\epsilon := \{\varepsilon_n\}_{n=1}^\infty \in \mathbb{E}$. Since $\widetilde{Z}_\epsilon(\hat{v};\hat{\eta})$ is in fact the sequential weak directional derivative of $\mathcal{Z}$ in X^1 on sequence ϵ (see Definition A.1.9), we will use notation $\bar{\delta}_w^\epsilon \mathcal{Z}(\hat{v};\hat{\eta})$ in place of $\widetilde{Z}_\epsilon(\hat{v};\hat{\eta})$. Moreover, for convenience, denote for a given $\varepsilon \neq 0$:

$$(\widetilde{y}^\varepsilon, \widetilde{\kappa}_1^\varepsilon, \ldots, \widetilde{\kappa}_J^\varepsilon) := \mathcal{Z}^\varepsilon(\hat{v};\hat{\eta}), \quad (\widetilde{y}, \widetilde{\kappa}_1, \ldots, \widetilde{\kappa}_J) := \bar{\delta}_w^\epsilon \mathcal{Z}(\hat{v};\hat{\eta}), \quad (\hat{y}, \hat{\kappa}_1, \ldots, \hat{\kappa}_J) := \mathcal{Z}(\hat{v})$$

Consider the weak form of the system (3.1) - (3.2) (see Definition 3.0.1) corresponding to $x_j := \hat{v}_j$ and the weak form of this system corresponding to $x_j := \hat{v}_j + \varepsilon\hat{\eta}_j$, for $j = 1, \ldots, J$ and for a given $\varepsilon \neq 0$. Subtract these weak forms and divide the resulting identities by ε. By the above introduced notation, we get that $(\widetilde{y}^\varepsilon, \widetilde{\kappa}_1^\varepsilon, \ldots, \widetilde{\kappa}_J^\varepsilon)$ is an element of X^2 satisfying the following conditions:

$$\widetilde{y}^\varepsilon(\,.\,,0) \equiv 0 \text{ in } L^2(\Omega), \qquad \widetilde{\kappa}_j^\varepsilon(0) = 0 \text{ for } j = 1, \ldots, J \tag{3.13}$$

$$\begin{aligned} \int_0^T \big\langle (\widetilde{y}^\varepsilon)', \phi \big\rangle \;+\; D\big(\nabla \widetilde{y}^\varepsilon, \nabla\phi\big)_{L^2(\Omega)} \;+ \\ -\;\bigg(\frac{F(\hat{v}+\varepsilon\hat{\eta})) - F(\hat{v}))}{\varepsilon} \;+\; \sum_{j=1}^J \frac{G_j(\hat{v}+\varepsilon\hat{\eta}) - G_j(\hat{v})}{\varepsilon}\,,\,\phi\bigg)_{L^2(\Omega)} dt \;=\; 0 \end{aligned} \tag{3.14}$$

$$\int_0^T \bigg(\beta_j(\widetilde{\kappa}_j^\varepsilon)' + \widetilde{\kappa}_j^\varepsilon - \frac{H_j(\hat{v}+\varepsilon\hat{\eta}) - H_j(\hat{v})}{\varepsilon} \bigg)\, \xi \, dt \;=\; 0 \tag{3.15}$$

for all $\phi \in L^2(0,T;H^1(\Omega))$ and all $\xi \in L^2(0,T)$, and where we have utilized the following definitions:

$$\begin{aligned} F(\widetilde{v})(x,t) \quad &:= \quad f\big(\mathcal{Z}_y(\widetilde{v})(x,t)\big) & &\text{a.e. on } Q_T \\ G_j(\widetilde{v})(x,t) \quad &:= \quad \mathcal{P}^{R,\Omega}\mathcal{T}_{\sigma_g}(\widetilde{v}_j)(x)\; \mathcal{Z}_{\kappa_j}(\widetilde{v})(t) \;=\; \Upsilon_{g_j}(\widetilde{v})(x)\; \mathcal{Z}_{\kappa_j}(\widetilde{v})(t) & &\text{a.e. on } Q_T \\ H_j(\widetilde{v})(t) \quad &:= \quad w_j\Big(\int_\Omega \mathcal{P}^{R,\Omega}\mathcal{T}_{\sigma_h}(\widetilde{v}_j)(x)\; (\mathcal{Z}_y(\widetilde{v})(x,t) - y^*(x,t))dx \Big) & & \\ &= \quad w_j\Big(\int_\Omega \Upsilon_{h_j}(\widetilde{v})(x)\; (\mathcal{Z}_y(\widetilde{v})(x,t) - y^*(x,t))dx \Big) & &\text{a.e. on } (0,T) \end{aligned}$$

for $\widetilde{v} \in V$ and $j = 1, \ldots, J$.

We intend to pass to the limit in identities (3.13) - (3.15), putting $\varepsilon = \varepsilon_n$ and sending n to ∞. The passage in the linear terms follows straight, by (3.12). We need to focus on the nonlinear terms appearing in the identity (3.14) and the identity (3.15). These are the terms associated with the difference quotients of F, of G_j and of H_j, for $j = 1, \ldots, J$.

The first term. Let us start with the term associated with the difference quotients of F, i.e.:

$$\int_0^T \Big(\frac{1}{\varepsilon_n} \left\{ F(\hat{v}+\varepsilon_n\hat{\eta})) - F(\hat{v})) \right\}, \phi \Big)_{L^2(\Omega)} dt \quad \text{where } \phi \in L^2(0,T;H^1(\Omega))$$

For the limit passage, we need to justify that $\frac{1}{\varepsilon_n}\left\{F(\hat{v}+\varepsilon_n\hat{\eta}))-F(\hat{v}))\right\}$ converges weakly in $L^2(0,T;L^2(\Omega))$, as n tends to ∞. But the weak convergence in $L^2(0,T;L^2(\Omega))$ is equivalent to the weak convergence in $L^2(Q_T)$. We will focus on investigating the latter. We will show that the stated weak convergence holds and that the weak limit is equal $f'(\hat{y})\bar{\delta}^{\epsilon}_{w}\mathcal{Z}_y(\hat{v};\hat{\eta})=f'(\hat{y})\widetilde{y}$.

Note, that $F(\hat{v})$ can be interpreted as $F(\hat{v})=\mathcal{N}_f\circ\mathcal{Z}_y(\hat{v})$ where $\mathcal{N}_f$ denotes the Nemytskii operator $\mathcal{N}_f$ associated with the function f. Therefore, the considered difference quotients of F converge weakly in $L^2(Q_T)$ to $\bar{\delta}^{\epsilon}_{w}(\mathcal{N}_f\circ\mathcal{Z}_y)(\hat{v};\hat{\eta})$, if the latter exists. Thus, we need to justify that $\bar{\delta}^{\epsilon}_{w}(\mathcal{N}_f\circ\mathcal{Z}_y)(\hat{v};\hat{\eta})$ exists in $L^2(Q_T)$ and equals $f'(\hat{y})\widetilde{y}$.

By Theorem 3.1.5 and by identification $L^{p_2}(0,T;L^{p_2}(\Omega))=L^{p_2}(Q_T)$, $\mathcal{Z}_y$ can be understood as $\mathcal{Z}_y:V\to L^{p_2}(Q_T)$. Moreover, by (3.12) and by the introduced notation, $\bar{\delta}^{\epsilon}_{w}\mathcal{Z}_y(\hat{v};\hat{\eta})$ exists in $L^{p_2}(Q_T)$ and equals $\widetilde{y}$.

By the assumption (B-3), it can be verified that f obeys the following growth condition

$$\sup_{s\in\mathbb{R}}\left|f(s)\right|/\left(1+\left|s\right|^{p_2/2}\right)<\infty$$

Therefore, by Theorem A.3.2 in Appendix A.3, $\mathcal{N}_f$ is well defined as $\mathcal{N}_f:L^{p_2}(Q_T)\to L^2(Q_T)$. By assumptions (B-3) and (E-1), the derivative f' exists and satisfies the following growth condition:

$$\sup_{s\in\mathbb{R}}\left|f'(s)\right|/\left(1+\left|s\right|^{(p_2/2)-1}\right)<\infty$$

Thus, by Theorem A.3.5 in Appendix A.3, the Nemytskii operator $\mathcal{N}_f$ is Fréchet differentiable from $L^{p_2}(Q_T)$ to $L^2(Q_T)$, with

$$D_F\mathcal{N}_f(p)(q)(x,t)=f'(p(x,t))q(x,t)\qquad\text{a.e. on } Q_T\text{, for } p,q\in L^{p_2}(Q_T)$$

By the above properties of $\mathcal{Z}_y$ and $\mathcal{N}_f$ and by the chain rule (see Theorems A.1.4 and A.1.10 in Appendix A.3), $\bar{\delta}^{\epsilon}_{w}(\mathcal{N}_f\circ\mathcal{Z}_y)(\hat{v};\hat{\eta})$ exists and

$$\bar{\delta}^{\epsilon}_{w}(\mathcal{N}_f\circ\mathcal{Z}_y)(\hat{v};\hat{\eta})=D_F\mathcal{N}_f(\mathcal{Z}_y(\hat{v}))\bar{\delta}^{\epsilon}_{w}\mathcal{Z}_y(\hat{v};\hat{\eta})=f'(\hat{y})\widetilde{y}\tag{3.16}$$

$$\frac{1}{\varepsilon_n}\left\{F(\hat{v}+\varepsilon_n\hat{\eta}))-F(\hat{v}))\right\}\ \rightharpoonup\ \bar{\delta}^{\epsilon}_{w}(\mathcal{N}_f\circ\mathcal{Z}_y)(\hat{v};\hat{\eta})\quad\text{in } L^2(Q_T)\tag{3.17}$$

The second term. Now, we proceed to the terms associated with the difference quotients of G_j, i.e.

$$\int_0^T\left(\frac{1}{\varepsilon_n}\left\{G_j(\hat{v}+\varepsilon_n\hat{\eta}))-G_j(\hat{v}))\right\},\phi\right)_{L^2(\Omega)}dt\quad\text{where } \phi\in L^2(0,T;H^1(\Omega))$$

for $j=1,\dots,J$. For the limit passage, we need to verify that $\frac{1}{\varepsilon_n}\left\{G_j(\hat{v}+\varepsilon_n\hat{\eta}))-G_j(\hat{v}))\right\}$ converges weakly in $L^2(0,T;L^2(\Omega))$, as $n\to\infty$, for $j=1,\dots,J$. We will use the fact that the weak convergence in $L^2(0,T;L^2(\Omega))$ and the weak convergence in $L^2(Q_T)$ are equivalent. We will show that the weak convergence in $L^2(Q_T)$ hold and that the weak limit equals $\varUpsilon_{g_j}(\hat{v})\widetilde{\kappa}_j+D_{G,w}\varUpsilon_{g_j}(\hat{v})(\hat{\eta})\hat{\kappa}_j$.

Term G_j, for $j=1,\dots,J$, can be understood as:

$$G_j(\hat{v})=I(\varUpsilon_{g_j}(\hat{v}),\mathcal{Z}_{\kappa_j}(\hat{v}))$$

where

$$I:L^2(\Omega)\times L^2(0,T)\longrightarrow L^2(Q_T),\qquad I(p,q)(x,t):=p(x)q(t)\quad\text{a.e. on } Q_T$$

for $p \in L^2(\Omega)$, $q \in L^2(0,T)$. Therefore, for $j = 1, \dots, J$, the considered difference quotients of G_j converge weakly in $L^2(Q_T)$ to $\bar{\delta}^\epsilon_w I \circ (\Upsilon_{g_j}, \mathcal{Z}_{\kappa_j})(\hat{v}; \hat{\eta})$, if this derivative exists. Thus, it is necessary to justify that the subject derivative indeed exists in $L^2(Q_T)$, and that it equals $\Upsilon_{g_j}(\hat{v})\widetilde{\kappa}_j + D_{G,w}\Upsilon_{g_j}(\hat{v})(\hat{\eta})\hat{\kappa}_j$.

By (3.12) and by the introduced notation, $\bar{\delta}^\epsilon_w \mathcal{Z}_{\kappa_j}(\hat{v}; \hat{\eta})$ exists in $L^2(0,T)$ and equals $\widetilde{\kappa}_j$.

By Lemma 3.1.4, the operator Υ_{g_j} is well defined and weakly Gâteaux differentiable from V to $L^2(\Omega)$.

Moreover, it is straightforward that $I(p,q)$ is measurable for arbitrary $p \in L^2(\Omega)$ and $q \in L^2(0,T)$ and, by Fubini theorem, belongs to $L^2(Q_T)$. Thus, I is well-defined. I is also bilinear and, again by Fubini theorem, bounded.

Hence, by the above properties of $\mathcal{Z}_{\kappa_j}$, Υ_{g_j} and I and by the product rule for Banach spaces (see Theorem A.1.5 in Appendix A.1), we infer that, for $j = 1, \dots, J$, there holds:

$$\begin{aligned} \bar{\delta}^\epsilon_w I \circ (\Upsilon_{g_j}, \mathcal{Z}_{\kappa_j})(\hat{v}; \hat{\eta}) &= I\left(\bar{\delta}^\epsilon_w \Upsilon_{g_j}(\hat{v}; \hat{\eta}), \mathcal{Z}_{\kappa_j}(\hat{v})\right) \; + \; I\left(\Upsilon_{g_j}(\hat{v}), \bar{\delta}^\epsilon_w \mathcal{Z}_{\kappa_j}(\hat{v}; \hat{\eta})\right) \\ &= D_G \Upsilon_{g_j}(\hat{v})(\hat{\eta})\hat{\kappa}_j \; + \; \Upsilon_{g_j}(\hat{v})\widetilde{\kappa}_j \end{aligned} \tag{3.18}$$

$$\frac{1}{\varepsilon_n}\left\{G_j(\hat{v} + \varepsilon_n \hat{\eta}) - G_j(\hat{v})\right\} \rightharpoonup \bar{\delta}^\epsilon_w I \circ (\Upsilon_{g_j}, \mathcal{Z}_{\kappa_j})(\hat{v}; \hat{\eta}) \quad \text{in } L^2(Q_T) \tag{3.19}$$

The third term. The remaining terms we need to investigate are the terms associated with the difference quotients of H_j, i.e. terms

$$\int_0^T \Big(\frac{1}{\varepsilon_n}\left\{H_j(\hat{v} + \varepsilon_n \hat{\eta})) - H_j(\hat{v}))\right\}\Big)\xi \, dt \quad \text{where } \xi \in L^2(0,T)$$

for $j = 1, \dots, J$. We require to justify that $\frac{1}{\varepsilon_n}\left\{H_j(\hat{v} + \varepsilon_n \hat{\eta})) - H_j(\hat{v}))\right\}$ converges weakly in $L^2(0,T)$, as n tends to ∞. We will prove that this weak convergence holds and that the weak limit in this convergence is equal

$$w_j'\big(\big(\Upsilon_{h_j}(\hat{v}) \, , \, \hat{y} - y^*\big)_{L^2(\Omega)}\big) \cdot \Big(\big(D_{G,w}\Upsilon_{h_j}(\hat{v})(\hat{\eta}) \, , \, \hat{y} - y^*\big)_{L^2(\Omega)} + \big(\Upsilon_{h_j}(\hat{v}) \, , \, \widetilde{y}\big)_{L^2(\Omega)}\Big)$$

Term H_j can be understood as:

$$\begin{aligned} H_j(\hat{v}) &= \mathcal{N}_{w_j} \circ I\left(\Upsilon_{h_j}(\hat{v}), \mathcal{Z}_y(\hat{v}) - y^*\right) \\ &= \mathcal{N}_{w_j} \circ I \circ \left(\Upsilon_{h_j}, i_{y^*} \circ \mathcal{Z}_y\right)(\hat{v}) \end{aligned}$$

where, for $j = 1, \dots, J$

$$\begin{array}{lll} i_{y^*} \colon L^{p_2}(0,T; L^2(\Omega)) \to L^{p_2}(0,T; L^2(\Omega)) & i_{y^*}(p) := p - y^* & \\ I \colon L^2(\Omega) \times L^{p_2}(0,T; L^2(\Omega)) \longrightarrow L^{p_2}(0,T) & I(q,r)(t) := (q(\,.\,), r(\,.\,,t))_{L^2(\Omega)} & \text{a.e. on } [0,T] \\ \mathcal{N}_{w_j} \colon L^{p_2}(0,T) \longrightarrow L^2(0,T) & \text{is the Nemytskii operator corresp. to } w_j & \end{array}$$

for $p, r \in L^{p_2}(0,T; L^2(\Omega))$, $q \in L^2(\Omega)$. Hence, for $j = 1, \dots, J$, the investigated difference quotients of H_j converge to $\bar{\delta}^\epsilon_w \mathcal{N}_{w_j} \circ I \circ \left(\Upsilon_{h_j}, i_{y^*} \circ \mathcal{Z}_y\right)(\hat{v}; \hat{\eta})$, if the latter exists. Thus, analysis of existence of this derivative is required. We will perform it now.

Note that $L^{p_2}(0,T; L^{p_2}(\Omega)) \subseteq L^{p_2}(0,T; L^2(\Omega))$ and $L^{p_2}(0,T; L^{p_2}(\Omega)) \hookrightarrow L^{p_2}(0,T; L^2(\Omega))$. Thus, $\mathcal{Z}_y$ is well defined with values in $L^{p_2}(0,T; L^2(\Omega))$. By the mentioned embedding and by (3.12), the derivative $\bar{\delta}^\epsilon_w \mathcal{Z}_y(\hat{v}; \hat{\eta})$ exists in $L^{p_2}(0,T; L^2(\Omega))$ and, by the introduced notation, equals $\widetilde{y}$.

Next, it follows straight that the operator i_{y^*} is well defined. It can be verified by the definition of the Fréchet differentiability, that the operator i_{y^*} is Fréchet differentiable with $D_F i_{y^*}(p)s = s$, for $p, s \in L^{p_2}(0,T; L^2(\Omega))$.

By Lemma 3.1.4, the operator Υ_{h_j} is well defined and weakly Gâteaux differentiable from V to $L^2(\Omega)$, for $j = 1, \dots, J$.

By Pettis theorem, $I(q,r)$ is measurable for arbitrary $q \in L^2(\Omega)$ and $r \in L^{p_2}(0,T; L^2(\Omega))$. Moreover, by the Fubini theorem and the Hölder inequality:

$$\begin{aligned} \big\|I(p,q)\big\|^{p_2}_{L^{p_2}(0,T)} \;&=\; \int_0^T \big|\big(p(\,.\,)\,,\, q(\,.\,,t)\big)_{L^2(\Omega)}\big|^{p_2}\,dt \\ \le \big\|p\big\|_2^{p_2} \int_0^T \big\|q(\,.\,,t)\big\|_2^{p_2}\,dt \;&=\; \big\|p\big\|_2^{p_2}\big\|q\big\|_{2,p_2}^{p_2} \end{aligned}$$

Hence, I is well defined. I is also bilinear and, by the above estimates, bounded.

By the assumption (B-4), it can be verified that w_j, for $j = 1, \dots, J$, satisfies the following growth condition

$$\sup_{s\in\mathbb{R}} \big|w_j(s)\big|/\big(1 + |s|^{p_2/2}\big) < \infty$$

Hence, by Theorem A.3.2 in Appendix A.3, $\mathcal{N}_{w_j}$ is well defined as $\mathcal{N}_f : L^{p_2}(0,T) \to L^2(0,T)$. Also, by assumptions (B-4) and (E-2), the derivative w_j' exists and:

$$\sup_{s\in\mathbb{R}} \big|w_j'(s)\big|/\big(1 + |s|^{(p_2/2)-1}\big) < \infty$$

Therefore, by Theorem A.3.5 in Appendix A.3, the Nemytskii operator $\mathcal{N}_{w_j}$ is Fréchet differentiable from $L^{p_2}(0,T)$ to $L^2(0,T)$, with

$$D_F\mathcal{N}_{w_j}(p)(q)(x,t) = w_j'(p(t))q(t) \qquad \text{a.e. on } (0,T), \text{ for } p,q \in L^{p_2}(0,T)$$

Having the above properties of $\mathcal{Z}_y$, i_{y^*}, Υ_{h_j}, I and $\mathcal{N}_{w_j}$, for $j = 1, \dots, J$, the chain rule and the product rule (see Theorems A.1.4 and A.1.5 in Appendix A.1) can be combined to infer that $\bar{\delta}^{\epsilon}_{w} H_j(\hat{v};\hat{\eta})$ exists and

$$\begin{aligned} \bar{\delta}^{\epsilon}_{w} H_{jk}(\hat{v};\hat{\eta}) \;&= \bar{\delta}^{\epsilon}_{w}\big(\mathcal{N}_{w_k} \circ I \circ \big(\Upsilon_{h_j}, i_{y^*} \circ \mathcal{Z}_y\big)\big)(\hat{v};\hat{\eta}) \\ &= w_k'\big(I\big(\Upsilon_{h_j}(\hat{v}), \mathcal{Z}_y(\hat{v}) - y^*\big)\big)\cdot \\ &\qquad \cdot\Big\{I\big(\bar{\delta}^{\epsilon}_{w}\Upsilon_{h_k}(\hat{v};\hat{\eta}), \mathcal{Z}_y(\hat{v}) - y^*\big) + I\big(\Upsilon_{h_k}(\hat{v}), \bar{\delta}^{\epsilon}_{w}\mathcal{Z}_y(\hat{v};\hat{\eta})\big)\Big\} \\ &= w_k'\big(\big(\Upsilon_{h_k}(\hat{v}), \hat{y} - y^*\big)_{L^2(\Omega)}\big)\cdot \\ &\qquad \cdot\Big\{\big(D_{G,w}\Upsilon_{h_k}(\hat{v})(\hat{\eta}), \hat{y} - y^*\big)_{L^2(\Omega)} + \big(\Upsilon_{h_k}(\hat{v}), \widetilde{y}\big)_{L^2(\Omega)}\Big\} \end{aligned} \tag{3.20}$$

$$\frac{1}{\varepsilon_n}\left(H_{jk}(\hat{v} + \varepsilon_n\hat{\eta}) - H_{jk}(\hat{v})\right) \;\rightharpoonup\; \bar{\delta}^{\epsilon}_{w} H_{jk}(\hat{v};\hat{\eta}) \quad \text{in } L^2(Q_T) \tag{3.21}$$

The analysis of the nonlinear terms is finished. Altogether, due to (3.12), (3.17), (3.19) and (3.21), we can pass with n to infinity in identities (3.13) - (3.15). Moreover, by (3.16), (3.18) and (3.20) we infer that the limit passage results in identities which correspond precisely to the definition of the weak solution of (3.9) - (3.10), with $(\widetilde{y}, \widetilde{\kappa}_1, \dots, \widetilde{\kappa}_J) = \bar{\delta}^{\epsilon}_{w}\mathcal{Z}(\hat{v};\hat{\eta})$ being the weak solution (see Definition 3.1.6). This concludes the proof of the theorem. ■

3.2 Optimization problem

In this section, we focus on the optimal targeting problem, announced in the beginning of Chapter 3. The main point of the present section is derivation of a formula for the Gâteaux differential of the cost functional (3.3). This formula allows to express necessary optimality conditions for the optimal targeting problem in an explicit way. Moreover, the subject formula was helpful to perform the numerical optimization experiments, described in Chapter 4. For completeness of our considerations, in this section we present also a simple result concerning existence of minimizers of the cost functional (3.3).

The present section is organized as follows. We begin with reformulating the cost functional (3.3) within a functional analysis framework, more convenient to work with. Next, in brief Section 3.2.1, we give a basic criterion concerning existence of minimizers for the cost functional. This criterion assumes compactness of the supports of the pattern functions σ_g and σ_h, entering the system (3.1) - (3.2). Restriction of compact supports may seem to be strong. Nevertheless, in Chapter 4, concerning numerical optimization experiments, we will use patterns functions with compact supports. Therefore results assuming compact supports of the pattern functions are sufficient for our purposes.

In Section 3.2.2, we proceed to the matter of differentiability of the cost functional. First, it is shown that the cost functional is Gâteaux differentiable. Next, we pass to characterizing the Gâteaux differential of the cost functional. Since, by definition, the Gâteaux differential of the cost functional in point $\hat{v} \in V$ is a bounded linear functional on V, it can be characterized as an element of $\Lambda^{\hat{v}} \in V^* = V$, dependent on v. The main theorem of Section 3.2.2 gives a formula for $\Lambda^{\hat{v}}$. The above results on differentiability of the cost functional require the operator $\mathcal{Z}$, defined in Section 3.1.2, to be weakly Gâteaux differentiable. Hence, these results inherit the assumptions guarantying weak Gâteaux differentiability of $\mathcal{Z}$, see Section 3.1.3.

In Section 3.2.1 and Section 3.2.2, the main results assume, in particular, that in the system (3.1) - (3.2) the function f is globally Lipschitz and y_0 belongs to $L^2(\Omega)$. In Section 3.2.3, we show how to generalize the main results of Section 3.2.1 and Section 3.2.2 to the case where f is locally Lipschitz only, with the condition (1.73) and where $y_0 \in L^\infty(\Omega)$. The results of Section 3.2.3 cover the case of the data utilized in the numerical optimization experiments described in Chapter 4.

Let us start. Note, that if the assumptions for the system (3.1) - (3.2) are such that the weak solution exists (i.e. $y \in L^2(Q_T)$, in particular), then the cost functional (3.3) can be identified with the cost functional $\mathcal{I}$, defined as follows:

$$\mathcal{I}\colon V \to \mathbb{R}, \qquad \mathcal{I}(\hat{v}) := \widetilde{\lambda} \big\| \mathcal{Z}_y^{T_0}(\hat{v}) - y^{*T_0} \big\|^2_{L^2(Q_T^{T_0})} \tag{3.22}$$

where parameters $\widetilde{\lambda} > 0$ and $T_0 \in (0, T)$ are given, $Q_T^{T_0}$ is defined as in the beginning of the present chapter and

$$\mathcal{Z}_y^{T_0} := \mathcal{P}^{R,T_0} \circ \mathcal{Z}_y, \qquad y^{*T_0} := \mathcal{P}^{R,T_0}(y^*) \tag{3.23}$$

We recall that, in the present chapter, the operator $\mathcal{P}^{R,T_0}$ is understood as $\mathcal{P}^{R,T_0}\colon L^2(Q_T) \to L^2(Q_T^{T_0})$. Conditions (3.22) - (3.23) are more convenient for analysis than the condition (3.3). Hence, since now until the end of Section 3.2, we will focus conditions (3.22) - (3.23) instead of the condition (3.3).

Having the above definition of $\mathcal{I}$, we formulate the optimization problem that we will focus on as:

$$\inf_{\hat{v} \in V} \mathcal{I}(\hat{v}) \tag{3.24}$$

3.2.1 Existence of local minimizers

In this brief section, we address the question concerning the existence of solutions to the problem (3.24). The following result is true:

Theorem 3.2.1 *In the system (3.1) - (3.2), let assumptions (B-1) - (B-5) be fulfilled, with additional restriction $K = J$. Assume also that at least one of the following is true:*

- *y^* fulfills the assumption (C-1) and functions w_k are bounded, for $j = 1, \dots, J$,*
- *y^* fulfills the assumption (C-2).*

and that σ_g, σ_h fulfill assumptions (F-1) and (F-3). Then, the optimization problem (3.24) attains at least one solution.

Proof. Let $dist_V$ denote the metric in the metric space V. By the assumption (F-3), $\mathcal{I}\colon V \to \mathbb{R}$ is constant on the set $\mathbb{E}^c$, being the complement in V of

$$\mathbb{E} = \Big\{ (x_1, \dots, x_J) : \; dist(x_j, \Omega) \leq C_{supp} \; j = 1, \dots, J \Big\}$$

where $C_{supp} = \max\{\mathrm{diam}(\mathrm{supp}(\sigma_g)), \mathrm{diam}(\mathrm{supp}(\sigma_h))\}$. Indeed, the operator $\varUpsilon$ is constant on $\mathbb{E}^c \subseteq V$ and hence $\mathcal{Z}_y = \mathcal{S}_y \circ \varUpsilon$ is constant on $\mathbb{E}^c$ and so $\mathcal{I}$ is.

On the other hand, our assumptions allow to apply Theorem 3.1.5 and conclude that $\mathcal{Z}\colon V \to X^2$ is continuous. By this, the component $\mathcal{Z}_y$ of $\mathcal{Z}$ is continuous when understood as $\mathcal{Z}_y\colon V \to L^2(Q_T)$. Hence, it can be verified that $\mathcal{Z}_y^{T_0}\colon V \to L^2(Q_T^{t^1})$ is continuous as well. The latter allows to infer the continuity of $\mathcal{I}\colon V \to \mathbb{R}$.

Moreover, due to the assumption (F-3), $\mathbb{E} \subset V$ is compact or empty. In the case when $\mathbb{E}$ is compact, $\mathcal{I}$, as a continuous functional, attains its minimum on $\mathbb{E}$ in some point $\bar{v} \in \mathbb{E}$. Then the minimal value of $\mathcal{I}$ on V is $\min\{\mathcal{I}(\hat{v}), \mathcal{I}(\bar{v})\}$ for an arbitrary $\hat{v} \in \mathbb{E}^c$. In the case of empty $\mathbb{E}$ the minimal value of $\mathcal{I}$ on V is $\mathcal{I}(\hat{v})$ for an arbitrary $\hat{v} \in \mathbb{E}^c$. ■

Remark. The proof of Theorem 3.2.1 is simple due to the restrictive the assumption (F-3). Nevertheless, the assumption (F-3) suffice to cover the data considered in the numerical optimization experiments described in Chapter 4. ▲

Remark. Dispensing the assumption (F-3) in Theorem 3.2.1 is not an obvious modification. This assumption allowed to reduce the optimization problem to the problem of existence of minimizers of $\mathcal{I}$ on a compact subset of V, what, along with the assumptions sufficient for the continuity of $\mathcal{I}$, immediately justified the desired result. Without the assumption (F-3), the methods of the proof of Theorem 3.2.1 do not reduce the problem to the problem of minimization on a compact subset of V. Hence, in this situation, the natural strategy would be to select a minimizing sequence, to justify its boundedness in V, to select a weakly convergent subsequence and next to justify the properties of $\mathcal{I}$ necessary for the limit passage on this subsequence. In situations of this kind, it is common that the boundedness of the minimizing sequence is concluded by the presence of some coercing term in the definition of a cost functional. Unfortunately, the definition of $\mathcal{I}$ does not contain any coercing term, allowing to obtain boundedness of the minimizing sequence. Thus, the mentioned strategy would be not straightforward to apply.

The above makes the problem of dispensing the assumption (F-3) in Theorem 3.2.1 interesting. However, we would like to focus rather on the problem of characterizing the solutions of problem (3.24) than on the problem of existence of its solutions. Therefore, we do not continue the investigation of the latter problem in the present work. ▲

3.2.2 The gradient of the cost functional

Section 3.2.2 is devoted to investigating the differentiability of $\mathcal{I}\colon V \to \mathbb{R}$ and deriving a characterization of its differential. The characterization of the differential of $\mathcal{I}$ is the main theorem of Section 3.2.2. Before deriving the announced characterization, we introduce an auxiliary system of equations, which we call *the adjoint system*. The idea of the proof of the main theorem consists in testing the solution of the linearized system (see Section 3.1.3) with the solution of the adjoint system, testing the solution of the adjoint system with the solution of the linearized system and comparing the results of these testings. Hence, both the adjoint system and the linearized system are essential for the proof of the main theorem of Section 3.2.2.

To be more precise, we aim in proving the Gâteaux differentiability of the cost functional $\mathcal{I}$, defined by (3.22) - (3.23), and representing its Gâteaux differential in the following form:

$$D_G\mathcal{I}(\hat{v})(\hat{\eta}) \;=\; \big(\Lambda^{\hat{v}}, \hat{\eta}\big)_V \tag{3.25}$$

for certain $\Lambda^{\hat{v}} \in V^* = V$. The element $\Lambda^{\hat{v}}$ in (3.25) is in fact the gradient of $\mathcal{I}$ and hence can be utilized to perform gradient-type optimization procedures. The characterization of $\Lambda^{\hat{v}}$, obtained below in the present section, was utilized in the numerical experiments described in Chapter 4.

Let us begin with some remarks on differentiability of $\mathcal{I}$.

Lemma 3.2.2 *Let the assumptions of Theorem 3.1.8 be fulfilled. Then, the cost functional $\mathcal{I}\colon V \to \mathbb{R}$, defined in (3.22) - (3.23), is Gâteaux differentiable and*

$$(D_G\mathcal{I})(\hat{v})(\hat{\eta}) \;=\; 2\widetilde{\lambda}\left(\mathcal{Z}_y^{T_0}(\hat{v}) - y^{*T_0}\,,\, D_{G,w}\mathcal{Z}_y^{T_0}(\hat{v})(\hat{\eta})\right)_{L^2(Q_T^{T_0})} \tag{3.26}$$

and $D_{G,w}\mathcal{Z}_y^{T_0}(\hat{v})(\hat{\eta})$ can be characterized as follows:

$$D_{G,w}\mathcal{Z}_y^{T_0}(\hat{v})(\hat{\eta}) \;=\; \mathcal{P}^{R,T_0}\big(D_{G,w}\mathcal{Z}_y(\hat{v})(\hat{\eta})\big) \tag{3.27}$$

Proof. First, note that, by (3.22) and (3.23), $\mathcal{I}$ can be understood as

$$\mathcal{I}(\hat{v}) = \widetilde{\lambda}\big\|\mathcal{P}^{R,T_0} \circ i_{y^*} \circ \mathcal{Z}_y\,(\hat{v})\big\|^2_{L^2(Q_T^{T_0})} \tag{3.28}$$

where $\widetilde{\lambda}$, T_0 and $Q_T^{T_0}$ are as in (3.22) - (3.23) and where i_{y^*} is defined by

$$i_{y^*}\colon L^2(Q_T) \to L^2(Q_T), \qquad i_{y^*}(p) := p - y^* \tag{3.29}$$

for $p \in L^2(Q_T)$.

