Tram Thi Ngoc Nguyen

Parameter identification for time-dependent inverse problems

DISSERTATION

submitted in fulfillment of the requirements for the degree of
Doktor der Technischen Wissenschaften

Alpen-Adria-Universität Klagenfurt
Fakultät für Technische Wissenschaften

Supervisor
Prof. Dr. Barbara Kaltenbacher
Universität Klagenfurt
Institut für Mathematik

First Evaluator
Prof. Dr. Fredi Tröltzsch
Technische Universität Berlin
Institut für Mathematik

Second Evaluator
Prof. Dr. Bernadette Hahn
Universität Würzburg
Institut für Mathematik

Klagenfurt, March 2020

Affidavit

I hereby declare in lieu of an oath that

- the submitted academic paper is entirely my own work and that no auxiliary materials have been used other than those indicated,
- I have fully disclosed all assistance received from third parties during the process of writing the thesis, including any significant advice from supervisors,
- any contents taken from the works of third parties or my own works that have been included either literally or in spirit have been appropriately marked and the respective source of the information has been clearly identified with precise bibliographical references (e.g. in footnotes),
- to date, I have not submitted this paper to an examining authority either in Austria or abroad and that
- when passing on copies of the academic thesis (e.g. in bound, printed or digital form), I will ensure that each copy is fully consistent with the submitted digital version.

I understand that the digital version of the academic thesis submitted will be used for the purpose of conducting a plagiarism assessment.

I am aware that a declaration contrary to the facts will have legal consequences.

Tram Thi Ngoc Nguyen m.p. Klagenfurt, March 2020

Acknowledgements

I would like to express my special appreciation to my supervisor Professor Barbara Kaltenbacher. Her wisdom, knowledge, guidance, support and patience have taught me to be a scientist. I am so honored to be one of your PhD students.

Particular thanks are given to evaluators Professor Fredi Tröltzsch and Professor Bernadette Hahn for carefully reading the manuscript and giving valuable feedback. I would also like to thank for their questions, which encourage me to widen my research from various perspectives.

In the course of Young Scientists Mentoring program, I spent a week at Saarland University, where I participated in many interesting discussions with Doctor Anna Wald. I truly thank her for being patient with my questions about physics of MPI.

I sincerely thank my mentor Elena as she regularly picks me up for lunch and talks to me, just to make sure everything is going well. I am deeply grateful for her caring.

My thanks are to my colleagues in Math department at University of Klagenfurt, whose intimacy and friendship make me feel home here. Thanks to Angelika for lending then letting me keep her guitar, which became a friend of mine in the past three years.

My little sister consistently helped me with grammar check throughout my dissertation. Her assistance is so helpful, although I have not followed her advice in editing the mathematical terminology :-).

And finally, this work is dedicated to my parents.

In solving a problem of this sort, the grand thing is to be able to reason backwards.

— Sherlock Holmes
A Study in Scarlet
Arthur Conan Doyle

Preface

Firstly, we explain the concept of an inverse problem. It begins with the notion that there is a direct problem, i.e., a well-posed problem. In other words, we have a sufficient knowledge of an object and of how it evolves in a mathematical model of a physical process. If the object is unknown, instead, some additional information of the direct problem is supplemented, we are in a position of dealing with an *inverse problem*. Mathematically, the inverse problem is stated as

$$\text{Determinine } x \text{ from given data } y: \qquad F(x) = y.$$

Many inverse problems involving *time-dependent partial differential equations* naturally arise and have essential applications, ranging from engineering via physics, geophysics, biology, ecology to economics. Applications include reconstruction of the interior of the human body from exterior electrical voltages, ultrasound, X-ray, magnetic measurements; recovery of geologic Earth structure from sea surface acquisition recording the arrival time of seismic waves; locating an unknown moving or navigated obstacles from acoustic or electromagnetic fields, etc. There exists an immense amount of literature on several aspects of inverse problem varying from abstract regularization theory to concrete applications. We selectively mention [31, 69, 62, 104, 56, 108, 41] for the abstract theory and [70, 97, 85, 73, 39, 90, 72, 30, 14, 24, 65] for investigations in real applications.

The mathematical model for those problems can be formulated as a state-space system in which the parameter is supposed to be identified from additional observations. *Parameter identification problems* deal with the reconstruction of unknown functions appearing as parameters (coefficients, sources, boundary values...) in systems of differential equations. Recently, parameter identification attracts much interest due to its vast applications in modern medicine. These problems are relatively new and challenging for two reasons: they are nonlinear, and they are improperly posed (ill-posed). A number of theoretical researches and numerical approaches are on the way to investigate these problems in various applied fields.

The principle theme of this thesis is coefficient identification in time-dependent PDEs. Although this subject has been studied extensively, many relevant questions are yet to be addressed, particularly in the context of nonlinear parameter identification:

Question 1. For the inverse problem of finding parameter x described by

$$F(x) := F(x, u(x)) = y \quad \text{where} \quad u \text{ depends on } x \text{ via a PDE},$$

the classical approach works with the operation equation $F(x) = y$ thus requires to construct the so-called parameter-to-state map $S : x \mapsto u$. We ask: Is there an alternative formulation that avoids the need of such conditional nonlinear map?

Question 2. As inverse problems are usually ill-posed, regularization methods are demanded to compute a good approximation to the true solution. These methods necessitate some structural conditions to be successfully implemented. The question arising here is: whether or not these conditions are fulfilled in real world physics problems?

Question 3. Lastly, can our work make a tangible contribution to the innovative applied science, especially in medical imaging?

In this spirit, we expose in the upcoming content our answer for each question, namely,

✧ **Chapter 2** refers to **Question 1.**

✧ **Chapter 3** refers to **Question 2.**

✧ **Chapters 5, 6** refer to **Question 3.**

We begin the thesis with **Chapter 1**, a review on theoretical preliminaries that are needed in the treatment of the whole thesis. **Chapter 4** introduces some basic concepts of Magnetic Particle Imaging in order to make the latter chapters more accessible.

The thesis reflects our work in:

- T. T. N. Nguyen, *Landweber-Kaczmarz for parameter identification in time-dependent inverse problems: All-at-once versus reduced version*, Inverse Problems, 35 (2019). Art. ID. 035009.

- B. Kaltenbacher, T. T. N. Nguyen, and O. Scherzer, *The tangential cone condition for some coefficient identification model problems in parabolic PDEs*, Springer volume on “Time-dependent Problems in Imaging and Parameter Identification”, (to appear).

- B. Kaltenbacher, T. T. N. Nguyen, A. Wald, and T. Schuster, *Parameter identification for the Landau-Lifshitz-Gilbert equation in magnetic particle imaging*, Springer volume on “Time-dependent Problems in Imaging and Parameter Identification”, (to appear).

- T. T. N. Nguyen, and A. Wald, *Numerical study of MPI: Parameter identification for the Landau-Lifshitz-Gilbert equation*, (on going).

Contents

Chapter 1

Theoretical preliminaries

In this chapter, we present some basic concepts of inverse problems leading to the theory of regularization methods in Section 1.1. Since the reconstruction in this thesis is mainly developed in Sobolev spaces, we collect in Section 1.2 some material facilitating the choice of appropriate function spaces: the Sobolev embeddings. In Section 1.3, we focus on existence of weak solutions to nonlinear parabolic PDEs, the essential equations studied in this research. Eventually, some well-known theorems from functional analysis will be recalled in Section 1.4 as they are prerequisite for the next chapters.

1.1 Inverse problems

1.1.1 Ill-posed problems

The terminology "inverse problem" indicates a problem that is inverse to a "forward problem" or "direct problem". The direct problem is formulated as the evaluation of a forward operator F acting on a known x in a space X. In the inverse problem, one tries to find the solution x from the equation $F(x) = y$ for given data y

- Direct problem: given x and F, evaluate $F(x)$
- Inverse problem: given y and F, solve $F(x) = y$ for x.

The formulation as an operator equation allows us to distinguish among finite, infinite-dimensional, linear or nonlinear problems.

A mathematical model is called "well-posed" in the sense of Hadamard if it satisfies the three following properties:

1. There exists a solution of the problem (existence).
2. There is at most one solution of the problem (uniqueness).

3. The solution depends continuously on the data (stability).

These requirements are not independent of each other. If a problem lacks of one of these three properties, we call it "ill-posed". While the direct problem is well-posed in most of the cases, the inverse problem is ill-posed due to its nature. The prototype of an inverse problem is an equation of the form

$$F(x) = y.$$

The ill-posed property challenges us in determining x from the above formulated equation since the solution may not exist (existence) or even if existing, information is missing if the problem has more than one solution (uniqueness). Moreover, if the stability criterion is violated, it is practically impossible to recover x because in reality, the data y always contains error from measurement and computation. Then, a decrease of the noise level does not ensure convergence of the computed solution to the true solution.

Therefore, special techniques, which are so-called *regularization*, and possibly additional priori information about the solution (e.g., source conditions) are required to successfully reconstruct a stable approximation of the solution.

1.1.2 Regularization theory

We consider the general model

$$F : X \to Y \qquad F(x) = y, \tag{1.1}$$

where F is a nonlinear operator acting between the Hilbert spaces (or Banach spaces) X and Y. If F is not surjective, it is reasonable to search for x such that $F(x)$ has minimal distance to y. Also we do not assume F to be injective; hence among the solutions, selecting the one with the minimal norm makes sense to the idea of a unique solution. We thus introduce the notions:

Definition 1.1.1. An element $x \in X$ is called

- *least-squares solution* of if (1.1)

$$\|F(x) - y\| = \min\{\|F(z) - y\| : z \in X\}$$

- *minimal-norm solution* of (1.1) if

$$\|x\| = \inf\{\|z\| : z \text{ is a least-squares solution of } (1.1)\}.$$

In general, existence of a least-squares solution is not guaranteed as the range of F might not be closed. This leads to the definition of the *Moore-Penrose generalized inverse* $F^\dagger$ in case of a linear operator F:

Definition 1.1.2. The *Moore-Penrose generalized inverse* is defined as

$$F^\dagger : \mathcal{D}(F^\dagger) := \mathcal{R}(F) + \mathcal{R}(F)^\perp \to \mathcal{R}(F^\dagger) := \mathcal{N}(F)^\perp$$

with $\mathcal{N}(F^\dagger) := \mathcal{R}(F)^\perp$.

Note that the Moore-Penrose generalized inverse is well-defined, thus it allows us to specify the unique minimal-norm solution

$$x^\dagger := F^\dagger y$$

for each $y \in \mathcal{D}(F^\dagger)$ in the linear case.