Note, that the definition of i_{y^*} in (3.29) makes sense under the assumption (E-3), since $L^{p_2}(0,T;L^2(\Omega)) \hookrightarrow L^2(Q_T)$. It also follows by the definition of the Fréchet differentiability that i_{y^*} is Fréchet differentiable and $D_F i_{y^*}(p)(q) = q$, for $p, q \in L^2(Q_T)$. Moreover, the operator $\mathcal{P}^{R,T_0}\colon L^2(Q_T) \to L^2(Q_T^{T_0})$ is linear and bounded and hence Fréchet differentiable with $D_F\mathcal{P}^{R,T_0}(p)(q) = \mathcal{P}^{R,T_0}(q)$, for $p, q \in L^2(Q_T)$ (see Observation A.1.7 in Appendix A.1). In addition, since the assumptions of Theorem 3.1.8 are fulfilled, the operator $\mathcal{Z}_y$ is weakly Gâteaux differentiable from V to $L^2(Q_T)$.

By the above remarks, by (3.28) and by Theorem A.1.4 and Observations A.1.6, A.1.8 in Appendix A.1, we conclude that the assertion holds. ■

Lemma 3.2.2 justifies the existence of $D_G\mathcal{I}$ and, by (3.26), gives certain characterization of the latter. Nevertheless, the subject characterization is not of form (3.25), being our aim. Thus, we now focus on deriving representation (3.25) of the differential of the cost functional $\mathcal{I}$.

The following system of equations, which we call *the adjoint system*, will be necessary for our purposes:

$$\begin{cases} -p_t - D\Delta p - f'(\hat{Y})p = (\hat{Y} - y^*)\mathbf{1}_{(T_0,T)} + \\ \qquad + \sum_{j=1}^{J} w_j'\Big(\int_\Omega \varUpsilon_{h_j}(\hat{v})(\hat{Y} - y^*)\,dx\Big)\varUpsilon_{h_j}(\hat{v})\,q_j \quad \text{on } Q_T \\ \dfrac{\partial p}{\partial n} = 0 \quad \text{on } \partial\Omega \times (0,T) \\ p(T,x) \equiv 0 \end{cases} \tag{3.30}$$

together with

$$\begin{cases} -\beta_1 q_1' + q_1 = \displaystyle\int_\Omega \varUpsilon_{g_1}(\hat{v})p\,dx & \text{on } [0,T] \\ \vdots & \vdots \\ -\beta_J q_J' + q_J = \displaystyle\int_\Omega \varUpsilon_{g_J}(\hat{v})p\,dx & \text{on } [0,T] \\ q_j(T) = 0 & \text{for } j = 1,\dots,J \end{cases} \tag{3.31}$$

where: Ω is a domain, $T > 0$, $Q_T := \Omega \times (0,T)$ and, for $j = 1,\dots,J$, $D, \beta_j > 0$ are given numbers, $f, w_j \colon \mathbb{R} \to \mathbb{R}$, $\hat{Y}, y^* \colon Q_T \to \mathbb{R}$ are given functions, $\hat{v} \in V$, $\varUpsilon_{g_j}$ and $\varUpsilon_{h_j}$ correspond to given $\sigma_g, \sigma_h \colon \mathbb{R}^{\mathbf{d}} \to \mathbb{R}$ (see (3.7) for the explanation of the latter correspondence), $T_0 \in (0,T)$ and $\mathbf{1}_{(T_0,T)} \colon (0,T) \to \mathbb{R}$ denotes the characteristic function of interval (T_0,T) (see *Notation conventions*). In the system (3.30) - (3.31), the unknown is the function $(p, q_1, \dots, q_J) \colon Q_T \to \mathbb{R}^{J+1}$.

Note, that if $(p, q_1, \dots, q_J)$ was a classical solution of the system (3.30) - (3.31), then $(\mathcal{P}^i_{Q_T}p, \mathcal{P}^i_T q_1, \dots, \mathcal{P}^i_T q_J)$, where $\mathcal{P}^i_{Q_T}$ and $\mathcal{P}^i_T$ are defined as in the beginning of the present chapter, would be a classical solution of the system (1.84) - (1.86) in Section 1.2.4, with

$$\begin{aligned} \mathbf{Y}(x,t) &:= 0, \\ \Theta_j(x,t) &:= 0, \\ \Xi_j &:= w_j'\Big(\int_\Omega \varUpsilon_{h_j}(\hat{v})\mathcal{P}^i_{Q_T}(\hat{Y} - y^*)\,dx\Big)\varUpsilon_{h_j}(\hat{v}), \\ \widetilde{f}(x,t,s) &:= f'(\mathcal{P}^i_{Q_T}(\hat{Y})(x,t))s + \\ &\quad + \mathcal{P}^i_{Q_T}(\hat{Y} - y^*)(x,t)\mathcal{P}^i_T\left(\mathbf{1}_{(T_0,T)}\right)(t), \\ \widetilde{w}_j(s) &:= s, \\ \mathbf{h}_j &:= \varUpsilon_{g_j}(\hat{v}), \end{aligned} \qquad \begin{aligned} \mathbf{Z}_j(t) &:= 1, \\ \widetilde{g}_j(x) &:= 0, \\ \widetilde{h}_j(x) &:= 0, \\ \widetilde{y}_0(x) &:= 0, \\ \widetilde{\kappa}_{j0} &:= 0, \end{aligned} \tag{3.32}$$

for $j = 1,\dots,J$, $x \in \Omega$, $t \in (0,T)$, $s \in \mathbb{R}$.

The above remark explains the motivation behind the following definition of weak solutions of (3.30) - (3.31), also involving the use of inverse time operators $\mathcal{P}^i_{Q_T}$ and $\mathcal{P}^i_T$:

Definition 3.2.3 *The element $(p, q_1, \dots, q_J) \in X^2$ is a weak solution of (3.30) - (3.31) if the element $(\mathcal{P}^i_{Q_T}p, \mathcal{P}^i_T q_1, \dots, \mathcal{P}^i_T q_J)$ is a weak solution of (1.84) - (1.86) with conditions (3.32) (see Definition 1.2.16).*

It is straightforward that if $(p, q_1, \dots, q_J) \in X^2$, then $(\mathcal{P}^i_{Q_T} p, \mathcal{P}^i_T q_1, \dots, \mathcal{P}^i_T q_J) \in X^2$. Thus, Definition 3.2.3 is meaningful.

With the above definition, we can justify the following existence and uniqueness result:

Lemma 3.2.4 *Let assumptions (B-1) - (B-4) be fulfilled, with additional restriction $K = J$. Let also assumptions (E-1) - (E-2) and (F-1) hold. Moreover, assume that $\hat{Y}, y^* \in L^2(0, T; L^2(\Omega))$.*

Then, the weak solution of the system (3.30) - (3.31) exists and is unique.

Proof. By Definition 3.2.3, it suffices to show, that the system (1.84) - (1.86) with conditions (3.32) has a unique weak solution in sense of Definition 1.2.16. For this end, it is enough to justify that the assumptions of Theorem 1.2.18 are fulfilled.

First, $\hat{Y} \in L^2(Q_T)$ and hence $\mathcal{P}^i_{Q_T}\hat{Y} \in L^2(Q_T)$. In particular, $\mathcal{P}^i_{Q_T}\hat{Y}$ is measurable. Moreover, f' is a Borel measurable function because it is the classical derivative of a continuous function (see assumptions (B-4) and (E-1)). Therefore, $f'(\mathcal{P}^i_{Q_T}(\hat{Y})(\,.\,,\,.\,))$ is measurable, as well as $f'(\mathcal{P}^i_{Q_T}(\hat{Y})(\,.\,,\,.\,))s$, for an arbitrary $s \in \mathbb{R}$. Also, $\mathcal{P}^i_{Q_T} y^*$ is measurable because, by our assumptions, $y^* \in L^2(Q_T)$ and hence $\mathcal{P}^i_{Q_T} y^* \in L^2(Q_T)$. This, along with the fact that $\mathcal{P}^i_{Q_T}\hat{Y}$ and $\mathcal{P}^i_{Q_T}\mathbf{1}_{(T_0,T)}$ are measurable, gives a conclusion that $\mathcal{P}^i_{Q_T}(\hat{Y} - y^*)\mathcal{P}^i_T(\mathbf{1}_{(T_0,T)})$ is measurable. Summing up the above remarks, we conclude that $\widetilde{f}$ defined in (3.32) is measurable, for an arbitrary $s \in \mathbb{R}$.

Second, by the assumption (B-3), f' is bounded. Therefore, it follows that $\widetilde{f}$ defined in (3.32) is Lipschitz continuous in s for a.e. $(x,t) \in Q_T$, with the Lipschitz constant independent of $(x,t) \in Q_T$.

Third, for $\widetilde{f}$ defined in (3.32), $\widetilde{f}(\,.\,,\,.\,,0) = \mathcal{P}^i_{Q_T}(\hat{Y} - y^*)\mathcal{P}^i_T(\mathbf{1}_{(T_0,T)})$, what belongs to $L^2(Q_T)$, since $\mathcal{P}^i_{Q_T}\hat{Y}, \mathcal{P}^i_{Q_T}\hat{y}^* \in L^2(Q_T)$.

Summing up the above, $\widetilde{f}$ defined in (3.32) fulfills the assumption (D-3).

Moreover, $\mathcal{P}^i_{Q_T}(\hat{Y} - y^*)$ belongs to $L^2(0, T; L^2(\Omega))$, hence it is strongly measurable. Therefore, by the Pettis theorem, $\widetilde{F}_j := \int_\Omega \Upsilon_{h_j}(\hat{v})\mathcal{P}^i_{Q_T}(\hat{Y} - y^*)\,dx$ understood as a real function of variable t is measurable, for $j = 1, \dots, J$. A the same time, w_j' is Borel measurable as a classical derivative of a continuous function (see assumptions (B-5) and (E-2)). Therefore, the function $w_j' \circ \widetilde{F}_j$ is measurable, for $j = 1, \dots, J$. The function $w_j' \circ \widetilde{F}_j$ is also bounded for $j = 1, \dots, J$, because w_j' is bounded, by the assumption (B-4). Thus, $w_j' \circ \widetilde{F}_j$ belongs to $L^\infty(0,T)$ for $j = 1, \dots, J$. Taking into account the latter and $\Upsilon_{h_j}(\hat{v}) \in L^2(\Omega)$, we conclude that Ξ_j defined in (3.32) fulfills the assumption (D-6).

The fact, that $\widetilde{y}_0$, $\widetilde{\kappa}_{j0}$, $\mathbf{Y}$, Θ_j, $\widetilde{w}_j$, $\mathbf{h}_j$ and $\mathbf{Z}_j$, for $j = 1, \dots, J$, fulfill assumptions (D-5) and (D-6), respectively, follows straight. Moreover, $\left(\widetilde{g}_j, \widetilde{h}_j\right)_{j=1,\dots,J} \in \widetilde{U}$.

To sum up, the assumptions of Theorem 1.2.18 are fulfilled and hence there exists a unique weak solution of the system (1.84) - (1.86) with conditions (3.32). ■

Now, we present the main theorem of Section 3.2.2, which gives a characterization of Gâteaux differential of the cost functional $\mathcal{I}$ in the form given in (3.25).

Theorem 3.2.5 *Let assumptions (B-1) - (B-5) be fulfilled, with additional restriction $K = J$. Assume also that at least one of the following is true:*

- *y^* fulfills the assumption (C-1) and functions w_j are bounded, for $j = 1, \dots, J$,*

- *y^* fulfills the assumption (C-2).*

Moreover, let assumptions (E-1) - (E-3) and (F-2) be fulfilled and let $\widetilde{\lambda} > 0$, $T_0 \in (0,T)$ *and* $\hat{v}, \hat{\eta} \in V$ *be given. Let also* $(\hat{y}, \hat{\kappa}_1, \ldots, \hat{\kappa}_J) = \mathcal{Z}(\hat{v})$ *and let* $(\widetilde{p}, \widetilde{q}_1, \ldots, \widetilde{q}_J)$ *be the weak solution of the system (3.30) - (3.31) corresponding to* $\hat{Y} := \hat{y}$.

Then, the cost functional $\mathcal{I}$*, defined in (3.22) - (3.23), is Gâteaux differentiable and its differential in point* $\hat{v}$ *in direction* $\hat{\eta}$ *is equal to* $D_G\mathcal{I}(\hat{v})(\hat{\eta}) = \left(\Lambda^{\hat{v}}, \hat{\eta}\right)_V$*, where* $\Lambda^{\hat{v}} \in V$ *is given by:*

$$\begin{aligned}\Lambda^{\hat{v}} &= \sum_{j=1}^{J} 2\widetilde{\lambda} \left(D_{G,w}\Upsilon_{g_j}(\hat{v})\right)^* \Big(\int_0^T \hat{\kappa}_j \widetilde{p}\, dt\Big) + \\ &+ \sum_{j=1}^{J} 2\widetilde{\lambda} \left(D_{G,w}\Upsilon_{h_j}(\hat{v})\right)^* \Big(\int_0^T w_j'\Big(\int_\Omega \Upsilon_{h_j}(\hat{v})(\hat{y}-y^*)\, dx\Big)\, (\hat{y}-y^*)\, \widetilde{q}_j\, dt\Big)\end{aligned} \tag{3.33}$$

The characterization of Gâteaux differential of the cost functional $\mathcal{I}$ given in Theorem 3.2.5 is not explicit, since the adjoint operators entering the formula (3.33) are not explicitly described. Hence, below we provide a theorem characterizing the latter operators.

Theorem 3.2.6 *Let the assumption (F-2) be fulfilled. Let also* $\hat{v} \in V$ *be given. Then, the adjoint operators* $\left(D_{G,w}\Upsilon_{g_j}(\hat{v})\right)^* : L^2(\Omega) \longrightarrow V$*, for* $j = 1, \ldots, J$*, are well defined and are characterized by the following formulas:*

$$\left(D_{G,w}\Upsilon_{g_j}(\hat{v})\right)^* \hat{F} = \Big(\underbrace{\mathbf{0}, \ldots, \mathbf{0}}_{j-1}, \underbrace{\left(D_{G,w}\mathcal{T}_{\sigma_g}(\hat{v}_j)\right)^* \mathcal{P}^{E,\Omega} \hat{F}}_{j\text{-th position}}, \underbrace{\mathbf{0}, \ldots, \mathbf{0}}_{J-j}\Big) \tag{3.34}$$

for $\hat{F} \in L^2(\Omega)$*, where* $\mathbf{0} \in \mathbb{R}^{\mathbf{d}}$ *and where the non-zero element on* j*-th position can be expressed by*

$$\left(D_{G,w}\mathcal{T}_{\sigma_g}(\hat{v}_j)\right)^* \mathcal{P}^{E,\Omega} \hat{F} = \Big(-\int_\Omega \hat{F}(z) \left(\mathcal{P}^{R,\Omega}\mathcal{T}_{\partial_i \sigma_g}(\hat{v}_j)\right)(z)\, dz\Big)_{i=1}^{\mathbf{d}} \tag{3.35}$$

The adjoint operators $\left(D_{G,w}\Upsilon_{h_j}(\hat{v})\right)^* : L^2(\Omega) \longrightarrow V$*, for* $j = 1, \ldots, J$*, are also well defined and are characterized by the same formulas, with* σ_g *replaced by* σ_h.

We recall that, in the present chapter, the particular operators entering the above formulas are understood as $\mathcal{P}^{E,\Omega} \colon L^2(\Omega) \to L^2(\mathbb{R}^{\mathbf{d}})$ and $\mathcal{T}_{\sigma_g}, \mathcal{T}_{\partial_i \sigma_g} \colon \mathbb{R}^{\mathbf{d}} \to L^2(\mathbb{R}^{\mathbf{d}})$.

Now, we present the proofs of Theorem 3.2.5 and Theorem 3.2.6.

Proof Theorem 3.2.5. The Gâteaux differentiability of $\mathcal{I}$ was already explained in Lemma 3.2.2 (note, that its assumptions are fulfilled in the present theorem). It remains to justify formulas characterizing the subject Gâteaux differential.

We will begin with justifying that the formula (3.33) is well-posed. For this end, note that the assumptions of Theorem 3.1.5 are fulfilled, hence $(\hat{y}, \hat{\kappa}_1, \ldots, \hat{\kappa}_J)$ in the assumptions of the present theorem is a well defined element of X^2. With this, assumptions of Lemma 3.2.4 are also fulfilled, hence $(\widetilde{p}, \widetilde{q}_1, \ldots, \widetilde{q}_J)$ in the assumptions is a well defined elements of X^2 as well. This, together with the Fubini theorem and the Hölder inequality, allows to justify that $\int_0^T \hat{\kappa}_j(t)\widetilde{p}(x,t)\, dt$, understood as a function of x, is a well defined element of $L^2(\Omega)$, hence it belongs to the domain of $\left(D_{G,w}\Upsilon_{g_j}(\hat{v})\right)^*$, for $j = 1, \ldots, J$. Similarly, we can find out that, for $j = 1, \ldots, J$, expression $\int_0^T w_j' \left(\int_\Omega \Upsilon_{h_j}(\hat{v})(\hat{y}-y^*)\, dx\right) (\hat{y}-y^*)\, \widetilde{q}_j dt$ belongs to the domain of $\left(D_{G,w}\Upsilon_{h_j}(\hat{v})\right)^*$, i.e. to $L^2(\Omega)$. Indeed, arguing as in the proof of Lemma 3.1.7, we get that $w_j' \left(\int_\Omega \Upsilon_{h_j}(\hat{v})(\hat{y}-y^*)\, dx\right)$ belongs to $L^\infty(0,T)$. This, along with $\hat{y} \in L^2(Q_T)$, $\widetilde{q}_j \in L^2(0,T)$, with assumptions for y^*, with the Fubbini theorem and with the Hölder inequality justifies the necessary. Thus, the formula (3.33) is meaningful.

Next, assumptions of Lemma 3.1.7 are fulfilled. Hence, the weak solution of (3.9) - (3.10) exists and is unique. Denote this weak solution as $(\widetilde{y}, \widetilde{\kappa}_1, \ldots, \widetilde{\kappa}_J)$. By Definition 3.1.6, it means that the identity in the part b) of Definition 1.2.16 is fulfilled with $y := \widetilde{y}$ and the identity in the part c) of Definition 1.2.16 is fulfilled with $\kappa_j := \widetilde{\kappa}_j$, with relations (3.11) utilized there. Since $X^2 \hookrightarrow L^2(0,T;H^1(\Omega))$, the element $(\widetilde{p}, \widetilde{q}_1, \ldots, \widetilde{q}_J)$ can serve as a test function in the referred identities, by putting $\phi := \widetilde{p}$ in the part b) and, for $j = 1, \ldots, J$, putting $\xi := \widetilde{q}_j$ in the part c) of Definition 1.2.16, with relations (3.11) applied there. Executing the above described substitutions and utilizing relations (3.11) in the subject identities, we get:

$$\begin{aligned}\int_0^T \langle \widetilde{y}_t, \widetilde{p} \rangle + D\big(\nabla\widetilde{y}, \nabla\widetilde{p}\big)_{L^2(\Omega)} + \Big(-f'(\hat{y})\widetilde{y} - \sum_{j=1}^J \varUpsilon_{g_j}(\hat{\upsilon})\widetilde{\kappa}_j \,,\, \widetilde{p}\Big)_{L^2(\Omega)} \,dt \;&= \\ = \;\int_0^T \Big(\sum_{j=1}^J D_{G,w}\varUpsilon_{g_j}(\hat{\upsilon})(\hat{\eta})\hat{\kappa}_j \,,\, \widetilde{p}\Big)_{L^2(\Omega)} \,dt&\end{aligned} \tag{3.36a}$$

$$\begin{aligned}\int_0^T \Big\{\beta_j\widetilde{\kappa}_j' + \widetilde{\kappa}_j - w_j'\Big(\int_\Omega \varUpsilon_{h_j}(\hat{\upsilon})(\hat{y} - y^*)\,dx\Big)\cdot\Big(\int_\Omega \varUpsilon_{h_j}(\hat{\upsilon})\widetilde{y}\,dx\Big)\Big\}\,\widetilde{q}_j\,dt \;&= \\ = \;\int_0^T w_j'\left(\int_\Omega \varUpsilon_{h_j}(\hat{\upsilon})(\hat{y} - y^*)\,dx\right)\cdot& \\ \cdot\left(\int_\Omega D_{G,w}\varUpsilon_{h_j}(\hat{\upsilon})(\hat{\eta})(\hat{y} - y^*)\,dx\right)\widetilde{q}_j\,dt \qquad &\text{for } j = 1, \ldots, J\end{aligned} \tag{3.36b}$$

Similarly, $\Big(\mathcal{P}^i_{Q_T}\widetilde{y}, \mathcal{P}^i_T\widetilde{\kappa}_1, \ldots, \mathcal{P}^i_T\widetilde{\kappa}_J\Big)$ can serve as a test function for weak solution $(\widetilde{p}, \widetilde{q}_1, \ldots, \widetilde{q}_J)$ of the system (3.30) - (3.31). More precisely, by Definition 3.2.3, in the identity in the part b) of Definition 1.2.16 we can put $y := \mathcal{P}^i_{Q_T}\widetilde{p}$, $\phi := \mathcal{P}^i_{Q_T}\widetilde{y}$ and, for $j = 1, \ldots, J$, in the identity in the part c) of Definition 1.2.16 we can put $\kappa_j := \mathcal{P}^i_T\widetilde{q}_j$, $\xi := \mathcal{P}^i_T\widetilde{\kappa}_j$, together with utilizing relations (3.32). Executing the above substitutions, utilizing relations (3.32) in the subject identities, integrating the time derivative terms by parts w.r.t. t (see Prop. 23.23 in [51] for the integration by parts formula for vector valued functions) and inverting the time direction by applying operators $\mathcal{P}^i_{Q_T}$ and $\mathcal{P}^i_T$, we get:

$$\begin{aligned}&\int_0^T \langle \widetilde{y}_t, \widetilde{p} \rangle + D\big(\nabla\widetilde{p}, \nabla\widetilde{y}\big)_{L^2(\Omega)} + \\ &\qquad + \;\Big(-f'(\hat{y})\widetilde{p} - \sum_{j=1}^J w_j'\Big(\int_\Omega \varUpsilon_{h_j}(\hat{\upsilon})(y - y^*)\,dx\Big)\,\varUpsilon_{h_j}(\hat{v})\,\widetilde{q}_j \,,\, \widetilde{y}\Big)_{L^2(\Omega)}\,dt \;= \\ &\quad = \;\int_0^T \big((\hat{y} - y^*)\mathbf{1}_{(T_0,T)} \,,\, \widetilde{y}\big)_{L^2(\Omega)}\,dt\end{aligned} \tag{3.37a}$$

$$\int_0^T \Big(\beta_j\widetilde{q}_j' + \widetilde{q}_j - \big(\varUpsilon_{g_j}(\hat{\upsilon}), \widetilde{p}\big)_{L^2(\Omega)}\Big)\,\widetilde{\kappa}_j\,dt \;=\; 0 \qquad \text{for } j = 1, \ldots, J \tag{3.37b}$$

Comparing (3.36) and (3.37), we observe that the sum of the left hand sides of (3.36) equals the sum of the left hand sides of (3.37). Hence, the sums of the right hand sides of (3.36) and of

(3.37) also equal. Thus, after changing the order of integration in these sums, we get:

$$\begin{aligned}\int_0^T \big((\hat{y}-y^*)\mathbf{1}_{(T_0,T)}\,,\,\widetilde{y}\big)_{L^2(\Omega)}\,dt \;&= \\ = \;&\sum_{j=1}^{J}\Big(\int_0^T \kappa_j\widetilde{p}\,dt\,,\,D_{G,w}\Upsilon_{g_j}(\hat{\upsilon})(\hat{\eta})\Big)_{L^2(\Omega)} + \\ &+ \sum_{j=1}^{J}\Big(\int_0^T w_j'\Big(\int_\Omega \Upsilon_{h_j}(\hat{\upsilon})(y-y^*)\,dx\Big)\,(y-y^*)\,\widetilde{q}_j\,dt\,,\,D_{G,w}\Upsilon_{h_j}(\hat{\upsilon})(\hat{\eta})\Big)_{L^2(\Omega)}\end{aligned} \tag{3.38}$$

Recall that $\widetilde{y}=\big(D_{G,w}\mathcal{Z}(\hat{\upsilon})(\hat{\eta})\big)_y = D_{G,w}\mathcal{Z}_y(\hat{\upsilon})(\hat{\eta})$ and $\hat{y}=\mathcal{Z}_y(\hat{\upsilon})$. By the definition of $\mathcal{Z}_y^{T_0}$ and y^{*T_0}, see (3.23), and by (3.27) in Lemma 3.2.2, we deduce that

$$\begin{aligned}\int_0^T \big((\hat{y}-y^*)\mathbf{1}_{(T_0,T)},\widetilde{y}\big)_{L^2(\Omega)}\,dt \;&=\; \int_{T_0}^T \Big(\mathcal{P}^{R,T_0}(\hat{y}-y^*)\,,\,\mathcal{P}^{R,T_0}\widetilde{y}\Big)_{L^2(\Omega)}\,dt \\ &= \Big(\mathcal{Z}_y^{T_0}(\hat{\upsilon})-y^{*T_0}\,,\,D_{G,w}\mathcal{Z}_y^{T_0}(\hat{\upsilon})(\hat{\eta})\Big)_{L^2(Q_T^{T_0})}\end{aligned} \tag{3.39}$$

Identities (3.38) and (3.39), by involving adjoint operators $\big(D_{G,w}\Upsilon_{g_j}(\hat{\upsilon})\big)^*$ and $\big(D_{G,w}\Upsilon_{h_j}(\hat{\upsilon})\big)^*$, for $j=1,\dots,J$, and by Lemma 3.2.2, justifies the assertion of Theorem 3.2.5. ■

Proof of Theorem 3.2.6. We will prove the assertion for operators $\big(D_{G,w}\Upsilon_{g_j}(\hat{\upsilon})\big)^*$. The case of operators $\big(D_{G,w}\Upsilon_{h_j}(\hat{\upsilon})\big)^*$ follows the same lines.

To prove the required, we repeat some arguments from the proof of Lemma 3.1.4. We observe that $\Upsilon_{g_j}=\mathcal{P}^{R,\Omega}\circ\mathcal{T}_{\sigma_g}\circ\mathcal{P}_j^{R,V}$, where the particular operators are understood as $\mathcal{P}^{R,\Omega}\colon L^2(\mathbb{R}^{\mathbf{d}})\to L^2(\Omega)$, $\mathcal{T}_{\sigma_g}\colon\mathbb{R}^{\mathbf{d}}\to L^2(\mathbb{R}^{\mathbf{d}})$ and $\mathcal{P}_j^{R,V}\colon V\to\mathbb{R}^{\mathbf{d}}$. Since $\sigma_g\in W^{1,2}(\mathbb{R}^{\mathbf{d}})$, we can apply Theorem A.4.5 to conclude that $\mathcal{T}_{\sigma_g}$ is weakly Gâteaux differentiable. Moreover, operators $\mathcal{P}^{R,\Omega}$ and $\mathcal{P}_j^{R,V}$ are linear and continuous. Thus, we can combine the above facts with Observation A.1.7, Observation A.1.11, Theorem A.1.4 and, for brevity, use identities $\hat{\upsilon}_j=\mathcal{P}_j^{R,V}(\hat{\upsilon})$ and $\hat{\eta}_j=\mathcal{P}_j^{R,V}(\hat{\eta})$ to get that Υ_{g_j} is weakly Gâteaux differentiable from V to $L^2(\Omega)$ and

$$D_{G,w}\Upsilon_{g_j}(\hat{\upsilon})(\hat{\eta}) \;=\; \mathcal{P}^{R,\Omega}\left(D_{G,w}\mathcal{T}_{\sigma_g}(\hat{\upsilon}_j)(\hat{\eta}_j)\right)$$

for arbitrary $\hat{\upsilon},\hat{\eta}\in V$.

In consequence, as operators $D_{G,w}\Upsilon_{g_j}(\hat{\upsilon})\colon V\to L^2(\Omega)$ are well defined for $j=1,\dots,J$, the adjoint operators also are well defined, what justifies the corresponding statement the second assertion of the theorem.

Next, we note that $\big(\mathcal{P}^{R,\Omega}\big)^*=\mathcal{P}^{E,\Omega}\colon L^2(\Omega)\to L^2(\mathbb{R}^{\mathbf{d}})$ and $\Big(\mathcal{P}_j^{R,V}\Big)^*=\mathcal{P}_j^{E,V}\colon\mathbb{R}^{\mathbf{d}}\to V$. Using this and the above derived representation of $D_{G,w}\Upsilon_{g_j}(\hat{\upsilon})(\hat{\eta})$, we conclude the following:

$$\begin{aligned}\Big(\hat{F}\,,\,D_{G,w}\Upsilon_{g_j}(\hat{\upsilon})(\hat{\eta})\Big)_{L^2(\Omega)} \;&=\; \Big(\hat{F}\,,\,\mathcal{P}^{R,\Omega}\left(D_{G,w}\mathcal{T}_{\sigma_g}(\hat{\upsilon}_j)\right)\hat{\eta}_j\Big)_{L^2(\Omega)} \;= \\ =\Big(\left(D_{G,w}\mathcal{T}_{\sigma_g}(\hat{\upsilon}_j)\right)^*\mathcal{P}^{E,\Omega}\hat{F}\,,\,\hat{\eta}_j\Big)_{\mathbb{R}^{\mathbf{d}}} \;&=\; \Big(\mathcal{P}_j^{E,V}\left(D_{G,w}\mathcal{T}_{\sigma_g}(\hat{\upsilon}_j)\right)^*\mathcal{P}^{E,\Omega}\hat{F}\,,\,\hat{\eta}\Big)_V\end{aligned}$$

Taking into account the definition of $\mathcal{P}_j^{E,V}$, the above justifies the formula (3.34).

Now, we are left to find the characterization of the adjoint of $D_{G,w}\mathcal{T}_{\sigma_g}(\hat{\upsilon}_j)(\,.\,)$, still not explicit above. Functions σ_g and σ_h satisfy the assumption (F-2), thus by the Theorem A.4.5 we have an explicit characterization of the differentials of $\mathcal{T}_{\sigma_g}$ and $\mathcal{T}_{\sigma_h}$ at our disposal. This helps us to achieve our goal:

$$\begin{aligned}
\Big(\mathcal{P}^{E,\Omega}\hat{F}\,,\,D_{G,w}\mathcal{T}_{\sigma_g}(\hat{\upsilon}_j)\hat{\eta}_j\Big)_{L^2(\mathbb{R}^{\mathbf{d}})} &= \Big(\mathcal{P}^{E,\Omega}\hat{F}\,,\,\big(-\mathcal{T}_{\nabla\sigma_g}(\hat{\upsilon}_j),\hat{\eta}_j\big)_{\mathbb{R}^{\mathbf{d}}}\Big)_{L^2(\mathbb{R}^{\mathbf{d}})} = \\
= \Big(\Big(-\int_{\mathbb{R}^{\mathbf{d}}}(\mathcal{P}^{E,\Omega}\hat{F})(z)\mathcal{T}_{\partial_i\sigma_g}(\hat{\upsilon}_j)(z)\,dz\Big)_{i=1}^{\mathbf{d}}\,,\,\hat{\eta}_j\Big)_{\mathbb{R}^{\mathbf{d}}} &= \\
= \Big(\Big(-\int_{\Omega}\hat{F}(z)\,\big(\mathcal{P}^{R,\Omega}\mathcal{T}_{\partial_i\sigma_g}(\hat{\upsilon}_j)\big)\,(z)\,dz\Big)_{i=1}^{\mathbf{d}}\,,\,\hat{\eta}_j\Big)_{\mathbb{R}^{\mathbf{d}}}
\end{aligned}$$

The above justifies the formula (3.35).

This concludes the proof of Theorem 3.2.6. ■

Thanks to Theorem 3.2.6, we can write the formula for $\Lambda^{\hat{\upsilon}} \in V$, asserted in Theorem 3.2.5, in a more explicit form:

Corollary 3.2.7 *Let assumptions imposed in Theorem 3.2.5 be fulfilled. Then, for $\hat{\upsilon} \in V$, the weak Gâteaux differential in $\hat{\upsilon}$ of the cost functional $\mathcal{I}$, defined in (3.22) - (3.23), exists and can be characterized by o $D_G\mathcal{I}(\hat{\upsilon})(\hat{\eta}) = \big(\Lambda^{\hat{\upsilon}},\hat{\eta}\big)_V$ for $\hat{\eta} \in V$, where $\Lambda^{\hat{\upsilon}} \in V$ is given by:*

$$\begin{aligned}
\big(\Lambda^{\hat{\upsilon}}_j\big)_i &= 2\widetilde{\lambda}\int_0^T\int_\Omega \hat{\kappa}_j\,\widetilde{p}\,\big(\mathcal{P}^{R,\Omega}\circ\mathcal{T}_{-\partial_i\sigma_g}\big)(\hat{\upsilon}_j)\,dx\,dt \\
&+ 2\widetilde{\lambda}\int_0^T\int_\Omega \widetilde{q}_j\,w_j'\Big(\int_\Omega \Upsilon_{h_j}(\hat{\upsilon})(\hat{y}-y^*)\,dx\Big)\,(\hat{y}-y^*)\,\big(\mathcal{P}^{R,\Omega}\circ\mathcal{T}_{-\partial_i\sigma_h}\big)(\hat{\upsilon}_j)\,dx\,dt
\end{aligned} \tag{3.40}$$

for $j = 1,\dots,J$, for $i = 1,\dots,\mathbf{d}$, where $(\hat{y},\hat{\kappa}_1,\dots,\hat{\kappa}_J)$ is the weak solution of the system (3.1) - (3.2) corresponding to $x_j := \hat{\upsilon}_j$, for $j = 1,\dots,J$, and $(\widetilde{p},\widetilde{q}_1,\dots,\widetilde{q}_J)$ is the weak solution of the system (3.30) - (3.31), corresponding to $\hat{y}$.

The above comes by combining formulas (3.33) and (3.34) - (3.35), changing the order of integration and noting that $-\mathcal{T}_{\partial_i\sigma}(\hat{\upsilon}_j)(x) = \mathcal{T}_{-\partial_i\sigma}(\hat{\upsilon}_j)(x)$ for a.e. $x \in \Omega$, for $\sigma = \sigma_g, \sigma_h$, $j = 1,\dots,J$, $i = 1,\dots,\mathbf{d}$.