Once the concept of the best-approximate solution has been clarified, we shall construct *regularization methods*:

Definition 1.1.3. Let $R_\alpha : Y \to X$ be a family of continuous operators and $\alpha : \mathbb{R}^+ \times Y \to I$ for some index set I. The pair (R_α, α) is called a regularization method for (1.1) if

$$\lim_{\delta \to 0} \left(\sup\{\|R_{\alpha(\delta, y^\delta)} y^\delta - x^\dagger\| : y^\delta \in Y, \|y^\delta - F(x^\dagger)\| \le \delta\} \right) = 0$$

and

$$\lim_{\delta \to 0} \left(\sup\{\alpha(\delta, y^\delta) : y^\delta \in Y, \|y^\delta - F(x^\dagger)\| \le \delta\} \right) = 0$$

for all $x^\dagger \in \mathcal{D}(F)$ and noisy data y^δ.

In the following, we list out some regularization methods, each of these methods relies on different theory:

Tikhonov regularization bases on minimization

Iterative Landweber regularization as a steepest decent method

Iterative Newton-type methods work on linearized problems

Conjugate gradient method builds a Krylov subspace.

Remark 1.1.4. In the context of this thesis, formulation of the iterative Landweber and Landweber-Kaczmarz will be investigated on general nonlinear parabolic models.

Bibiliographical remark 1.1.5. For an insight into regularization theory for inverse problems, we refer to, e.g., [31, 69, 62, 108, 41].

1.1.3 Landweber-Kaczmarz iteration

Most of the iterative methods for regularizing the ill-posed inverse problem (1.1) base on the transformed fixed-point equation

$$x = x - F'(x)^*(F(x) - y^\delta),$$

which results from minimizing the least square misfit functional

$$\|F(x) - y^\delta\|^2 \to \min!$$

by taking steps in the direction of the negative of the gradient of the function. This motivates a fixed-point iteration known as the *Landweber iteration* [49]

$$x_{k+1} = x_k - \mu_k F'(x_k)^*(F(x_k) - y^\delta) \qquad k \in \mathbb{N}_0 \tag{1.2}$$

starting from an initial guess x_0, and the iterative process is driven toward the steepest descent.

In case of having a collection of operators $F = (F_0, \dots, F_{n-1}) : \bigcap_{i=0}^{n-1} \mathcal{D}(F_i) \subset X \to Y^n$ as well as data $y^\delta = (y_0^\delta, \dots, y_{n-1}^\delta)$, the Landweber-Kaczmarz method reads as

$$x_{k+1} = x_k - \mu_k F'_{j(k)}(x_k)^\star(F_{j(k)}(x_k) - y^\delta_{j(k)}) \qquad k \in \mathbb{N}_0 \tag{1.3}$$

with $j(k) = k - n\lfloor k/n \rfloor$, where $\lfloor k \rfloor$ is the largest integer lower or equal to k. It is recognizable that the Landweber-Kaczmarz method applies the Landweber iteration cyclically.

In (1.2) and (1.3), $F'(x)$ is the derivative of F at x and $F'(x)^*$ is its adjoint. F' does not necessarily need to be the Fréchet or Gâteaux derivative of F, it is just required to be some linear operator that is uniformly bounded in a neighborhood of the initial guess (see (1.9)).

Discrepancy principle

[62] suggests to terminate the Landweber iteration according to a *discrepancy principle*:

Definition 1.1.6.

$$\|F(x_{k_*}) - y^\delta\| \leq \tau\delta < \|F(x_k) - y^\delta\| \qquad 0 \leq k < k_* \tag{1.4}$$

for an appropriately chosen positive constant τ.

For the Landweber-Kaczmarz method, we define:

Definition 1.1.7. Let

$$x_{k+1} = x_k - w_k h_k, \qquad \text{where} \qquad h_k = \mu_k F'_{j(k)}(x_k)^*(F_{j(k)}(x_k) - y^\delta_{j(k)}) \tag{1.5}$$

and

$$w_k = \begin{cases} 1 & \text{if } \|F_{j(k)}(x_k) - y^\delta_{j(k)}\| \geq \tau\delta_{j(k)} \\ 0 & \text{otherwise} \end{cases} \tag{1.6}$$

then the iterations stop at the first index k_*, when $w_k = 0$ is reached in a full cycle

$$w_{k_*-i} = 0, \quad i = 0, \dots, n-1 \qquad \text{and} \qquad w_{k_*-n} = 1. \tag{1.7}$$

Here the discrepancy principle is combined with a loping strategy.

Since data y^δ is always contaminated by noise, this is a reasonable stopping rule as we require the residual $\|F_{j(k)}(x_k) - y^\delta_{j(k)}\|$ to be of the order of the data error via some sufficiently large $\tau > 2$ rather than hoping for residual smaller than δ_k, the noise level in the $j(k)$-th equation.

In (1.4) and (1.6), the constant τ in is chosen subject to the *tangential cone condition* (1.8). Additionally, the step size μ in (1.5) is derived from uniform boundedness of the derivative of the forward operator specified by (1.9).

Basic assumptions

Assumption 1.1.8.

1. The *tangential cone condition*

$$\begin{aligned} &\|F(\tilde{x}) - F(x) - F'(x)(\tilde{x} - x)\|_Y \leq c_{tc}\|F(\tilde{x}) - F(x)\|_Y \\ &\forall x, \tilde{x} \in \mathcal{B}_\rho(x_0) \end{aligned} \tag{1.8}$$

   holds for some small $c_{tc} < \frac{1}{2}$.

2. $F'(x)$ is uniformly bounded

$$\|F'(x)\|_{L(X,Y)} \leq C \qquad \forall x \in \mathcal{B}_\rho(x_0)) \tag{1.9}$$

   for some constant C.

These conditions [49] are imposed locally in $\mathcal{B}_\rho(x_0) \subseteq X$, the ball of center x_0 and radius ρ, to ensure convergence of the iterative regularization methods. They also guarantee that the iterates $x_k, 0 \leq k \leq k_*$ remain in $\mathcal{D}(F)$, which makes the iterative methods well-defined.

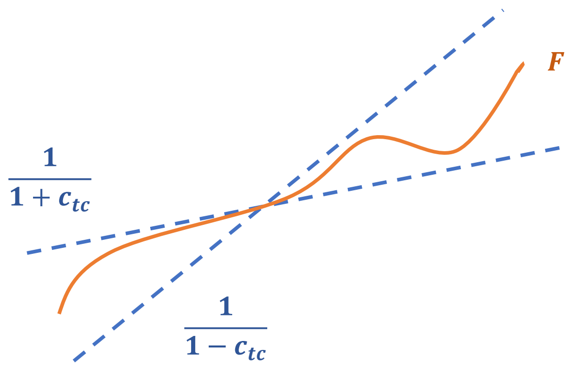

Figure 1.1: Tangential cone condition.

From (1.8), it follows immediately with the triangle inequality that

$$\frac{1}{1+c_{tc}}\|F'(x)(\tilde{x}-x)\| \leq \|F(\tilde{x})-F(x)\| \leq \frac{1}{1-c_{tc}}\|F'(x)(\tilde{x}-x)\|$$

for all $x, \tilde{x} \in \mathcal{B}_\rho(x_0)$. Graphically, the graph of $F(\tilde{x})$ for all $\tilde{x}$ in a neighborhood of point x should lie within the double cone, whose two generatrix lines are made of the tangent at x with slopes being scaled by the factor $\frac{1}{1+c_{tc}}$ and $\frac{1}{1-c_{tc}}$, respectively (Figure 1.1).

Note that (1.8) even implies unique existence of the minimal-norm solution:

Proposition 1.1.9. *Let (1.8) be fulfilled and (1.1) be solvable in $\mathcal{B}_\rho(x_0)$. Then a unique minimal-norm solution exists. It is characterized as the solution $x^\dagger$ of $F(x) = y$ in $\mathcal{B}_\rho(x_0)$ satisfying the condition*

$$x^\dagger - x_0 \in \mathcal{N}(F'(x^\dagger))^\perp.$$

Proof. [62, Proposition 2.1].

Remark 1.1.10. The tangential cone condition (1.8) will be examined for a series of time dependent benchmark inverse problems in Chapter 3.

Convergence analysis

Theorem 1.1.11. *Let the assumptions (1.8), (1.9) hold and let $k_* = k_*(\delta, y^\delta)$ be chosen according to the stopping rule (1.4). Moreover, we assume that $F(x) = y$ is solvable in in $\mathcal{B}_\rho(x_0)$. Then the Landweber iterates x_{k_*} converge to a solution of $F(x) = y$. If $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ for all $x \in \mathcal{B}_\rho(x^\dagger)$, then x_{k_*} converges to $x^\dagger$ as $\delta \to 0$.*

Proof. [49], [62, Theorem 2.6]

Theorem 1.1.12. *Let the assumptions* (1.8)*,* (1.9) *hold and let* $k_* = k_*(\delta, y^\delta)$ *be chosen according to the stopping rule* (1.6)*,* (1.7)*. Moreover, we assume that* $F(x) = y$ *is solvable in in* $\mathcal{B}_\rho(x_0)$*. Then the Landweber-Kaczmarz iterates* x_{k_*} *converge to a solution of* $F(x) = y$*. If* $\mathcal{N}(F_j'(x^\dagger)) \subset \mathcal{N}(F_j'(x))$ *for all* $x \in \mathcal{B}_\rho(x^\dagger), j = 1 \ldots, n-1$*, then* x_{k_*} *converges to* $x^\dagger$ *as* $\delta \to 0$*.*

Proof. [75], [62, Theorem 3.26].

Bibiliographical remark 1.1.13. For a more comprehensive look at this method (also other iterative methods) inlcluding convergence rates, we refer to [62, 45, 46, 75].

1.2 Sobolev embeddings

Definition 1.2.1. (Sobolev space) [32, Section 5.2.2]
The *Sobolev space*

$$W^{k,p}(\Omega)$$

consists of all locally integral functions $u : \Omega \to \mathbb{R}$ such that for each multiindex α with $|\alpha| \leq k, D^\alpha u$ exists in the weak sense and belongs to $L^p(\Omega)$.
The standard norm on $W^{k,p}(\Omega)$ is defined by

$$\begin{aligned}
\|u\|_{W^{k,p}(\Omega)} &:= \left(\sum_{|\alpha| \leq k} \int_\Omega |D^\alpha u|^p \, dx \right)^{\frac{1}{p}} \qquad 1 \leq p < \infty \\
\|u\|_{W^{k,\infty}(\Omega)} &:= \operatorname*{ess\,sup}_\Omega \sum_{|\alpha| \leq k} |D^\alpha u|^p.
\end{aligned}$$

Remark 1.2.2.
For $1 \leq p \leq \infty$: $W^{k,p}(\Omega)$ are Banach spaces.
For $1 < p < \infty$: $W^{k,p}(\Omega)$ are reflexive spaces.
For $p = 2$: $H^k(\Omega) := W^{k,p}(\Omega)$ are Hilbert spaces.