Remark. Note that the formula (3.40) is explicit enough to approximate it with numerical methods. Indeed, for a given σ_g and σ_h, functions $\mathcal{T}_{-\partial_i\sigma_g}$, $\mathcal{T}_{-\partial_i\sigma_h}$ and $\Upsilon_{h_j}(\hat{\upsilon})$, entering (3.40), can be expressed explicitly by their definitions. Thus, assuming that one is able to find numerical approximations of solutions $(\hat{y},\hat{\kappa}_1,\dots,\hat{\kappa}_J)$ and $(\widetilde{p},\widetilde{q}_1,\dots,\widetilde{q}_J)$, the formula (3.40) can be approximately evaluated with a use of numerical integration methods. ▲

Formulating necessary optimality condition is a usual step towards characterizing the solutions of a considered optimization problem. A first choice necessary optimality condition is frequently the generalization of the Fermat condition for multidimensional sets given in Theorem A.2.1 in Appendix A.2). Applying the latter requires the knowledge on the Gâteaux differential of the cost functional. In the case of optimization problem (3.24), we can use Theorem A.2.1 along with the characterization of $D_G\mathcal{I}$, provided by Corollary 3.2.7, to obtain the following necessary optimality condition:

Corollary 3.2.8 *Let the assumptions of Corollary 3.2.7 hold. If $\hat{\upsilon} \in V$ solves the optimization problem (3.24) then condition*

$$\big(\Lambda^{\hat{\upsilon}},\hat{\nu}-\hat{\upsilon}\big)_V \geq 0 \qquad \forall_{\hat{\nu}\in V}$$

is fulfilled, for $\Lambda^{\hat{\upsilon}}$ as in Corollary 3.2.7.

3.2.3 Generalizations for locally Lipschitz reactive term

In the present section, we prove results for optimization problem (3.24) under assumptions different that those utilized in the main results of 3.2.1 and Section 3.2.2. In the results of the latter sections, it was assumed for the system (3.1) - (3.2), in particular, that f is Lipschitz and that $y_0 \in L^2(\Omega)$. Below, we will change the Lipschitz continuity of f to local Lipschitz continuity plus the growth condition given in (1.73) and we will change the assumption for y_0 to $y_0 \in L^\infty(\Omega)$. Moreover, we will require higher integrability of the pattern function σ_g.

Below, we justify analogues of the previously proven theorems concerning existence of minimizers for the cost functional $\mathcal{I}$ (Theorem 3.2.1) and the characterization of its gradient (Theorem 3.2.5), but with the above mentioned modifications in the assumptions.

The purpose of the present section is the following. In Chapter 4 of the present work, we describe numerical simulations for optimization problem (3.24). The subject simulations involved data assuming locally Lipschitz f satisfying the condition (1.73) and $y_0 \in L^\infty(\Omega)$. For this reason, we aimed in deriving analytical results covering the case of the data utilized in the simulations. Hence the below content.

The proofs presented below, in their essence, consist in reducing optimization problem (3.24) with locally Lipschitz f obeying (1.73) to optimization problem (3.24) with globally Lipschitz f. Since for globally Lipschitz f the existence of minimizers and the formula for the gradient of the cost functional are already known (Theorem 3.2.1 and Theorem 3.2.5), the mentioned reduction will imply the necessary results.

For the proof of Theorem Theorem 3.2.5, the theorem on the differentiability of the state operator $\mathcal{Z}$, associated with globally Lipschitz and differentiable f, was crucial. The reduction approach in the present section allow to avoid direct analysis of differentiability of the state operator $\mathcal{Z}$ associated with locally Lipschitz f.

In Section 3.2.3, we proceed as follows. We begin with introducing some definitions and notations which will be necessary in the sequel. Next, we formulate simple results concerning existence and uniqueness of the weak solutions for the case of the modified assumptions for f, y_0 and σ_g mentioned above. The subject existence and uniqueness results concern the system (3.1) - (3.2), the system (3.30) - (3.31) and certain associated systems, which will be defined below for technical reasons. Eventually, we proceed to proving analogues of Theorem 3.2.1 and Theorem 3.2.5 for the modified assumptions for f, y_0 and σ_g .

Let us proceed to formulation of the necessary definitions.

For continuous $f\colon \mathbb{R} \to \mathbb{R}$, for $n > 0$ it possible to define the following function $f^n\colon \mathbb{R} \to \mathbb{R}$:

$$\begin{aligned} &f^n(s) := f(s) && \text{for } s \in (-n, n) \\ &f^n(s) := f(-(n+1)) && \text{for } s < -(n+1) \\ &f^n(s) := f(n+1) && \text{for } s > n+1 \end{aligned} \tag{3.41}$$

and

$$\begin{cases} f^n \text{ is linear on } [-(n+1), -n], \text{ linear on } [n, n+1] \text{ and} \\ f^n(-(n+1)) := f(-(n+1)) \qquad f^n(n+1) := f(n+1) \\ \qquad f^n(-n) := f(-n) \qquad\qquad\quad f^n(n) := f(n) \end{cases} \tag{3.42}$$

If, in addition, $f'(s)$ exists for all $s \in \mathbb{R}$, it is meaningful to define f^n by the condition (3.41) and by the following condition instead of (3.42):

$$\begin{cases} f^n \text{ is 3rd degree polynomial on } [-(n+1),-n]\text{, 3rd deg. pol. on } [n,n+1] \text{ and} \\ \begin{array}{rr} f^n(-(n+1)) := f(-(n+1)) & f^n(n+1) := f(n+1) \\ f^n(-n) := f(-n) & f^n(n) := f(n) \\ f^{n\prime}(-(n+1)) := f'(-(n+1)) & f^{n\prime}(n+1) := f'(n+1) \\ f^{n\prime}(-n)) := f'(-n) & f^{n\prime}(n) := f'(n) \end{array} \end{cases} \tag{3.43}$$

The following observations are straightforward:

Observation 3.2.9 *If $f\colon \mathbb{R} \to \mathbb{R}$:*

- *is continuous, then f^n defined by (3.41) and (3.42) is so, for all $n > 0$.*
- *is differentiable in every point of $\mathbb{R}$, then f^n defined by (3.41) and (3.43) is so, for all $n > 0$.*
- *is locally Lipschitz, then f^n defined by (3.41) and (3.42) as well as f^n defined by (3.41) and (3.43) are globally Lipschitz, for all $n > 0$.*
- *obeys the condition (1.73) with constant C_f, then f^n defined by (3.41) and (3.42) as well as as well as f^n defined by (3.41) and (3.43) also obey the condition (1.73), with the same C_f, for all positive n such that $n + 1 \geq C_f$.*

In the present content, we still assume that $\mathcal{Z}\colon V \to X^2$ and $\Upsilon\colon V \to U$ are defined as in Section 3.1.2 and $\mathcal{I}\colon V \to \mathbb{R}$ is defined by conditions (3.22) - (3.23). However, in the below considerations, it will be convenient to have the following additional notation. Assume that arbitrary functions $f^n\colon \mathbb{R} \to \mathbb{R}$ are given, for all $n > 0$. Then, for $n > 0$:

- The system (3.1) - (3.2) with f^n instead of f will be denoted by $\big((3.1)\text{ - }(3.2)\big)^n$.
- The system (3.30) - (3.31) with $f^{n\prime}$ instead of f' will be denoted by $\big((3.30)\text{ - }(3.31)\big)^n$.
- By $\mathcal{Z}^n$, where
$$\mathcal{Z}^n = (\mathcal{Z}^n_y, \mathcal{Z}^n_{\kappa_1}, \ldots, \mathcal{Z}^n_{\kappa_J})\colon\ V \longrightarrow X^2$$
we will understand the operator assigning the weak solution of $\big((3.1)\text{ - }(3.2)\big)^n$ to a given $\hat{v} \in V$, assuming assignment $x_j := \hat{v}_j$ for $j = 1, \ldots, J$ in $\big((3.1)\text{ - }(3.2)\big)^n$.
- By $\mathcal{I}^n\colon V \to \mathbb{R}$ we will understand the cost functional given by (3.22) - (3.23), with $\mathcal{Z}^n_y$ instead of $\mathcal{Z}_y$.

Now, we pass to existence facts for systems (3.1) - (3.2), (3.30) - (3.31), $\big((3.1)\text{ - }(3.2)\big)^n$ and $\big((3.30)\text{ - }(3.31)\big)^n$. The following facts are corollaries from earlier considerations in the present work:

Corollary 3.2.10 *In the system (3.1) - (3.2), let assumptions (B-1), (B-2) and (B-4) be fulfilled, with additional restriction $K = J$. Moreover, assume that*

- *f is Locally Lipschitz continuous and obeys the condition (1.73), for some constant $C_f > 0$,*
- *$y_0 \in L^\infty(\Omega)$ and $\kappa_{j0} \in \mathbb{R}$ for $j = 1, \ldots, J$,*

- *$\sigma_g \in L^{s_1}(\mathbb{R}^{\mathbf{d}})$ and $\sigma_h \in L^2(\mathbb{R}^{\mathbf{d}})$, where $s_1 \geq \max\{2, \frac{\mathbf{d}}{2}\}$.*

Let also at least one of the below conditions hold:

- *y^* is as in (C-1) and functions w_j are bounded, for $j = 1, \ldots, J$,*
- *y^* is as in (C-2).*

Then, there exist a unique weak solution of the system (3.1) - (3.2). In consequence, the operator $\mathcal{Z}\colon V \to X^2$ and the cost functional $\mathcal{I}\colon V \to \mathbb{R}$ are well defined.

Proof. To prove Corollary 3.2.10, note that the system (3.1) - (3.2) is a particular case of the system (0.1) - (0.3), with $K = J$ and with $\big(g_j, h_j, \alpha_{jk}\big)_{j,k} := \varUpsilon(\hat{v})$. Hence, by Theorem 1.2.14 and Theorem 1.2.15, we obtain the assertion. ■

Corollary 3.2.11 *Let the assumptions of Corollary 3.2.10 be fulfilled. Let functions f^n for $n > 0$ be given by (3.41) and (3.42). Then, for $n > 0$, there exist a unique weak solution of the system $\big((3.1)$ - $(3.2)\big)^n$. In consequence, the operator $\mathcal{Z}^n\colon V \to X^2$ and cost functional $\mathcal{I}^n\colon V \to \mathbb{R}$ are well defined, for $n > 0$.*

If, in addition, $f'(s)$ exist for all $s \in \mathbb{R}$, then the above assertion holds also for functions f^n given by (3.41) and (3.43), for $n > 0$.

Above, the assumption that f' exists everywhere is necessary only to guarantee that f^n is well defined for $n > 0$, in the case where f^n is defined by conditions (3.41) and (3.43).

Proof. First, consider the case of y^* is as in (C-1) and bounded functions w_j, $j = 1, \ldots, J$. For f as assumed in Corollary 3.2.10, functions f^n are Lipschitz, for both f^n defined by (3.41) and (3.42) and f^n defined by (3.41) and (3.43) (see Observation 3.2.9). Thus, one can verify that the system $\big((3.1)$ - $(3.2)\big)^n$ meets the assumptions of Corollary 1.2.8 with f^n instead of f, regardless on the variant of f^n. Hence, the assertion follows by Corollary 1.2.8.

The case of y^* is as in (C-2) follows exactly the same lines, with the use of Corollary 1.2.9 instead of the use of Corollary 1.2.8. ■

Corollary 3.2.12 *In the system (3.30) - (3.31), let assumptions (B-1), (B-2), (B-4) be fulfilled, with additional restriction $K = J$ and assume that $f\colon \mathbb{R} \to \mathbb{R}$ is Locally Lipschitz continuous and obeys the condition (1.73), for some constant $C_f > 0$. Assume also that (E-1) - (E-2) and (F-1) hold. Moreover, assume that $\hat{Y} \in L^\infty(Q_T)$ and $y^* \in L^2(0, T; L^2(\Omega))$.*

Then, the weak solution of the system (3.30) - (3.31) exists and is unique.

Corollary 3.2.13 *Let the assumptions of Corollary 3.2.12 be fulfilled. Let functions f^n for $n > 0$ be given by (3.41) and (3.43). Then, for $n > 0$, there exist a unique weak solution of the system $\big((3.30)$ - $(3.31)\big)^n$.*

We have formulated Corollary 3.2.12 prior to Corollary 3.2.13, for the sake of more readable presentation. But technically, Corollary 3.2.13 should be proven first.

Proof of Corollary 3.2.13. For f as assumed in Corollary 3.2.12, functions f^n as assumed in Corollary 3.2.13 are Lipschitz and differentiable (see Observation 3.2.9). Thus, for $n > 0$, the system $\big((3.30)$ - $(3.31)\big)^n$ obeys assumptions of Lemma 3.2.4 with f^n instead of f. Hence, by Lemma 3.2.4, the assertion follows. ■

Proof of Corollary 3.2.12. Let f^n be given by (3.41) and (3.43), for $n > 0$. If suffices to show that arbitrary weak solution of (3.30) - (3.31) is a weak solution of $\big((3.30)$ - $(3.31)\big)^n$, for certain n, and that arbitrary weak solution of $\big((3.30)$ - $(3.31)\big)^n$ is a weak solution of (3.30) - (3.31). Having this, the assertion follows by Corollary 3.2.13.

Chose $\widetilde{n} > \|\hat{Y}\|_{L^\infty(Q_T)}$. By the condition (3.41), we have

$$(f)'(\hat{Y}(x,t)) \;=\; (f^{\widetilde{n}})'(\hat{Y}(x,t)) \qquad \text{for a.e. } (x,t) \in Q_T$$

Thus, by Definition 3.2.3, every weak solution of (3.30) - (3.31) is a weak solution of $\big((3.30)$ - $(3.31)\big)^{\widetilde{n}}$ and every weak solution of $\big((3.30)$ - $(3.31)\big)^{\widetilde{n}}$ is a weak solution of (3.30) - (3.31). This closes the proof. ■

We proceed to the key part of Section 3.2.3. The below statements, which are the main statements of Section 3.2.3, rely strongly on Theorem 1.2.13.

Theorem 3.2.14 *Let the system (3.1) - (3.2) fulfill the assumptions of Theorem 3.2.1, except the assumptions concerning f, y_0 and σ_g. For f, y_0 and σ_g, we make the following assumptions*

- *f is locally Lipschitz continuous and obeys the condition (1.73) for certain constant C_f,*
- *$y_0 \in L^\infty(\Omega)$,*
- *σ_g obeys assumptions (F-1) and (F-3) and, in addition, $\sigma_g \in L^{s_1}(\mathbb{R}^{\mathbf{d}})$ for certain $s_1 \geq \frac{\mathbf{d}}{2}$.*

Then, the optimization problem (3.24) attains at least one solution.

Proof. Let functions f^n be defined by (3.41) and (3.42), for $n > 0$.

Let $\hat{v} \in V$. Denote $(y, \kappa_1, \ldots, \kappa_J) = \mathcal{Z}(\hat{v})$ (what is well defined, see Corollary 3.2.10). By the definition of $\mathcal{Z}$, $(y, \kappa_1, \ldots, \kappa_J)$ is the weak solution of the system (3.1) - (3.2) corresponding to $x_j := \hat{v}_j$, $j = 1, \ldots, J$.

The system (3.1) - (3.2) with $x_j := \hat{v}_j$, $j = 1, \ldots, J$ is a particular case of the system (0.1) - (0.3), with $K = J$ and with $\big(g_j, h_j, \alpha_{jk}\big)_{j,k} := \Upsilon(\hat{v})$. By the assumptions presently imposed for the system (3.1) - (3.2), Theorem 1.2.13 with $\hat{u} := \Upsilon(\hat{v})$ can be applied to the system (3.1) - (3.2) to conclude that:

$$\|y\|_{L^\infty(Q_T)} \;\leq\; C_0 \tag{3.44}$$

where C_0 stands for the constant from the estimate (1.78) in Theorem 1.2.13. C_0 depends in particular on constants denoted in Theorem 1.2.13 as C_g and R^U. Since we assume $\hat{u} := \Upsilon(\hat{v})$, one can check that, to apply Theorem 1.2.13, it suffices to set

$$C_g := \|\sigma_g\|_{L^{s_1}(\mathbb{R}^{\mathbf{d}})}, \qquad R^U := J\big(\|\sigma_g\|^2_{L^2(\mathbb{R}^{\mathbf{d}})} + \|\sigma_h\|^2_{L^2(\mathbb{R}^{\mathbf{d}})} + 1\big)$$

for arbitrary $\hat{v} \in V$. Other quantities on which C_0 depends (which are indicated in Theorem 1.2.13) also are independent of $\hat{v} \in V$. Hence, having chosen C_g and R^U as above, C_0 in (3.44) is independent of $\hat{v} \in V$ as well.

Note that the assumption $\sigma_g \in L^2(\mathbb{R}^{\mathbf{d}}) \cap L^{s_1}(\mathbb{R}^{\mathbf{d}})$ is essential above because of the assumptions for the integrability of $\hat{u}_{g_j}$ imposed in Theorem 1.2.13 (in the present case, $\hat{u}_{g_j} := \Upsilon_{g_j}(\hat{v}) = \sigma_g(\,.\, - \hat{v}_j)|_\Omega$). Theorem 1.2.13 requires $\hat{u}_{g_j} \in L^{\max\{2,\mathbf{d}/2\}}(\Omega)$ at least, for $j = 1, \ldots, J$. Moreover, Theorem 1.2.13 requires $y_0 \in L^\infty(\Omega)$, thus the latter also is necessary.

Choose $\widetilde{n} > C_0$. By the condition (3.41) and by (3.44) we see that

$$f^{\widetilde{n}}(y(x,t)) \;=\; f(y(x,t)) \qquad \text{for a.e. } (x,t) \in Q_T \tag{3.45}$$

for arbitrary $\hat{v} \in V$. Thus, inserting the above into the definition of the weak solution (see Definition 3.0.1) we find that $(y, \kappa_1, \dots, \kappa_J)$ is also the weak solution of $\big((3.1)\text{ - }(3.2)\big)^{\widetilde{n}}$, corresponding to $x_j := \hat{v}_j$, for $j = 1, \dots, J$ (which exists and is unique, see Corollary 3.2.11). Therefore, $\mathcal{Z}(\hat{v}) = \mathcal{Z}^{\widetilde{n}}(\hat{v})$ and, in consequence,

$$\mathcal{I}(\hat{v}) \;=\; \mathcal{I}^{\widetilde{n}}(\hat{v}) \qquad \text{for all } \hat{v} \in V$$

Now, note that for $\mathcal{I}^{\widetilde{n}}$, and hence for $\mathcal{I}$, the existence of minimizers follows by Theorem 3.2.1. Indeed, by the assumption concerning f, functions f^n are Lipschitz (see Observation 3.2.9). Thus, one may verify that the system $\big((3.1)\text{ - }(3.2)\big)^n$ obeys assumptions of Theorem 3.2.1, for all $n > 0$, in particular for $n := \widetilde{n}$. Application of Theorem 3.2.1 concludes the proof.

Above, assumptions (F-1) and (F-3) are essential because Theorem 3.2.1 also requires them. ■

Theorem 3.2.15 *Let the system (3.1) - (3.2) fulfill the assumptions of Theorem 3.2.5, except the assumptions concerning f, y_0 and σ_g. For f, y_0 and σ_g, we make the following assumptions:*

- *f is locally Lipschitz continuous, obeys the condition (1.73) for certain constant C_f and obeys the assumption (E-1),*
- *$y_0 \in L^\infty(\Omega)$,*
- *σ_g obeys the assumption (F-2) and, in addition, $\sigma_g \in L^{s_1}(\mathbb{R}^{\mathbf{d}})$ for certain $s_1 \geq \frac{\mathbf{d}}{2}$.*

Then, the cost functional $\mathcal{I}$, defined in (3.22) - (3.23), is Gâteaux differentiable and its differential in point $\hat{v}$ in direction $\hat{\eta}$ is equal to $D_G\mathcal{I}(\hat{v})(\hat{\eta}) = \big(\Lambda^{\hat{v}}, \hat{\eta}\big)_V$, where $\Lambda^{\hat{v}} \in V$ is given by the formula (3.33).

PROOF. In the present proof, the following notation will be convenient. For $n > 0$, let $(3.33)^n$ denote the formula (3.33) with the following modifications:

- $(\hat{y}, \hat{\kappa}_1, \dots, \hat{\kappa}_J)$ is replaced by $(\hat{y}^n, \hat{\kappa}_1^n, \dots, \hat{\kappa}_J^n) = \mathcal{Z}^n(\hat{v})$,
- $(\widetilde{p}, \widetilde{q}_1, \dots, \widetilde{q}_J)$ is replaced by $(\widetilde{p}^n, \widetilde{q}_1^n, \dots, \widetilde{q}_J^n)$ being the weak solution of the system $\big((3.30)\text{ - }(3.31)\big)^n$ corresponding to $\hat{Y} := \hat{y}^n$.

In the proof, we assume that functions f^n are defined by (3.41) and (3.43), for $n > 0$.

Let $\hat{v} \in V$. Assume that $(y, \kappa_1, \dots, \kappa_J) \in X^2$ is the weak solution of the system (3.1) - (3.2), corresponding to $x_j := \hat{v}_j$, $j = 1, \dots, J$ (which exists and is unique, see Corollary 3.2.10).

By the same argument as in the proof of Theorem 3.2.14, the estimate (3.44) hold, with constant C_0 independent of $\hat{v} \in V$. Note in particular that deriving (3.44) required Theorem 1.2.13 and that the present assumptions concerning integrability of σ_g are sufficient to apply Theorem 1.2.13. Moreover, Theorem 1.2.13 requires $y_0 \in L^\infty(\Omega)$, thus the latter also is utilized here.

Let $\widetilde{n} > C_0$. By (3.44), by the condition (3.41) and by the choice of $\widetilde{n}$, we have (3.45), independently on the choice of $\hat{v} \in V$. Hence, inserting (3.45) into the definition of the weak solution (see Definition 3.0.1), $(y, \kappa_1, \dots, \kappa_J)$ is the weak solution of the system $\big((3.1)\text{ - }(3.2)\big)^{\widetilde{n}}$

(which exists and is unique, see Corollary 3.2.11), for all $\hat{v} \in V$. In consequence, $\mathcal{I}(\hat{v}) = \mathcal{I}^{\widetilde{n}}(\hat{v})$, for all $\hat{v} \in V$.

Functions f^n are Lipschitz continuous and $f^{n\prime}(s)$ exists for all $s \in \mathbb{R}$ (see Observation 3.2.9). Thus, it can be verified that the system (3.1) - (3.2) fulfills the assumption of Theorem 3.2.5, for all $n > 0$, in particular for $n := \widetilde{n}$. Therefore, by Theorem 3.2.5 we conclude that $\mathcal{I}^{\widetilde{n}}$ is Gâteaux differentiable and for all $\hat{v}, \hat{\eta} \in V$ we have $D_G\mathcal{I}^{\widetilde{n}}(\hat{v})(\hat{\eta}) = \left(\Lambda^{\hat{v}}_n, \hat{\eta}\right)_V$, where $\Lambda^{\hat{v}}_n \in V$ is given by the formula $(3.33)^{\widetilde{n}}$. Since $\mathcal{I}^{\widetilde{n}} = \mathcal{I}$, $\mathcal{I}$ also is Gâteaux differentiable and $D_G\mathcal{I}(\hat{v})(\hat{\eta}) = \left(\Lambda^{\hat{v}}_n, \hat{\eta}\right)_V$, for $\hat{v}, \hat{\eta} \in V$.

Above, the assumption (F-2) is essential because Theorem 3.2.5 also requires it.

The proof will be closed once we show that $\Lambda^{\hat{v}}_{\widetilde{n}} = \Lambda^{\hat{v}}$ for $\widetilde{n}$ as above, for $\hat{v} \in V$. Comparing formulas (3.33) and $(3.33)^{\widetilde{n}}$, which define $\Lambda^{\hat{v}}$ and $\Lambda^{\hat{v}}_{\widetilde{n}}$ respectively, we see that we need to justify the following, for all $\hat{v} \in V$:

- $\left(\hat{y}^{\widetilde{n}}, \hat{\kappa}^{\widetilde{n}}_1, \ldots, \hat{\kappa}^{\widetilde{n}}_J\right) = (\hat{y}, \hat{\kappa}_1, \ldots, \hat{\kappa}_J)$, where $(\hat{y}, \hat{\kappa}_1, \ldots, \hat{\kappa}_J) := \mathcal{Z}(\hat{v})$,
- $\left(\widetilde{p}^{\widetilde{n}}, \widetilde{q}^{\widetilde{n}}_1, \ldots, \widetilde{q}^{\widetilde{n}}_J\right) = (\widetilde{p}, \widetilde{q}_1, \ldots, \widetilde{q}_J)$, where $(\widetilde{p}, \widetilde{q}_1, \ldots, \widetilde{q}_J)$ is the weak solution of the system (3.30) - (3.31) corresponding to $\hat{Y} := \hat{y}$.

Equality $\left(\hat{y}^{\widetilde{n}}, \hat{\kappa}^{\widetilde{n}}_1, \ldots, \hat{\kappa}^{\widetilde{n}}_J\right) = (\hat{y}, \hat{\kappa}_1, \ldots, \hat{\kappa}_J)$ follows by showing that, for $\widetilde{n}$ as assumed, a weak solution of $\left((3.1) - (3.2)\right)^{\widetilde{n}}$ is a weak solution of (3.1) - (3.2) corresponding to $x_j := \hat{v}$, $j = 1, \ldots, J$. But we have already shown above that a weak solution of (3.1) - (3.2) is a weak solution of $\left((3.1) - (3.2)\right)^{\widetilde{n}}$. The opposite follows immediately, since we have the existence and uniqueness results for both systems (see Corollary 3.2.10 and Corollary 3.2.11).

To justify equality $\left(\widetilde{p}^{\widetilde{n}}, \widetilde{q}^{\widetilde{n}}_1, \ldots, \widetilde{q}^{\widetilde{n}}_J\right) = (\widetilde{p}, \widetilde{q}_1, \ldots, \widetilde{q}_J)$, we proceed as follows. We need to show that $\left(\widetilde{p}^{\widetilde{n}}, \widetilde{q}^{\widetilde{n}}_1, \ldots, \widetilde{q}^{\widetilde{n}}_J\right)$ is in fact the weak solution of (3.30) - (3.31) corresponding to $\hat{Y} := \hat{y}$. But it follows with arguments similar to the above ones. By $\hat{y}^{\widetilde{n}} = \hat{y}$ (already proven), by (3.44), by (3.41) and by the choice of $\widetilde{n}$, we have

$$f^{\widetilde{n}\prime}(\hat{y}^{\widetilde{n}}(x,t)) \;=\; f'(\hat{y}(x,t)) \qquad \text{for a.e. } (x,t) \in Q_T$$

The above along with $\hat{y}^{\widetilde{n}} = \hat{y}$ yields the necessary.

The proof of Theorem 3.2.15 is complete. ■

From Theorem 3.2.15 and Theorem 3.2.6, we can derive an analogue of Corollary 3.2.7:

Corollary 3.2.16 *Let the assumptions of Theorem 3.2.15 be fulfilled. Then, the cost functional $\mathcal{I}$, defined in (3.22) - (3.23), is Gâteaux differentiable and its differential in point $\hat{v}$ in direction $\hat{\eta}$ is equal to $D_G\mathcal{I}(\hat{v})(\hat{\eta}) = \left(\Lambda^{\hat{v}}, \hat{\eta}\right)_V$, where $\Lambda^{\hat{v}} \in V$ is given by the formula (3.40).*

The above follows, as in the case of Corollary 3.2.7, by applying formulas (3.33), (3.34) and (3.35), changing the integration order and observing that $-\mathcal{T}_{\partial_i\sigma}(\hat{v}_j)(x) = \mathcal{T}_{-\partial_i\sigma}(\hat{v}_j)(x)$ holds for a.e. $x \in \Omega$, for $\sigma = \sigma_g, \sigma_h$, for $j = 1, \ldots, J$ and for $i = 1, \ldots, \mathbf{d}$.

Chapter 4

Optimal targeting problem — numerical prototypes

In this chapter, we describe numerical experiments for the optimal targeting problem, announced in §2 of *Introduction*. We will base on the mathematically more precise formulation of the subject problem given in Section 3.2. We will thus identify the optimal targeting problem with the optimization problem (3.24), consisting in minimization of cost functional $\mathcal{I}$ (defined by conditions (3.22) - (3.23)).

In Chapter 3, we have already answered the question concerning possibility of solving optimization problem (3.24) (Theorem 3.2.1, Theorem 3.2.14), as well as given the characterization of the solutions (Corollary 3.2.8). Now, we are going to focus on the matter of numerical construction of the solutions.

Therefore, in the present chapter, the main point of our interest is the matter of choice of optimization algorithms proper to attack optimization problem (3.24). Thus, we test a few optimization methods to check how their performance varies with changes of parameters and functions entering the definition of cost functional $\mathcal{I}$ or the system (3.1) - (3.2).

Cost functional $\mathcal{I}$ depends on the control parameter (i.e. the targetings of the control and measurement devices actions), which parametrizes the feedback law (i.e. the algorithm of computing the response functions) in thermostat control mechanism (see *Introduction* for details). Consider the case of T_0 being close to T in the definition of cost functional $\mathcal{I}$ (see (3.22) - (3.23)). This determines a cost functional encoding idea of measuring the gap between the process state and reference state in the neighborhood of the terminal time T (see the remarks in §2 of *Introduction*). The latter gap can serve as a natural measure of the efficiency of the thermostat control mechanism. Hence, the problem of minimization of cost functional $\mathcal{I}$ with T_0 close to T is consistent with one the general ideas of the present work, which is to optimize the feedback law in the thermostat control mechanism in order to improve its efficiency (see the beginning of *Introduction*). For this reason, in the present chapter we are particularly interested in the case of T_0 close to T.

Other point of our interest was the independence of the optimization results on the initial state of the controlled process, described in the system (3.1) - (3.2) by y_0, in the case of T_0 close to T. To explain our motivations, consider the model with an open-loop control described by the sole equation (3.1) (without (3.2)), where the user is responsible for the choice of both functions g_j, characterizing the control devices actions, and the power functions κ_j. It follows by intuition that the optimal choice of κ_j perhaps depends on the initial state y_0 (regardless of whether T_0 is close to T in the definition of $\mathcal{I}$ or not). Therefore, the independence of solutions of the optimal targeting problem on the initial state of the process would be an advantage of the thermostat control mechanism, at least in comparison to the mentioned system with an open-loop control (see also the general ideas described in the beginning of *Introduction*). Hence, during our experiments, we have made an attempt to verify whether the subject independence indeed exists or not.

By the results of Chapter 2, we may expect that, in certain cases, the alleged independence on y_0 of the solutions of the optimal targeting problem can be true. Indeed, in the simulations described in Chapter 2 we observed that in some (but not all) situations the process controlled by thermostats stabilizes near to the same state, independently on the initial state y_0 of the process. In other words, the process states achieved near to the terminal time T were very similar, regardless on y_0. For this kind of situations, the cost functional $\mathcal{I}$ with T_0 close to T can vary insignificantly under changes of y_0, because such $\mathcal{I}$ captures only the data concerning the process near to the terminal time T. Hence, the minimal points for $\mathcal{I}$ with T_0 close to T also can vary insignificantly under changes of y_0.

The optimization algorithms utilized in our experiments were gradient-based algorithms — the steepest descent method and the nonlinear conjugate gradient method, implemented in the Polak-Ribière mode with certain modification. The latter method was used in two variants: one with a periodic reset of the algorithm every N_r iterations, with N_r equal to the dimension of the optimization space; the other without the periodic reset. Each of the methods involves computing the gradient of the cost functional. In our experiments, the gradient was computed basing on the characterization given in Corollary 3.2.16. The latter characterization involves solving systems (3.1) - (3.2) and (3.30) - (3.31). Besides, computing the value of the cost functional $\mathcal{I}$ also involves solving the system (3.1) - (3.2). For solving numerically these two systems, we employed the finite element method for discretization in space, the implicit Euler schemes for discretization in time and the Picard iterations method for treating the nonlinear terms entering the system (3.1) - (3.2).

To compare performance of particular optimization algorithms, we in fact compare the number of iterations necessary to approximate a solution of (3.24) when using a given algorithm with a given stop criterion. Thus, by saying that performance of a given optimization algorithm was better (worse) in situation A than in situation B we mean that the number of iterations of the algorithm in situation A was lower (higher) than in situation B.

The results of the experiments suggest that average performance of the steepest descent method for optimization problem (3.24) vary with changes of the parameter T_0, entering the definition of the cost functional $\mathcal{I}$ (average, in a sense to be clarified later). Setting T_0 close to T resulted in more iterations of the algorithm than for $T_0 = 0$ (Section 4.4.1 and Section 4.4.2). In this sense, problem (3.24) with T_0 close to T is more difficult than with $T = 0$. Nevertheless, changing the optimization algorithm to nonlinear conjugate gradient with reset leveled the mentioned difference in the average performance (Section 4.4.2).

We have also tested behavior of the nonlinear conjugate gradient method with reset under changes of the time horizon T in the system (3.1) - (3.2). We observed that lengthening the time horizon T also resulted in inferior average performance of the optimization algorithm (Section 4.4.3). This happened despite the nonlinear conjugate method with reset was successful in leveling the performance differences for changes of the parameter T_0.

To sum up, the average performance of the optimization algorithms changed when varying both T_0 and T. However, for changes of T_0, the differences in the average performance was observed for the steepest descent method and disappeared when using the nonlinear conjugate gradient method with reset.

As mentioned, the case of T_0 being close to T is particularly interesting for us. In this case, the experiments results suggest that when lengthening the time horizon of the system (3.1) -

(3.2), the optimization procedure output becomes more independent of the initial condition in the latter the system (Section 4.4.3). This confirms our expectations, described above.

However, lengthening the time interval increases computational cost for numerical treatment of problem (3.24). Indeed, assuming that the time step in the numerical scheme remains the same, the cost of solving the system (3.1) - (3.2) increases as the time horizon becomes longer. Each evaluation of the cost functional $\mathcal{I}$ requires solving the system (3.1) - (3.2), thus the computational cost of searching for minimums of $\mathcal{I}$ grows as the computational cost of solving (3.1) - (3.2) grows. Therefore, it is expensive computational task to solve optimization problem (3.24) and obtain results independent of y_0, because it is necessary to choose long time horizon T. Moreover, as mentioned, lengthening the time interval in our experiments resulted in higher number of iterations, what made the computational task even more expensive.