Theorem 1.2.3. *(Sobolev embedding)*
Let $\Omega \subset \mathbb{R}^d$ *be a bounded Lipschitz domain. Moreover, let* $1 \leq p < \infty$*, and let* k *be a nonnegative integer. Then the following embeddings exist and are continuous:*

- *for* $kp < d : W^{k,p}(\Omega) \hookrightarrow L^q(\Omega)$ *if* $1 \leq q \leq \frac{dp}{d-kp}$
- *for* $kp = d : W^{k,p}(\Omega) \hookrightarrow L^q(\Omega)$ *if* $1 \leq q < \infty$
- *for* $kp > d : W^{k,p}(\Omega) \hookrightarrow C(\overline{\Omega})$*.*

Proof. [32, Sections 5.6, 5.7]

Remark 1.2.4. The following embeddings will be frequently used

- for $d = 1$: $H^1(\Omega) \hookrightarrow C(\overline{\Omega})$
- for $d = 2$: $H^1(\Omega) \hookrightarrow L^q(\Omega)$ if $1 \leq q < \infty$
- for $d = 3$: $H^1(\Omega) \hookrightarrow L^6(\Omega), H^2(\Omega) \hookrightarrow C(\overline{\Omega})$.

The evolution problems in the next chapters, besides the space variable, involve also the time variable t, which takes a special role in reflecting the mathematical analysis. We here present a few useful assertions about abstract function spaces in a finite time interval $I := [0, T]$, whose values are in Banach spaces.

Definition 1.2.5. [32, Section 9.3.2]

(i) Given Banach space V, the space

$$L^p(I; V)$$

consists of all strongly measurable functions $u : I \to V$ with

$$\|u\|_{L^p(I;V)} := \left(\int_I \|u(t)\|_V^p \, dt \right)^{\frac{1}{p}} \qquad 1 \leq p < \infty$$
$$\|u\|_{L^\infty(I;V)} := \operatorname*{ess\,sup}_{t \in I} \|u(t)\|_V.$$

(ii) Given Banach space V, the space

$$C(I; V)$$

comprises all continuous functions $u : I \to V$ with

$$\|u\|_{C(I;V)} := \max_{t \in I} \|u(t)\|_V.$$

Definition 1.2.6. (Sobolev-Bochner space) [100, Section 7.1]
For V_1 a Banach space and V_2 a locally convex Banach space, $V_1 \subseteq V_2$, let us define the *Sobolev-Bochner space*

$$W^{1,p,q}(I; V_1, V_2) := \{u \in L^p(I; V_1); \frac{du}{dt} \in L^q(I; V_2)\} \qquad 1 \leq p, q \leq \infty,$$

which is equipped with the norm

$$\|u\|_{W^{1,p,q}(I;V_1,V_2)} := \|u\|_{L^p(I;V_1)} + \left\| \frac{du}{dt} \right\|_{L^q(I;V_2)}.$$

Remark 1.2.7. For $p = q = 2$, if V_1, V_2 are Hilbert spaces then so is $W^{1,2,2}(I; V_1, V_2)$.

A basic abstract setting for evolution problems relies on a *Gelfand triple* construction. It means that H is a Hilbert space identified with its dual: $H \cong H^*$; the embedding $V \hookrightarrow H$ is continuous and dense; consequently, $H \hookrightarrow V^*$ continuously.

Theorem 1.2.8. *(Embedding with Gelfand triple)*
Let $V \subseteq H \cong H^* \subseteq V^*$ be a Gelfand triple, and let $p^* = p/(p-1)$ be the conjugate exponent to p. Then $W^{1,p,p^*}(I; V, V^*) \hookrightarrow C(I; H)$ continuously.
Moreover, the following by-parts integration formula holds for any $u, v \in W^{1,p,p^*}(I; V, V^*)$ and any $t_1, t_2 \in I$

$$(u(t_2), v(t_2)) - (u(t_1), v(t_1)) = \int_{t_1}^{t_2} \left\langle \frac{du(t)}{dt}, v(t) \right\rangle_{V^*,V} + \left\langle u(t), \frac{dv(t)}{dt} \right\rangle_{V,V^*} dt.$$

Proof. [100, Lemma 7.2].

Remark 1.2.9. We will often employ the embedding $W^{1,2,2}(I; V, V^*) \hookrightarrow C(I, H)$.

Bibiliographical remark 1.2.10. For this section, we make reference to [32, 100, 109].

1.3 Evolution problems: Existence and uniqueness of solutions

In this section, we focus on the evolution governed by the abstract initial-value problem on a finite time interval

$$\frac{du}{dt} + F(t, u(t)) = f(t) \quad \text{for a.e. } t \in [0, T], \qquad u(0) = u_0, \tag{1.10}$$

where $f \in L^2(0, T; V^*), u_0 \in H, F : (0, T) \times V \to V^*$ is a nonlinear *Carathéodory mapping* (see below), and $V \subseteq H \cong H^* \subseteq V^*$ is a Gelfand triple with separable reflexive Banach space V.

Definition 1.3.1. (Carathéodory mapping) [109, Section 4.3], [100, Section 1.3]

A mapping $\varphi : I \times \mathbb{R}^j \to \mathbb{R}^k$ is a *Carathéodory mapping* if $\varphi(\cdot, u) : I \to \mathbb{R}^k$ is measurable for all $u \in \mathbb{R}^j$ and $\varphi(t, \cdot) : \mathbb{R}^j \to \mathbb{R}^k$ is continuous for a.e. $t \in I$.

Definition 1.3.2. (Nemytskii operator) [109, Section 4.3], [100, Section 1.3]
Let $I \subset \mathbb{R}^d$ be a bounded and measurable set, and let $\varphi := \varphi(t, u) : I \times \mathbb{R}^j \to \mathbb{R}^k$. The mapping given by

$$\Phi(u) = \varphi(\cdot, u(\cdot)),$$

which assigns to a function $u : I \to \mathbb{R}^j$ the function $z : I \to \mathbb{R}^k, z(t) = \Phi(t, u(t))$, is called a *Nemytskii operator* or *superposition operator*.

Remark 1.3.3. Matching to (1.10), we have $I := (0,T)$, V and V^* are respectively in place of $\mathbb{R}^j$ and $\mathbb{R}^k$, thus $\varphi := F : (0,T) \times V \to V^*$. This means F in (1.10) induces the Nemytkii operator (for which without causing any confusion, we use the same notation)

$$[F(\cdot, u)](t) = F(t, u(t)).$$

Example 1.3.4. The following mapping $u(\cdot) \to F(u)(\cdot)$ defines the Nemytskii operator

$$[F(u)](t) = (u(t))^3,$$

that occurs in Chapter 2, Section 2.7.

Our plan is to build a weak solution to (1.10) using a *Galerkin approximation* and introducing essential conditions that are necessary for the method.

Abstract Galerkin approximation

As V is separable, from the sequence of finite-dimensional subspaces

$$\forall k \in \mathbb{N} : V_k \subset V_{k+1} \subset V \quad \text{and} \quad \bigcup_{k\in\mathbb{N}} V_k \text{ is dense in } V,$$

we define the Galerkin approximation $u_k \in W^{1,2,2}(0,T; V_k, V_k^*)$ *(Construct approximate solution)* by

$$\forall v \in V_k, \forall^{\text{a.e.}} t \in [0,T] : \left\langle \frac{du(t)}{dt}, v \right\rangle_{V_k^*, V_k} + \langle F(t, u_k(t)), v\rangle_{V_k^*, V_k} = \langle f(t), v\rangle_{V_k^*, V_k}$$
$$u_k(0) = \langle u_0, v\rangle_{V_k^*, V_k}.$$

If $\{u_k\}_{k\in\mathbb{N}}$ exist *(Existence of approximate solutions)* and are uniformly bounded *(A-priori estimate)*, then passing to limits *(Limit passage)* we have $u_k \rightharpoonup u$, a solution to the initial-value problem (1.10).

We shall not give a comprehensive explanation of constructing the Galerkin approximation for weak solutions to (1.10), which is referred to [100]. Instead, we confine ourselves to summarizing some unique existence results of such solutions.

Theorem 1.3.5. *(Unique existence)*
Assume that

1. *for almost all $t \in (0,T)$, the mapping $F(t,.)$ is* pseudomonotone*, i.e., $F(t,.)$ is bounded and*

$$\left.\begin{aligned} \limsup_{k\to\infty} \langle F(t,u_k), u_k - u\rangle_{V^*,V} \le 0 \\ u_k \rightharpoonup u \end{aligned}\right\} \Rightarrow \begin{cases} \forall v \in V : \langle F(t,u), u - v\rangle_{V^*,V} \\ \le \liminf\limits_{k\to\infty} \langle F(t,u_k), u_k - v\rangle_{V^*,V} \end{cases}$$

2. *F is* semi-coercive*, i.e.,*

$$\forall v \in V, \forall^{a.e.} t \in (0,T) : \langle F(t,v), v\rangle_{V^*,V} \geq c_0 |v|_V^2 - c_1(t)|v|_V - c_2(t)\|v\|_H^2$$

for some $c_0 > 0, c_1 \in L^2(0,T), c_2 \in L^1(0,T)$ and some seminorm $|.|_V$ satisfying $\forall v \in V : \|v\|_V \leq c_{|.|}(|v|_V + \|v\|_H)$ for some $c_{|.|} > 0$

3. *the growth condition holds*

$$\exists \gamma \in L^2(0,T), \hbar : \mathbb{R} \to \mathbb{R} \text{ increasing} : \|F(t,v)\|_{V^*} \leq \hbar(\|v\|_H)(\gamma(t) + \|v\|_V)$$

4. *F satisfies a condition for uniqueness of the solution, e.g.,*

$$\forall u, v \in V, \forall^{a.e.} t \in (0,T) : \langle F(t,u) - F(t,v), u - v\rangle_{V^*,V} \geq -\rho(t)\|u - v\|_H^2$$

for some $\rho \in L^1(0,T)$.

Then (1.10) has a unique solution $u \in W^{1,2,2}(0,T;V,V^*)$.

Proof. [100, Theorems 8.27, 8.31].

Remark 1.3.6. These conditions will be verified for a class of inverse source problems in Chapter 2, Section 2.8.2.

Remark 1.3.7. [100, Chapters 8.1-8.5] analyzes the existence theory for nonlinear evolution equations in the more abstract framework $W^{1,p,p^*}(0,T;V,V^*)$ meaning general Lebesgue spaces instead of L^2 with respect to time as we are considering here.

Observation 1.3.8. In case F is linear, its boundedness implies the growth condition with $\hbar = \|F(t,\cdot)\|$, and semi-coercivity implies uniqueness condition with $\rho = c_2$ for $c_1 \leq 0$. The assumption on pseudomonotonicity to ensure week convergence of $F(\cdot, u_k)$ to $F(\cdot, u)$ when the approximation solution sequence u_k converges weakly to u, can be replaced by weak continuity of F which holds for F being linear bounded .