In fact, in our experiments, the computational time necessary to approximate a solution of $\mathcal{I}$ for long time interval was impractically long. Reduction of this time would be a desired result. In Section 4.4.4, we propose some possible strategies for reduction of optimization procedures computational cost, which can be tested in the future experiments.

Chapter 4 is divided into two parts: 1) the part for specification of utilized parameters, optimization methods and numerical schemes (Section 4.1, Section 4.2 and Section 4.3, respectively) and 2) the part devoted to description of results of optimization procedures performed with the use of these parameters, methods and schemes (Section 4.4). In Section 4.4.4, concluding the second part, we propose refinements for the optimization algorithms and numerical schemes described in Section 4.2 and Section 4.3.

4.1 Structural assumptions

Below, we describe structural assumptions concerning optimization problem (3.24), which were imposed for simulations described in Section 4.4. This assumptions specify the parameters necessary to determine the cost functional $\mathcal{I}$, defined by (3.22) - (3.23), was the target of our optimization experiments.

Let us begin with the assumptions concerning the system (3.1) - (3.2), defining which is necessary for defining the cost functional $\mathcal{I}$. Basically, our intention was to operate with assumptions analogous to those described in Section 2.1. However, some of the assumptions imposed there needed modifications before employing them here.

To be more precise, in the system (3.1) - (3.2) we assume that $\mathbf{d} = 2$, that domain Ω is given as in (2.7) and that reactive term f is given as in (2.8). Note that both Ω and f chosen by us fit the assumptions of Corollary 3.2.16.

At the same time, we cannot reuse the assumptions described in Section 2.1 for pattern functions σ_g, σ_h and switching functions w_j, $j = 1, \ldots, J$, for the below reasons:

1. Concerning the pattern functions σ_g and σ_h, note that if they obey the formula (2.3) from Section 2.1, then they are not elements of $W^{1,2}(\mathbb{R}^{\mathbf{d}})$. In particular, for pattern functions as in (2.3), partial derivatives $\partial_i \sigma_g$ and $\partial_i \sigma_h$, for $i = 1, \ldots, \mathbf{d}$, are not well defined. Simultaneously, Corollary 3.2.16 assumes $\sigma_g, \sigma_h \in W^{1,2}(\mathbb{R}^{\mathbf{d}})$. The gradient formula (3.40), asserted in Corollary 3.2.16, also involves the partial derivatives of σ_g and σ_h for $j = 1, \ldots, J$. Thus, the subject gradient formula fails if the pattern functions are given by (2.3). In consequence, the formula (2.3) cannot be applied in the present context, because, as mentioned in the beginning of Chapter 4, we intend to use the gradient characterization asserted in Corollary 3.2.16.

2. Concerning the switching functions w_j, $j = 1, \dots, J$, note that the formula (2.9) defines non-differentiable w_j. Simultaneously, the differentiability of the switching functions w_j is assumed in Corollary 3.2.16. Thus, Corollary 3.2.16 fails to hold if the switching functions are given by (2.9). Hence, the formula (2.9) cannot be utilized here, because we intend to utilize the gradient characterization given in Corollary 3.2.16.

To deal with the above difficulties, we impose the following assumptions for pattern functions σ_g, σ_h and switching functions w_j, $j = 1, \dots, J$:

1. We have chosen the below pattern functions to be utilized in experiments described in Section 4.4:

$$\sigma_g(x) = \begin{cases} C_g & \text{on } B(0, r_{\sigma,1}) \\ 0 & \text{on } (B(0, r_{\sigma,2}))^c \\ \text{radially linear} & \text{otherwise} \end{cases} \qquad \sigma_h(x) = \begin{cases} C_h & \text{on } B(0, r_{\sigma,1}) \\ 0 & \text{on } (B(0, r_{\sigma,2}))^c \\ \text{radially linear} & \text{otherwise} \end{cases} \tag{4.1}$$

   for certain $r_{\sigma,2} > r_{\sigma,1}$, and $C_g, C_h > 0$. Note, that the pattern functions given in (4.1) can be understood as a regularization of the pattern functions given in (2.3) — putting $r_{\sigma,2} = r_\sigma$, one can observe that σ_g given in (4.1) tends in $L^2(\mathbb{R}^{\mathbf{d}})$ to σ_g given in (2.3) as $r_{\sigma,1} \to r_{\sigma,2}$, and the same holds for σ_h.

   With the pattern functions as in (4.1), Lemma 3.1.4 guarantees weak Gâteaux differentiability of the associated operators Υ_{g_j} and Υ_{h_j}, for $j = 1, \dots, J$. Moreover, for σ_g and σ_h as in (4.1), the weak directional derivatives $\partial_i \sigma_g$ and $\partial_i \sigma_h$, for $i = 1, \dots, \mathbf{d}$ are well defined. Hence, the formula asserted by Corollary 3.2.16 is well defined.

2. For experiments described in Section 4.4, we have chosen switching functions being smoothed versions of the switching functions given in (2.9). Smoothing with second order polynomials was performed.

   The details of the smoothing procedure which was applied are as follows. Choose constants $C_{smooth} \in [0, 1]$ and $L_w < 0$. Define the function

$$w_{aux,1}(s) := L_w s$$

   Denote by s^+_{smooth} the point where $w_{aux,1}$ achieves value $-C_{smooth}$ and by s^-_{smooth} the point where $w_{aux,1}$ achieves value $+C_{smooth}$. Define also p_+, p_- as second degree polynomials of one variable determined by the following conditions:

$$\begin{aligned} p_+(s^+_{smooth}) &= w_{aux,1}(s^+_{smooth}) = -C_{smooth} \\ p'_+(s^+_{smooth}) &= w'_{aux,1}(s^+_{smooth}) = L_w \\ \min_{\mathbb{R}}(p_+) &= -1 \end{aligned}$$

$$\begin{aligned} p_-(s^-_{smooth}) &= w_{aux,1}(s^-_{smooth}) = C_{smooth} \\ p'_-(s^-_{smooth}) &= w'_{aux,1}(s^-_{smooth}) = L_w \\ \max_{\mathbb{R}}(p_-) &= 1 \end{aligned}$$

   Denote by s_{max} the maximizer of p_- and by s_{min} the minimizer of p_+. Note that points s^+_{smooth}, s^-_{smooth}, s_{max} and s_{min} are determined by the choice of constants C_{smooth} and L_w. Explicit formulas for these points can be derived, if necessary. We do not present the latter formulas here only for brevity reasons.

Having this, we define the following function w_j, for $j = 1, \dots, J$, being a spline of functions $+1$, p_-, $w_{aux,1}$, p_+, -1:

$$w_j(s) = H_w w_{aux,2}(s) \qquad w_{aux,2}(s) = \begin{cases} +1 & \text{on } (-\infty, s_{max}] \\ p_-(s) & \text{on } (s_{max}, s^-_{smooth}] \\ w_{aux,1}(s) & \text{on } (s^-_{smooth}, s^+_{smooth}) \\ p_+(s) & \text{on } [s^+_{smooth}, s_{min}) \\ -1 & \text{on } [s_{min}, +\infty) \end{cases} \tag{4.2}$$

for certain $H_w > 0$. In the experiments described in Section 4.4, we have assumed the switching functions in the system (3.1) - (3.2) to be given by (4.2).

Since the points s^+_{smooth}, s^-_{smooth}, s_{max} and s_{min} are determined by constants C_{smooth} and L_w, functions w_j, $j = 1, \dots, J$ defined in (4.2) are determined by the choice of constants L_w, H_w and C_{smooth}.

One can verify that functions w_j defined by (4.2) belong to $C^1(\mathbb{R})$, for $j = 1, \dots, J$. Thus, Corollary 3.2.16 is valid if they are utilized as the switching functions in the system (3.1) - (3.2).

As in Section 2.1, we assume that the value of C_h is determined by the relation (2.10), for certain $C_{switch} > 0$. The meaning of the constant C_{switch} was explained in Section 2.1, thus we do not repeat this explanation here.

Remark. In Section 2.1, for deriving the relation (2.10), the points in which the switching functions achieved the extremal values (more precisely, the closest to $s = 0$ points in which w_j attains a global extremum) were essential. For the switching functions w_j considered there (see the formula (2.9)), the subject points were $\pm 1/|L_w|$. Here, with w_j defined as in (4.2), the extremal values are achieved in different points, above denoted as s_{max} and s_{min}. Thus, to be puristic, we should derive an analog of the relation (2.10) one more time, accounting the new switching functions having new extremal points, if we wanted to preserve the idea lying behind the constant C_{switch}, explained in Section 2.1. Nevertheless, for simplicity, we decided to neglect the effects inferred by the shift of the extremal points caused by the change of the switching functions. ▲

Now, since we assume that C_h is determined by the relation (2.10) we substitute the pattern function σ_h to the subject relation and find out that C_h can be expressed more explicitly by:

$$C_h = \left(\frac{\pi}{3} \left|L_w\right| C_{switch} \left((r_{\sigma,1})^2 + r_{\sigma,1} r_{\sigma,2} + (r_{\sigma,2})^2\right)\right)^{-1} \tag{4.3}$$

In addition, we make the following assumption for the parameter $\widetilde{\lambda}$ in the definition of the cost functional $\mathcal{I}$:

$$\widetilde{\lambda} = (T - T_0)^{-1} \tag{4.4}$$

where T_0 is the parameter entering the definition of the cost functional $\mathcal{I}$.

To sum up, for Ω given by (2.7), the switching function w_j as in (4.2), pattern functions σ_g and σ_h as in the formula (4.1) and C_h as in the formula (4.3), the system (3.1) - (3.2) is uniquely determined by the choice of the following functions and parameters:

$$y_0, \kappa_{10}, \dots, \kappa_{J0}, \quad y^*$$

$$T, \quad D, \beta_1, \dots, \beta_J, \quad J, \quad x_1, \dots, x_J, \quad r_{\sigma,1}, r_{\sigma,2}, \quad C_g, C_{switch}, L_w, H_w, C_{smooth}$$

With the above indicated conditions and with $\widetilde{\lambda}$ as in (4.4), cost functional $\mathcal{I}$ is fully determined by specification of the above listed functions and parameters and, additionally, by specification of the parameter T_0.

4.2 Optimization methods

We describe now optimization methods utilized for solving optimization problem (3.24). All experiments described in Section 4.4 base on the below described methods.

Generally, two methods were employed: the steepest descent method and the nonlinear conjugate gradient method (described and extensively commented e.g. in [38] or [7]). The second of these two was considered in two variants — one with reset of the algorithm every N_r iterations, for a given natural N_r, the other without the reset. Below, we describe these methods in more detail.

For convenience, we use the following notation in the present section. Let $F\colon I \to \mathbb{R}$ be a given function, where $I = [0, b]$ or $I = [0, b)$, with $b \in \mathbb{R}^+ \cup \{+\infty\}$. By $\text{minn}_{s\in I}F(s)$ we understand the problem of finding the local minimum of F which is the closest to origin point $s = 0$. Note that the solution of $\text{minn}_{s\in I}F(s)$ can be different than the global minimum of F, even if the global minimum exists.

SD method. By the steepest descent method (SD method, in brief), we understand the following algorithm:

1. Choose $\hat{v}^0 \in V$. Set $n = 0$.
2. If the stop criterion (to be described below) is fulfilled, then terminate. Else:
    (a) Compute $r^n := -\nabla\mathcal{I}(\hat{v}^n)$. Set $d^n := r^n$.
    (b) Find $s_n \in [0, 1]$ solving 1-D minimization problem $\text{minn}_{s\in[0,1]}\mathcal{I}(\hat{v}^n + sd^n)$.
    (c) Assign $\hat{v}^{n+1} := \hat{v}^n + s_n d^n$.
    (d) Increment n and repeat step 2.

CG method. By the nonlinear conjugate gradient method (CG method, in brief), we understand the following algorithm:

1. Choose $\hat{v}^0 \in V$. Set $n = 0$. Set $d^{-1} := \mathbf{0} \in V$.
2. If the stop criterion (to be described below) is fulfilled, then terminate. Else:
    (a) Compute $r^n := -\nabla\mathcal{I}(\hat{v}^n)$.
    (b) Compute coefficient ϱ_n (to be described below) and set $d^n := r^n + \varrho_n d^{n-1}$
    (c) Find $s_n \in [0, 1]$ solving 1-D minimization problem $\text{minn}_{s\in[0,1]}\mathcal{I}(\hat{v}^n + sd^n)$.
    (d) Assign $\hat{v}^{n+1} := \hat{v}^n + s_n d^n$.
    (e) Increment n and repeat step 2.

To complete the above specifications, we need to describe the stop criterion and coefficient ϱ_n.

Stop criterion. In our experiments, we terminated further execution of the optimization algorithms if $n = N_{opt}$, for a given natural N_{opt}, or if $n \geq 1$ and the last computed s_n satisfied $s_n = 0$.

Coefficient ϱ_n. Various choices of coefficient ϱ_n are possible for the nonlinear conjugate gradient method (see [38, Chap.5.2] or [7, p.329]). Our choice of the subject coefficient involved the Polak-Ribière concept (presented e.g. in the latter references):

$$\varrho^{PR} := \|r^n\|_V^{-2}(r^n, r^n - r^{n-1})_V$$

with some modifications, concerning the reset of the algorithm. More precisely, in each simulation described in Section 4.4, one of the following methods for computing ϱ_n was involved:

- *Method 1.* If $n = 0$, set $\varrho_n = 0$, for consistency. For $n \geq 1$, set $\varrho_n := \varrho^{PR}$ and next, if $\varrho_n \leq 0$, reset CG algorithm, i.e. assign $\varrho_n := 0$.

- *Method 2.* If $n = 0$, set $\varrho_n = 0$, for consistency. For $n \geq 1$, set $\varrho_n := \varrho^{PR}$ and next:

  1. If $\varrho_n \leq 0$, reset CG algorithm, i.e. assign $\varrho_n := 0$.
  2. For a given $N_r \in \mathbb{N}$, if there was no reset in last N_r iterations, i.e. in iterations $n - N_r + 1, n - N_r + 2, \dots, n$, of CG algorithm, then reset the algorithm, i.e. assign $\varrho_n := 0$.

  In the experiments described in Section 4.4, value $N_r = 2J$ was always used, whenever *Method 2.* was utilized, where J is the same as in the system (3.1) - (3.2).

We will use the following terminology:

- **CG-r method** is the CG method without reset every N_r iterations, i.e. the CG method with *Method 1.* for choosing coefficient ϱ_n.

- **CG+r method** is the CG method with reset every N_r iterations, i.e. the CG method with *Method 2.* for choosing coefficient ϱ_n.

Remark. Resetting the algorithm if coefficient ϱ^{PR} occurs to be negative is necessary because, if this is the case, the vector $r^n + \varrho^{PR} d^{n-1}$ can be not a descent direction (see [38, p.122-123]). Resetting the algorithm every N_r iterations also is a common practice, with the usual choice of N_r equal to the dimension of V (see [38, p.124]). The latter remark suggests $N_r = 2J$ in our case, as assumed above. ▲

Remark. In the above described methods we solve 1-D problems of the form $\text{minn}_{s\in[0,1]}\mathcal{I}(\hat{v} + s\hat{d})$, for certain $\hat{v}, \hat{d} \in V$, not just $\min_{s\in[0,1]}\mathcal{I}(\hat{v} + s\hat{d})$. On level of general ideas it means that we intend to extract the local minimum of $\mathcal{I}(\hat{v} + \,.\,\hat{d})$ which is situated closest to the point $s = 0$. This serves to keep the iteration points $\hat{v}^1, \hat{v}^2, \hat{v}^3, \dots$ in the same „valley" in the graph of $\mathcal{I}$ in which the initial point $\hat{v}^0$ lays. ▲

To sum up, we specify the optimization algorithm by the choice of: 1) the initial point $\hat{v}^0 \in V$, 2) the parameter N_{opt} and 3) the optimization method (SD, CG-r or CG+r).

4.3 Numerical methods

Here, we describe numerical schemes for performing the optimization methods described in Section 4.2. These schemes were utilized in experiments described in Section 4.4, whenever the subject optimization methods were involved.

By the specifications given in Section 4.2, we see that performing the subject methods requires a method for evaluating the cost functional $\mathcal{I}$, a method for computing its gradient and a method of solving the 1-D optimization problem. The base for the first two methods are the definition of $\mathcal{I}$ given in (3.22) - (3.23) and the gradient formula (3.40), asserted in Corollary 3.2.16. Both the formula (3.22) - (3.23) and the gradient formula (3.40) depend on the weak solution of the system (3.1) - (3.2). Moreover, the gradient formula (3.40) require the weak solution of the system (3.30) - (3.31). Hence, in total, to perform the subject optimization methods, we need methods for:

1) computing the solutions of the system (3.1) - (3.2) and the system (3.30) - (3.31),

2) computing the gradient of $\mathcal{I}$ in a given point,

3) computing the value of $\mathcal{I}$ in a given point,

4) solving 1-D optimization problem $\text{minn}_{s\in[0,1]}\mathcal{I}(\hat{v}+s\hat{d})$, for suitable $\hat{v},\hat{d}\in V$.

In the experiments described in Section 4.4, each of the above subproblems was solved approximately, by use of numerical methods. Thus, in fact, in our experiments, we have treated problem (3.24) not with the SD or CG methods itself, but numerical approximations of these methods. Below, we describe the numerical schemes which were utilized for solving subproblems 1) - 4), whenever solving these subproblems was necessary during execution of the SD or CG methods in our experiments.

4.3.1 Main system and adjoint system

Now, we describe numerical methods utilized in the experiments described in Section 4.4 for solving systems (3.1) - (3.2) and (3.30) - (3.31). The below methods were utilized in the experiments whenever it was necessary to solve the mentioned systems.

For discretization in space, the finite element method was used for both systems. The triangulation of Ω utilized for the finite element method was as in Figure 2.1 in Section 2.2 (recall that we assumed Ω to be given for our experiments by (2.7)). The finite element space considered in our experiments was the space of continuous functions, linear on each element of the triangulation.

For discretization in time for the system (3.1) - (3.2), we employed implicit Euler scheme and, for discretization in time for the system (3.30) - (3.31), backward implicit Euler scheme was applied. In both cases, the discretization of the time interval $[0,T]$ assumed uniform distribution of the time discretization points.

Moreover, the nonlinear terms in the system (3.1) - (3.2) were treated with the use of Picard iterations method.

Now, let us give a more detailed description of the above sketched numerical schemes for (3.1) - (3.2) and (3.30) - (3.31). Below, we assume that $\hat{v}\in V$ is given and that $x_1,\ldots,x_J$ in the system (3.1) - (3.2) are determined by $x_j := \hat{v}_j$, for $j=1,\ldots,J$.

Similarly as in Chapter 2, denote:

$N+1$ — the number of triangulation mesh vertexes along each spatial direction (i.e., the triangulation has $(N+1)^2$ vertexes),

$M+1$ — the number of time discretization points in interval $[0,T]$,

N_{Picard} — the number of Picard iterations applied in each time step to treat the nonlinear terms appearing in (3.1) - (3.2).

Denote also $\tau_M := M^{-1}$ and $\tau_N := N^{-1}$.

In addition, denote the triangulation presented in Figure 2.1 in Section 2.2 by Ω_N, denote the space of functions continuous on Ω_N and linear on each element of the triangulation by $P_1(\Omega_N)$ and denote vectors of standard „hat" basis of $P_1(\Omega_N)$ by ϕ_n, for $n = 1, \ldots, (N+1)^2$.

Remark. Two implicit Euler schemes are mentioned above: the „usual" one and a scheme which we have called backward implicit Euler scheme. By the backward implicit Euler scheme for the differential equation $-\dot{\mathbf{x}} = F(\mathbf{x}, t)$ on $[0, T]$, with the terminal condition $\mathbf{x}(T) = \widetilde{\mathbf{x}}$, we mean the following scheme:

$$\mathbf{x}^M = \widetilde{\mathbf{x}}, \quad \mathbf{x}_m - \mathbf{x}_{m+1} = \tau_M F(\mathbf{x}_m, t_m)$$

for $t_m = m\tau_M$, $m = 0, 1, \ldots, M-1$, where M and τ_M are as above. The „usual" implicit Euler scheme is a common scheme, hence we do not define it here. ▲

The system (3.1) - (3.2) is treated with the same numerical scheme as the system (2.5) - (2.6) in Section 2.1, with $g_j := \Upsilon_{g_j}(\hat{v})$ and $h_j := \Upsilon_{h_j}(\hat{v})$. More precisely, the output of the numerical scheme for (3.1) - (3.2) is exactly the function $(Y_N, k_{1,N}, \ldots, k_{J,N})$ defined in Section 2.1, assuming that we put $g_j := \Upsilon_{g_j}(\hat{v})$ and $h_j := \Upsilon_{h_j}(\hat{v})$ in the system (2.5) - (2.6). We treat such $(Y_N, k_{1,N}, \ldots, k_{J,N})$ as a function approximating the weak solution of (2.5) - (2.6).

The above referred scheme for approximating the weak solution of the system (3.1) - (3.2) was employed in the experiments described in Section 4.4 whenever computing the value of the cost functional $\mathcal{I}$ or computing its gradient was necessary (recall that both of these involve the weak solution of the system (3.1) - (3.2)).

Note that the above numerical scheme for (3.1) - (3.2) involves matrices $\mathbb{M}_N$ and $\mathbb{A}_N$, defined in Section 2.2.

The system (3.30) - (3.31) is treated with numerical methods which are analogous as the methods applied for the system (3.1) - (3.2). Nevertheless, since the algebraic form of both systems differ, below we describe the numerical scheme for the system (3.30) - (3.31) in more detail.

First, for a given function $F\colon \Omega \to \mathbb{R}$, let $[F]_N$ and $\overrightarrow{F}$ be defined as in Section 2.2. Recall also that $\overrightarrow{F}=\overrightarrow{[F]_N}$.

We use the following discretization in space for the system (3.30) - (3.31). Put $g_j := \Upsilon_{g_j}(\hat{v})$ and $h_j := \Upsilon_{h_j}(\hat{v})$ for $j = 1, \ldots, J$. In the system (3.30) - (3.31), we insert $[g_j]_N$, $[h_j]_N$, $[\hat{Y}]_N$ and $[y^*]_N$ instead of $\Upsilon_{g_j}(\hat{v})$, $\Upsilon_{h_j}(\hat{v})$, $\hat{Y}$ and y^*, respectively. For the subject modification of the system (3.30) - (3.31), we approximate its solution by the solution of the following variational problem:

$$\begin{cases} -\frac{d}{dt}\big(p_N, \phi\big)_{L^2(\Omega_N)} + D\big(\nabla p_N, \nabla\phi\big)_{L^2(\Omega_N)} - \Big([f'(\hat{Y})]_N p_N\,,\,\phi\Big)_{L^2(\Omega_N)} = \\ \qquad = \Big([\hat{Y}]_N - [y^*]_N\,,\,\phi\Big)_{L^2(\Omega_N)} \mathbf{1}_{(T_0,T)} + \\ \qquad + \sum_{j=1}^{J} w_j'\Big(\Big([h_j]_N\,,\,[\hat{Y}]_N - [y^*]_N\Big)_{L^2(\Omega_N)}\Big)\,\big([h_j]_N, \phi\big)_{L^2(\Omega_N)} q_{j,N} \qquad \begin{matrix}\forall_{\phi\in P_1(\Omega_N)}\\ \text{on } [0,T]\end{matrix} \\ \dfrac{\partial p_N}{\partial n} = 0 \qquad \text{on } \partial\Omega_N \times (0,T) \\ p_N(T) = \mathbf{0} \end{cases} \tag{4.5}$$

together with

$$\begin{cases} -\beta_1 \frac{d}{dt} q_{1,N} + q_{1,N} = \big([g_1]_N, p_N\big)_{L^2(\Omega_N)} & \text{on } [0,T] \\ \vdots & \vdots \\ -\beta_J \frac{d}{dt} q_{J,N} + q_{J,N} = \big([g_J]_N, p_N\big)_{L^2(\Omega_N)} & \text{on } [0,T] \\ q_{j,N}(T) = 0 \quad \forall_{j=1,\ldots,J} \end{cases} \tag{4.6}$$

where $\mathbf{0} \in P_1(\Omega_N)$, $p_N(t) \in P_1(\Omega_N)$ and $q_{j,N}(t) \in \mathbb{R}$, for $j = 1,\ldots,J$, $t \in [0,T]$ and where the desired solution is $(p_N, q_{1,N},\ldots,q_{J,N})$. One may note, that $f'([\hat{Y}]_N)$ is not necessarily in $P_1(\Omega)$. For this reason, we define the above variational problem by inserting $[f'(\hat{Y})]_N$ term and not $f'([\hat{Y}]_N)$ term. Note moreover that term $(\nabla y_N, \nabla \phi_N)_{L^2(\Omega_N)}$ in the system (4.5) - (4.6) is well defined, because $P_1(\Omega_N) \subseteq H^1(\Omega_N)$ (see Theorem 2.1.1. in [13]).

Remark. The sets Ω and Ω_N are equal. Nonetheless, similarly as in the case of the system (2.12) - (2.13) in Section 2.2, in (4.5) - (4.6) we use notation „Ω_N" instead of „Ω" to stress that we are considering a space discretization of original the system (3.30) - (3.31). ▲

As mentioned in Section 2.2, for given $F, G \in P_1(\Omega)$, we can write:

$$(F,G)_{L^2(\Omega)} = (\vec{F})^T \mathbb{M}_N \, \vec{G}, \qquad (\nabla F, \nabla G)_{L^2(\Omega)} = (\vec{F})^T \mathbb{A}_N \, \vec{G}$$

where matrices $\mathbb{M}_N$ and $\mathbb{M}_N$ are defined as in Section 2.2. One can verify that in addition the following hold for a.e. $t \in [0,T]$:

$$\Big([f'(\hat{Y}(\,.\,,t)]_N F(\,.\,)\,,\, G(\,.\,)\Big)_{L^2(\Omega)} = (\vec{F})^T \mathbb{C}_N(t)\, \vec{G}$$

where matrix $\mathbb{C}_N(t)$ is defined by:

$$\mathbb{C}_N(t) = \left(\int_{\Omega_N} [f'(\hat{Y}(x,t)]_N \phi_m(x) \phi_n(x)\, dx \right)_{n,m=1}^{(N+1)^2}$$

Using the above remarks, we transform the system (4.5) - (4.6) to the matrix form:

$$\begin{cases} -\frac{d}{dt} \mathbb{M}_N \, \vec{p_N} \; + D\mathbb{A}_N \, \vec{p_N} \; - \mathbb{C}_N \, \vec{p_N} = \\ \qquad = \; \mathbb{M}_N \Big(\overrightarrow{[\hat{Y}]_N} - \overrightarrow{[y^*]_N} \Big) \mathbf{1}_{(T_0,T)} \; + \\ \qquad + \; \sum_{j=1}^{J} w'_j \Big(\overrightarrow{[h_j]_N}^{\,T} \, \mathbb{M}_N \Big(\overrightarrow{[\hat{Y}]_N} - \overrightarrow{[y^*]_N} \Big) \Big) \mathbb{M}_N \, \overrightarrow{[h_j]_N} \, q_{j,N} & \forall_{\phi \in P_1(\Omega_N)} \\ & \text{on } [0,T] \\ \vec{p_N} \, (T) = \mathbf{0} \end{cases} \tag{4.7}$$

together with

$$\begin{cases} -\beta_1 \frac{d}{dt} q_{1,N} + q_{1,N} = \overrightarrow{[g_1]_N}^{\,T} \, \mathbb{M}_N \, \vec{p_N} & \text{on } [0,T] \\ \vdots & \vdots \\ -\beta_J \frac{d}{dt} q_{J,N} + q_{J,N} = \overrightarrow{[g_J]_N}^{\,T} \, \mathbb{M}_N \, \vec{p_N} & \text{on } [0,T] \\ q_{j,N}(T) = 0 \quad \forall_{j=1,\ldots,J} \end{cases} \tag{4.8}$$

where $\mathbf{0} \in \mathbb{R}^{(N+1)^2}$ and where the desired solution is $(\vec{p_N}, q_{1,N},\ldots,q_{J,N})$.

Next, we approximate the solution of (4.7) - (4.8) by use of backward implicit Euler scheme, basing on $M+1$ time discretization points, uniformly distributed in $[0,T]$. Denote the subject approximate solution of (4.7) - (4.8) by $(\overrightarrow{P_N}, \hat{Q}_{1,N}, \ldots, \hat{Q}_{J,N})$. This approximate solution is a function defined in time discretization points, $t = m\tau_M$, $m = 0, 1, \ldots, M$, with values in $\mathbb{R}^{(N+1)^2} \times \mathbb{R}^J$.

Basing on the latter, we construct a function $(P_N, Q_{1,N}, \ldots, Q_{J,N})$, defined in the time discretization points $t = m\tau_M$, $m = 0, 1, \ldots, M$ and taking values in $P_1(\Omega_N) \times \mathbb{R}^J$ in the following way. We put $P_N(t) := \sum_{n=1}^{(N+1)^2} (\overrightarrow{P_N}\ (t))_n \phi_n$ and $Q_{j,N}(t) := \hat{Q}_{j,N}(t)$, for $j = 1, \ldots, J$, for $t = m\tau_M$, $m = 0, 1, \ldots, M$.

The scheme for numerical solving the system (3.30) - (3.31) is finished by obtaining the function $(P_N, Q_{1,N}, \ldots, Q_{J,N})$, described above. In other words, we treat $(P_N, Q_{1,N}, \ldots, Q_{J,N})$ as an approximation of the weak solution of (3.30) - (3.31).

The above scheme for solving (3.30) - (3.31) was utilized in our experiments, with $\hat{Y} = Y_N$, whenever computing the gradient of the cost functional $\mathcal{I}$ was necessary (recall that the gradient of $\mathcal{I}$ depends on the weak solution of (3.30) - (3.31)).

Note that the numerical scheme for (3.1) - (3.2), described above, is uniquely determined by the choice of the parameter N (determining the finite element space), the parameter M (determining the time discretization) and the parameter N_{Picard} (determining the Picard iterations method for treating the nonlinear terms in (3.1) - (3.2)). Moreover, for a given $\hat{Y}$, the above described scheme for (3.30) - (3.31) is determined by choice of N and M.

For use in our experiments, matrices $\mathbb{M}_N$ and $\mathbb{A}_N$ were assembled, similarly as in the numerical scheme described in Section 2.2, by explicit computing the integrals appearing in the definitions of the subject matrices (no numerical integration was used). The matrix $\mathbb{C}_N(t)$, for $t \in [0,T]$, was assembled with help of the function `quad` of the GNU Octave package, being a function for numerical integration.

4.3.2 Evaluating the cost functional

Below, we describe a numerical scheme for evaluation of the cost functional $\mathcal{I}$. The scheme was utilized in experiments described in Section 4.4 whenever it was necessary in the optimization methods involved in the subject experiments (see Section 4.2). We still assume that, for a given $F \colon \Omega \to \mathbb{R}$, the definitions of $[F]_N$ and $\overrightarrow{F}$ are as in Section 2.2.

For a given $\hat{v}$, the scheme for approximating the value $\mathcal{I}(\hat{v})$, defined by conditions (3.22) - (3.23), is as follows.

First, we use the described in Section 4.3.1 numerical scheme for obtaining a numerical solution of the system (3.1) - (3.2), with $x_j := \hat{v}_j$, for $j = 1, \ldots, J$. Let $(Y_N, k_{1,N}, \ldots, k_{J,N})$ denote this numerical solution.

Second, we perform integration with respect to space in time discretization points, i.e. we evaluate $\widetilde{E}_m := \big\|Y_N(\,.\,, t_m) - [y^*]_N(\,.\,, t_m)\big\|_2^2$ for $t_m = m\tau_M$, $m = 0, 1, \ldots, M$. To do it, we use the below formula, which is true by the relation (2.14):

$$\widetilde{E}_m = \int_{\Omega_N} \big|Y_N(x, t_m) - [y^*]_N(x, t_m)\big|^2\, dx = \Big(\overrightarrow{Y_N}\ (t_m) - \overrightarrow{y^*}\ (t_m)\Big)^T \mathbb{M}_N \Big(\overrightarrow{Y_N}\ (t_m) - \overrightarrow{y^*}\ (t_m)\Big)$$

for t_m as above, for $m = 0, 1, \ldots, M$.

Third, we integrate with respect to time on interval (T_0, T). However, now we dispose only certain values $\widetilde{E}_m$ for time discretization points. To integrate on interval (T_0, T), we need to extend this values to some function given on the whole interval. For this end, we assume the piecewise linear behavior of the function in question. More precisely, we construct a piecewise linear function $\widehat{E}\colon [0,T] \to \mathbb{R}$ by assigning $\widehat{E}(t_m) := \widetilde{E}_m$ for $t_m := m\tau_M$, $m = 0,\dots,M$ and

$$\widehat{E}(t) := \frac{t_{m+1}-t}{\tau_M}\widehat{E}(t_m) + \frac{t-t_m}{\tau_M}\widehat{E}(t_{m+1}) \tag{4.9}$$

for $t \in (t_m, t_{m+1})$, $m = 0,\dots,M-1$.

We intend to compute integral $\int_{T_0}^{T} \widehat{E}(t)\,dt$. We apply the trapezoidal quadrature to compute the subject integral, with nodes of the quadrature being the time discretization points $t_0,\dots,t_M$ plus the down limit of integration (if T_0 is not amongst the time discretization points). Since $\widehat{E}$ is continuous on $[0,T]$ and linear on each of intervals spanned by two neighboring nodes of the quadrature, the subject quadrature returns the exact value of the integral $\int_{T_0}^{T} \widehat{E}(t)\,dt$.

The numerical scheme for evaluation of $\mathcal{I}(\hat{v})$ is finished by obtaining, with the above means, integral $\int_{T_0}^{T} \widehat{E}(t)\,dt$. In other words, we assume that the value of the subject integral approximate the value of $\mathcal{I}(\hat{v})$.

4.3.3 Computing the gradient

Below, we describe a numerical scheme for computing an approximation of the gradient of $\mathcal{I}$. The scheme consists in approximate evaluating the formula (3.40), asserted in Corollary 3.2.16. The subject scheme was utilized in the experiments described in Section 4.4 whenever the employed optimization procedures (described in Section 4.2) required computing the gradient of $\mathcal{I}$.

For brevity, we will use the following notation for a part of the terms entering the formula (3.40): $\widetilde{T}^{\sigma}_{i,j} := \left(\mathcal{P}^{R,\Omega} \circ \mathcal{T}_{-\partial_i\sigma}\right)(\hat{v}_j)$, for $\sigma = \sigma_g, \sigma_h$ and for $j = 1,\dots,J$, $i = 1,\dots,\mathbf{d}$. Denote also $h_j := \Upsilon_{h_j}(\hat{v})$, for $j = 1,\dots,J$.

Assume that $\hat{v} \in V$ is given. The scheme for computing $\nabla\mathcal{I}(\hat{v})$ is as follows.

Keep in mind that we intend to approximately evaluate the formula (3.40), which, by Corollary 3.2.16, characterizes the gradient of $\mathcal{I}$.

First, we use the described in Section 4.3.1 numerical scheme for obtaining an approximate solution of the system (3.1) - (3.2). Denote this approximate solution by $(Y_N, k_{1,N},\dots,k_{J,N})$. Having this, we use the described in Section 4.3.1 numerical scheme for gaining an approximate solution of (3.30) - (3.31), with $\hat{Y} = Y_N$. Denote the latter approximate solution by $(P_N, Q_{1,N},\dots,Q_{J,N})$.

Remark. A consistency problem may seem to occur. Namely, $\hat{Y}$ is a function defined on $[0,T]$ with values in $L^2(\Omega)$ and Y_N is defined only in points $t_m \in [0,T]$, for $t_m = m\tau_M$, $m = 0,1,\dots,M-1$, where M and τ_M are as in Section 4.3.1, with values in $P_1(\Omega_N) \subseteq L^2(\Omega)$. This makes the above assignment $\hat{Y} = Y_N$ meaningless. To resolve this obstacle, one may attempt to extend Y_N to the whole interval $[0,T]$, e.g. by linear interpolation, before making the assignment. But in fact, this is not necessary, because the numerical scheme for solving (3.30) - (3.31), given in Section 4.3.1, utilizes only the information on $\hat{Y}$ in points t_m as above. Hence, an arbitrary extension of Y_N to whole $[0,T]$ is good, but also irrelevant at the same time. ▲

We intend to approximate the value of the formula (3.40), with Ω_N, Y_N, $k_{j,N}$, P_N, $Q_{j,N}$, $[y^*]_N$, $[h_j]$, $[\widetilde{T}^{\sigma_g}_{i,j}]_N$ and $[\widetilde{T}^{\sigma_h}_{i,j}]_N$ instead of Ω, $\hat{y}$, $\hat{\kappa}_j$, $\widetilde{p}$, $\widetilde{q}_j$, y^*, h_j, $\widetilde{T}^{\sigma_g}_{i,j}$ and $\widetilde{T}^{\sigma_h}_{i,j}$, respectively, for $j = 1,\dots,J$.

Thus, second, we perform integration w.r.t. space in time discretization points. More precisely, we evaluate the following:

$$\begin{aligned}
\widetilde{E}_{1,m} &:= k_{j,N}(t_m)\int_{\Omega_N} P_N(x,t_m)\,[\widetilde{T}^{\sigma_g}_{i,j}]_N(x)\,dx\\
\widetilde{E}_{2,m} &:= w'_j\Big(\int_{\Omega_N}\big(Y_N(x,t_m)-[y^*]_N(x,t_m)\big)\,[h_j]_N(x)\,dx\Big)\\
\widetilde{E}_{3,m} &:= Q_{j,N}(t_m)\widetilde{E}_{2,m}\int_{\Omega_N}\big(Y_N(x,t_m)-[y^*]_N(x,t_m)\big)\,[\widetilde{T}^{\sigma_h}_{i,j}]_N(x)\,dx
\end{aligned}$$

for $t_m = m\tau_M$, $m = 0,1,\ldots,M$. To compute the above integrals, we use the following identities, being true due to (2.14):

$$\begin{aligned}
\int_{\Omega_N} P_N(x,t_m)\,[\widetilde{T}^{\sigma_g}_{i,j}]_N(x)\,dx &= \Big(\vec{P_N}\,(t_m)\Big)^T\mathbb{M}_N\,\vec{\widetilde{T}^{\sigma_g}_{i,j}}\\
\int_{\Omega_N}\big(Y_N(x,t_m)-[y^*]_N(x,t_m)\big)\,[h_j]_N(x)\,dx &= \Big(\vec{Y_N}\,(t_m)-\vec{y^*}\,(t_m)\Big)^T\mathbb{M}_N\,\vec{h_j}\\
\int_{\Omega_N}\big(Y_N(x,t_m)-[y^*]_N(x,t_m)\big)\,[\widetilde{T}^{\sigma_h}_{i,j}]_N(x)\,dx &= \Big(\vec{Y_N}\,(t_m)-\vec{y^*}\,(t_m)\Big)^T\mathbb{M}_N\,\vec{\widetilde{T}^{\sigma_h}_{i,j}}
\end{aligned}$$

Third, we define the following function $\widehat{E}^{\nabla}\colon [0,T]\to\mathbb{R}$ and integrate it on interval (T_0,T). For $t_m = m\tau_M$, $m = 0,1,\ldots,M$ we put $\widehat{E}^{\nabla}(t_m) := \widetilde{E}_{1,m} + \widetilde{E}_{3,m}$. For $t\in(t_m,t_{m+1})$, $m = 0,1,\ldots,M-1$, we put $\widehat{E}^{\nabla}(t)$ to be equal the value implied by the linear interpolation of values of $\widehat{E}^{\nabla}$ in points t_m and t_{m+1}. More precisely, $\widehat{E}^{\nabla}(t)$ is defined by the formula (4.9), with $\widehat{E}$ replaced by $\widehat{E}^{\nabla}$.

For computing integral $\int_{T_0}^{T}\widehat{E}^{\nabla}(t)\,dt$, we use the trapezoidal quadrature, with $M+1$ nodes, coinciding with the $M+1$ time discretization points $t_0,\ldots,t_M$. Since the integrand $\widehat{E}^{\nabla}$ is continuous on $[0,T]$ and linear on each interval spanned by two neighboring quadrature nodes, the subject quadrature returns the exact value of $\int_{T_0}^{T}\widehat{E}^{\nabla}(t)\,dt$.

We assume that integral $\int_{T_0}^{T}\widehat{E}^{\nabla}(t)\,dt$ approximates the value of $(\Lambda^{\hat{v}}_j)_i$ in Corollary 3.2.16, for $j = 1,\ldots,J$, $i = 1,\ldots,\mathbf{d}$. This gives approximation of $\nabla\mathcal{I}(\hat{v})$, because $\nabla\mathcal{I}(\hat{v}) = \Lambda^{\hat{v}}$. Hence, the numerical scheme for computing the gradient of $\mathcal{I}$ in $\hat{v}$ is finished by evaluating the above integral.

4.3.4 1-D optimization

Now, we describe a method for approximate solving 1-D optimization problem $\mathrm{minn}_{s\in[0,1]}\mathcal{I}(\hat{v}+s\hat{d})$, entering the optimization methods described in Section 4.2 with suitable $\hat{v},\hat{d}\in V$. The method was utilized whenever solving the 1-D problem was necessary in the experiments described in Section 4.4.

A method for approximating the solution of the 1-D optimization problem will be called *line search procedure*. Moreover, denote $\widetilde{\mathcal{I}}(s) := \mathcal{I}(\hat{v}+s\hat{d})$. We will call $\widetilde{\mathcal{I}}$ the target function.

The precise description of our line search procedure for solving problem $\mathrm{minn}_{s\in[0,1]}\widetilde{\mathcal{I}}(\hat{v}+s\hat{d})$, for a given $\hat{v},\hat{d}\in V$ is as follows:

1. Initialization: we set $N_{ls} := 10$, define the search interval $I_{ls} := [0,1]$ and define the set of evaluation points $P_{ls} = \{\widetilde{s}_i = i/N_{ls}\colon\ i = 0,1,\ldots,N_{ls}\}$.

2. We approximate values $\widetilde{\mathcal{I}}(\widetilde{s}_i)$, for $i = 0,1,\ldots,N_{ls}$, using the numerical scheme for evaluating the cost functional $\mathcal{I}$ described in Section 4.3.2.

3. We choose local minimums, i.e. points $\widetilde{s}_i \in P_{ls}$ fulfilling $\widetilde{\mathcal{I}}(\widetilde{s}_i) \leq \widetilde{\mathcal{I}}(\widetilde{s}_{i-1})$ and $\widetilde{\mathcal{I}}(\widetilde{s}_i) \leq \widetilde{\mathcal{I}}(\widetilde{s}_{i+1})$ (or one of these inequalities, if $\widetilde{s}_i$ is the extremal point of I_{ls}). Amongst these local minimums, we choose the one which is situated closest the point $s = 0$. Denote this minimum by $\widetilde{s}$ and its index in P_{ls} by $\widetilde{i}$ (i.e. $\widetilde{i}$ is the index such that $\widetilde{s}_{\widetilde{i}} = \widetilde{s}$).

4. We verify whether the stop criterion is fulfilled or not. If yes — then we terminate the line search algorithm and return point $\widetilde{\widetilde{s}} := \widetilde{s}$. The stop criterion is as follows: verify whether $\widetilde{s}_1 - \widetilde{s}_0 \leq R_{ls}$, where R_{ls} is given. Note that, since the points $\widetilde{s}_0, \ldots, \widetilde{s}_{N_{ls}}$ are uniformly distributed in I_{ls}, we can alternatively verify the inequality $\widetilde{s}_{i+1} - \widetilde{s}_i \leq R_{ls}$ for an arbitrary $i = 1, \ldots, N_{ls}$.

   In the experiments described in Section 4.4, we have always used the value $R_{ls} = 0.001$.

5. We determine a new search interval and a new set of evaluation points in the following way:

   (a) If $\widetilde{s} = \widetilde{s}_0$, set $I_{ls} := [\widetilde{s}_0, \widetilde{s}_1]$ and $P_{ls} := \{\widetilde{s}_0, \frac{1}{2}(\widetilde{s}_0 + \widetilde{s}_1), \widetilde{s}_1\}$ (3 new evaluation points).

   (b) If $\widetilde{s} = \widetilde{s}_{N_{ls}}$, set $I_{ls} := [\widetilde{s}_{N_{ls}-1}, \widetilde{s}_{N_{ls}}]$ and $P_{ls} := \{\widetilde{s}_{N_{ls}-1}, \frac{1}{2}(\widetilde{s}_{N_{ls}-1} + \widetilde{s}_{N_{ls}}), \widetilde{s}_{N_{ls}}\}$ (3 new evaluation points).

   (c) If neither of the above two cases hold, set $I_{ls} := [\widetilde{s}_{\widetilde{i}-1}, \widetilde{s}_{\widetilde{i}+1}]$ and
   $P_{ls} := \{\widetilde{s}_{\widetilde{i}-1}, \frac{1}{2}(\widetilde{s}_{\widetilde{i}-1} + \widetilde{s}_{\widetilde{i}}), \widetilde{s}_{\widetilde{i}}, \frac{1}{2}(\widetilde{s}_{\widetilde{i}} + \widetilde{s}_{\widetilde{i}+1}), \widetilde{s}_{\widetilde{i}+1}\}$ (5 new evaluation points).

   Set $N_{ls} := \#P_{ls}$. Relabel the points of set P_{ls} as $\widetilde{s}_0, \ldots, \widetilde{s}_{N_{ls}}$.

6. Go to the step 2.

The line search procedure is terminated by determining the point above denoted as $\widetilde{\widetilde{s}}$. In other words, we assume that $\widetilde{\widetilde{s}}$ approximates the solution of problem $\operatorname{minn}_{s\in[0,1]}\widetilde{\mathcal{I}}(\hat{v} + s\hat{d})$, for a given $\hat{v}, \hat{d} \in V$. The above procedure for solving problem $\operatorname{minn}_{s\in[0,1]}\widetilde{\mathcal{I}}(\hat{v} + s\hat{d})$ was always utilized whenever solving this kind of problem was necessary in the experiments described in Section 4.4.

Remark. Assume that the function $\mathcal{I}(\hat{v}^n + \,.\,d^n)$ is sufficiently regular for convergence of the line search procedure to the real solution of $\operatorname{minn}_{s\in[0,1]}\mathcal{I}(\hat{v}^n + sd^n)$, for $\hat{v}^n$ and d^n being as in the optimization methods described in Section 4.2. Compare the stop criterion imposed in the optimization methods and the stop criterion in the above line search procedure. The stop criterion for the optimization methods is fulfilled if the line search procedure returns $\widetilde{\widetilde{s}} = 0$. This happens if $\widetilde{s} = 0$ and $\widetilde{s}_1 - \widetilde{s}_0 \leq R_{ls}$. Thus, due to our assumption, one may conclude that the stop criterion for the optimization methods is fulfilled if the „real step length” s_n, i.e. the real solution of $\operatorname{minn}_{s\in[0,1]}\mathcal{I}(\hat{v}^n + sd^n)$, is lesser than R_{ls}. ▲

Remark. The general idea of the above line search procedure can be explained in the following way. The subject procedure consists of two stages. In the first stage, we perform the uniform line search, with $N_{ls} + 1$ *evaluation points*, for a given natural N_{ls}. The uniform line search results in reducing the initial *search interval* $[0, 1]$ to some new shorter search interval. Next, in the second stage, we run iteratively a bisection-like line search on the new search interval. The stage of uniform line search consists simply in an additional iteration with many $(N_{ls} + 1)$ evaluation points, placed at the beginning of the whole line search procedure. The bisection-like line search stage is realized by all subsequent iterations. ▲

Remark. The motivation behind the usage of the above composite line search method, consisting of two stages, is as follows. Recall that we intend to solve the minimization problem of extracting the local minimum on $[0, 1]$ being the closest to $s = 0$. The uniform line search utilizes more evaluation points in one iteration than the bisection-like line search. Hence, in the first iteration, we use the uniform line search to reduce the risk that we will loose essential information on the geometry of the target function $\widetilde{\mathcal{I}}$ on the initial search interval. This increases the chance that we select a consecutive search interval which contains the the minimum of $\widetilde{\mathcal{I}}$ which is the closest to $s = 0$. Next, after choosing the new search interval, which is significantly shorter than the initial one, we switch to the bisection-like line search because it is superior to the uniform line search in terms of speed. ▲

Remark. We use name „bisection-like line search", not „bisection line search", because the latter is usually used for other algorithm. We have not found the description of the above bisection-like line search method in publications, thus we could not establish the proper name of the subject line search method. The source in which we have encountered the description of the subject method is the lecture script [34] (in Polish). ▲

4.4 Results of simulations

In this section, we describe results of our experiments concerning attempts to find numerically an approximate solutions of optimization problem (3.24). All below described simulations based on one of the optimization methods specified in Section 4.2. The numerical schemes which were utilized for implementing these methods are described in Section 4.3. The assumptions concerning problem (3.24) were as in Section 4.1.

In Section 4.4.1, we compare the results of the SD method, for two different parameters T_0 in the cost functional $\mathcal{I}$ and two different process initial states y_0 in the system (3.1) - (3.2). The results suggest that the performance of the SD method is poorer for the parameter T_0 close to T. Moreover, a dependence of the optimization output on y_0 is observed for T_0 close to T, what is opposite to our expectations (explained in the beginning of Chapter 4).

In Section 4.4.2, we vary not only T_0 and y_0, but also the reference state y^* in the system (3.1) - (3.2). Moreover, the simulations are performed with the use of three optimization algorithms: SD, CG-r and CG+r. The results confirm further that the average performance of the SD method varies as T_0 varies (average, in a sense of both the mean and the median of number of iterations). Nevertheless, the difference in the average performance vanquishes when switching from the SD method to the CG+r method. Basing on the results, we conclude that the CG+r method is most appropriate for our optimization problem.

In Section 4.4.3, we compare results of the CG+r method for the optimization problem with T_0 close to T, for varying values of the parameter T and for a varying initial state y_0. The results suggest that the average performance of the CG+r method changes with changes of the time interval, determined by the parameter T. However, it is also observed that lengthening the time interval resulted with greater independence on y_0 of the optimization output. Due to our general motivations, see the beginning of Chapter 4, we prefer situations exhibiting the latter effect, thus simulations with rather long time horizon are interesting for us. Nevertheless the long time interval makes the optimization procedures more time consuming. Hence, in Section 4.4.4, we propose some possible refinements to our optimization procedures, to test in the future experiments.

All below described experiments were performed with the use of the GNU Octave software (version 3.6.4) and computer cluster Halo2 (a machine of Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw). Halo2 processors are AMD

Quad-Core Opteron processors with architecture x86_64 „Barcelona". No parallelization was used, each optimization procedure run using one processing core.

Two types of plots are contained in the present section: 1) plots of scalar functions defined on domain Ω (e.g. the initial state y_0 or the reference state y^* in the system (3.1) - (3.2)) and 2) plots of particular configurations of the control and measurement devices. Conventions for both mentioned types of plots are analogous as the conventions described in Section 2.3.

By *the configuration of the control and measurement devices* we mean, similarly as in Section 2.3, the choice of the supports of functions g_j and h_j, for $j = 1, \ldots, J$, characterizing the control and measurement devices actions. Here, these are functions $\mathcal{P}^{R,\Omega}\mathcal{T}_{\sigma_g}(x_j)$ and $\mathcal{P}^{R,\Omega}\mathcal{T}_{\sigma_h}(x_j)$ in system (3.1) - (3.2), with $x_j := \hat{v}_j$ for $j = 1, \ldots, J$, where $\hat{v} \in V$ is a given control parameter.

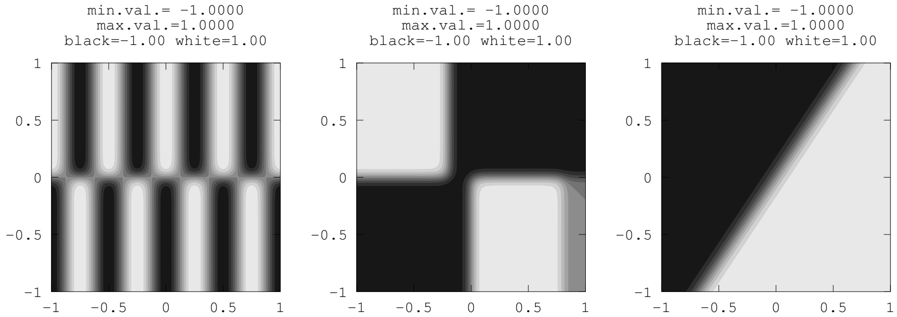
Note that, due to specific assumptions for the pattern functions (see (4.1)), the visualization of the supports of the functions characterizing the devices actions give characterization of points $x_1, \ldots, x_J$ (up to permutation). In consequence, one can retrieve the control parameter $\hat{v} \in V$ basing on the mentioned visualizations of supports.

In all experiments described in the present section, initial states y_0 and reference states y^* for the system (3.1) - (3.2) were chosen from the set of three particular variants, presented in Figure 4.1. In description of each experiment, we will specify explicitly which variants were used. Figure 4.1 presents the same plots as Figure 2.3 in Section 2.3 but we place it here again, for completeness and convenience. The formulas determining the functions plotted in Figure 4.1 are:

$$\hat{y}(x_1, x_2) \;=\; \cos\big(4\pi x_1\big) \cdot \Big(1 - 2\big(1 + e^{30\,x_2}\big)^{-1}\Big) \tag{4.10}$$

$$\begin{aligned} \hat{y}(x_1, x_2) \;=\; & -1 + \Big(2\big(1 + e^{-30\,x_1}\big)^{-1} - \big(1 + e^{-30(x_1 - 0.8)}\big)^{-1}\Big) \cdot \big(1 + e^{30\,x_2}\big)^{-1} + \\ & + \; 2\big(1 + e^{30(x_1 + 0.2)}\big)^{-1} \cdot \big(1 + e^{-30\,x_2}\big)^{-1} \end{aligned} \tag{4.11}$$

$$\hat{y}(x_1, x_2) \;=\; 1 - 2\big(1 + e^{-15\frac{3\sqrt{13}}{13}(x_2 - 1.5\,x_1)}\big) \tag{4.12}$$



(a) Variant 1.

(b) Variant 2.

(c) Variant 3.

Figure 4.1: Variants of the initial state y_0 and the reference state y^* utilized in the experiments described in Section 4.4. The plots present scalar functions defined on the considered $\mathbb{R}^2$ domain. The formulas determining the plotted functions are (4.10) for Fig. 4.1a, (4.11) for Fig. 4.1b and (4.12) for Fig. 4.1c.

Moreover, in all experiments, it was assumed that the number of the control and measurement devices equals twenty ($J = 20$). In addition, for each experiment experiment, the configuration of control and measurement devices used as a start configuration for the optimization algorithms was as in Figure 4.2.

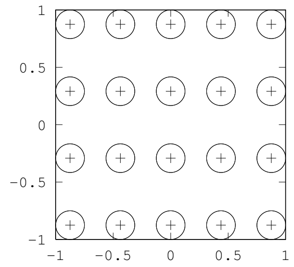


Figure 4.2: The start configuration of control and measurement devices for optimization procedures utilized in the experiments described in Section 4.4. In other words, the plot characterizes the control parameter $\hat{v}^0 \in V$, utilized in the descriptions in Section 4.2.

Also, in each of the below described experiments the following data were used. The parameters concerning the system (3.1) - (3.2) (see Section 4.1 for explanation of the parameters meaning) were:

$$
\begin{array}{llll}
D = 0.03 & r_{\sigma,2} = 1/8 & C_{switch} = 0.2 & C_{smooth} = 0.9 \\
\beta_j = 1 \ \forall_{j=1,\ldots,J} & r_{\sigma,1} = 0.6 \cdot r_{\sigma,2} & L_w = -10 & \\
\kappa_{j0} = 0 \ \forall_{j=1,\ldots,J} & C_g = 16/\pi & H_w = 10 &
\end{array}
$$

Other parameters, i.e. the parameter T (concerning the system (3.1) - (3.2), see Section 4.1), parameters N, M, N_{Picard} (concerning the numerical scheme, see Section 4.3) and the parameter N_{opt} (concerning the stop of optimization algorithms, see Section 4.2) will be specified below, in the descriptions of particular experiments. The choice of the optimization procedures (SD method, CG-r method or CG+r method) also will be specified there.

4.4.1 Experiment 1 — various initial conditions and cost functionals

This experiment served for comparing the behavior of the SD method for optimization problem (3.24), for two different parameters T_0, entering the definition of the cost functional $\mathcal{I}$. One of the considered values of T_0 correspond to the concept of the cost functional that consists in measuring the gap between the reference state and the evolution of the process on the whole time interval of the experiment, $[0, T]$. The other value of T_0 corresponds to the cost functional concept that consists in measuring the subject gap only in the neighborhood of the terminal time T.

The second of the above cost functional concepts fits our main motivation, described in the beginning of Chapter 4, which is the problem of choosing the targeting of the devices actions w.r.t. the task of bringing the process state possibly close to the reference state at the terminal time T. In this case of the cost functional, it is desired that the optimization procedure will return results being independent of the initial state y_0 (see the explanation in the beginning of Chapter 4). Unfortunately, the latter occurs to be not true, at least with the data employed in the present experiment. Below, we suggest some possible solutions to this situation.

Despite the fact that we are interested in the cost functional with measurement concentrated close to terminal time, the comparison with the other mentioned type of the cost functional (measurement distributed over the whole $[0, T]$) also is interesting. This comparison, as we will see below, can suggest that the SD method applied in the investigated optimization problem differs in the its performance depending on the chosen parameter T_0.

In the presently considered experiment, the time horizon for the system (3.1) - (3.2) was $T = 2$.

The reference state y^* was assumed to be as in Figure 4.1c.

The following parameters for the numerical scheme were assumed: $N = 80$, $N_{Picard} = 2$ and $M = 100$.

The applied optimization algorithm was SD method, described in Section 4.2, with $N_{opt} = 1000$.

Four simulations were performed, corresponding to two variants of the initial state y_0 and two values of the cost functional parameter T_0. The subject two choices of y_0 were corresponding to the functions plotted in Figure 4.1a and Figure 4.1b (we call it variant 1. and variant 2., respectively). The two considered values of T_0 were $T_0 = 0$ and $T_0 = 0.9T$.

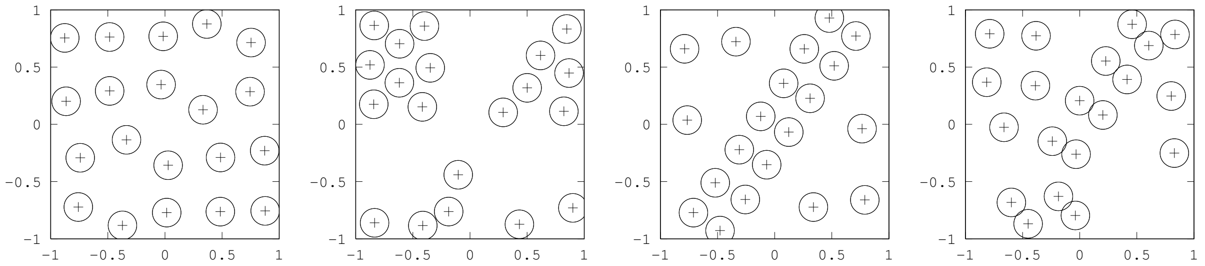
Simulation	Iterations	Initial cost	Terminal cost
y_0 variant 1, $T_0 = 0$	39	0.918962	0.720012
y_0 variant 2, $T_0 = 0$	68	1.780059	0.981571
y_0 variant 1, $T_0 = 0.9T$	118	0.109127	0.017851
y_0 variant 2, $T_0 = 0.9T$	**1000**	0.232079	0.020284

Table 4.1: Performance of optimization procedures considered in Section 4.4.1, for two variants of the initial state y_0 and two values of the parameter T_0 considered in the subject section. Column „Iterations" informs how many iterations of the optimization procedure (see integer n in the description of the SD and CG methods, given in Section 4.2) were performed before the procedure fulfilled the stop criterion. If the optimization procedure was terminated due to the condition $n = N_{opt}$ and not $s_n = 0$, (see the specification of the stop criterion, given in Section 4.2), the number of iteration is given with bold font. The last two columns present the values of the cost functional at start of an optimization procedure and after the optimization procedure terminated. In other words, values $\mathcal{I}(\hat{v}^0)$ and $\mathcal{I}(\hat{v}^n)$, for n corresponding to the stop iteration, are presented there (with $\hat{v}^i$ being as in the description of SD and CG methods given in Section 4.2).

Table 4.1 compares the performance of the SD method in the four considered simulations. A grater number of iterations was necessary to fulfill the stop criterion for simulations concerning $T_0 = 0.9T$. In particular, in the simulation concerning $T_0 = 0.9T$ and variant 2. of y_0, the SD method failed to stop in one thousand iterations. This is greatly worse result that in the case of the other three simulations. One can pose a hypothesis that worse performance of the SD method for $T_0 = 0.9T$ is a general rule. In Section 4.4.2, we will make a further step towards verification of the subject hypothesis.

Now, let us take a look at the devices configurations obtained by the here considered optimization procedures.

The two simulations with $T_0 = 0$ differ only with the variant of y_0. Comparing Figures 4.3a and 4.3b we see that the result of these simulation varies strongly. The meaning of the optimization problem (3.24) with the parameter $T_0 = 0$ entering the cost functional can be explained as follows. The problem is to adjust the configuration of the devices in a manner that results in quick reduction of the difference between the initial state of the process and the reference state. In other words, the difference between y_0 and y^* is crucial and hence the dependence on y_0 of the subject two simulations results could be expected. In addition, one may compare Figures 4.3a and 4.3b with Figures 4.4a and 4.4b, respectively. If one merged the corresponding figures pairwise, it could be noted, that the obtained targeting of the devices actions coincide with the



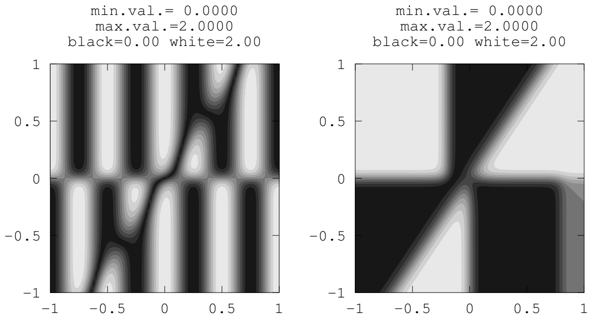
(a) y_0 variant 1., $T_0 = 0$, iter. 39.

(b) y_0 variant 2., $T_0 = 0$, iter. 68.

(c) y_0 variant 1., $T_0 = 0.9T$, iter. 118.

(d) y_0 variant 2., $T_0 = 0.9T$, iter. 1000.

Figure 4.3: Configurations of the control and measurement devices actions, obtained by optimization procedures addressed in Section 4.4.1, for two variants of the initial state y_0 and two values of the parameter T_0 considered in the subject section. Values of the parameter T_0 and the variants of the initial state (corresponding to the functions plotted Figure 4.1) are indicated in the figures. Each plot presents the configuration corresponding to the terminal iteration of the subject optimization procedures (see column „Iterations” in Table 4.1).



(a) $|y_0 - y^*|$ for variant 1. of y_0.

(b) $|y_0 - y^*|$ for variant 2. of y_0.

Figure 4.4: The function $|y_0 - y^*|$, for y^* being as assumed in Section 4.4.1 and for two variants of y_0 considered in the subject section. Fig. 4.4a corresponds to the case of y_0 being as in Fig. 4.1a and Fig. 4.4b corresponds to y_0 as in Fig. 4.1b.

light fields in the plots of difference $|y_0 - y^*|$. It means that the optimization procedure has located the control and measurement devices actions there where the subject difference was the greatest.

The two simulations corresponding to $T_0 = 0.9T$ also differ only with the variant of y_0. However, this time we expect a looser dependence of the results on y_0. The latter expectation can be justified with reasoning as already mentioned in the introduction to Chapter 4. Let us recall it. Most of the data considered in the subject simulations is as in Section 2.3.2 and Section 2.3.3. There, the process occurred to stabilize in the neighborhood of certain time-invariant state, independent of y_0. Therefore, one could expect that in the present simulations the process also may stabilize near certain y_0-independent, time-invariant state. If this was the case, then the values of the cost functional would not differ significantly under changes of y_0, because for $T_0 = 0.9T$ the cost functional accounts only the behavior of the process near the terminal time, where the process evolves independently on y_0. In consequence, minimal points of the cost functional also would depend on y_0 insignificantly.

The results returned by the SD method for the case $T_0 = 0.9T$ deny part of the above expectations. Comparing Figures 4.3c and 4.3d shows that the obtained configurations of the devices differ for the two considered variants of y_0. The difference between the two patterns is not that big as in the case of Figures 4.3a and 4.3b. Nevertheless, depending on particular accuracy requirements, the match between the patterns in Figures 4.3c and 4.3d can be considered to be not enough accurate.

Several hypotheses concerning the latter observations, concerning the dependence of the optimization results of y_0 in case $T_0 = 0.9T$, can be posed. In particular, the following ones seem to be natural:

(a) the above hypotheses concerning the stabilization near to a time-invariant state independent of y_0 are false,

(b) the time interval of the model in the presently considered simulations was too short for the process to get close enough to the time-invariant state,

(c) the optimization procedure was not accurate enough to approximate the minimum of the cost functional with sufficient precision (it is possible because in the simulation concerning variant 2. of y_0 and $T_0 = 0.9T$ the optimization procedure stopped due to a large number of iterations, not due to a short step length — see Table 4.1).

We will touch part of the above hypotheses in the forthcoming sections.

4.4.2 Experiment 2 — comparing optimization methods

In the below described experiment, we compare performance of the SD method with performance of CG methods (more precisely, the CG-r and CG+r methods), for optimization problem (3.24). The simulations were performed for varying initial states y_0, varying reference states y^*, entering the system (3.1) - (3.2), and varying values of the parameter T_0, entering the cost functional $\mathcal{I}$.

The aims of the experiment were threefold. First, we wanted to get further verification of the observations made in Section 4.4.1, that the performance of the SD method for the optimization problem with $T_0 = 0.9T$ is inferior to the case of $T_0 = 0$. This objective is realized by performing more simulations, using the SD method, for both cases of T_0. Second, we posed a particular aim to verify whether the CG methods are more appropriate for our optimization problem in the case of $T_0 = 0.9T$, which is particularly interesting for us (see the introduction to Chapter 4). Third, we wanted to compare the results obtained in Section 4.4.1 for the case $T_0 = 0.9T$ with the use of the SD method with results obtained in the same case with the use of the CG methods. This serves for investigating the reasons of the dependence of the optimizations results on y_0, what was observed in Section 4.4.1. A discussion concerning the three introduced objectives will be conducted below.

In the presently considered experiment, the time horizon for the system (3.1) - (3.2) was $T = 2$.

The following parameters for the numerical scheme were assumed: $N = 80$, $N_{Picard} = 2$ and $M = 100$ (i.e. $\tau_M = M^{-1} = 0.02$).

The stop criterion parameter for the optimization methods was $N_{opt} = 1000$.

54 simulations were performed, corresponding to different variants of: the initial state y_0, the reference state y^*, the cost functional parameter T_0 and the optimization method. Three choices

of y_0, three choices of y^*, two choices of T_0 and three choices of the optimization methods were considered, what gives $3 \times 3 \times 2 \times 3 = 54$ different data configurations. Hence 54 simulations.

The three considered variants of y_0 were corresponding to the three functions, plotted in Figure 4.1a, Figure 4.1b and Figure 4.1c (we call it variant 1., variant 2. and variant 3., respectively). The three variants of y^* also were corresponding to these three functions. The two values of T_0 taken into account were $T_0 = 0$ and $T_0 = 0.9T$. The three optimization methods were: 1) SD method, 2) CG-r method and 3) CG+r method (see Section 4.2 for explanation of these methods).

According to the above, four of the simulations described here are exactly those described in Section 4.4.1 (the simulations with variant 3. of y^* and with the use of the SD method). Nevertheless, we attach the result of the subject four simulations here, for more convenient comparison with other results.