Alternatively, there are possibilities to obtain pseudomonotonicity from coercivity straightforwardly, also from monotonicity and continuity, or from strong continuity [36, Lemma 6.7], which fit very well for the linear case .

Theorem 1.3.9. *(Regularity in the autonomous case)*
Let $F : V \to V^*$ be pseudomonotone, semi-coercive, and let

$$\begin{aligned}
&u_0 \in V\\
&f \in L^2(0,T;H)\\
&F = F_1 + F_2 \quad \text{with} \quad F_1 = \varphi', \quad \varphi : V \to \mathbb{R} \text{ convex}\\
&\varphi(v) \geq c_0\|v\|_V^2 - c_1\|v\|_H^2, \qquad \|F_2(v)\|_H \leq C(1 + \|v\|_V)
\end{aligned}$$

for some $c_0 > 0$. Then there exists a solution $u \in W^{1,\infty,2}(0,T;V,H)$ to (1.10) in the autonomous case, i.e., F is independent of t.

Proof. [100, Theorems 8.16].

Remark 1.3.10. An extension to regularity in the non-autonomous case will be discussed in Chapter 2, Section 2.8.1.

Bibiliographical remark 1.3.11. In addition to [100], existence and uniqueness theory for concrete linear and quasilinear parabolic problems can be found, e.g., in the books [32, 81, 95].

Remark 1.3.12. An alternative approach to study the existence theory for evolution problems on Banach spaces is through the framework of *semigroups*. Several parabolic PDEs can be realized within this framework by combing the semigroup theory, e.g, from [32, 96] with elliptic results in [82]. Chapter 3 gives some existence proofs relying on this method.

1.4 Fundamental theorems

Theorem 1.4.1. *(Riesz Representation Theorem) [69, Theorem A.22]*
Let X be a Hilbert space. For every $x \in X$, the functional $f_x(y) := (y, x), y \in X$ defines a linear bounded mapping from X to $\mathbb{K}$, i.e., $f_x \in X^*$. Furthermore, for every $f \in X^*$ there exists exactly one $x \in X$ with $f(y) = (y, x), \forall y \in X$ and

$$\|f\|_{X^*} := \sup_{y \neq 0} \frac{|f(y)|}{\|y\|_X} = \|x\|_X.$$

Theorem 1.4.2. *(Dominated convergence theorem) [32, Appendix E, Theorem 4]*
Assume that the functions $\{f_m\}_{m=1}^{\infty}$ are integrable and $f_m \to f$ a.e. Suppose also that $|f_m| \leq g$ a.e. for some integrable function g. Then

$$\int_{\mathbb{R}^d} f_m \, dx \to \int_{\mathbb{R}^d} f \, dx.$$

Theorem 1.4.3. *(Contraction Mapping Principle) [69, Theorem A.59]*
Let $K \subset X$ be a closed subset of the Banach space X and $T : K \to K$ be a contraction*, i.e.,*

$$\exists c < 1 : \|T(x) - T(y)\| \leq c\|x - y\| \qquad \forall x, y \in K.$$

Then T has a unique fixed-point *$\tilde{x}$, i.e., $T(\tilde{x}) = \tilde{x}$.*

Now, let U, V denote real Banach spaces and $\mathcal{U}$ be an open subset of U. We define:

Definition 1.4.4. (Gâteaux derivative) [109, Section 2.6]
Let $F : \mathcal{U} \subset U \to V$, $u \in \mathcal{U}$ and $h \in U$ be given. Suppose there exists an operator $A \in L(U, V)$ such that

$$\lim_{t \to 0} \frac{F(u + th) - F(u)}{t} = Ah \qquad \forall h \in U.$$

Then F is said to be *Gâteaux differentiable* at u, and A is referred to the *Gâteaux derivative* of F at u.

Moreover:

Definition 1.4.5. (Fréchet derivative) [109, Section 2.6]
F is said to be *Fréchet differentiable* at u if

$$\frac{\|(F(u+h)-F(u)-Ah\|_V}{\|h\|_U} \to 0 \quad \text{as} \quad \|h\|_U \to 0.$$

The operator A is then called the *Fréchet derivative* of F at u.

Theorem 1.4.6. *(Chain rule) [109, Section 2.6]*
Let $F : \mathcal{U} \subset U \to V$ and $G : V \to Z$ be Fréchet differentiable at $u \in \mathcal{U}$ and at $F(u) \in V$, respectively. Then the composition $G \circ F : U \to Z$ is Fréchet differentiable at u, and

$$(G \circ F)'(u) = G'(F(u))F'(u).$$

Remark 1.4.7. (Chain rule 1)
The chain rule is attainable for Gâteaux differentiability if additionally, we have F continuous, the derivative $u \mapsto G'(u)v$ continuous and $u \mapsto G'(u)$ locally uniformly bounded.

Proof. From

$$\begin{aligned}
&(G{\circ}F)'(u)h - G'(F(u))F'(u)h \\
&\quad= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[G(F(u+\epsilon h)-G(F(u))\right] - G'(F(u))F'(u)h \\
&\quad= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big[G(F(u+\epsilon h)-G(F(u))-G'(F(u))(F(u+\epsilon h)-F(u))\Big] \\
&\qquad+\lim_{\epsilon\to 0}\Big[G'(F(u))\Big(\frac{F(u+\epsilon h)-F(u)}{\epsilon}-F'(u)h\Big)\Big] \\
&\quad= A_\epsilon + B_\epsilon,
\end{aligned}$$

we see, as $\epsilon \to 0$,

$$B_\epsilon = 0 \quad \text{if} \quad \begin{cases} G'(F(u)) & \text{is linear bouned} \\ F & \text{is Gâteaux differentiable} \end{cases},$$

and

$$\begin{aligned}
A_\epsilon &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\int_0^\lambda G'\big(F(u)+\lambda(F(u+\epsilon h)-F(u))\big)-G'\big(F(u)\big)\,d\lambda\,\big(F(u+\epsilon h)-F(u)\big)\\
&= \lim_{\epsilon\to 0}\int_0^\lambda G'\big(F(u)+\lambda(F(u+\epsilon h)-F(u))\big)-G'\big(F(u)\big)\,d\lambda\,\left(\frac{F(u+\epsilon h)-F(u)}{\epsilon}\right)\\
&= \lim_{\epsilon\to 0}\int_0^\lambda G'\big(F(u)+\lambda(F(u+\epsilon h)-F(u))\big)-G'\big(F(u)\big)\,d\lambda\,F'(u)h\\
&+ \lim_{\epsilon\to 0}\int_0^\lambda G'\big(F(u)+\lambda(F(u+\epsilon h)-F(u))\big)-G'\big(F(u)\big)\,d\lambda\left(\frac{F(u+\epsilon h)-F(u)}{\epsilon}-F'(u)h\right)\\
&= 0 \quad \text{if} \quad \begin{cases} F & \text{is Gâteaux differentiable and continuous}\\ G & \text{is Gâteaux differentiable}\\ u\mapsto G'(u)v & \text{is continuous}\\ u\mapsto G'(u) & \text{is locally uniformly bounded}\end{cases}.
\end{aligned}$$

Here we apply Lebesgue's Dominated Convergence Theorem. □

Remark 1.4.8. (Chain rule 2)
If G is Fréchet differentiable, F is Gâteaux differentiable and continuous then the chain rule holds.
In this case, one can write

$$\begin{aligned}
A_\epsilon &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\,o\big(\|F(u+\epsilon h)-F(u)\|\big)\\
&= \lim_{\epsilon\to 0}\frac{\|F(u+\epsilon h)-F(u)\|}{\epsilon}\,\lim_{\epsilon\to 0}O\big(\|F(u+\epsilon h)-F(u)\|^\alpha\big)\\
&= 0
\end{aligned}$$

for some $\alpha > 0$.

A helpful inequality will be used in some energy estimates:

Theorem 1.4.9. *(Grönwall's inequality) [28, Theorem 1, Corollary 3]*
Let x, Φ, χ be real continuous functions defined on $[a, b]$, and $\chi(t) \geq 0, \forall t \in [a, b]$. Assume that

$$x(t) \leq \Phi(t) + \int_a^t \chi(s)x(s)\,ds \qquad \forall t \in [a, b].$$

Then

$$x(t) \leq \Phi(t) + \int_a^t \Phi(s)\chi(s)\exp\left(\int_s^t \chi(\xi)\,d\xi\right)ds \qquad \forall t \in [a, b].$$

Moreover, if Φ is constant, we have

$$x(t) \leq \Phi\exp\left(\int_a^t \chi(\xi)\,d\xi\right) \qquad \forall t \in [a, b].$$

Part I

ALL-AT-ONCE AND REDUCED SETTINGS

Chapter 2

Parameter identification in time-dependent inverse problems

In this chapter, we consider a general time-space system, whose model operator and observation operator are locally Lipschitz continuous, over a finite time horizon and parameter identification by using Landweber-Kaczmarz regularization. The problem is investigated in two different modeling settings: An *all-at-once* and a *reduced* version, together with two observation scenarios: continuous and discrete observations. Segmenting the time line into several subintervals leads to the idea of applying the Kaczmarz method. A loping strategy is incorporated into the method to yield the loping Landweber-Kaczmarz iteration.

The chapter is outlined as follows: Section 2.1 states the motivation. Section 2.2 introduces the general model for the state-space problem. In the next two sections, we present the all-at-once and reduced formulations. Section 2.5 compares the two modeling settings and discusses time-dependent parameter identification. Section 2.6 is dedicated to deriving the algorithm and its convergence. Section 2.7 examines a class of inverse source problems. Finally, we conclude the work and sketch some ideas for potential research.

2.1 Challenge with cubic nonlinearity

As a motivating prototype example, we consider parameter identification from the following system

$$\dot{u} = \Delta u - u^3 + \theta \qquad (t,x) \in (0,T)\times\Omega \tag{2.1}$$

$$u(0) = u_0 \qquad x \in \Omega \tag{2.2}$$

$$y = \mathcal{C}u \qquad (t,x) \in (0,T)\times\Omega \tag{2.3}$$

with $\Omega \subset \mathbb{R}^d$. $\dot{u}$ denotes the first order time derivative of u and the right hand side includes the nonlinearity $\Phi(u) := u^3$. This equation is equipped with the initial condition (2.2) and possibly further Dirichlet or Neumann boundary conditions on $(0,T) \times \partial\Omega$. In (2.3), the measured data y is obtained from a linear observation, this means $\mathcal{C}$ is a linear operator. In this evolution system, u and θ are two unknowns.

Parabolic PDEs with cubic power nonlinearity arise in many applications, and we selectively mention some falling into this category

- $\Phi(u) = u(1-u^2)$: Ginzburg-Landau equations of superconductivity [15], Allen-Cahn equation for phase separation process in a binary metallic alloy [1, 92], Newell-Whitehead equation for convection of fluid heated from below [38].
- $\Phi(u) = u^2(1-u)$: Zel'dovich equation in combustion theory [38].
- $\Phi(u) = u(1-u)(u-\alpha), 0 < \alpha < 1$: Fisher's model for population genetics [95], Nagumo equation for bistable transmission lines in electric circuit theory [89].

To the best of our knowledge, the state-of-the-art research on this type of problem in Sobolev space framework is limited to the power $\gamma \leq 1 + \frac{4}{d}, d \geq 3$, where d is the space dimension [58]. The reason for this constraint is that, the growth conditions used in proving well-definedness and differentiability of the forward operators prevent higher nonlinearity. Inspired by [58], in this study, staying in the Sobolev space framework, we aim at increasing the nonlinearity to the power of 3, or even higher, in order to be compatible with those applications. For this purpose, we plan to construct appropriate function spaces for the forward operators and impose on them relevant assumptions. The main assumption we rely on is basically the local Lipschitz continuity condition. Recently, some authors [110, 50] also employ local Lipschitz continuity as the key ingredient for the research on backward parabolic problems, but rather in a semigroup framework than in the Sobolev space framework, e.g., of [100], which we rely on to get even somewhat weaker conditions.

The parabolic equation occurring in problem (2.1)-(2.3) is semilinear. In the book by Tröltzsch [109], well-posedness and differentiability of the control-to-state map are investigated for optimal control of semilinear parabolic equations. In this context, we wish to point to the series of papers [20, 21, 22] by Casas, Ryll and Tröltzsch, where they consider optimal control problems for semilinear parabolic PDEs with cubic nonlinearities. In particular, an issue addressed also there was the fact that the solution spaces have to be chosen appropriately for proving differentiability of the control-to -state map; it was (a subset of) $L^\infty((0,T) \times \Omega)$ there, which was probably also chosen in view of the fact that state constraints were taken into account. Motivated by this work, we discuss here these relevant questions for the parameter-to-state map of semilinear diffusion systems in the context of inverse source problems. Beyond increasing nonlinearity as compared to [58], this study provides explicit formulas for Hilbert space adjoints that makes our method more computationally efficient, and Gâteaux as well as Fréchet derivatives obtained in this work yield more knowledge on the properties of the nonlinear model and on the convergence performance of iterative methods.

Being more realistic than the model (2.1)-(2.3) with observation on all of $(0, T)$, in practice the unknown parameter is recovered from experimental techniques that in some cases, limit the measurement only to some particular time points, see, e.g., [97, 13] for just two out of many examples from material science and system biology, respectively. Therefore, beside the continuous observation, we also desire to cover the discrete observation case in this study.

2.2 Abstract nonlinear parameter identification problems

We consider the following state-space system

$$\dot{u}(t) = f(t, u(t), \theta) \quad t \in (0, T) \tag{2.4}$$

$$u(0) = u_0(\theta) \tag{2.5}$$

$$y(t) = g(t, u(t), \theta) \quad t \in (0, T) \tag{2.6}$$

or

$$y_i = g_i(u(t_i), \theta) \qquad i = 1 \dots n \tag{2.7}$$

on $\Omega \subset \mathbb{R}^d$, where f is a nonlinear function and additional observation data y or y_i are obtained from continuous or discrete measurement as in (2.6) or (2.7), respectively. In the general case, g, g_i may be linear or nonlinear. In particular, observations may be partial only, such as boundary traces on $\partial\Omega$ or a part of it.

The model operator and observation operators map between the function spaces

$$f : (0, T) \times V \times \mathcal{X} \to V^* \tag{2.8}$$

$$g : (0, T) \times V \times \mathcal{X} \to Y \quad \text{or} \quad g_i : V \times \mathcal{X} \to Y, \tag{2.9}$$

where $\mathcal{X}, Y, H$ and V are separable Hilbert spaces and $V \subseteq H \subseteq V^*$ form a Gelfand triple. Moreover, we assume that f, g meet the Caratheodory mapping condition (see Chapter 1, Definition 1.3.1).

The initial condition is supposed to map to the sufficiently smooth image space

$$u_0 : \mathcal{X} \to V \tag{2.10}$$

as a condition to attain some regularity results for the solution to the problem (2.4)-(2.5).

Between V and V^*, the Riesz isomorphism

$$I : V^* \to V, \quad \langle u^*, Iv^* \rangle_{V^*,V} = (u^*, v^*)_{V^*}$$

and

$$\tilde{I} : V^* \to V, \quad (\tilde{I}u^*, v)_V = \langle u^*, v \rangle_{V^*,V}$$

are used to derive the adjoints. $\tilde{I}$ as defined above exists as one can choose, for example, $\tilde{I} = D^{-1}$, where D is the Riesz isomorphism

$$D : V \to V^*, \quad \langle Du, v \rangle_{V^*,V} = (u, v)_V .$$

Here, $(.,.)$ and $\langle .,. \rangle$ with the subscripts indicate the inner products and the dual parings, respectively . The notations D and I refer to the spatial differential and integration operators in the context of elliptic differential equations. We also distinguish the superscript *, the Banach space adjoint, and *, the Hilbert space adjoint, which is an ingredient for the iterative methods considered here.

For fixed θ, f and g as defined above induce Nemytskii operators (see Chapter 1, Definition 1.3.2) on the function space $\mathcal{U}$. This function space will be according to which problem setting is being dealt with, i.e., all-at-once or reduced setting. However, they map into the same image space $\mathcal{W}$ and observation space $\mathcal{Y}$

$$\mathcal{W} = L^2(0,T;V^*), \qquad \mathcal{Y} = L^2(0,T;Y), \tag{2.11}$$

which are Hilbert spaces. Therefore, we can investigate the problem in the Hilbert space framework, provided that the corresponding argument spaces of the forward operators in the two settings are Hilbert spaces as well.

We now discuss the state space $\mathcal{U}$. The function space proposed in [58] $\mathcal{U} = W^{1,2,2}(0,T;V,V^*) := \{u \in L^2(0,T;V) : \dot{u} \in L^2(0,T;V^*)\}$ with $V = H_0^1(\Omega)$ is not suitable for the cubic nonlinearity in (2.1) Also later on, for differentiablity of the forward operator relying on local Lipschitz continuity (see Propositions 2.3.1 and 2.4.2), we need functions in $\mathcal{U}$, whose values are in V, to be essentially bounded in $(0,T)$. For this reason, $L^\infty(0,T;V)$ appears to be an appropriate choice for the state space.

In this work, we restrict ourselves to the Hilbert space framework to avoid additional technicalities in general Banach spaces. The convergence analysis for regularization theory in Banach space, e.g., in [104, 67, 103] requires the preimage space to be reflexive and thus would not be applicable with the choice $\mathcal{U} = L^\infty(0,T;V)$. These facts together with the observation $H^1(0,T) \hookrightarrow L^\infty(0,T)$ render $\mathcal{U} = H^1(0,T;V)$ as in (2.18) a suitable candidate. This state space will be used for the all-at-once setting in Section 2.3.

Independently, in the reduced setting in Section 2.4, the state space just plays the role of an intermediate space since the forward operator is formulated to map directly between the parameter space and the observation space: $F : \mathcal{X} \to \mathcal{Y}, F(\theta) = y$. Hence, as long as $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces, we can stay with Hilbert space regularization theory even though $\tilde{\mathcal{U}} = W^{1,2,2}(0,T;V,V^*) \cap L^\infty(0,T;V)$ defined in (2.31) is a non-Hilbert state space.

In the context of optimal control for reaction-diffusion equations, the group of Casas, Ryll and Tröltzsch published several results focusing on the Schlögl and FitzHugh-Nagumo system [20, 21, 22], where they prove differentiablity of the control-to-state map also basing on local Lipschitz continuity and a prerequisite L^∞-approach for the cubic nonlinear reaction term. In this concrete inverse source problem (2.1)-(2.3), the parameter-to-state map appears to be relevant with the control-to-state map in the distributed control problem. These authors establish energy estimates in $W^{1,2,2}(0,T;V,V^*)\cap L^\infty((0,T)\times\Omega)$ [109, Theorem 5.9], [20, Lemma 2.4]. We, however in this chapter, evaluate them in $W^{1,2,2}(0,T;V,V^*)\cap L^\infty(0,T;V)$, as we consider here the model operator, which later induces the Nemytskii operator, $f(\theta):(0,T)\times V$ rather than there $f=\Delta-(\cdot)^3:(0,T)\times\Omega$. Besides, techniques to prove well-definedness of the parameter- (control)-to-state map are different.

In reality, we do not have access to the exact data. The experimental data always contains some noise of which we assume to know just the noise level. The noise perturbing the system is present both on the right hand side of the model equation and in the observation, and is denoted respectively by w^δ and z^δ. When the measurement is a collection of data at discrete observation time points, the corresponding noise added to each observation is z_i^δ. Altogether, we can formulate the noisy system

$$\dot{u}(t)=f(t,u(t),\theta)+w^\delta(t)\quad t\in(0,T) \tag{2.12}$$

$$u(0)=u_0(\theta) \tag{2.13}$$

$$y^\delta(t)=g(t,u(t),\theta)+z^\delta(t)\quad t\in(0,T) \tag{2.14}$$

or

$$y_i^\delta=g_i(u(t_i),\theta)+z_i^\delta\qquad i=1\dots n. \tag{2.15}$$

Correspondingly, those additive noise terms live in the function spaces

$$w^\delta\in\mathcal{W},\qquad z^\delta\in\mathcal{Y},\qquad z_i^\delta\in Y\qquad i=1\dots n \tag{2.16}$$

and are supposed to satisfy

$$\|w^\delta\|_\mathcal{W}\le\delta_w,\qquad \|z^\delta\|_\mathcal{Y}\le\delta_z,\qquad \|z_i^\delta\|_Y\le\delta_i\qquad i=1\dots n$$

with the noise levels $\delta_w,\delta_z,\delta_i>0$.

In this chapter, we investigate both Landweber and Landweber-Kaczmarz methods. The key difference between these two regularization methods is that in the Landweber method, we use one fixed forward operator; while in the Landweber-Kaczmarz method, different forward operators are applied in a cyclic manner (Chapter 1, Section 1.1.3). This feature makes Kaczmarz methods especially attractive for problems with large datasets or high dimensional model operator. In particular, Kaczmarz method does not need to finish operating the whole system to update one iteration like in the Landweber, instead, it just successively sweeps through each of the equations in the system; and that all the equations build a so-called a collection of the forward operators. Motivated by the Kaczmarz scheme, in order to construct a collection of forward operators, we either segment the time line into several subintervals or use the discrete observations. The Landweber-Kaczmarz method relying those strategies are particularly studied in Remarks 2.3.5, 2.3.6, 2.3.7 for the all-at-once setting and in Remark 2.38 for the reduced setting; visualization for these methods is shown in the numerical section.