Simulation	Iterations			Ratio	
	SD	CG-r	CG+r	$\frac{\text{CG-r}}{\text{SD}}$	$\frac{\text{CG+r}}{\text{SD}}$
y_0 var. 1, y^* var. 3, $T_0 = 0$	39	58	58	1.4872	1.4872
y_0 var. 2, y^* var. 3, $T_0 = 0$	68	106	97	1.5588	1.4265
y_0 var. 3, y^* var. 3, $T_0 = 0$	188	139	176	0.7394	0.9362
y_0 var. 1, y^* var. 2, $T_0 = 0$	261	49	52	0.1877	0.1992
y_0 var. 2, y^* var. 2, $T_0 = 0$	558	76	82	0.1362	0.1470
y_0 var. 3, y^* var. 2, $T_0 = 0$	**1000**	139	168	0.1390	0.1680
y_0 var. 1, y^* var. 1, $T_0 = 0$	184	52	50	0.2826	0.2717
y_0 var. 2, y^* var. 1, $T_0 = 0$	179	92	87	0.5140	0.4860
y_0 var. 3, y^* var. 1, $T_0 = 0$	106	79	122	0.7453	1.1509
Mean	*287.0*	*87.8*	*99.1*	*0.6434*	*0.6970*
Median	*184.0*	*79.0*	*87.0*	*0.5140*	*0.4860*
y_0 var. 1, y^* var. 3, $T_0 = 0.9T$	118	64	52	0.5424	0.4407
y_0 var. 2, y^* var. 3, $T_0 = 0.9T$	**1000**	255	279	0.2550	0.2790
y_0 var. 3, y^* var. 3, $T_0 = 0.9T$	211	77	77	0.3649	0.3649
y_0 var. 1, y^* var. 2, $T_0 = 0.9T$	212	96	84	0.4528	0.3962
y_0 var. 2, y^* var. 2, $T_0 = 0.9T$	250	89	95	0.3560	0.3800
y_0 var. 3, y^* var. 2, $T_0 = 0.9T$	384	125	112	0.3255	0.2917
y_0 var. 1, y^* var. 1, $T_0 = 0.9T$	**1000**	60	82	0.0600	0.0820
y_0 var. 2, y^* var. 1, $T_0 = 0.9T$	526	35	35	0.0665	0.0665
y_0 var. 3, y^* var. 1, $T_0 = 0.9T$	**1000**	137	42	0.1370	0.0420
Mean	*522.3*	*104.2*	*95.3*	*0.2845*	*0.2603*
Median	*384.0*	*89.0*	*82.0*	*0.3255*	*0.2917*

Table 4.2: Performance of the optimization methods considered in Section 4.4.2, for three variants of the initial state y_0, three variants of the reference state y^* and two values of the parameter T_0 considered in the subject section. The meaning of column „Iterations" and the notation concerning the stop criterion (the bold font entries) are as in the case of Table 4.1. Column „Ratio" presents, for each simulation, the ratio of iteration numbers concerning indicated optimization methods (with rounding to 4 significant digits). The mean values given in the latter column refer to the mean of the ratio values, not to the ratio of the mean numbers of iterations in the preceding columns. The analogous convention concerns the median values.

Consider the data presented in Table 4.2. First, observe that both the mean and the median of the number of iterations necessary for SD method to stop are much greater in the case of $T_0 = 0.9T$ than in the case of $T_0 = 0$. Basing on the subject result, one may suspect that the SD method has worse performance in the case $T_0 = 0.9T$ (in the sense of the expected value or of the median). This is consistent with the preliminary observation concerning the behavior of the SD method, contained in Section 4.4.1.

Next, compare the performance of the SD method with the performance of the two considered CG methods. In the case of $T_0 = 0$, we observe that the mean of the reduction of the number of iterations necessary to achieve the stop criterion when using one of the CG methods instead of the SD method is over 30% (see column „Ratio" in Table 4.2). The median of the reduction is about 50%. In the case of $T_0 = 0.9T$, both the mean and the median of the reduction are significantly greater and take value about 70%.

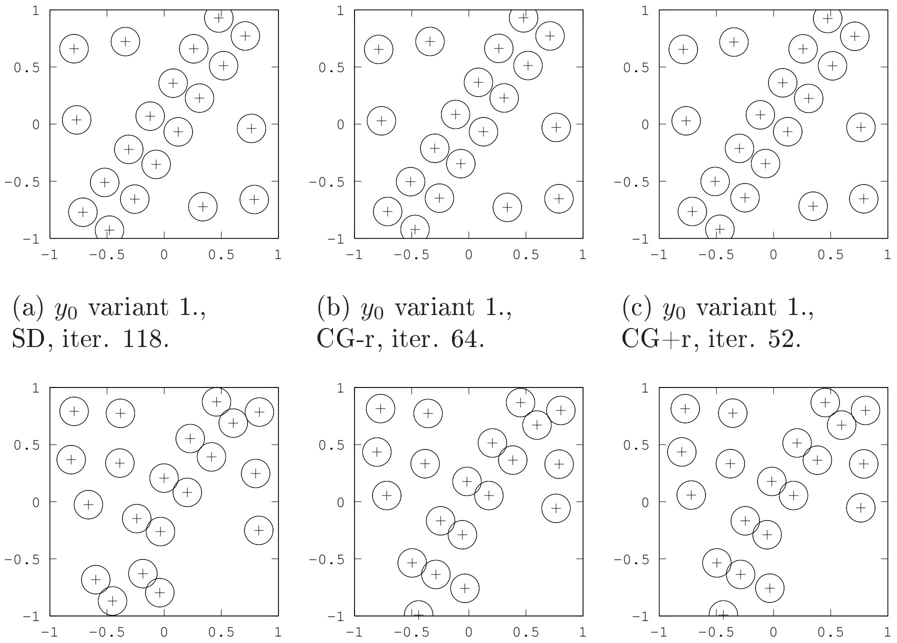
In addition, we remark that for the CG methods the optimization procedures never stopped due to achieving a large number of iterations, equal N_{opt}. For these methods, the stop reason was always a short step length (for the description of the stop criterion, see Section 4.2). Note however an interesting particularity that in the case of $T_0 = 0$ there were two situations where the SD method was in advantage to the CG methods, in sense of number of iterations (variants 1. and 2. of y_0 with variant 3. of y^*), while in the case of $T_0 = 0.9T$ the CG methods always behaved better than the SD method.

Another interesting observation is that both the mean and the median of the number of iterations for the CG+r method were similar both for $T_0 = 0$ and for $T_0 = 0.9T$. For the SD method, this is not true. In the case of the CG-r method, the differences in the mean and the median of the number of necessary iterations occurring in comparison of $T_0 = 0$ and $T_0 = 0.9T$ cases also were small (in comparison to the SD method), but not that small as in the case of the CG+r method. It looks like the performance of the CG+r method, in sense of the mean and the median, was most immune to the change of the parameter T_0, among the considered methods.

To sum up the above observations, the SD method seems to have statistically worse performance in the case of $T_0 = 0.9T$ than in the case of $T_0 = 0$ (in the sense of the mean and the median of the number of iterations). This difference in the behavior of the optimization method is leveled by switching to the CG+r method. In both cases ($T_0 = 0$ and $T_0 = 0.9T$), switching to one of the CG methods was a fruitful step. Nevertheless, the benefits of switching to the CG methods were considerably higher in the case $T_0 = 0.9T$.

Among the three proposed optimization methods, the method that seems to be in favor for our purposes is the CG+r method. It was most immune to changes of the cost functional (in the sense of the mean and the median of the number of iterations). In this sense, the performance of this method is most predictable. Moreover, in the case of $T_0 = 0.9T$, which is the case of our interest, its performance is statistically the best among the proposed methods (in the sense of the mean and the median). Besides, applying a reset procedure in the nonlinear conjugate gradient method seems to be a standard approach, at least in a part of the literature concerning this method.

In addition to the above observations, we focus for the moment on the simulations concerning the case of variant 3. of y^* and $T_0 = 0.9T$. This case was one of the subjects of Section 4.4.1, with conclusion that the dependence of the optimization results on the initial state y_0 can be observed. In simulations described in Section 4.4.1, SD method was used. In the simulation concerning variant 2. of y_0 and $T_0 = 0.9T$, it stopped due to large number of iterations, not due to short step length (see Table 4.1). Therefore the obtained approximation of local minimum of the cost functional $\mathcal{I}$ could be of low quality. Now, we can compare the results described in Section 4.4.1 with optimization results obtained by CG-r and CG+r methods. For the latter methods the optimization procedure always stopped due to the short step length (see Table 4.2).



(a) y_0 variant 1., SD, iter. 118.

(b) y_0 variant 1., CG-r, iter. 64.

(c) y_0 variant 1., CG+r, iter. 52.

(d) y_0 variant 2., SD, iter. 1000.

(e) y_0 variant 2., CG-r, iter. 255.

(f) y_0 variant 2., CG+r, iter. 279.

Figure 4.5: Configurations of the control and measurement devices concerning variant 3. of y^* and $T_0 = 0.9T$, obtained as a result of simulations described in Section 4.4.2, for two variants of the initial state y_0 and three optimization methods. The optimization methods and the variants of the initial state (corresponding to the functions plotted Figure 4.1) are indicated in the figures. Each plot presets the configuration corresponding to the final iteration of the subject optimization procedures (see column „Iterations" in Table 4.2). Figures 4.5a and 4.5d are the same as Figures 4.3c and 4.3d, respectively, but we place them here for more convenient comparison of optimization methods.

For this reason, we assume that, for variant 2. of y_0 and $T_0 = 0.9T$, the approximation of the local minimums of $\mathcal{I}$ obtained be the subject methods is of higher quality than the approximation obtained in Section 4.4.1. Thus, comparison of the results can serve for verifying the hypothesis that the dependence on y_0 observed in Section 4.4.1 was a consequence of poor quality of the optimization procedures output (see hypothesis (c) in the concluding part of the latter section).

Comparing particular plots presented in Figure 4.5, one may observe that for CG-r and CG+r methods dependence of the optimization output on y_0 also takes place, similarly as in the case of the SD method. This, under the assumption that the quality of the optimization results is acceptably high for the CG-r and CG+r methods, stays against the hypothesis that sole optimization output quality was responsible for the dependence on y_0 observed in simulations described in Section 4.4.1.

4.4.3 Experiment 3 — various initial conditions and time horizons

In the present section, we compare results of the CG+r method applied to optimization problem (3.24), for two different initial states y_0 and for the time horizon parameter T greater that in Section 4.4.1 and Section 4.4.2. The cost functional considered in the below described experiment correspond to the idea of measurement of the gap between the process and the reference state in the neighborhood of the terminal time T.

The aim of the below described experiment was further attempt to verify hypotheses concerning the dependence of the optimization results on y_0, observed in the experiments described in Section 4.4.1 and Section 4.4.2. As we will see below, lengthening the time interval results in considerably higher immunity of the optimization problem to the changes of y_0. This supports hypothesis (b), formulated as one of the conclusions of Section 4.4.1.

As a side result, we observe that the number of CG+r iterations necessary for time horizons parameters T considered here is higher that in the previous sections, for $T = 2$.

In the presently considered experiment, the reference state y^* for the system (3.1) - (3.2) was assumed to be as the function plotted in Figure 4.1c.

The following parameters for the numerical scheme were assumed: $N = 80$, $N_{Picard} = 2$ and $\tau_M = 0.02$.

The applied optimization algorithm was CG+r method, described in Section 4.2.

The stop criterion parameter for the optimization methods was $N_{opt} = 600$.

Four simulations were performed, corresponding to two different variants of the initial state y_0 and two different variants of the time horizon parameter T. The subject two variants of y_0 were as the functions plotted in Figure 4.1a and Figure 4.1b (we call it variant 1. and variant 2., respectively). The two values of the parameter T were $T = 4$ and $T = 6$.

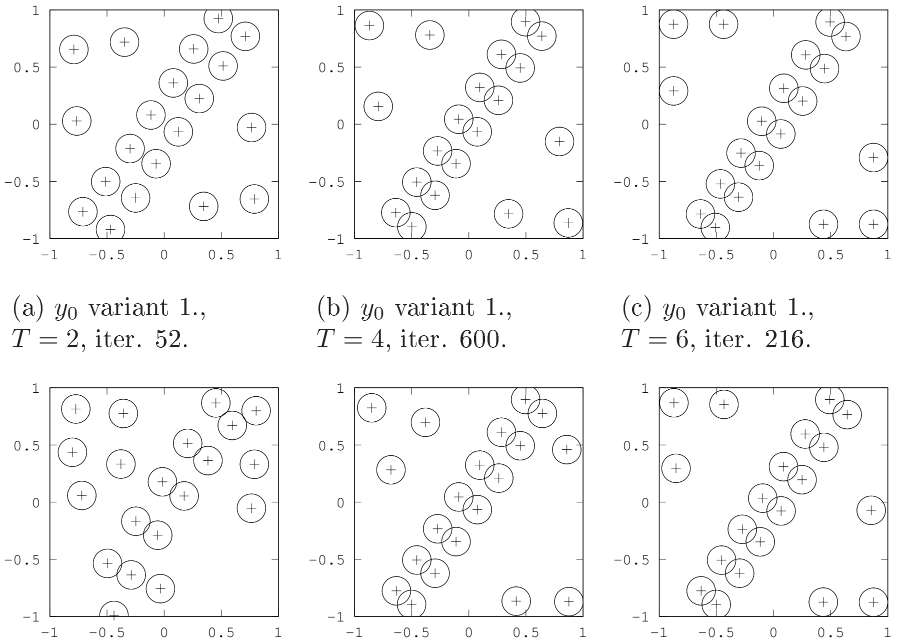
Simulation	Iterations	Initial cost	Terminal cost
y_0 variant 1, $T = 4$	**600**	0.078051	0.007050
y_0 variant 2, $T = 4$	468	0.083608	0.007149
y_0 variant 1, $T = 6$	216	0.076831	0.006993
y_0 variant 2, $T = 6$	**600**	0.077166	0.007088

Table 4.3: Behavior of optimization procedures considered in Section 4.4.3, for two variants of the initial condition and two values of the parameter T considered in the subject section. The meaning of particular columns and the notation concerning the stop criterion (the bold font entries) are as in the case of Table 4.1.

Note that the time step length τ_M is the same as in the previous experiments, however the time horizon is longer and hence the number of the time steps M in the time discretization is greater as well. This makes the computational time necessary to perform one iteration of an optimization algorithm greater than it was the case in the previous experiments. This is the reason for which we have reduced the value of the parameter N_{opt} to 600 (in the previous experiments, we considered $N_{opt} = 1000$).

The use of the CG+r method instead of the SD method also serves for reducing the computational effort, since, by the previous results, CG+r has performance superior to SD and more predictable than CG-r, in the sense of the mean and the median of the number of iterations (see Section 4.4.2). Nevertheless, comparison with previously described results shows that the numbers of iterations in the presently considered simulations, with $T = 4$ or $T = 6$ (see Table 4.3), are higher that the numbers of iterations for analogous simulations with $T = 2$ (i.e. those simulations in Table 4.2 which concern variant 3. of y^* and $T_0 = 0.9T$ and which use the CG+r method). This allows to pose a hypothesis that the performance of the CG+r method for optimization problem (3.24) varies with changes of T.

Speaking at the level of general ideas, results of previous experiments may suggest that the difficulty of the optimization problem (3.24) varies with changes of T_0 (because the performance of the SD method varies, see Section 4.4.2), while the here presented results may suggest that the difficulty changes also with changes of T (because the performance of the CG+r method changes). Nevertheless, it is worth recalling that the differences in the performance of the CG+r method were not present when changing the parameter T_0, in opposite to changes of T.



(a) y_0 variant 1., $T = 2$, iter. 52.

(b) y_0 variant 1., $T = 4$, iter. 600.

(c) y_0 variant 1., $T = 6$, iter. 216.

(d) y_0 variant 2., $T = 2$, iter. 279.

(e) y_0 variant 2., $T = 4$, iter. 468.

(f) y_0 variant 2., $T = 6$, iter. 600.

Figure 4.6: Configurations of the control and measurement devices, obtained by optimization procedures addressed in Section 4.4.3, for two variants of the initial state y_0 and two values of the parameter T considered in the subject section. Values of the parameter T and the variants of the initial state (corresponding to the functions plotted Figure 4.1) are indicated in the figures. Each plot presets the configuration corresponding to the final iteration of the subject optimization procedures (see column „Iterations" in Table 4.3). Figures 4.6a and 4.6d concern simulations described in Section 4.4.2 and are the same as Figures 4.5c and 4.5f, respectively, but we place them here for more convenient comparison.

Now, we will compare the optimization output obtained by the here considered simulations with the output obtained in the simulations described in the previous sections.

Two of the simulations described in Section 4.4.2 differ with the simulations described here only with time horizon T (these are the simulations considered there which concern variant 3. of y^* and $T_0 = 0.9T$ and which use the CG+r method). In Section 4.4.2, for the subject two simulations, shorter time horizon, $T = 2$, was considered. At the same time, dependence of the optimization results on the initial state y_0 was observed. Now, we can compare the results concerning $T = 2$, described in Section 4.4.2 (Figures 4.6a and 4.6d), with the results concerning longer time horizon (Figures 4.6b, 4.6c, 4.6e and 4.6f).

First, we can observe as the difference between the optimization output for distinct y_0 variants decreases when lengthening the time horizon T. The difference between the results obtained for $T = 4$ (Figures 4.6b and 4.6e) are visible smaller that the differences for $T = 2$ (Figures 4.6a and

4.6d). Still, some difference can be observed also for $T = 4$. Comparing the results concerning $T = 4$ with the results concerning $T = 6$ (Figures 4.6c and 4.6f), we observe further growth of similarity between the optimization results obtained for the two considered variants of y_0.

Second, as an additional observation, note that in all of Figures 4.6b, 4.6c, 4.6e and 4.6f, a strong visual dependence of the results with the reference state y^* (Figure 4.1c) is visible. This is expressed by concentration of the devices actions near the diagonal-like line, associated with the reference state y^* (Figure 4.1c) and by symmetry of the actions targeting with respect to the subject line. In particular, for variant 2. of y_0 this dependence seems to be clearer for $T = 4$ and $T = 6$ (Figures 4.6e and 4.6f) than in the case of $T = 2$ (Figure 4.6d). The level of symmetry visible in Figures 4.6e and 4.6f is higher than in Figure 4.6d.

To sum up, the use of a longer time interval resulted in leveling the dependence on y_0, observed in Section 4.4.2 for the simulations associated with $T_0 = 0.9T$ and y^* as in Figure 4.1c. Recall also that changing the optimization method from SD to CG+r did not bring this kind of results (see Section 4.4.2). These observations seem to confirm hypothesis (b), formulated for SD method in the concluding part of Section 4.4.1.

Moreover, looser the dependence on y_0 of the optimization results was, the stronger dependence on y^* was visible.

4.4.4 Technical remarks

We now give some technical remarks concluding the present chapter.

First, we have not conducted the convergence analysis of the optimization procedures applied in our experiments. Below, we will comment which additional steps would be necessary in the convergence analysis.

Second, as indicated in the present chapter, our experiments for numerical treatment of the optimization problem (3.24) were a rather heavy computational effort. At the same time, we are particularly interested in performing the optimization experiments for long time horizons (because it resulted in reduced dependence of the results on the initial condition, see Section 4.4.3), what makes the the experiments even more time consuming.

To be precise, for simulations described in Section 4.4.1, with $T = 2$, the mean time of single iterations was about 500 sec. For simulations with $T = 4$ or $T = 6$, described in Section 4.4.3, the mean iteration time was even longer (about 850 and 1200 sec, respectively). This made the latter simulations impractically long, because they required hundreds of iterations because achieving the stop criterion (see Table 4.3).

Thus, below we comment on certain possibilities of reducing the computational time necessary in numerical treatment of the optimization problem (3.24).

Remark. The above information concerning computational time for a single iteration is not precise because, unfortunately, we have not saved timestamps concerning each particular iteration during our experiments. ▲

Convergence analysis

We begin with remarks concerning the convergence of the optimization procedures utilized in our experiments.

It can be shown (see [24] or Chapters 3.2 and 5.2 in [38]) that, under appropriate conditions, the convergence of the SD and CG methods, described in Section 4.2, to a stationary point takes place in the following sense:

- $\lim_{n\to\infty}\left\|\nabla\mathcal{I}(\hat{v}^n)\right\|_V \longrightarrow 0$ for the SD method.
- $\liminf_{n\to\infty}\left\|\nabla\mathcal{I}(\hat{v}^n)\right\|_V \longrightarrow 0$ for the CG method.

Roughly speaking, the above mentioned appropriate conditions concern:

- The regularity of the cost functional $\mathcal{I}$. It should be differentiable in the classical sense (i.e., in the Fréchet sense, not only Gˆteaux) and its gradient should be Lipschitz continuous (see [24] or Chapters 3.2 and 5.2 in [38]).
- The line search procedure. It should return exact solution of 1-D optimization problem (see Theorem 2.1 and Theorem 4.3 in [24]), or it should fulfill so-called Wolfe conditions in the case of the SD and CG+r algorithms (see Theorem 3.2 and subsequent remarks in [38]) or the Wolfe conditions plus so-called sufficient descent condition in the case of the CG-r algorithm (see Corollary 4.4 in [24]).
- Besides, the results in [24] require the set of points with values of the cost functional equal below the value of the start point (call it the level set) to be bounded.

In our work, we have not investigated the Lipschitz continuity of $\nabla\mathcal{I}$ nor we have addressed the matter of boundedness of the level set.

In the optimization procedures in our simulations, the aim of the line search procedure (see Section 4.3.4) was to approximate the exact solutions of the 1-D optimization problem. This may seem to be reasonable to approximate the exact solutions, since, in view of the above remarks, they are sufficient for the convergence results. Nevertheless, despite the exact solutions are sufficient, their close approximations not need to be such. The referred above results require either exact solutions or so-called Wolfe conditions with, possibly, so-called sufficient descent condition. In general, the exact solutions, as well as their approximations, do not necessarily obey the Wolfe conditions. Thus, however the line search procedure proposed in Section 4.3.4 worked properly, for the convergence analysis it may convenient to change the line search procedure for a procedure obeying the Wolfe conditions and the sufficient descent condition.

Moreover, the numerical schemes applied for solving the 1-D problem base on inexact evaluation of $\mathcal{I}$ (see Section 4.3.2) and inexact evaluation of d^n, caused by inexact evaluation of the gradient of $\mathcal{I}$ (see Section 4.3.3). Indeed, the vector d^n in the 1-D problem depends strongly on the gradient of $\mathcal{I}$, both for the SD and for the CG method. Thus, for the convergence analysis of the optimization procedures, it would be required to investigate the influence of the latter effects to the convergence.

Summing up, to investigate the convergence of the real optimization procedures applied in our simulations (which are merely approximations of the ideal SD and CG methods, described in Section 4.2), it would be necessary to:

- Prove results on Lipschitz continuity of $\nabla\mathcal{I}$.
- Propose a line search procedure obeying the Wolfe conditions and the sufficient descent condition.
- Answer the questions concerning the convergence of the numerical scheme concerning the evaluation of $\mathcal{I}$ (Section 4.3.2) and the evaluation of the gradient (Section 4.3.3).
- Solve the rather technical problem of guaranteeing that the level set is bounded, if one wants to base on the results of [24].

Since we have not performed analysis of the above points in the present work, we leave the question concerning the convergence of the SD and CG algorithms applied in our experiments open.

Possible oscillations near the stationary points

Roughly speaking, in our situation, possible refinements concerning the reduction of the computational time of the numerical optimization experiments can be grouped into two categories. One of them is the group of refinements focusing on reducing the computational time of a single iteration of the optimization procedures, the other one is the group of refinements serving for reducing the expected number of iterations. The below remark concerns concerns the latter group of refinements.

Taking a look at Tables 4.1, 4.2 and 4.3, one can observe that the number of iterations necessary to reach the point fulfilling the stop criterion varies strongly for particular simulations. For some simulations, the number of iterations was particularly high, e.g. for simulations with long time interval, described in Section 4.4.3. Thus one can pose a hypothesis, which we do not verify here, that in these simulations the optimization procedure was oscillating close to the stationary point for many iterations before reaching the stop criterion. Here, by oscillations me mean consecutive iterations of the optimization procedures which bring no significant changes of the values of the cost functional nor of the control parameter. This kind of oscillations is certainly an undesired effect, making the computational time significantly longer.

We propose two strategies of refining the optimization procedures applied in our experiments. The subject strategies can be tested in future experiments and, if the alleged oscillations indeed were present in our experiments, can result in improved performance of the optimization algorithms. These strategies are:

- Use a stronger stop criterion. In our simulations, the stop criterion was probably quite weak, in sense that strong conditions have to be fulfilled to trigger the stop criterion. It is tempting to propose a stop criterion which detects the moment when the optimization procedure does not make significant progress anymore, or when the oscillations begin. Nevertheless, due to variety of possibilities which could be considered in this context, we do not continue with this issue here.

- Apply the Newton method combined with one of the SD or CG methods. This idea is not new. It is known that the Newton method, if starting sufficiently close to the stationary point, converges to this point quickly (see Theorem 3.5, p. 44 in [38]). Thus, a reasonable optimization procedure can be to start with the SD or CG method and switch to the Newton method when a proper switching criterion is triggered. Further proposition is to use some quasi-Newton method instead of the Newton method itself, to avoid the necessity of computing the Hessian and dealing with conditions sufficient for second order differentiability of $\mathcal{I}$.

Note also that the inaccuracies in computing the gradient of $\mathcal{I}$, which were mentioned above in the context of the convergence analysis, also can be related with the alleged oscillations of the optimization procedures near the stationary points. Small perturbations of the gradient near the local minimum can influence the convergence of gradient optimization algorithms, however this is also merely a hypothesis. To conclude, if the oscillations indeed are present in our simulations, then, besides the above proposed strategies, it may be worthwhile to consider possibilities of improving the accuracy of the numerical schemes concerning the evaluation of the gradient of $\mathcal{I}$.

Reduction to the stationary problem

There are also certain directions of development which can help to reduce the computational time of a single iteration in our optimization procedures. In this context, we propose the following strategy, which in fact consists in replacing the optimization problem (3.24) with other optimization problem, potentially requiring less computational power.

The strategy is to reduce the system (3.1) - (3.2) to a stationary model, not involving the time variable. Having this, one can define an alternative cost functional, basing on the gap between the solution of the stationary model and the reference state. New optimization problem would be to minimize the new cost functional.

Computing a numerical solution of the stationary model should be less time consuming than computing the numerical solution of (3.1) - (3.2). In our simulations, the main effort in every iteration of the optimization procedures concerned solving the system (3.1) - (3.2) multiple times. Hence, a single iteration of the new optimization problem would be probably much less time consuming.

From mathematical point of view, applying this approach would require the analysis of the new optimization problem itself, consisting of the steps analogous to those in the present work, as the existence and uniqueness results, stability analysis and results concerning differentiability of the state operator and of the cost functional.

On the level of general ideas, the new optimization problem approximates the original optimization problem with T_0 close to T, under the condition that the dynamical system associated with (3.1) - (3.2) posses a one point attracting set. Therefore, this approach is possible but demands, besides the analysis of the new optimization problem itself, the analysis of large time behavior of the system (3.1) - (3.2), involving in particular analysis of the attracting sets. This analysis probably would be not trivial because, as remarked in Section 2.3.4, in certain situations the attracting set, if exists, probably is bigger than one point. In consequence, a non-obvious problem of characterizing those parameters and functions entering the system (3.1) - (3.2) for which a one point attracting set exists would be faced during the large time behavior analysis.

Numerical schemes with an improved integration method

Next, we would like to give a more extensive comment on the numerical schemes concerning the evaluation of the cost functional $\mathcal{I}$ (see Section 4.3.2) and its gradient (Section 4.3.3). The subject schemes return inexact values, what, as already remarked above, can both have consequences for the analysis of convergence of the numerical optimization procedures and cause the hypothetical oscillations of the numerical procedures.

The schemes for evaluation of the cost functional and its gradient give inexact values, for multiple reasons. First, for a given $\hat{v} \in V$, the values $\mathcal{I}(\hat{v})$ and $\nabla\mathcal{I}(\hat{v})$ are computed basing not on the weak solutions of systems (3.1) - (3.2) and (3.30) - (3.31), but on the approximate solutions of these system, obtained by the methods described in Section 4.3.1. Second, the time integrals of the approximate solutions or their transformations, appearing both in the definition of $\mathcal{I}(\hat{v})$ and in the formula characterizing $\nabla\mathcal{I}(\hat{v})$, are computed inexactly. Being puristic, the time integrals of the approximate solutions are not even defined because we assumed that the approximate solutions are defined only in the time discretization points.

Thus, let us propose an alternative approach concerning numerical schemes for the cost functional and its gradient, which can be tested in the future experiments. We still assume that the structural assumptions presented in Section 4.1 hold. The alternative approach can be sketched a follows:

1. Define approximate solution with continuous time for the system (3.1) - (3.2) as a piecewise linear extension to $[0, T]$ of the approximate solution with discrete time, defined for the system (3.1) - (3.2) in Section 4.3.1 for time discretization points $t_0, t_1, \ldots, t_M$. Let this linear extension be chosen such that it is linear on each interval (t_m, t_{m+1}), $m = 0, 1, \ldots, M-1$. This makes the approximate solution with continuous time unique. The approximate solution with continuous time for the system (3.30) - (3.31) is defined analogously, basing on the approximate solution with discrete time, defined for the system (3.30) - (3.31) in Section 4.3.1.

2. To approximate the cost functional, we evaluate the formula $\int_{T_0}^{T}\int_{\Omega_N}\left|Y_N - [y^*]_N\right|^2$, where Y_N is the first component of the approximate solution with continuous time for the system (3.1) - (3.2). Below, the subject formula will be called *the modified cost formula*. The notation $[y^*]_N$ has meaning as in Section 4.3.2.

   To evaluate the modified cost formula, we proceed as follows. First, define the function $\widetilde{E}\colon [0, T] \to \mathbb{R}$ with the formula analogous as the formula for $\widetilde{E}_m$ in Section 4.3.2, but with $t \in [0, T]$ instead of discrete points $t_0, t_1, \ldots, t_M$. Note, that the function $\widetilde{E}$ is piecewise parabolic and continuous, as a product of two piecewise linear continuous functions. Next, compute the time integral from T_0 to T of $\widetilde{E}$ using the parabolic quadrature with nodes coinciding with the time discretization points. Such quadrature gives the exact value of integral $\int_{T_0}^{T}\widetilde{E}$, because $\widetilde{E}$ is piecewise parabolic and continuous. Hence, by the definition of $\widetilde{E}$, it is also an exact value of the modified cost formula.

3. To approximate the gradient, we proceed analogously. We base on the formula (3.40). In the subject formula, we substitute approximate solutions with continuous time instead of the real solutions and $P_1(\Omega_N)$ approximations of other functions instead of the functions itself. Let us call the result *the modified gradient formula*. We treat the modified gradient formula as an approximation of the gradient of the cost functional.

   To evaluate the modified gradient formula, we define functions $\widetilde{E}_1, \widetilde{E}_2, \widetilde{E}_3\colon [0, T] \to \mathbb{R}$ analogously as $\widetilde{E}_{1,m}$, $\widetilde{E}_{2,m}$ and $\widetilde{E}_{3,m}$ in Section 4.3.3, but with $t \in [0, T]$ instead of discrete points $t_0, t_1, \ldots, t_M$. We also define $\widehat{E}^{\nabla} := \widetilde{E}_1 + \widetilde{E}_3$. Next, we integrate $\widehat{E}^{\nabla}$ with respect to time. Now, a difference with the step concerning evaluation of the cost functional occurs because $\widehat{E}^{\nabla}$ is not piecewise parabolic, in opposite to $\widetilde{E}$. Observing the structure of $\widetilde{E}_1$, $\widetilde{E}_2$, $\widetilde{E}_3$ one can note that:

   - $\widetilde{E}_1$ is piecewise parabolic and continuous as a product of two piecewise linear continuous functions.
   - $\widetilde{E}_2$ is piecewise parabolic and continuous as a composition of two piecewise linear continuous functions (due to our structural assumptions, w'_j is piecewise linear). Nevertheless, the nodes of $\widetilde{E}_2$ do not coincide with the time discretization points.
   - $\widetilde{E}_3$ is piecewise polynomial of fourth order and continuous, as a product of $\widetilde{E}_2$ and two piecewise linear continuous functions. The nodes of $\widetilde{E}_3$ do not coincide with the time discretization points.
   - $\widehat{E}^{\nabla}$ is piecewise fourth order polynomial and continuous, with nodes not coinciding with the time discretization points.

   As a result, to compute the integral $\int_{T_0}^{T}\widehat{E}^{\nabla}$ using fourth order polynomials quadrature, with more elaborate choice of the quadrature nodes, depending on w'_j. Deriving the exact algebraic formulas is possible but to complicated to do it here. Nevertheless — the obtained value is the exact value of $\int_{T_0}^{T} \widehat{E}^{\nabla}$ and hence, by the definition of $\widehat{E}^{\nabla}$, also the exact value of the modified gradient formula.

Numerous advantages of the above proposed schemes for evaluating the cost functional and its gradient can be indicated. First, in comparison to the schemes presented in Section 4.3.2 and Section 4.3.3, we have better control on the output of the numerical schemes, because the above proposed schemes compute exact values of concrete formulas, i.e. of the modified cost formula and the modified gradient formula.

Second, it is tempting, and may possible, to prove that the modified gradient formula in fact characterizes the exact gradient of the cost functional given by the modified cost formula. Such statement, if proven, would have interesting consequences, in particular for the analysis of convergence of the numerical optimization procedures to the stationary points of the cost functional $\mathcal{I}$. In contrary to the convergence analysis for the numerical schemes given in Section 4.3 (see the remarks above in the present section), here, it wouldn't be necessary to prove that the gradient approximation is fine enough. It would be sufficient to prove the convergence of the SD and CG methods for the cost functional associated with the modified gradient formula and then to prove the convergence of the latter cost functional to the original cost functional $\mathcal{I}$. To sum up, if the modified gradient formula was the gradient of the modified cost formula, then it would be possible to apply „first discretize then optimize" approach instead of „first optimize then discretize".