This study does not attempt to research uniqueness of the exact solution $(u^\dagger, \theta^\dagger)$. Nevertheless, we refer to the book [62], which exposes some general results for this important question based on the tangential cone condition together with the assumption of a trivial null space of $F'(x^\dagger)$ in some neighborhood of $x^\dagger$. To verify the later condition in some concrete time-space problems, one can find detailed discussions, e.g., in the book by Isakov [56].

2.3 All-at-once formulation

In this section, we recast the system into a form which allows solving both state u and parameter θ simultaneously. This formulation shows practical benefits, that will be discussed in detail in Section 2.5.1. We define the forward operator

$$\mathbb{F} : \mathcal{U} \times \mathcal{X} \to \mathcal{W} \times V \times \mathcal{Y}, \qquad \mathbb{F}(u, \theta) = \begin{pmatrix} \dot{u} - f(., u, \theta) \\ u(0) - u_0(\theta) \\ g(., u, \theta) \end{pmatrix}, \tag{2.17}$$

then the system can be written as the nonlinear operator equation

$$\mathbb{F}(u, \theta) = \mathbb{Y} = (0, 0, y).$$

This fits into the Hilbert space framework by setting the function space for the state u to

$$\mathcal{U} = H^1(0, T; V) \tag{2.18}$$

and $\mathcal{W}, \mathcal{Y}$ are as in (2.11). On the space $\mathcal{U}$, we employ

$$(u, v)_{\mathcal{U}} = \int_0^T (\dot{u}(t), \dot{v}(t))_V \, dt + (u(0), v(0))_V \tag{2.19}$$

as an inner product, which induces a norm being equivalent to the standard norm $\sqrt{\int_0^T \|u(t)\|_V^2 + \|\dot{u}(t)\|_V^2 dt}$. This is the result of well-posedness in $\mathcal{U}$ of the first order ODE $\dot{u}(t) = f(t), t \in (0, T), u(0) = u_0$ for $f \in L^2(0, T; V), u_0 \in V$.

The operator $\mathbb{F}$ is well-defined by additionally imposing boundedness and local Lipschitz continuity on the functions inducing the Nemytskii operators. Differentiability of $\mathbb{F}$ also follows from the latter condition.

Proposition 2.3.1. *Let the Caratheodory mappings f, g be:*

(A1) Gâteaux differentiable with respect to their second and third arguments for almost all $t \in (0, T)$ with linear and continuous derivatives

(A2) locally Lipschitz continuous in the sense

$$\begin{aligned}&\forall M \geq 0, \exists L(M) \geq 0, \forall^{a.e.} t \in (0,T):\\ &\qquad \|f(t,v_1,\theta_1) - f(t,v_2,\theta_2)\|_{V^*} \leq L(M)(\|v_1 - v_2\|_V + \|\theta_1 - \theta_2\|_{\mathcal{X}}),\\ &\qquad\qquad \forall v_i \in V, \theta_i \in \mathcal{X} : \|v_i\|_V, \|\theta_i\|_{\mathcal{X}} \leq M, i = 1,2\end{aligned}$$

and satisfy the boundedness condition, i.e., $f(.,0,0) \in \mathcal{W}$.
The same is assumed to hold for g. Moreover, let u_0 also be Gâteaux differentiable.

Then $\mathbb{F}$ defined by (2.17) is Gâteaux differentiable on $\mathcal{U} \times \mathcal{X}$ and its derivative is given by

$$\mathbb{F}'(u,\theta) = \begin{pmatrix} (\frac{d}{dt} - f'_u(.,u,\theta)) & -f'_\theta(.,u,\theta) \\ \delta_0 & -u'_0(\theta) \\ g'_u(.,u,\theta) & g'_\theta(.,u,\theta) \end{pmatrix}. \tag{2.20}$$

Proof. We show Gâteaux differentiability of f at an arbitrary point $(u,\theta) \in \mathcal{U} \times \mathcal{X}$. Since f is Gâteaux differentiable and its derivative $(u,\theta) \mapsto f'(t,u,\theta)[v,\xi]$ is continuous by (A1), without loss of generality we consider the direction (v,ξ) lying in a unit ball, $\epsilon \in (0,1]$ and have

$$\frac{1}{\epsilon}\|f(.,u+\epsilon v,\theta+\epsilon\xi) - f(.,u,\theta) - \epsilon f'_u(.,u,\theta)v - \epsilon f'_\theta(.,u,\theta)\xi\|_{\mathcal{W}} = \left(\int_0^T r_\epsilon(t)^2 dt\right)^{\frac{1}{2}},$$

where

$$\begin{aligned} r_\epsilon(t) = \Bigg\| \int_0^1 \Big(&(f'_u(t,u(t)+\lambda\epsilon v(t),\theta+\lambda\epsilon\xi) - f'_u(t,u(t),\theta))v(t) \\ &+ (f'_\theta(t,u(t)+\lambda\epsilon v(t),\theta+\lambda\epsilon\xi) - f'_\theta(t,u(t),\theta))\xi \Big) d\lambda \Bigg\|_{V^*}. \end{aligned} \tag{2.21}$$

From local Lipschitz continuity and Gâteaux differentiability of f, we deduce, by choosing $M = \max\{\sqrt{2T}, \sqrt{2}\}(\|u\|_{\mathcal{U}} + 1) + \|\theta\|_{\mathcal{X}} + 1$,

$$\begin{aligned}\|f'_u(t,u(t),\theta)w\|_{V^*} &= \lim_{\epsilon\to 0}\left\|\frac{f(t,u(t)+\epsilon w,\theta) - f(t,u(t),\theta)}{\epsilon}\right\|_{V^*}\\ &\leq L(M)\|w\|_V.\end{aligned}$$

Continuity of the embedding $\mathcal{U} \hookrightarrow C(0,T;V)$ is invoked above. Indeed, for any $t \in (0,T)$

$$\|u(t)\|_V \leq \int_0^T \|\dot{u}(s)\|_V ds + \|u(0)\|_V \leq \max\{\sqrt{2T}, \sqrt{2}\}\|u\|_{\mathcal{U}}.$$

As an immediate consequence

$$\|f'_u(t, u(t) + \lambda\epsilon v(t), \theta + \lambda\epsilon\xi)\|_{V\to V^*} \le L(M)$$
$$\|f'_\theta(t, u(t) + \lambda\epsilon v(t), \theta + \lambda\epsilon\xi)\|_{\mathcal{X}\to V^*} \le L(M)$$

for any $\lambda \le 1, \epsilon \le 1$. Substituting all the operator norm estimates into $r_\epsilon(t)$, we obtain

$$r_\epsilon(t) \le 2L(M)\left(\|v(t)\|_V + \|\xi\|_{\mathcal{X}}\right) := \bar{r}(t), \tag{2.22}$$

which is a square integrable function. Since f is continuously differentiable, $r_\epsilon \to 0$ as $\epsilon \to 0$ for almost all every $t \in (0,T)$, applying Lebesgue's Dominated Convergence Theorem yields Gâteaux differentiability of f. Differentiability of g is analogous.

Obviously, continuity of the derivative $(u,\theta) \mapsto f'(u,\theta)[v,\xi]$ inherits from the operator inducing it

$$\begin{aligned}
&\epsilon_n := \|f'(u_n,\theta_n)[v,\xi] - f'(u,\theta)[v,\xi]\|_{\mathcal{W}^*}\\
&= \left(\int_0^T \|\big(f'_u(t,u_n,\theta_n) - f'_u(t,u,\theta)\big)v(t) + \big(f'_\theta(t,u_n,\theta_n) - f'_\theta(t,u,\theta)\big)\theta\|^2_{V^*}\,dt\right)^{1/2}\\
&\le \sup_{0\le t\le T}\|f'(t,u_n,\theta_n) - f'(t,u,\theta)\|_{V\times\mathcal{X}\to V^*\times V^*}\left[\left(\int_0^T \|v(t)\|_V^2\,dt\right)^{1/2} + \sqrt{T}\|\xi\|_{\mathcal{X}}\right].
\end{aligned}$$

Then $\epsilon_n \to 0$ as $n \to \infty$. Also, local uniform boundedness of $f'(\cdot,\cdot)$ is obvious. The same holds for g.

Also, well-definedness of f, g can be deduced from local Lipschitz continuity and the boundedness condition. □

Remark 2.3.2. If f, g and u_0 are Fréchet differentiable then so is $\mathbb{F}$.
This assertion follows from the observation

$$\mathbb{F} \text{ is Fréchet differentiable if } \sup_{\substack{\|v\|_{\mathcal{U}}\le 1\\ \|\xi\|_{\mathcal{X}}\le 1}} \int_0^T r_\epsilon^2(t)dt \to 0 \text{ as } \epsilon \to 0$$

for r_ϵ as in (2.21) (and likewise with g in place of f), or

$$\begin{aligned}
\sup_{\substack{\|v\|_{\mathcal{U}}\le 1\\ \|\xi\|_{\mathcal{X}}\le 1}} \int_0^T r_\epsilon^2(t)dt &\le \int_0^T \sup_{\substack{\|v\|_{\mathcal{U}}\le 1\\ \|\xi\|_{\mathcal{X}}\le 1}} r_\epsilon^2(t)dt\\
&\le \int_0^T \sup_{\substack{\|v(t)\|_V\le c\\ \|\xi\|_{\mathcal{X}}\le 1}} r_\epsilon^2(t)dt \to 0 \text{ as } \epsilon \to 0
\end{aligned}$$

with $c := \max\{\sqrt{2T}, \sqrt{2}\}$.
If f, g are Fréchet differentiable, this convergence is attainable due to

$$\begin{aligned}\sup_{\substack{\|v(t)\|_V \le c\\ \|\xi\|_{\mathcal{X}} \le 1}} r_\epsilon^2(t) &\le \sup_{\substack{\|w\|_V \le c\\ \|\xi\|_{\mathcal{X}} \le 1}} r_\epsilon^2(t) \to 0 \text{ as } \epsilon \to 0\\ \sup_{\substack{\|v(t)\|_V \le c\\ \|\xi\|_{\mathcal{X}} \le 1}} r_\epsilon^2(t) &\le \sup_{\substack{\|v(t)\|_V \le c\\ \|\xi\|_{\mathcal{X}} \le 1}} 2L(M)\left(\|v(t)\|_V + \|\xi\|_{\mathcal{X}}\right) \le 2L(M)(1+c).\end{aligned}$$

Remark 2.3.3.