Moreover, if the alleged oscillations of our optimization procedures (mentioned above in this section) were in fact caused by inaccuracies concerning the gradient of $\mathcal{I}$, then switching to the above proposed „first discretize then optimize" approach could be a remedy to the oscillations matter.

Nevertheless, the proposed approach has also its drawbacks. The numerical scheme necessary for exact evaluation of the modified gradient formula depends on the function w_j'. For example, reasoning as above we find that if w_j' was piecewise a polynomial of order three, then the quadrature necessary for exact evaluation of $\int_{T_0}^{T} \widehat{E}^{\nabla}$ would rise from four to five. If w_j' was not a polynomial at all, then a question arises how to choose a proper quadrature for integral $\int_{T_0}^{T} \widehat{E}^{\nabla}$. Thus, if one wanted to apply this approach, one would face the problem of automatic choice of quadrature during the implementation of the optimization procedures. This problem, however interesting, could cause problems both at the algorithmic level and at the level of code implementation, which would become more cumbersome than it was the case in our situation.

Concluding remarks

Below, we comment on certain issues which were not investigated in the present work. We indicate certain problems concerning the model (0.1) - (0.3), introduced in §1 of *Introduction*, or the optimal targeting problem, introduced in §2 of *Introduction*, which were not solved in the preceding chapters. We also comment on possibilities of refining the model (0.1) - (0.3) or the setting of the optimal targeting problem itself.

The below questions remain open in the present work and can be investigated in the future:

- In Section 2.3.4, we have indicated some observations concerning the large time behavior of the model (0.1) - (0.3). We have posed certain hypotheses, basing on the effects observed in the numerical results. One of them was that the structure of the alleged attracting set of the dynamical system associated with the model (0.1) - (0.3) significantly depends on the parameters of the model. It would be desired to confirm the subject hypotheses by analytical proofs. In particular, it would be interesting to characterize those parameters entering system (0.1) - (0.3) for which the controlled process tends to some time-invariant state, independent of the initial condition. In other words, we are interested in those model parameters, for which a one-point attracting set exists.

  The existence of a one-point attracting set would mean that, in the model (0.1) - (0.3), the efficiency of the thermostat control mechanism, understood as the distance between the process state and the reference state for large times, is insensitive to the changes of the initial state of the process. The insensitivity to the changes of the initial condition is one of the hypothetical advantages of the controls involving the automatic corrections idea (see *Introduction*), as .e.g. the thermostat control mechanism. Hence, the characterization of those parameters of the model for which the latter property holds would be a desired result.

- Neither for the numerical schemes described in Chapter 2 nor for the ones described in Chapter 4 we have performed the convergence analysis. Therefore, from the mathematical point of view, the convergence analysis is one of the natural fields for the further research.

  In Section 4.4.4, certain steps necessary for the analysis of the convergence of the optimization procedures utilized in the experiments described in Chapter 4 are indicated.

- The simulations concerning the optimal targeting problem, described in Chapter 4, were rather time consuming. In Section 4.4.4, we have indicated some possibilities of reducing the computational time of generating approximate solutions of the optimal targeting problem. One of the aims of the future research can be to test a part of the subject possibilities.

  In particular, performing optimization procedures based on the „first discretize then optimize" approach, proposed in Section 4.4.4, and comparing the results with the results described in Chapter 4 could be an interesting experiment. In some of the simulations described in Chapter 4, the optimization procedures needed a particularly large number of iterations to stop. One of the hypotheses concerning the latter effect, indicated in Section 4.4.4, is that they it is related with the inaccuracies in the numerical scheme for the evaluation of the gradient of the cost functional. The first discretize then optimize approach, as explained in Section 4.4.4, should eliminate the problem of inaccuracies of the numerical scheme for evaluation of the gradient of the cost functional. Therefore, the comparison of the results obtained with the latter approach and the results described in Chapter 4 can help to answer the questions concerning the reasons behind the mentioned effect of the large number of iterations.

- We have not investigated the sensitivity of the effectiveness of the thermostat control mechanism in the model (0.1) - (0.3) to perturbations of the model itself, i.e. of the diffusion coefficient D or the reactive term f in the main equation (0.1) (here, we understand effectiveness as in *Introduction*, see comment a), page x). Insensitivity to perturbations of the model is one of alleged advantages of the automatic corrections mechanism, indicated in the beginning of *Introduction*. A further investigation can concern also the sensitivity of the solutions of the optimal targeting problem to the changes of the subject parameters.

- In the beginning of Section 1.2, we have indicated that Lipschitz continuous switching functions in the system (0.1) - (0.3) can be utilized to approximate the case of discontinuous switching functions, as $-sgn$, which are not allowed directly by the analytical results of the present work. At the same time, the results of Section 1.1, concerning the modified system (1.1) - (1.3), allowed certain multivalued switching functions containing $-sgn$. Thus, it would be interesting to investigate the convergence of the solutions of the system (0.1) - (0.3) with Lipschitz switching functions approximating $-sgn$ to a solution of the modified system (1.1) - (1.3) with appropriate multivalued switching functions containing $-sgn$. The subject convergence was not analyzed in the present work and can be an aim for further investigations.

Besides the above indicated technical problems, one may consider to refine the model considered in the present work, as well as introduce changes to the optimal targeting problem. In this scope, we point out the following possibilities:

- The model (0.1) - (0.3) assumes that a process described by a reaction-diffusion equation is controlled by thermostats. Not all real-world phenomenas which are the subject of the control theory in PDEs can be described this way. The references given in §3 of *Introduction* present examples of the models with thermostat control mechanism in which a state equation (or system of equations) other that the scalar reaction-diffusion equation is considered.

  Hence, one of the generalizations of the content of the present work can consist in assuming a more general state equation or equations to be controlled by thermostats. Generalizing further, one can try to implement the thermostat control mechanism for abstract dynamical systems and indicate which properties of the subject systems are essential for deriving results similar to the here presented ones (as the existence of solutions, the stability, the differentiability of the cost functional).

- The considered thermostat control mechanism, basing on which we formulate the optimal targeting problem, also can be refined. The model (0.1) - (0.3) involves a thermostat control mechanism with assumes no hysteresis in the work of the switching mechanism (see the remarks on possible variants of thermostat control mechanism in §3 of *Introduction*). In the present work, the latter assumption was imposed for the sake of the simplicity of the investigated mathematical model. Nevertheless, a thermostat control mechanism with hysteresis would be more realistic, since in real world perfectly immediate reaction to observed changes is not possible. In fact, a big part of the mathematical references given in §3 of *Introduction* address models involving thermostat control mechanism with hysteresis (however, none of those works focus on the optimal targeting problem).

- One may consider also certain modifications in the optimal targeting problem as well. In the present work, we assumed that the number of the control devices equals the number of the measurement devices and, moreover, that the control and measurement devices are pairwise coupled (see §2 of *Introduction*). By coupling of the devices, we mean the assumption that their actions has pairwise the same targeting in space and a given control device responds to the data collected by the coupled measurement device with weight equal 1. These assumptions were imposed to exclude the problem of the choice of weights from our research. Nevertheless, in certain applications, the problem of the choice of the weights in thermostat control mechanisms seems to be natural and should not be excluded from the setting of the optimization problem.

  For example, one may consider the situation of the hyperthermia cancer therapy, described in §3 of *Introduction*. As mentioned there, the temperature in the patient tissues can be measured by magnetic resonance imaging and the energy can be applied by control devices transmitting or electromagnetic waves. In the setting of thermostat control mechanism, the actions of the magnetic resonance can be interpreted as a dense mesh of small measurement spots of fixed location. However, the user is permitted to calibrate the control devices and, in consequence, to manipulate the targeting of their actions in space. In this situation, it is not natural to assume that the weights entering the thermostat control mechanism are given. Thus, to handle the situation of the above type, one could define a new optimization problem, taking into account the problem of the choice of both the targeting of the control devices actions and the weights, assuming that the actions of the measurement devices have fixed targeting.

Appendix A

Auxiliary theorems

A.1 Differentiability in Banach spaces

The below definitions of directional derivative, Gâteaux derivative and Fréchet derivative are equivalent as those in [50], Chap. 4. The notions of the weak derivatives introduced in this Section bases on [4], Chap. 1, Sec. 4. In addition, [50] provides the proofs (or techniques for the proofs) for most of facts and theorems presented below.

Definition A.1.1 *Let $T : X \to Y$ be an operator between two Banach spaces. For $\hat{u}, \hat{v} \in X$, we call $\delta T(\hat{u}; \hat{v}) \in Y$ (or $\delta_w T(\hat{u}; \hat{v}) \in Y$) the directional derivative (or the weak directional derivative, respectively) of T in point $\hat{u} \in X$ in direction $\hat{v} \in X$ if*

$$\begin{aligned} \delta T(\hat{u}; \hat{v}) &= \lim_{\varepsilon \to 0} \frac{T(\hat{u} + \varepsilon \hat{v}) - T(\hat{u})}{\varepsilon} \\ \Big\langle \hat{\phi} \,,\, \delta_w T(\hat{u}; \hat{v}) \Big\rangle_{Y^*,Y} &= \lim_{\varepsilon \to 0} \Big\langle \hat{\phi} \,,\, \frac{T(\hat{u} + \varepsilon \hat{v}) - T(\hat{u})}{\varepsilon} \Big\rangle_{Y^*,Y} \quad \forall_{\hat{\phi} \in Y^*} \end{aligned} \tag{A.1}$$

The operator $\delta T(\hat{u}; \,.\,)$ (or $\delta_w T(\hat{u}; \,.\,)$), acting on X is called the variation (or the weak variation, respectively) of T in point $\hat{u} \in X$.

Definition A.1.2 *If the (weak) variation in point $\hat{u}$ is a bounded linear operator from X to Y, then we say that T is (weakly) Gâteaux differentiable in $\hat{u}$ and we define the (weak) Gâteaux derivative of T in $\hat{u}$ respectively as*

$$\begin{aligned} D_G T(\hat{u}) &:= \delta T(\hat{u}; \,.\,) \\ D_{G,w} T(\hat{u}) &:= \delta_w T(\hat{u}; \,.\,) \end{aligned} \tag{A.2}$$

Definition A.1.3 *We say that $D_F T(\hat{u}) \in L(X, Y)$ (or $D_{F,w} T(\hat{u}) \in L(X, Y)$) is the Fréchet derivative (or the weak Fréchet derivative, respectively) in point $\hat{u} \in X$ if*

$$\begin{aligned} \lim_{\hat{v} \to 0} \frac{T(\hat{u} + \hat{v}) - T(\hat{u}) - D_F T(\hat{u}) \hat{v}}{\|\hat{v}\|_X} &= 0 \\ \lim_{\hat{v} \to 0} \Big\langle \hat{\phi}, \frac{T(\hat{u} + \hat{v}) - T(\hat{u}) - D_{F,w} T(\hat{u}) \hat{v}}{\|\hat{v}\|_X} \Big\rangle_{Y^*,Y} &= 0 \quad \forall_{\hat{\phi} \in Y^*} \end{aligned} \tag{A.3}$$

Note, that by the above definition the existence of a directional derivative implies the existence of the weak directional derivative. An analogous relation holds between the notion of the Gâteaux differntiability and the weak Gâteaux differntiability and between the Fréchet differntiability and the weak Fréchet differntiability.

Theorem A.1.4 (The chain rule) *Let X_1, X_2 and X_3 be Banach spaces and let $T_1 : X_1 \to X_2$, $T_2 : X_2 \to X_3$. Suppose, that:*

1. *T_1 has the (weak) directional derivative in point $\hat{u} \in X_1$ in direction $\hat{v} \in X_1$,*
2. *T_2 is Fréchet differentiable (at least in point $T_1(\hat{u})$).*

Then the composite operator $T_2 \circ T_1$ has the (weak) directional derivative in point $\hat{u} \in X_1$ in direction $\hat{v} \in X_1$ and it can be expressed respectively as:

$$\begin{aligned} \delta(T_2 \circ T_1)(\hat{u}; \hat{v}) &= (D_F T_2)(T_1(\hat{u}))\delta T_1(\hat{u}; \hat{v}) \\ \delta_w(T_2 \circ T_1)(\hat{u}; \hat{v}) &= (D_F T_2)(T_1(\hat{u}))\delta_w T_1(\hat{u}; \hat{v}) \end{aligned} \tag{A.4}$$

The proof is very similar to the proof of Proposition 4.10 in [50].

Note, that Theorem A.1.4 implies that if T_1 is (weakly) Gâteaux differentiable and T_2 is Fréchet differentiable then the superposition $T_2 \circ T_1$ is (weakly) Gâteaux differentiable and the chain rule holds.

Theorem A.1.5 (The product rule) *Let X_1, $X_{2,1}$, $X_{2,2}$ and X_3 be Banach spaces, let $T_1\colon X_1 \to X_{2,1}$, $T_2\colon X_1 \to X_{2,2}$, $\hat{B} : X_{2,1} \times X_{2,2} \to X_3$ and denote $H(\hat{u}) := \hat{B}(T_1(\hat{u}), T_2(\hat{u}))$. Fix $\hat{u}, \hat{v} \in X_1$. We make the following assumptions:*

1. *B is bilinear and bounded,*
2. *T_i has the (weak) directional derivative in point $\hat{u}$ in direction $\hat{v}$, for $i = 1, 2$.*

Then H also has the (weak) directional derivative in point $\hat{u}$ in direction $\hat{v}$ and it can be expressed respectively as:

$$\begin{aligned} \delta H(\hat{u}; \hat{v}) &= \hat{B}(\delta T_1(\hat{u}; \hat{v}), T_2(\hat{u})) \;+\; \hat{B}(T_1(\hat{u}), \delta T_2(\hat{u}; \hat{v})) \\ \delta_w H(\hat{u}; \hat{v}) &= \hat{B}(\delta_w T_1(\hat{u}; \hat{v}), T_2(\hat{u})) \;+\; \hat{B}(T_1(\hat{u}), \delta_w T_2(\hat{u}; \hat{v})) \end{aligned} \tag{A.5}$$

The assertion follows as in the proof of Proposition 4.11 in [50].

Observation A.1.6 *Note, that for $Y = \mathbb{R}$ in Definition A.1.2 the weak Gâteaux differentiability becomes equivalent to the Gâteaux differentiability. For this reason, if we set in Theorem A.1.4 Y as $\mathbb{R}$ and T_1 as a weakly Gâteaux differentiable operator, then we get that the superposition $T_2 \circ T_1$ is not only weakly Gâteaux differentiable but also Gâteaux differentiable and the chain rule holds.*

Observation A.1.7 *Every bounded linear operator $T\colon X \to Y$ acting between two Banach spaces X and Y is Fréchet differentiable and its Fréchet differential in an arbitrary point is equal to the operator itself, i.e. $D_F T(\hat{u})(\hat{v}) = T(\hat{v})$ for all $\hat{u}, \hat{v} \in X$.*

Observation A.1.8 *Let H be a real Hilbert space with norm $\|\,.\,\|_H$ and scalar product $(\,.\,,\,.\,)_H$. Let the operator $T\colon H \to \mathbb{R}$ be defined by $T(\hat{u}) := \|u\|_H^2$. Then, T is Fréchet differentiable and*

$$D_F T(\hat{u})\hat{v} \;=\; 2\,(\hat{u},\hat{v})_H \qquad \forall \hat{u},\hat{v} \in H \tag{A.6}$$

The Observations A.1.6 and A.1.7 follow straight while the Observation A.1.8 is an exercise involving direct application of the derivative definition: first, we calculate the directional derivatives to obtain the characterization of the Gâteaux derivative of T (see, e.g., [45, p.57]) and then we estimate the reminder of the linearization to show, that the Gâteaux derivative is in fact the Fréchet derivative of T as well.

If the convergence in (A.1) in Definition A.1.1 holds only for some sequence $\{\varepsilon_n\}_{n=1}^{\infty}$, where $\varepsilon_n \neq 0$, $\varepsilon_n \to 0$ as $n \to \infty$, then it is meaningful to pose a question: are the chain rule and the product rule still true? In the latter context, the below notion will be convenient for the sake of brevity:

Definition A.1.9 *For an operator $T : X \to Y$, point $\hat{u} \in X$, direction $\hat{v} \in X$ and a sequence $\epsilon := \{\varepsilon_n\}_{n=1}^{\infty}$, $\varepsilon_n \to 0$ as $n \to \infty$, if the difference quotients $\frac{1}{\varepsilon_n}\left(T(\hat{u}+\varepsilon_n\hat{v}) - T(\hat{u})\right)$ are (weakly) convergent as $n \to \infty$ then we call the limit the sequential (weak) directional derivative on the sequence ϵ and denote it $\bar{\delta}^{\epsilon}T(\hat{u};\hat{v})$ (or $\bar{\delta}^{\epsilon}_{w}T(\hat{u};\hat{v})$, respectively).*

Theorem A.1.10 *Assume that $\epsilon := \{\varepsilon_n\}_{n=1}^{\infty}$, $\varepsilon_n \to 0$ as $n \to \infty$. The following modifications of Theorems A.1.4 and A.1.5 are true:*

1. *In Theorem A.1.4, if we replace the assumption on the existence of the (weak) directional derivative of T_1 by an assumption of the existence of the sequential (weak) directional derivative of T_1 on the sequence ϵ, then the assertion of the theorem holds in the sequential version, i.e. the sequential (weak) directional derivative of $T_2 \circ T_1$ on the sequence ϵ exists and it can be expressed respectively as:*

$$\begin{aligned} \bar{\delta}^{\epsilon}(T_2 \circ T_1)(\hat{u};\hat{v}) \;&=\; (D_F T_2)(T_1(\hat{u}))\bar{\delta}^{\epsilon}T_1(\hat{u};\hat{v}) \\ \bar{\delta}^{\epsilon}_{w}(T_2 \circ T_1)(\hat{u};\hat{v}) \;&=\; (D_F T_2)(T_1(\hat{u}))\bar{\delta}^{\epsilon}_{w}T_1(\hat{u};\hat{v}) \end{aligned} \tag{A.7}$$

2. *In Theorem A.1.5, if we replace the assumption on the existence of the (weak) directional derivatives of T_i, $i = 1, 2$ by an assumption of the existence of the sequential (weak) directional derivatives of T_i, $i = 1, 2$ on the sequence ϵ, then the assertion of the theorem holds in the sequential version, i.e. the sequential (weak) directional derivative of H on the sequence ϵ exists and it can be expressed respectively as:*

$$\begin{aligned} \bar{\delta}^{\epsilon}H(\hat{u};\hat{v}) \;&=\; \hat{B}(\bar{\delta}^{\epsilon}T_1(\hat{u};\hat{v}), T_2(\hat{u})) \;+\; \hat{B}(T_1(\hat{u}), \bar{\delta}^{\epsilon}T_2(\hat{u};\hat{v})) \\ \bar{\delta}^{\epsilon}_{w}H(\hat{u};\hat{v}) \;&=\; \hat{B}(\bar{\delta}^{\epsilon}_{w}T_1(\hat{u};\hat{v}), T_2(\hat{u})) \;+\; \hat{B}(T_1(\hat{u}), \bar{\delta}^{\epsilon}_{w}T_2(\hat{u};\hat{v})) \end{aligned} \tag{A.8}$$

The proof of this theorem in fact consists in analyzing the proofs of Theorems A.1.4 and A.1.5 and noting that the above modification is possible.

For the superposition of two operators $T_2 \circ T_1$, Theorem A.1.4 implies that the chain rule is correct if we assume the Fréchet differentiability of T_2 and the (weak) Gâteaux differentiability of T_1. In the converse situation, namely assuming only the (weak) Gâteaux differentiability of T_2, the chain rule is not true, even if the inner operator T_1 is Fréchet differentiable. However, there is a particular case in which we can get the chain rule for (weakly) Gâteaux differentiable T_2:

Observation A.1.11 *Let X_1, X_2 and X_3 be Banach spaces, let $T_1 : X_1 \to X_2$, $T_2 : X_2 \to X_3$ and let $\hat{u} \in X_1$. Suppose, that:*

1. *T_1 is a continuous linear operator,*
2. *T_2 is (weakly) Gâteaux differentiable (at least in point $T_1(\hat{u})$).*

Then the composite operator $T_2 \circ T_1$ is (weakly) Gâteaux differentiable in point $\hat{u} \in X_1$ and it can be expressed respectively as:

$$\begin{aligned} D_G(T_2 \circ T_1)(\hat{u})(\hat{v}) &= (D_G T_2)(T_1(\hat{u}))(T_1(\hat{v})) &= (D_G T_2)(T_1(\hat{u}))(D_F T_1(\hat{u})(\hat{v})) \\ D_{G,w}(T_2 \circ T_1)(\hat{u})(\hat{v}) &= (D_{G,w} T_2)(T_1(\hat{u}))(T_1(\hat{v})) &= (D_{G,w} T_2)(T_1(\hat{u}))(D_F T_1(\hat{u})(\hat{v})) \end{aligned} \tag{A.9}$$

Proof. The proof follows immediately. Let us check the difference quotient in point $\hat{u} \in X_1$ in direction $\hat{v} \in X_1$:

$$\varepsilon^{-1}\Big(T_2\big(T_1(\hat{u} + \varepsilon\hat{v})\big) - T_2\big(T_1(\hat{u})\big)\Big) = \varepsilon^{-1}\Big(T_2\big(T_1(\hat{u}) + \varepsilon T_1(\hat{v})\big) - T_2\big(T_1(\hat{u})\big)\Big)$$

what tends to the (weak) directional derivative of T_2 in point $T_1(\hat{u})$ in direction $T_1(\hat{v})$ when $\varepsilon \to 0$. If T_2 is weakly Gâteaux differentiable then the above suffices to verify the asserted formulas. ■

A.2 Optimality conditions for differentiable functionals

Having introduced the notion of derivatives in Banach spaces and their basic properties, we can link this theory to the theory of optimization and formulate the optimality criterion, generalizing the Fermat's necessary condition for existence of minimum of a real function of one real variable:

Theorem A.2.1 *Let $C \subset D \subset X$ where X is a real Banach space, C is a nonempty and convex subset and D is an open subset of X containing C. Let also $T : D \to \mathbb{R}$ be a Gâteaux differentiable functional. Then the necessary condition for $\bar{u} \in X$ to solve the optimization problem $\inf_{\hat{u} \in C} T(\hat{u})$ if that the following condition is fulfilled in $\bar{u}$:*

$$D_G T(\bar{u})(\hat{w} - \bar{u}) \geq 0 \qquad \forall_{\hat{w} \in C} \tag{A.10}$$

For the proof, see Lemma 2.21 in [45].

A.3 Nemytskii operators

Below, we present a short part of the theory of Nemytskii operators, necessary in the present work. We do not need the theory of Nemytskii operators in its full generality. Our attention is restricted to autonomous Nemytskii operators acting on functions defined on Lebesgue-measurable subsets of $\mathbb{R}^n$ of bounded measure. A reader interested in the more general theory is referred to [2] or [17]. Actually, the below facts concerning Nemytskii operators are based on the content of Chapters 6 and 7 in [17].

In general, for a set $\mathbb{A}$ and a function $F\colon \mathbb{A}\times\mathbb{R}\to\mathbb{R}$, *the Nemytskii operator* associated with F, denote it $\mathcal{N}_F$, is the operator acting from the set of real functions on $\mathbb{A}$ to itself defined by the following condition:

$$\mathcal{N}_F(\hat{u})(x) := F\big(x,\hat{u}(x)\big) \qquad \text{for } x\in\mathbb{A}, \text{ for } \hat{u}\colon \mathbb{A}\to\mathbb{R}$$

We are interested in the situation of F being a function of only one variable, $F\colon\mathbb{R}\to\mathbb{R}$. In this case, the operator $\mathcal{N}_F$ is called *autonomous Nemytskii operator* and can be expressed as:

$$\mathcal{N}_F(\hat{u}) := F\circ\hat{u} \qquad \text{for } \hat{u}\colon \mathbb{A}\to\mathbb{R}$$

The Nemytskii operator $\mathcal{N}_F$ is often considered to act between Lebesgue spaces $L^{s_1}(\mathbb{A})$ and $L^{s_2}(\mathbb{A})$, for some exponents s_1 and s_2. However, to understand $\mathcal{N}_F$ this way, we need to remember that elements of the Lebesgue spaces are not the functions, but equivalence classes of the relation of being equal a.e. If $\hat{v}=\hat{w}$ a.e. on $\mathbb{A}$ and F is measurable, then $\{x\in\mathbb{A}\colon\ F\circ\hat{v}\neq F\circ\hat{w}\}\subseteq\{x\in\mathbb{A}\colon\ \hat{v}\neq\hat{w}\}$ and hence $F\circ\hat{v}=F\circ\hat{w}$ a.e. on $\mathbb{A}$. Thus the following definition is meaningful:

Definition A.3.1 *Let $F\colon\mathbb{R}\to\mathbb{R}$ be a measurable function, $\mathbb{A}$ be a measure space and $s_1,s_2\in[1,\infty]$. Assume that an operator $\mathcal{N}_F : L^{s_1}(\mathbb{A})\to L^{s_2}(\mathbb{A})$ is defined by the formula $\mathcal{N}_F(\hat{u})=[F\circ\mathbf{\hat{u}}]$, where $[\,.\,]$ denotes the equivalence class of the relation of being equal a.e. on $\mathbb{A}$, $\hat{u}$ is understood as and equivalence class, subject to the latter relation, and $\mathbf{\hat{u}}\in\hat{u}$. Then, $\mathcal{N}_F$ is called autonomous Nemytskii operator.*

Below, we will give conditions, under which the autonomous Nemytskii operators are well defined as operators form a Lebesgue space to a Lebesgue space. Besides, we will formulate continuity and differentiability criterion in the Lebesgue spaces. For this end, we will present certain results from [17]. Book [2] also addresses the matter of well-posedness, continuity and differentiability of Nemytskii operators. But there contained results are formulated in different fashion than in [17] and frequently are not direct equivalents of the results from [17] on which we base.

The theory of Nemytskii operators in its full generality is not necessary in this work. It will be sufficient, if we restrict our attention to the case where $\mathbb{A}=\mathbb{E}$ for certain $\mathbb{E}\subset\mathbb{R}^n$ of finite Lebesgue measure, for a given $n\in\mathbb{N}\setminus\{0\}$.

Theorem A.3.2 *Let $\mathbb{E}$ be a Lebesgue-measurable subset of $\mathbb{R}^n$ of finite measure and let $F:\mathbb{R}\to\mathbb{R}$. Assume also that $1\le s_1\le\infty$, $1\le s_2<\infty$ and that F is measurable and satisfies the following growth condition:*

$$\sup_{s\in\mathbb{R}}\ \big|F(s)\big|/\big(1+|s|^{s_1/s_2}\big)\ <\ \infty$$

Then $\mathcal{N}_F$ is well defined as an operator from $L^{s_1}(\mathbb{E})$ into $L^{s_2}(\mathbb{E})$. Moreover, $\mathcal{N}_F$ is bounded (i.e. is bounded on bounded sets).

This is the particular case of Theorem 7.13, part a) in [17].

Remark. In [17], a condition of so-called universal measurability of a function is utilized in the formulation of the part a) in Theorem 7.13. Nevertheless, in the case of finite, complete measure spaces, the notions of universally measurable functions and measurable functions coincide (see the remarks on pp. 337 of [17]). This helps to apply the result from [17] for measurable functions, as in the present case. ▲

Theorem A.3.3 *If, in Theorem A.3.2, we additionally assume that the function F is continuous and $s_1<\infty$, then the autonomous Nemytskii operator $\mathcal{N}_F$ is continuous from $L^{s_1}(\mathbb{E})$ to $L^{s_2}(\mathbb{E})$.*

The above is a consequence of the previous theorem and Theorem 7.19 in [17].

Remark. In [17], the notions of Carathéodory function and Shragin functions are used in the formulation of Theorem 7.19. Nevertheless, continuous functions are Carathéodory functions (by definition, see pp. 341 therein) and Carathéodory functions are Shragin functions (pp. 341 therein). This helps to apply the result from [17] for measurable functions, as in the present case. ▲

Now, we proceed to differential properties of Nemytskii operators acting between Lebesgue spaces. For this purpose, the notion of the multiplication operator will be useful. For $\hat{u} \in L^{s_0}(\mathbb{E})$, *the multiplication operator* $\mathcal{M}_{\hat{u}}$ is defined as

$$\mathcal{M}_{\hat{u}}(\hat{v})(x) = \hat{u}(x)\hat{v}(x) \qquad \text{for a.e. } x \in \mathbb{E}, \text{ for } \hat{v} \in L^{s_1}(\mathbb{E})$$

Remark. To be precise, in the above setting, multiplication operators act not on functions but on equivalence classes in the relation of being equal a.e. Thus, the puristic definition of $\mathcal{M}_{\hat{u}}$ should base on formula $\mathcal{M}_{\hat{u}}(\hat{v}) = [\hat{\mathbf{u}}\hat{\mathbf{v}}]$ where $\hat{\mathbf{u}} \in \hat{u}$, $\hat{\mathbf{v}} \in \hat{v}$, $\hat{u}$ and $\hat{v}$ are understood as equivalence classes and $[\,.\,]$ is as in Definition A.3.1. ▲

Observation A.3.4 *For given* $1 \leq s_2 < s_1 < \infty$, $\mathcal{M}_{\hat{u}}(\hat{v})$ *belongs to* $L^{s_2}(\mathbb{E})$*, assuming that* $\hat{u} \in L^{s_0}(\mathbb{E})$ *with* $s_0 = s_1 s_2/(s_1 - s_2)$.

This follows by the Hölder inequality (for a more explicit proof, see Lemma 7.37 in [17]). Thus, given s_0, s_1, s_2 as above and $\hat{u} \in L^{s_0}(\mathbb{E})$, the operator $\mathcal{M}_{\hat{u}}$ is a well defined operator from $L^{s_1}(\mathbb{E})$ to $L^{s_2}(\mathbb{E})$.

Theorem A.3.5 *Let* $\mathbb{E}$ *be a Lebesgue-measurable subset of* $\mathbb{R}^n$ *of finite measure and let* $F : \mathbb{R} \to \mathbb{R}$*. Assume also that* F' *exists everywhere on* $\mathbb{R}$ *and that the numbers* $1 \leq s_2 < s_1 < \infty$ *are given. Then the autonomous Nemytskii operator* $\mathcal{N}_F$ *is everywhere Fréchet differentiable from* $L^{s_1}(\mathbb{E})$ *to* $L^{s_2}(\mathbb{E})$ *if and only if* F' *satisfies the following growth condition:*

$$\sup_{s \in \mathbb{R}} \left|F'(s)\right| / \left(1 + |s|^{(s_1/s_2)-1}\right) \; < \; \infty \tag{A.11}$$

If this is the case, then the Fréchet differential of $\mathcal{N}_F$ *in a point* $\hat{u} \in L^{s_1}(\mathbb{E})$ *on a direction* $\hat{v} \in L^{s_1}(\mathbb{E})$ *is given by*

$$D_F \mathcal{N}_F(\hat{u})\hat{v} = \mathcal{M}_{F' \circ \hat{u}}(\hat{v}) \tag{A.12}$$

or more directly

$$(D_F \mathcal{N}_F(\hat{u})\hat{v})(x) = F'(\hat{u}(x))\hat{v}(x) \qquad \textit{for a.e. } x \in \mathbb{E}$$

For the proof, see Proposition 7.45 in [17].

Remark. By the assumption $\hat{u} \in L^{s_1}(\mathbb{E})$ and the growth condition (A.11), one can verify that $F' \circ \hat{u} \in L^{s_0}(\mathbb{E})$, for s_0 as in Observation A.3.4. Hence, in view of Observation A.3.4, the differential of $\mathcal{N}_F$, characterized by the formula (A.12), is a well defined operator from $L^{s_1}(\mathbb{E})$ to $L^{s_2}(\mathbb{E})$. ▲

A.4 Translation operators

This section concerns translation operators defined as follows:

Definition A.4.1 *Assume that $F\colon \mathbb{R}^n \to \mathbb{R}^l$, for some $l,n \in \mathbb{N}\setminus\{0\}$. We define the translation operator $\mathcal{T}_F$ associated with F as*

$$\mathcal{T}_F(x) := F(\,.\,-x)$$

We want to investigate properties of the translation operators understood as $\mathcal{T}_F\colon \mathbb{R}^n \to (L^s(\mathbb{R}^n))^l$ for some exponent s. This forces both F and $\mathcal{T}_F(x)$ for $x\in\mathbb{R}^n$ to be elements of $(L^s(\mathbb{R}^n))^l$ and hence the above definition in the latter context should be understood in the „almost everywhere" sense, i.e. the operator $\mathcal{T}_F : \mathbb{R}^n \to (L^s(\mathbb{R}^n))^l$ acts into equivalence classes of functions in the relation of being equal a.e. in $\mathbb{R}^n$ rather than into functions, where F also is an equivalence class in this relation. This is straight forward that for $\mathbf{F}_1,\mathbf{F}_2\in F$ there holds $[\mathbf{F}_1(\,.\,-x)] = [\mathbf{F}_2(\,.\,-x)]$, where $[\,.\,]$ denotes the equivalence class of the subject relation corresponding to a given element, hence it is possible to pose the definition of the translation operator correctly.

For brevity of notation of vector spaces associated with operator $\mathcal{T}_F$, in this section we focus on the case of $F\colon\mathbb{R}^n\to\mathbb{R}$. Also, the following notation will be valid in the present section:

$$\mathcal{T}_F^\varepsilon(x;y) := \varepsilon^{-1}\big(\mathcal{T}_F(x+\varepsilon y) - \mathcal{T}_F(x)\big) \quad \text{for } x,y\in\mathbb{R}^n$$

We do not claim that the below results are new, but we have not found suitable facts concerning the translation operators defined as above in the literature.