- The local Lipschitz condition (A2) is weaker than the one used in [50], as we only have the V^*-norm on the left hand side of the Lipschitz condition.
- Note that differentiablity of $\mathbb{F}$ can be interpreted on the stronger image space $C(0,T;V^*)$ since $r(.)$ is not only square integrable but also uniformly bounded with respect to time (provided that local Lipschitz continuity is fulfilled for all $t \in (0,T)$ instead of for almost all $t \in (0,T)$).
- Another idea could be a weakening of the Lipschitz continuity condition to

  $$\begin{aligned}&\forall M \ge 0, \exists L(M) \ge 0, \gamma \in L^2(0,T), \forall^{a.e.} t \in (0,T):\\ &\qquad \|f(t,v_1,\theta_1) - f(t,v_2,\theta_2)\|_{V^*} \le L(M)\gamma(t)(\|v_1 - v_2\|_V + \|\theta_1 - \theta_2\|_{\mathcal{X}}),\\ &\qquad\qquad \forall v_i \in V, \theta_i \in \mathcal{X} : \|v_i\|_V, \|\theta_i\|_{\mathcal{X}} \le M, i = 1,2,\end{aligned}$$

  then square integrability in time can be transferred from $\|v(.)\|_V$ to $\gamma(.)$.

We now derive the Hilbert space adjoint for $\mathbb{F}'(u,\theta)$.

Proposition 2.3.4. *The Hilbert space adjoint of $\mathbb{F}'(u,\theta)$ is given by*

$$\begin{aligned}&\mathbb{F}'(u,\theta)^\star : \mathcal{W} \times V \times \mathcal{Y} \to \mathcal{U} \times \mathcal{X}\\ &\mathbb{F}'(u,\theta)^\star = \begin{pmatrix} (\frac{d}{dt} - f'_u(.,u,\theta))^\star & \delta_0^\star & g'_u(.,u,\theta)^\star \\ -f'_\theta(.,u,\theta)^\star & -u'_0(\theta)^\star & g'_\theta(.,u,\theta)^\star \end{pmatrix},\end{aligned} \tag{2.23}$$

where

$$
\begin{aligned}
&g'_u(.,u,\theta)^\star z = u^z \\
&g'_\theta(.,u,\theta)^\star z = \int_0^T g'_\theta(t,u(t),\theta)^* z(t)dt \\
&\delta_0^\star h = u^h \\
&u'_0(\theta)^\star = u'_0(\theta)^* \tilde{I}^{-1} \\
&\big(\frac{d}{dt} - f'_u(.,u,\theta)\big)^\star w = u^w + \int_0^t \tilde{I}^* I w(s)ds \\
&f'_\theta(.,u,\theta)^\star w = \int_0^T f'_\theta(t,u(t),\theta)^* I w(t)dt,
\end{aligned}
$$

u^z, u^w, u^h *solve*

$$
\begin{cases}
\ddot{u}^z(t) = -\tilde{I} g'_u(t,u(t),\theta)^* z(t) \quad t \in (0,T) \\
\dot{u}^z(T) = 0, \quad \dot{u}^z(0) - u^z(0) = 0
\end{cases}
$$

$$
\begin{cases}
\ddot{u}^w(t) = \tilde{I} f'_u(t,u(t),\theta)^* I w(t) \quad t \in (0,T) \\
\dot{u}^w(T) = 0, \quad \dot{u}^w(0) - u^w(0) = 0
\end{cases}
$$

$$
\begin{cases}
\ddot{u}^h(t) = 0 \quad t \in (0,T) \\
\dot{u}^h(T) = 0, \quad \dot{u}^h(0) - u^h(0) = -h,
\end{cases}
$$

and

$$
\begin{aligned}
&f'_u(t,u(t),\theta)^* : V^{**} = V \to V^*, \quad && f'_\theta(t,u(t),\theta)^* : V^{**} = V \to \mathcal{X}^* = \mathcal{X} \\
&g'_u(t,u(t),\theta)^* : Y^* = Y \to V^*, \quad && g'_\theta(t,u(t),\theta)^* : Y^* = Y \to \mathcal{X}^* = \mathcal{X} \\
&u'_0(\theta)^* : V^* \to \mathcal{X}^* = \mathcal{X}
\end{aligned}
$$

are the respective Banach and Hilbert space adjoints.

Proof. The adjoints of the derivatives with respect to u can be established as follows.

For any $z \in \mathcal{Y}$ and $v \in \mathcal{U}$

$$
\begin{aligned}
&(z, g_u'(., u, \theta)v)_{\mathcal{Y}} \\
&\quad = \int_0^T (z(t), g_u'(t, u(t), \theta)v(t))_Y \, dt \\
&\quad = \int_0^T \langle g_u'(t, u(t), \theta)^* z(t), v(t) \rangle_{V^*, V} \, dt \\
&\quad = \int_0^T \left(\tilde{I} g_u'(t, u(t), \theta)^* z(t), v(t) \right)_V dt \\
&\quad = \int_0^T (-\ddot{u}^z(t), v(t))_V \, dt \\
&\quad = \int_0^T (\dot{u}^z(t), \dot{v}(t))_V \, dt + (u^z(0), v(0))_V - (\dot{u}^z(T), v(T))_V + (\dot{u}^z(0) - u^z(0), v(0))_V \\
&\quad = (u^z, v)_{\mathcal{U}} \, .
\end{aligned}
$$

For any $w \in \mathcal{W}$ and $v \in \mathcal{U}$

$$
\begin{aligned}
&\left(w, (\frac{d}{dt} - f_u'(., u, \theta))v \right)_{\mathcal{W}} \\
&\quad = \int_0^T (w(t), -f_u'(t, u(t), \theta)v(t))_{V^*} \, dt + \int_0^T \left(w(t), \frac{d}{dt} v(t) \right)_{V^*} dt \\
&\quad = \int_0^T \langle -f_u'(t, u(t), \theta)^* I w(t), v(t) \rangle_{V^*, V} \, dt + \int_0^T \left(I w(t), \tilde{I} \frac{d}{dt} v(t) \right)_V dt \\
&\quad = \int_0^T \left(-\tilde{I} f_u'(t, u(t), \theta)^* I w(t), v(t) \right)_V dt + \int_0^T \left(\tilde{I}^* I w(t), \frac{d}{dt} v(t) \right)_V dt \\
&\quad = \int_0^T (-\ddot{u}^w(t), v(t))_V \, dt + \int_0^T \left(\frac{d}{dt} \int_0^t \tilde{I}^* I w(s) ds, \frac{d}{dt} v(t) \right)_V dt \\
&\quad = \int_0^T (\dot{u}^w(t), \dot{v}(t))_V \, dt + (u^w(0), v(0))_V \\
&\qquad\qquad + \int_0^T \left(\frac{d}{dt} \int_0^t \tilde{I}^* I w(s) ds, \frac{d}{dt} v(t) \right)_V dt + \left(\int_0^0 \tilde{I}^* I w(s) ds, v(0) \right)_V \\
&\quad = \left(u^w + \int_0^t \tilde{I}^* I w(s) ds, v \right)_{\mathcal{U}} .
\end{aligned}
$$

For any $h \in V$ and $v \in \mathcal{U}$

$$\begin{aligned}(h, \delta_0 v)_V &= \int_0^T \big(-\ddot{u}^h(t), v(t)\big)_V \, dt + (h, v(0))_V \\ &= \int_0^T \big(\dot{u}^h(t), \dot{v}(t)\big)_V \, dt + \big(u^h(0), v(0)\big)_V + \big(h + \dot{u}^h(0) - u^h(0), v(0)\big)_V \\ &= \big(u^h, v\big)_{\mathcal{U}} .\end{aligned}$$

The three remaining adjoints are straightforward. For any $\xi \in \mathcal{X}, z \in \mathcal{Y}$

$$\begin{aligned}(g'_\theta(\cdot, u, \theta)\xi, z)_{\mathcal{Y}} &= \int_0^T (g'_\theta(t, u(t), \theta)\xi, z(t))_Y \, dt = \int_0^T (\xi, g'_\theta(t, u(t), \theta)^* z(t))_{\mathcal{X}} \, dt \\ &= \left(\xi, \int_0^T g'_\theta(t, u(t), \theta)^* z(t) \, dt\right)_{\mathcal{X}} \\ &= (\xi, g'_\theta(\cdot, u, \theta)^\star z)_{\mathcal{X}} .\end{aligned}$$

For any $\xi \in \mathcal{X}, w \in \mathcal{W}$

$$\begin{aligned}(f'_\theta(\cdot, u, \theta)\xi, w)_{\mathcal{W}} &= \int_0^T (f'_\theta(t, u(t), \theta)\xi, w(t))_{V^*} \, dt = \int_0^T \langle f'_\theta(t, u(t), \theta)\xi, Iw(t)\rangle_{V^*,V} \, dt \\ &= \int_0^T (\xi, f'_\theta(t, u(t), \theta)^* Iw(t))_{\mathcal{X}} \, dt = \left(\xi, \int_0^T f'_\theta(t, u(t), \theta)^* Iw(t) \, dt\right)_{\mathcal{X}} \\ &= (\xi, f'_\theta(\cdot, u, \theta)^\star w)_{\mathcal{X}} .\end{aligned}$$

For any $\xi \in X, v \in V$

$$(u'_0(\theta)\xi, v)_V = \langle u'_0(\theta)\xi, \tilde{I}^{-1} v\rangle_{V,V^*} = (\xi, u'_0(\theta)^* \tilde{I}^{-1} v)_{\mathcal{X}} .$$

□

Remark 2.3.5. For the Kaczmarz approach relying on time segmentation, the idea is dividing the time line $(0, T)$ into several subintervals, making up a collection of model operators as well as observation data. The system is therefore constructed in the following way

$$\mathbb{F}_0(u, \theta) = \begin{pmatrix} (\dot{u} - f(., u, \theta))\,|_{(0,\tau_1)} \\ u(0) - u_0(\theta) \\ g(., u, \theta)|_{(0,\tau_1)} \end{pmatrix} \tag{2.24}$$

$$\mathbb{F}_j(u, \theta) = \begin{pmatrix} (\dot{u} - f(., u, \theta))\,|_{(\tau_j,\tau_{j+1})} \\ g(., u, \theta)|_{(\tau_j,\tau_{j+1})} \end{pmatrix} \quad j = 1 \ldots n-1 \tag{2.25}$$

for $0 = \tau_0 < \tau_1 < \ldots \tau_{n-1} < \tau_n = T$. The function space setting for the Landweber-Kaczmarz method is

$$\mathbb{F}_0 : \mathcal{U} \times \mathcal{X} \to \mathcal{W}_0 \times V \times \mathcal{Y}_0, \qquad \mathbb{F}_j : \mathcal{U} \times \mathcal{X} \to \mathcal{W}_j \times \mathcal{Y}_j \quad j = 1 \ldots n-1$$

with

$$\mathcal{W}_j = L^2(\tau_j, \tau_{j+1}; V^*), \quad \mathcal{Y}_j = L^2(\tau_j, \tau_{j+1}; Y) \tag{2.26}$$

thus

$$\mathbb{F}'(u,\theta)_j^\star = \begin{pmatrix} \left(\left(\frac{d}{dt} - f'_u(.,u,\theta)\right)|_{(\tau_j,\tau_{j+1})}\right)^\star & \delta_0^\star & (g'_u(.,u,\theta)|_{(\tau_j,\tau_{j+1})})^\star \\ -(f'_\theta(.,u,\theta)|_{(\tau_j,\tau_{j+1})})^\star & -u'_0(\theta)^\star & (g'_\theta(.,u,\theta)|_{(\tau_j,\tau_{j+1})})^\star \end{pmatrix}, \tag{2.27}$$

where the terms in the middle column $\delta_0^\star, -u'_0(\theta)^\star$ are present only in $\mathbb{F}'(u,\theta)_0^\star$.