Theorem A.4.2 *Let $s\in[1,\infty)$ and $F\in L^s(\mathbb{R}^n)$. Then the operator $\mathcal{T}_F\colon\mathbb{R}^n\to L^s(\mathbb{R}^n)$ is uniformly continuous.*

Proof. The translation in a Lebesgue space is a norm conserving operation, hence if $\mathcal{T}_F$ is continuous in one point then it is continuous in every point of $\mathbb{R}^n$ with the same modulus of continuity. Therefore it is enough to verify the continuity of $\mathcal{T}_F$ to get the uniform continuity. This can be done by verifying the continuity of $\mathcal{T}_F$ for $F\in C_c(\mathbb{R}^n)$ and subsequently by approximating arbitrary $F\in L^s(\mathbb{R}^n)$ with functions from $C_c(\mathbb{R}^n)$. This reasoning is realized e.g. in the proof of [1, Th. 2.32]. ■

Lemma A.4.3 *Let $F\in W^{1,s}(\mathbb{R}^n)$, $s\in[1,\infty)$ and $x,y\in\mathbb{R}^n$. Then $\big\|\mathcal{T}_F^\varepsilon(x;y)\big\|_s \le \big|\,y\,\big|_{s'}\big\|\nabla F(x)\big\|_s$ for all $\varepsilon\neq 0$, where s' is the Hölder conjugate of s.*

Proof. The proof rely on reasoning utilized in the proof of [21, Chap. 5.8.2, Th. 3]. However, the above Theorem is formulated slightly different than the one in [21] hence we present the proof below.

Begin with the case of $F\in C^1(\mathbb{R}^n)$. Denote by $\mathbf{e}_i$ the i-th vector of the canonical base in $\mathbb{R}^n$. Then:

$$F(x+\varepsilon\mathbf{e}_i) - F(x) \;=\; \int_0^\varepsilon \partial_i F(x+t\mathbf{e}_i)\,dt \;=\; \varepsilon\int_0^1 \partial_i F(x+t\varepsilon\mathbf{e}_i)\,dt$$

Now we can write:

$$\begin{aligned}\big\|\mathcal{T}_F^\varepsilon(x;\mathbf{e}_i)\big\|_s^s \;&\le\; \int_{\mathbb{R}^n}\left(\int_0^1\big|\partial_i F(x+t\varepsilon\mathbf{e}_i)\big|\,dt\right)^s dx\\ &\le\; \int_0^1\int_{\mathbb{R}^n}\big|\partial_i F(x+t\varepsilon\mathbf{e}_i)\big|^s\,dt\,dx = \int_0^1\big\|\partial_i F\big\|_s^s\,dt = \big\|\partial_i F\big\|_s^s\end{aligned} \tag{A.13}$$

Fix $x, y \in \mathbb{R}^n$. Note that for arbitrary $y \in \mathbb{R}^n$:

$$\mathcal{T}_F^\varepsilon(x,y) = \sum_{i=1}^{n} y_i \mathcal{T}_F^{\varepsilon y_i}(\mathbf{x}_i, \mathbf{e}_i)$$

where $\mathbf{x}_i := x$ for $i = 1$ and $\mathbf{x}_i := \mathbf{x}_{i-1} + y_i$ for $i = 2, \ldots, n$. By the above, by (A.13) and by Hölder inequality for sequences we have:

$$\|\mathcal{T}_F^\varepsilon\|_s \;\leq\; \sum_{i=1}^{n} |y_i| \|\mathcal{T}_F^{\varepsilon y_i}(\mathbf{x}_i; \mathbf{e}_i)\|_s \;\leq\; \sum_{i=1}^{n} |y_i| |\partial_i F|s \;\leq\; \big|\, y \,\big|_{s'} \|\nabla F\|_s$$

$C^1(\mathbb{R}^n)$ functions are dense in $W^{1,s}(\mathbb{R}^n)$ for $s \in [1, \infty)$, see [1, Th. 3.17], hence we infer that the above holds also for all $F \in W^{1,s}(\mathbb{R}^n)$. ■

As a consequence, we can prove sufficient conditions for the Lipschitz continuity and the weak Gâteaux differentiability of $\mathcal{T}_F$.

Theorem A.4.4 *Let $F \in W^{1,s}(\mathbb{R}^n)$, $s \in [1, \infty)$. Then the operator $\mathcal{T}_F \colon \mathbb{R}^n \to L^s(\mathbb{R}^n)$ is globally Lipschitz continuous.*

Theorem A.4.4 is a direct consequence of Lemma A.4.3.

Theorem A.4.5 *Let $F \in W^{1,s}(\mathbb{R}^n)$, $s \in (1, \infty)$. Then $\mathcal{T}_F \colon \mathbb{R}^n \to L^s(\mathbb{R}^n)$ is weakly Gâteaux differentiable and its weak Gâteaux differential in point $x \in \mathbb{R}^n$ in direction $y \in \mathbb{R}^n$ is given by*

$$\begin{aligned} \big(D_{G,w}\mathcal{T}_F(x)(y)\big)(z) \;&=\; -\,D_F F(z-x)y \;= \\ &=\; -\,\big(\nabla F(z-x), y\big)_{\mathbb{R}^n} \;=\; -\big(\mathcal{T}_{\nabla F}(x)(z), y\big)_{\mathbb{R}^n} \end{aligned} \tag{A.14}$$

for a.e. $z \in \mathbb{R}^n$

Proof. Note, that translations commute with differentiation, hence it suffices to verify the assertion for $x = \mathbf{0}$ — if the difference quotients converge weakly to $-\left(\mathcal{T}_{\nabla F}(\mathbf{0}), y\right)_{\mathbb{R}^n}$ then the translated by x difference quotients converge weakly to $-\left(\mathcal{T}_{\nabla F}(x), y\right)_{\mathbb{R}^n}$.

For $x = \mathbf{0}$ and for $\phi \in C_c^\infty(\mathbb{R}^n)$

$$\begin{aligned} &\int_{\mathbb{R}^n} \mathcal{T}_F^\varepsilon(\mathbf{0}; y)(z)\,\phi(z)\,dz \;=\; \int_{\mathbb{R}^n} \varepsilon^{-1}\big(F(z-\varepsilon y) - F(z)\big)\,\phi(z)\,dz \;= \\ =\; &\int_{\mathbb{R}^n} F(z)\,\varepsilon^{-1}\big(\phi(z+\varepsilon y) - \phi(z)\big)\,dz \;\xrightarrow{\varepsilon}\; \int_{\mathbb{R}^n} F(z)\big(\nabla\phi(z), y\big)_{\mathbb{R}^n}\,dz \;= \\ =\; -&\int_{\mathbb{R}^n} \big(\nabla F(z), y\big)_{\mathbb{R}^n}\phi(z)\,dz \;=\; -\int_{\mathbb{R}^n} \big(\mathcal{T}_{\nabla F}(\mathbf{0})(z), y\big)_{\mathbb{R}^n}\phi(z)\,dz \end{aligned}$$

Moreover, $C_c^\infty(\mathbb{R}^n)$ is dense in $L^{s'}(\mathbb{R}^n)$ (see [1, par. 2.30]) and due to Lemma A.4.3 the difference quotients $\mathcal{T}_F^\varepsilon(\mathbf{0}; y)$ are bounded w.r.t. ε in $L^s(\mathbb{R}^n)$. Therefore, the above convergence holds also for all $\phi \in L^{s'}(\mathbb{R}^n)$ what concludes the proof. ■

Example. Theorem A.4.4 together with Theorem A.4.5 give a big class of functions F for which the associated translation operator $\mathcal{T}_F$ is both Lipschitz and weakly Gâteaux differentiable. However, an example of $F \in L^s(\mathbb{R}^n)$ for which $\mathcal{T}_F$ is Lipschitz continuous but not weakly Gâteaux differentiable can be easily indicated. For instance, take into consideration $F(x) := \mathbf{1}_{B(0,r)}(x)$ with given radius $r > 0$ and the space $L^2(\mathbb{R}^n)$. It can be verified that the Lipschitz continuity of

$\mathcal{T}_F$ in $L^2(\mathbb{R}^n)$ is true. At the same time, it is straightforward to check in the case of $n = 1$, that $\mathcal{T}_F$ is not weakly Gâteaux differentiable. To see it, one can check that the difference quotients of $\mathcal{T}_F$ for $n = 1$ are unbounded in $L^2(\mathbb{R})$, hence they cannot be weakly convergent, what contradicts the weak Gâteaux differentiability. ▲

A.5 Multivalued mappings

This short section mostly bases on concepts concerning multivalued mappings presented in [4]. The Reader is referred there for more detailed theory of multivalued mappings.

A multivalued mapping from set $\mathbb{A}_1$ to set $\mathbb{A}_2$ is a function with values in the set of subsets of $\mathbb{A}_2$. A given multivalued mapping can be understood both as an usual function from $\mathbb{A}_1$ to $2^{\mathbb{A}_2}$ or as a generalization of usual function from $\mathbb{A}_1$ to $\mathbb{A}_2$. In the below definitions and facts, the second of these two interpretations is exploited. However defining a multivalued mapping F from $\mathbb{A}_1$ to $\mathbb{A}_2$ we prefer to use notation $F\colon \mathbb{A}_1 \to 2^{\mathbb{A}_2}$ in order to emphasize that F is not an usual function from $\mathbb{A}_1$ to $\mathbb{A}_2$.

For a given multivalued mapping $F\colon \mathbb{A}_1 \to 2^{\mathbb{A}_2}$, we denote by $G(F)$ its graph, defined by

$$G(F) := \bigcup_{\omega \in \mathbb{A}_1} \big\{(\omega, F(\omega)) \subset \mathbb{A}_1 \times \mathbb{A}_2\big\} = \big\{(\omega_1, \omega_2) \in \mathbb{A}_1 \times \mathbb{A}_2 \colon \omega_2 \in F(\omega_1)\big\}$$

Thus we understand $G(F)$ as a subset of $\mathbb{A}_1 \times \mathbb{A}_2$ and not as a subset of $\mathbb{A}_1 \times 2^{\mathbb{A}_2}$.

For convenience of notation, for a multivalued mapping $F\colon \mathbb{A}_1 \to 2^{\mathbb{A}_2}$ as above and for a given subset $\widetilde{\mathbb{A}} \subseteq \mathbb{A}_1$ we denote by $F|_{\widetilde{\mathbb{A}}}$ the restriction of F to $\widetilde{\mathbb{A}}$.

Moreover, still keeping the above meaning of $\mathbb{A}_1$, $\mathbb{A}_2$, $\widetilde{\mathbb{A}}$ and F, we denote:

$$F(\widetilde{\mathbb{A}}) = \bigcup_{\omega \in \widetilde{\mathbb{A}}} F(\omega)$$

Basing on the above notation, we define the superposition of two multivalued mappings in the following way. Let sets $\mathbb{A}_1$, $\mathbb{A}_2$ and $\mathbb{A}_3$ be given and let $F_1\colon \mathbb{A}_1 \to 2^{\mathbb{A}_2}$ and $F_2\colon \mathbb{A}_2 \to 2^{\mathbb{A}_3}$ be multivalued mappings. We denote $F_2 \circ F_1(\omega) = F_2(F_1(\omega))$ for all $\omega \in \mathbb{A}_1$.

If $\mathbb{A}_1$ and $\mathbb{A}_2$ are topological spaces, a notion of continuity can be defined for a multivalued mapping $F\colon \mathbb{A}_1 \to 2^{\mathbb{A}_2}$. Below, for simplicity, we restrict our attention to the case where both $\mathbb{A}_1$ and $\mathbb{A}_2$ are Banach spaces.

Definition A.5.1 *For two Banach spaces X and Y, a multivalued mapping $T : X \longrightarrow 2^Y$ is said to be bounded on X if and only if there exists $R > 0$ such, that $T(\hat{x}) \subseteq B(0, R)$ for all $\hat{x} \in X$.*

Definition A.5.2 *For two Banach spaces X and Y, a multivalued mapping $T : X \longrightarrow 2^Y$ is said to be upper semicontinuous in $\hat{x} \in X$ if and only if for every neighborhood $O \subseteq Y$ of $T(\hat{x})$, there exists a neighborhood $U \subseteq X$ of $\hat{x}$ such that $T(\hat{z}) \subset O$ for $\hat{z} \in U$. T is said to be upper semicontinuous if it is upper semicontinuous for all $\hat{x} \in X$.*

Definition A.5.3 *For two Banach spaces X and Y, a multivalued mapping $T : X \longrightarrow 2^Y$ is said to be lower semicontinuous in $\hat{x} \in X$ if and only if for every $\hat{y} \in T(\hat{x})$ and every neighborhood $O \subseteq Y$ of $\hat{y}$, there exists a neighborhood $U \subseteq X$ of $\hat{x}$ such that $T(\hat{z}) \cap O \neq \emptyset$ for $\hat{z} \in U$. T is said to be lower semicontinuous if it is lower semicontinuous for all $\hat{x} \in X$.*

Definition A.5.4 *For two Banach spaces X and Y, a multivalued mapping $T : X \longrightarrow 2^Y$ is said to be continuous in $\hat{x} \in X$ if and only if it is both upper and lower semicontinuous in $\hat{X}$. T is said to be continuous if it is continuous for all $\hat{x} \in X$.*

If the values of T in the above definitions are singletons, then T can be understood as a usual single-valued operator between Banach spaces. Note that in this case, the property of upper semicontinuity in Definition A.5.2 reduces to the definition of continuity of T. The same observation holds for the notion of lower semicontinuity of multivalued mappings in Definition A.5.3. Thus the upper semicontinuity an the lower semicontinuity of a multivalued mapping is a property that is stronger that the upper semicontinuity of a usual single-valued operator.

The following two examples of multivalued mappings are as in [4, p. 109, Ch. 3 Sec. 1] and illustrate the differences between the notion of upper semicontinuity and lower semicontinuity of multivalued mappings. Let $F_1, F_2 : \mathbb{R} \to 2^{\mathbb{R}}$ be defined by

$$F_1(s) = \begin{cases} 0 & \text{for } s \in \mathbb{R} \setminus \{0\} \\ [-1,1] & \text{for } s = 0 \end{cases} \qquad F_2(s) = \begin{cases} [-1,1] & \text{for } s \in \mathbb{R} \setminus \{0\} \\ 0 & \text{for } s = 0 \end{cases}$$

It is straightforward that F_1 is upper semicontinuous and not lower semicontinuous. At the same time, F_2 is lower semicontinuous but not upper semicontinuous.

Now, by the below proposition, we will indicate more examples of upper semicontinuous mappings:

Proposition A.5.5 *For a given single-valued function $F\colon \mathbb{R} \to \mathbb{R}$, define $\vec{F}(s) := \lim_{r \to s^-} F(r)$, $\overset{\leftarrow}{F}(s) := \lim_{r \to s^+} F(r)$, $\widetilde{F}_{\min}(s) := \min \left\{ \vec{F}(s), \overset{\leftarrow}{F}(s) \right\}$ and $\widetilde{F}_{\max}(s) := \max \left\{ \vec{F}(s), \overset{\leftarrow}{F}(s) \right\}$ for $s \in \mathbb{R}$. If F is such that $\vec{F}(s)$ and $\overset{\leftarrow}{F}(s)$ are well defined for all $s \in \mathbb{R}$, then the multivalued mapping $\widetilde{F}\colon \mathbb{R} \to 2^{\mathbb{R}}$ given by*

$$\widetilde{F}(s) = [\widetilde{F}_{\min}(s), \widetilde{F}_{\max}(s)] \qquad \textit{for } s \in \mathbb{R} \tag{A.15}$$

is upper semicontinuous.

Proof. For convenience, for $\varepsilon > 0$ and for $\mathbb{A} \subseteq \mathbb{R}$, we denote by $\mathbb{A}_\varepsilon$ the ε-neighborhood of $\mathbb{A}$, i.e. the set $\{s \in \mathbb{R}\colon dist_{\mathbb{R}}(s, \mathbb{A}) < \varepsilon\}$, where $dist_{\mathbb{R}}$ denote the distance in the metric space $\mathbb{R}$.

Step 1. Fix $s_0 \in \mathbb{R}$ and $\varepsilon > 0$. It suffices to show that there exists $\delta > 0$ such that, for s satisfying $|s_0 - s| < \delta$, there holds $\widetilde{F}(s) \subset \left(\widetilde{F}(s_0)\right)_\varepsilon$. The latter inclusion is equivalent to

$$\begin{aligned} \sup \widetilde{F}(s) &< \sup \widetilde{F}(s_0) + \varepsilon \\ \inf \widetilde{F}(s) &< \inf \widetilde{F}(s_0) - \varepsilon \end{aligned} \tag{A.16}$$

For $s = s_0$ the above is trivial. We will focus on the case $s > s_0$. The case $s < s_0$ can be treated analogously.

Step 2. Let the number $\bar{\varepsilon} > 0$ be fixed. Then, by definition of $\overset{\leftarrow}{F}$, there exists $\delta_1 > 0$ such that

$$\left|\overset{\leftarrow}{F}(s_0) - F(s)\right| < \bar{\varepsilon} \qquad \text{for } s_0 < s < s_0 + \delta_1 \tag{A.17}$$

Inequality (A.17) means that the values of F belong to certain interval for s sufficiently close to s_0. From this we infer that the limits of values of F remain in the closure of the latter interval, hence:

$$\left|\overset{\leftarrow}{F}(s_0) - \overset{\leftarrow}{F}(s)\right| \le \bar{\varepsilon} < 2\bar{\varepsilon} \qquad \text{for } s_0 < s < s_0 + \delta_1 \tag{A.18}$$

By triangle inequality, (A.17) and (A.18) imply that:

$$\big|\overleftarrow{F}(s) - F(r)\big| < 3\bar{\varepsilon} \qquad \text{for } s_0 < r, s < s_0 + \delta_1 \tag{A.19}$$

Next, by definition of $\overrightarrow{F}$, for a given s there exists $\delta_2 > 0$ such that

$$\big|\overrightarrow{F}(s) - F(r)\big| < \bar{\varepsilon} \qquad \text{for } s_0 < r < s < s_0 + \delta_1,\ \big|r - s\big| < \delta_2 \tag{A.20}$$

Now, let r and s satisfy conditions $s_0 < r < s < s_0 + \delta_1$, $\big|r - s\big| < \delta_2$. The difference $\overleftarrow{F}(s) - \overrightarrow{F}(s)$ can be represented as $\overleftarrow{F}(s) - F(r) + F(r) - \overrightarrow{F}(s)$. The latter representation together with the triangle inequality, (A.19) and (A.20) yields

$$\big|\overleftarrow{F}(s) - \overrightarrow{F}(s)\big| < 4\bar{\varepsilon} \qquad \text{for } s_0 < s < s_0 + \delta_1 \tag{A.21}$$

Step 3. Take $\bar{\varepsilon} = \varepsilon/6$ and $s > s_0$. Choose δ_1 as in the previous step if the proof. Using (A.18) and (A.21), we obtain:

$$\begin{aligned} \sup \widetilde{F}(s) &= \max\Big\{\overrightarrow{F}(s), \overleftarrow{F}(s)\Big\} < \max\Big\{\overleftarrow{F}(s) + 4\bar{\varepsilon}, \overleftarrow{F}(s)\Big\} \\ &< \max\Big\{\overleftarrow{F}(s_0) + 2\bar{\varepsilon} + 4\bar{\varepsilon}, \overleftarrow{F}(s_0) + 2\bar{\varepsilon}\Big\} = \overleftarrow{F}(s_0) + 6\bar{\varepsilon} = \overleftarrow{F}(s_0) + \varepsilon \end{aligned} \tag{A.22}$$

for $s_0 < s < s_0 + \delta_1$. In the same manner we get

$$\inf \widetilde{F}(s) > \overleftarrow{F}(s_0) - \varepsilon \tag{A.23}$$

Step 4. If $\overleftarrow{F}(s_0) \geq \overrightarrow{F}(s_0)$, then (A.22) with (A.23) imply

$$\begin{aligned} \sup \widetilde{F}(s) &< \overleftarrow{F}(s_0) + \varepsilon && = \sup \widetilde{F}(s_0) + \varepsilon \\ \inf \widetilde{F}(s) &> \overleftarrow{F}(s_0) - \varepsilon \geq \overrightarrow{F}(s_0) - \varepsilon && = \inf \widetilde{F}(s_0) - \varepsilon \end{aligned}$$

and (A.16) is proven. If $\overleftarrow{F}(s_0) \geq \overrightarrow{F}(s_0)$, then (A.16) can be proven analogously. The proof is complete. ■

Proposition A.5.5, by the mapping $F \mapsto \widetilde{F}$, gives a method of assigning an unique upper semicontinuous multivalued mapping to a given function, satisfying respective assumptions. For example, for $F(s) = -sgn(s)$ Proposition A.5.5 can be applied and the formula $\widetilde{F}(s) = [\widetilde{F}_{\min}(s), \widetilde{F}_{\max}(s)]$ in the statement of the proposition gives a multivalued mapping

$$\widetilde{F}(s) = \begin{cases} +1 & \text{for } s < 0 \\ [-1, +1] & \text{for } s = 0 \\ -1 & \text{for } s > 0 \end{cases}$$

Other important notion concerning multivalued mappings is monotonicity and maximal monotonicity:

Definition A.5.6 *Let H be a Hilbert space and let a multivalued mapping $T\colon H \to 2^H$ be given. We say that T is monotone if and only if*

$$(x_1 - x_2, y_1 - y_2)_H \geq 0 \quad \textit{for all } (x_1, y_1), (x_2, y_2) \in G(T)$$

We say that T is maximal monotone if and only if there is no monotone multivalued mapping $\widetilde{T}\colon H \to 2^H$ such that $G(T) \subsetneq G(\widetilde{T})$.

The below facts emphasize properties of maximal monotone multivalued mappings. Note in particular, that the first of the below two propositions indicates a big class of upper semicontinuous multivalued mappings, extending the collection of examples of upper semicontinuity already given above.

Proposition A.5.7 *Let H be a Hilbert space and let M be its compact subset. A maximal monotone multivalued mapping $T\colon H \to 2^M$ is upper semicontinuous.*

Proof. First, by [4, Prop. 3, Ch. 6, Sec. 7] we can infer that the graph of a maximal monotone multivalued mapping is sequentially closed. Since Hilbert spaces are metric, it means that $G(T)$ is closed. Next, by [4, Coro. 9, Ch.3, Sec. 1], in particular multivalued mappings on Hilbert spaces with closed graph and with values in a compact set are upper semicontinuous. This justifies the desired assertion. ■

Proposition A.5.8 *Let H be a Hilbert space and let $T\colon H \to 2^H$ be a maximal monotone multivalued mapping. Then values of T are closed and convex.*

For justification of Proposition A.5.8, see [4, Prop. 3, Ch. 6, Sec. 7].

Maximal monotone mappings do not need to have nonempty values. For instance, consider $T\colon \mathbb{R} \to \mathbb{R}$ defined by $T(s) = \ln(s)$ for $s > 0$, $T(s) = \emptyset$ otherwise. The mapping is monotone and, by a simple verification, maximal. At the same time, it has infinitely many empty values. But, assuming that a maximal monotone mapping is bounded, the possibility of empty values can be excluded for mappings $T\colon \mathbb{R} \to \mathbb{R}$:

Proposition A.5.9 *Let $T\colon \mathbb{R} \to \mathbb{R}$ be maximal monotone and bounded. Then, $T(s)$ are nonempty for all $s \in \mathbb{R}$.*

Proof. We will justify the assertion by contradiction. Namely, assume that there exists $s_0 \in \mathbb{R}$ such that $T(s_0) = \emptyset$. We will prove that T can be extended to s_0 in a manner preserving the monotonicity, what will contradict the maximality of T.

Since T is bounded, the infimum of the values being „on the right" of s_0 (i.e. the number $\inf \bigcup_{s>s_0} T(s)$) is finite. Denote it as C_R. Similarly, the supremum of the values being „on the left" of s_0, denote it C_L, is finite. It follows straight that $C_L \leq C_R$. Otherwise, a contradiction to monotonicity would be implied. In consequence, the set $[C_L, C_R]$ is nonempty (here, in the case of $C_L = C_R$, we interpret the latter set as the singleton $\{C_R\}$).

Now, simply note that by assigning the value $[C_L, C_R]$ to the point s_0 we obtain an extension of T which is monotone. This follows straight by the definition of monotonicity (Definition A.5.6) and definitions of C_R and C_L. The maximality of T has been contradicted, what concludes the proof. ■

We do not claim that the results given in Lemma A.5.5 and Lemma A.5.9 are new, however we do not know a suitable literature reference for the subject statements.

Appendix B

Index of figures and tables

Figures:

Tables:

Appendix C

Index of theorems

Below, general theorems utilized in the theoretical chapters of the present work are listed, together with list of pages where the theorems were formulated or necessary. However, the list may omit some occurrences of the indexed theorems. Also, the list omits theorems of lesser importance. If we write *„passim”*, then it means that a given theorem was used in too many places to list them here, or certain of its occurrences were not of significant importance, or it was used implicitly, without explicit reference in the text.

Bibliography

[1] Adams, R. A., and Fournier, J. J. F. *Sobolev Spaces*, second ed. Pure and Applied Mathematics (Amsterdam), 140. Elsevier/Academic Press, Amsterdam, 2003.

[2] Appell, J., and Zabrejko, P. P. *Nonlinear Superposition Operators.* Cambridge Tracts in Mathematics, 95. Cambridge University Press, 1990.

[3] Arendt, W., Batty, C., Hieber, M., and Neubrander, F. *Vector-valued Laplace Transforms and Cauchy Problems*, second ed. Monographs in Mathematics, 96. Birkhäuser/Springer Basel AG, Basel, 2011.

[4] Aubin, J. P., and Ekeland, I. *Applied Nonlinear Analysis.* Pure and Applied Mathematics. John Wiley & Sons, Inc., New York, 1984. A Wiley-Interscience Publication.

[5] Bagagiolo, F. Automatic control via preisach hysteresis of a filtration model with capillarity. *Adv. Math. Sci. Appl. 17*, 1 (2007), 1–22.

[6] Barbu, V. *Nonlinear semigroups and differential equations in Banach spaces.* Editura Academiei Republicii Socialiste România, Bucharest, 1976. Translated from the Romanian.

[7] Bazaraa, M., Sherali, H., and Shetty, C. *Nonlinear Programming, Theory and Algorithms*, second (1993) ed. John Wiley & Sons, Inc., New York-Chichester-Brisbane-Toronto-Singapore, 1979.

[8] Behnia, B., Suthar, M., and Webb, A. G. Closed-loop feedback control of phased-array microwave heating using thermal measurements from magnetic resonance imaging. *Concepts in Magnetic Resonance (Magnetic Resonance Engineering) 15*, 1 (2002), 101–110.

[9] Bohnenblust, H. F., and Karlin, S. On a theorem of ville. In *Contributions to the Theory of Games*, H. W. Kuhn and A. W. Tucker, Eds. Princeton University Press, 1950, pp. 155–160. Vol. I.

[10] Brokate, M., and Friedman, A. Optimal design for heat conduction problems with hysteresis. *SIAM J. Control Optim. 27*, 4 (1989), 697–717.

[11] Campbell, S., and Macki, W. Control of the temperature at one end of a rod. *Mathematical and Computer Modelling 32* (2000), 825–842.

[12] Cavaterra, C., and Colombo, F. Automatic control problems for reaction-diffusion systems. *J. Evol. Equ 2*, 2 (2002), 241–273.

[13] Ciarlet, P. *The Finite Element Method for Elliptic Problems.* North-Holland Publishing Co., Amsterdam-New York-Oxford, 1978. Studies in Mathematics and its Applications, Vol. 4.

[14] Coddington, E. A., and Levinson, N. *Theory of ordinary differential equations.* McGraw-Hill Book Company, Inc., New York-Toronto-London, 1955.

[15] Colli, P., Grasselli, M., and Sprekels, J. Automatic control via thermostats of a hyperbolic stefan problem with memory. *Appl. Math. Optim. 39* (1999), 229–255.

[16] Das, S. K., Clegg, S. T., and Samulski, T. V. Computational techniques for fast hyperthermia temperature optimization. *Medical Physics 26*, 2 (1999), 319–328.

[17] Dudley, R. M., and Norvaiša, R. *Concrete functional calculus.* Springer Monographs in Mathematics. Springer, New York, 2011.

[18] Dudziuk, G. The model of closed-loop control by thermostats: properties and numerical simulations. Published on arXiv.org, 2nd version (2013), URL: `http://arxiv.org/abs/1305.5442v2`.

[19] Dudziuk, G., and Niezgódka, M. Closed-loop control of a reaction-diffusion system. *Adv. Math. Sci. Appl. 21*, 2 (2011), 383–402.

[20] Ekeland, I., and Temam, R. *Convex Analysis and Variational Problems.* North-Holland, Amsterdam, 1976.

[21] Evans, L. C. *Partial differential equations*, second ed. Graduate Studies in Mathematics, 19. American Mathematical Society, Providence, RI, 2010.

[22] Filippov, A. F. *Differential equations with discontinuous righthand sides*, vol. 18 of *Mathematics and its Applications (Soviet Series).* Kluwer Academic Publishers Group, Dordrecht, 1988. Translated from the Russian.

[23] Friedman, A., and Hoffmann, K.-H. Control of free boundary problems with hysteresis. *SIAM J. Control Optim. 26* (1988), 42–55.

[24] Gilbert, J. C., and Nocedal, J. Global convergence properties of conjugate gradient methods for optimization. *SIAM J. Optimization 2*, 1 (1992), 21–42.

[25] Glashoff, K., and Sprekels, J. An application of Glicksberg's theorem to set-valued integral equations arising in the theory of thermostats. *SIAM J. Math. Analysis 12* (1981), 477–486.

[26] Glashoff, K., and Sprekels, J. The regulation of temperature by thermostats and set-values integral equations. *J. Integral Eqns 4* (1982), 95–112.

[27] Glicksberg, I. L. A further generalization of the Kakutani fixed point theorem, with application to Nash equilibrium points. *Proceedings of the American Mathematical Society 3*, 1 (1952), 170–174.

[28] Götz, I. G., Hoffmann, K.-H., and Meirmanov, A. M. Periodic solutions of the Stefan problem with hysteresis-type boundary conditions. *manuscripta mathematica 87* (1993), 179–199.

[29] Graselli, M. Automatic control of the temperature in thermoviscoelasticity. *Math. Meth. Appl. Sci. 22* (1999), 1147–1464.

[30] Gurevich, P., and Jäger, W. Parabolic problems with the Preisach hysteresis operator in boundary conditions. *J. Differential Equations 247* (2009), 2966–3010.

[31] Gurevich, P., Jäger, W., and Skubachevskii, A. On periodicity of solutions for thermocontrol problems with hysteresis-type switches. *SIAM J. Math. Anal. 41* (2009), 733–752.

[32] Gurevich, P., and Tihhomirov, S. Symmetric periodic solutions of parabolic problems with discontinuous hysteresis. *J Dyn Diff Equat 23* (2011), 923–960.

[33] Hoffmann, K.-H., Niezgódka, M., and Sprekels, J. Feedback control via thermostats of multidimensional two-phase stefan problems. *Nonlinear Anal.: Theory, Meth. and Appl. 15* (1990), 955–976.

[34] Jakubczyk, B. Lecture Notes on Optimization II. Lecture script for optimization courses conducted by Bronisław Jakubczyk from Institute of Mathematics of Polish Academy of Sciences. Written in Polish. Original title „Wykłady z Optymalizacji II". Version from 20/01/2004. First version typed in TEX by Karolina Napierała.

[35] Kowalski, M. E., Behnia, B., Webb, A. G., and Jin, J.-M. Optimization of electromagnetic phased-arrays for hyperthermia via magnetic resonance temperature estimation. *IEEE Transactions on Biomedical Engineering 49*, 11 (2002), 1229–1241.

[36] Kowalski, M. E., and Jin, J.-M. A temperature-based feedback control system for electromagnetic phased-array hyperthermia: theory and simulation. *Physics in Medicine and Biology 48* (2003), 633–651.

[37] Ladyženskaja, O. A., Solonnikov, V. A., and Ural'ceva, N. N. *Linear and quasilinear equations of parabolic type.* Translations of Mathematical Monographs, Vol. 23. American Mathematical Society, Providence, R.I., 1968. Translated from the Russian by S. Smith.

[38] Nocedal, J., and Wight, S. *Numerical Optimization*, second ed. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006.

[39] Pawłow, I. *Analysis and Control of Evolution Multi-Phase Problems with Free Boundaries.* Ossolineum Publishing House, WrocĹ,aw, Poland, 1987. Habilitation thesis, System Research Institute of Polish Academy of Sciences.

[40] Robinson, J. C. *Infinite-Dimensional Dynamical Systems.* Cambridge University Press, 2001.

[41] Rudin, W. *Real and complex analysis*, third ed. McGraw-Hill Book Co., New York, 1987.

[42] Salomir, R., et al. Hyperthermia by MR-guided focused ultrasound: Accurate temperature control based on fast MRI and a physical model of local energy deposition and heat conduction. *Magnetic Resonance in Medicine 43* (2000), 342–347.

[43] Showalter, R. *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations.* Mathematical surveys. Amer Mathematical Society, 1997.

[44] Simon, J. Compact sets in the space $L^p(0,T;B)$. *Ann. Mat. Pura Appl. (4) 146* (1987), 65–96.

[45] Tröltzsch, F. *Optimal control of partial differential equations: theory, methods and applications.* Graduate Studies in Mathematics, v.112. American Mathematical Society, Providence, RI, 2010. translated from the 2005 German original by Sprekels J.

[46] van der Zee, J. Heating the patient: a promising approach? *Annals of Oncology 13*, 8 (2002), 1173–1184.

[47] Weihrauch, M., Wust, P., et al. Adaptation of antenna profiles for control of MR guided hyperthermia (HT) in a hybrid MR-HT system. *Medical Physics 34*, 12 (2007), 4717–4725.

[48] Wust, P., et al. Hyperthermia in combined treatment of cancer. *The Lancet Oncology 3* (2002), 487–497.

[49] Yosida, K. *Functional Analysis*, second ed. Springer-Verlag, Berlin - Heidelberg - New York, 1966.

[50] Zeidler, E. *Nonlinear Functional Analysis and its Applications. I. Fixed-Point Theorems.* Springer-Verlag, New York, 1986. translated from the German original by Wadsack P. R.

[51] Zeidler, E. *Nonlinear Functional Analysis and its Applications. II/A. Linear Monotone Operators.* Springer-Verlag, New York, 1990. translated from the German original by the author and Boron L. F.

[52] Zeidler, E. *Nonlinear Functional Analysis and its Applications. II/B. Nonlinear Monotone Operators.* Springer-Verlag, New York, 1990. translated from the German original by the author and Boron L. F.