Remark 2.3.6. In (2.23), since the adjoints of the derivatives with respect to u solve the linear conventional ODEs of second order, we can write them in their explicit forms

$$\begin{aligned} u^z(t) &= \int_0^T (t+1)\tilde{I}g'_u(s,u(s),\theta)^* z(s)ds - \int_0^t (t-s)\tilde{I}g'_u(s,u(s),\theta)^* z(s)ds \\ \tilde{u}^w(t) &:= u^w(t) + \int_0^t \tilde{I}^* I w(s)ds \\ &= -\int_0^T (t+1)\tilde{I}f'_u(s,u(s),\theta)^* I w(s)ds + \int_0^t (t-s)\tilde{I}f'_u(s,u(s),\theta)^* I w(s)ds \\ &\quad + \int_0^t \tilde{I}^* I w(s)ds \\ u^h(t) &= h \end{aligned}$$

for $t \in (0,T)$.

Incorporating the Kaczmarz scheme in Remark 2.3.5, we modify the adjoints accordingly

$$\begin{aligned} u^z(t) &= \int_{\tau_{j(k)}}^{\tau_{j(k+1)}} (t+1)\tilde{I}g'_u(s,u(s),\theta)^* z(s)ds \\ &\quad - \int_{\min\{t,\tau_{j(k)}\}}^{\min\{t,\tau_{j(k+1)}\}} (t-s)\tilde{I}g'_u(s,u(s),\theta)^* z(s)ds \\ \tilde{u}^w(t) &= -\int_{\tau_{j(k)}}^{\tau_{j(k+1)}} (t+1)\tilde{I}f'_u(s,u(s),\theta)^* I w(s)ds \\ &\quad + \int_{\min\{t,\tau_{j(k)}\}}^{\min\{t,\tau_{j(k+1)}\}} (t-s)\tilde{I}f'_u(s,u(s),\theta)^* I w(s)ds + \int_{\min\{t,\tau_{j(k)}\}}^{\min\{t,\tau_{j(k+1)}\}} \tilde{I}^* I w(s)ds \\ \mathbf{u}^h(t) &= h \end{aligned}$$

for $t \in (0,T)$ with $j(k) = k - n\lfloor k/n \rfloor$, and also

$$
\begin{aligned}
&g'_\theta(.,u,\theta)^\star z = \int_{\tau_j}^{\tau_{j(k+1)}} g'_\theta(t,u(t),\theta)^* z(t)dt \\
&u'_0(\theta)^\star = u'_0(\theta)^* I^{-1} \\
&f'_\theta(.,u,\theta)^\star w = \int_{\tau_j}^{\tau_{j(k+1)}} f'_\theta(t,u(t),\theta)^* I w(t)dt.
\end{aligned}
$$

Remark 2.3.7. We now analyze the case of discrete measurement. Let $\{t_i\}_{i=1\ldots n}$ be the discrete observation time points, the system is now defined by

$$
\mathbb{F} : \mathcal{U} \times \mathcal{X} \to \mathcal{W} \times V \times Y^n, \qquad (u,\theta) \mapsto \begin{pmatrix} (\dot{u} - f(.,u,\theta)) \\ u(0) - u_0(\theta) \\ g_1(u,\theta) \\ \vdots \\ g_n(u,\theta) \end{pmatrix} \tag{2.28}
$$

with $g_i(u,\theta) = g(u,\theta)(t_i) = g(t_i, u(t_i), \theta), i = 1 \ldots n$, according to the definition of the Nemytskii operator. Differentiability of the Nemytskii operator g_i could be inferred from differentiablity of the operator inducing it without the need of a local Lipschitz continuity condition.

The Hilbert space adjoint of $\mathbb{F}'(u,\theta)$ then takes the following form

$$
\mathbb{F}'(u,\theta)^\star = \begin{pmatrix} (\frac{d}{dt} - f'_u(.,u,\theta))^\star & \delta_0^\star & g'_{iu}(u,\theta)^\star \\ -f'_\theta(.,u,\theta)^\star & -u'_0(\theta)^\star & g'_{i\theta}(u,\theta)^\star \end{pmatrix} \tag{2.29}
$$

in which the adjoint of $g'_u(u,\theta)$ mapping from the observation space Y^n to $\mathcal{U}$ is

$$
\begin{aligned}
(z, g'_u(u,\theta)v)_{Y^n} &= \sum_{i=1}^n (z_i, g'_{iu}(u,\theta)v(t_i))_Y = \sum_{i=1}^n \left(\tilde{I} g'_{iu}(u,\theta)^* z_i, v(t_i) \right)_V \\
&= (u^z, v)_{\mathcal{U}},
\end{aligned}
$$

where

$$
\begin{cases} \ddot{u}^z(t) = 0 \quad t \in (0,T) \\ \dot{u}^z(0) - u^z(0) = 0, \quad \dot{u}^z(t_i) = \tilde{I} g'_{iu}(u,\theta)^* z_i \quad i = 1 \ldots n \end{cases}
$$

hence

$$
u^z = \sum_{i=1}^n u_i^z \qquad \text{with} \qquad u_i^z = \begin{cases} \tilde{I} g'_{iu}(u,\theta)^* z_i(t+1) & t \le t_i \\ \tilde{I} g'_{iu}(u,\theta)^* z_i(t_i+1) & t > t_i. \end{cases}
$$

If integrating into the Kaczmarz scheme, on every subinterval of index i we get $u^z = u_i^z$, so each equation in the system corresponds to one measurement.

Remark 2.3.8. Remarks 2.3.6 and 2.3.7 show that the choice of the state space $\mathcal{U}$ here provides an explicit formula for the Hilbert space adjoint $\mathbb{F}'(u,\theta)^\star$. This enables us to speed up the computations in Landweber, Landweber-Kaczmarz and possibly in Newton-type methods.

2.4 Reduced formulation

In this section, we formulate the system into one operator mapping directly from the parameter space to the observation space. To this end, we introduce the parameter-to-state map

$$S : \mathcal{X} \to \tilde{\mathcal{U}}, \qquad \text{where} \qquad u = S(\theta) \text{ solves } (2.4) - (2.5),$$

then the forward operator for the reduced setting can be expressed as

$$F : \mathcal{X} \to \mathcal{Y}, \qquad \theta \mapsto g(\cdot, S(\theta), \theta) \tag{2.30}$$

and the inverse problem of recovering θ from y is

$$F(\theta) = y.$$

In order to apply the Hilbert space analysis of Landweber and Landweber-Kazcmarz schemes in this reduced setting, it suffices to choose $\mathcal{X}$ and $\mathcal{Y}$ as Hilbert spaces, while the state space $\tilde{\mathcal{U}}$ may be a Banach space, an option that we actually make use of here

$$\tilde{\mathcal{U}} = W^{1,2,2}(0,T;V,V^*) \cap L^\infty(0,T;V). \tag{2.31}$$

On the other hand, $\tilde{\mathcal{U}}$ is only required for the observation operator g to be well-defined, thus when proving differentiablity of F, we just need to evaluate the image of $S'(\theta)\xi$ in $\mathcal{D}(g) = L^2(0,T;V)$ (although the actual state space $\tilde{\mathcal{U}}$ is much stronger). With the same reason, derivation for the Banach space adjoints can be carried out between spaces $L^2(0,T;V), L^2(0,T;V^*), L^2(0,T;Y)$ and $\mathcal{X}$.

To ensure existence of the parameter-to-state map, we assume that the operator f meets the conditions in the following Assumption 2.4.1 (see Chapter 1, Theorem 1.3.5 for expression) and the conditions (R1) from Proposition 2.4.2. Moreover, we wish to emphasize that, comparing to assumptions in Theorem 1.3.5, the restrictive growth condition on F does not need to be assumed here.

Assumption 2.4.1. Let $\theta \in \mathcal{X}$. Assume that

(S1) for almost all $t \in (0,T)$, the mapping $-f(t,.,\theta)$ is pseudomonotone

(S2) $-f(.,.,\theta)$ is semi-coercive

(S3) f satisfies a condition for uniqueness of the solution.

Proposition 2.4.2. *Assume the model operator f can be decomposed into the form*

$$-f = f_1 + f_2 + f_3$$

with

$$\begin{aligned}
&f_3 : (0,T) \times \mathcal{X} \to H \\
&f_2 : (0,T) \times V \times \mathcal{X} \to H \\
&f_1 : (0,T) \times V \times \mathcal{X} \to V^*.
\end{aligned}$$

Additionally, let the following conditions be fulfilled

(R1) the mappings f_1, f_2, f_3 satisfy

$$\begin{aligned}
&f_3 \in L^2(0,T;H) \text{ and continuous w.r.t. } \theta \\
&\|f_2(t,v,\theta_\epsilon)\|_H \le c_2^\theta(\gamma(t) + \|v\|_V) \text{ for some } \gamma \in L^2(0,T) \\
&f_1 = \varphi'_v \\
&\varphi : [0,T) \times V \times \mathcal{X} \to \mathbb{R} \text{ convex w.r.t. } v \text{ and continuous} \\
&\varphi(t,v,\theta_\epsilon) \ge c_0^\theta \|v\|_V^2 - c_1^\theta \|v\|_H^2 \\
&\|\varphi'_t(t,v,\theta_\epsilon)\|_H \le \tilde{c}_2^\theta(\tilde{\gamma}(t) + \|v\|_V^2) \text{ for some } \tilde{\gamma} \in L^1(0,T)
\end{aligned}$$

for all θ_ϵ in some neighborhood of θ in $\mathcal{X}$, with some $c_0^\theta > 0, 2c_1^\theta T < 1/2$.
If $\varphi : D(\varphi) \to \mathbb{R}^+, c_1^\theta$ does not need to be sufficiently small.

(R2) f is Gâteaux differentiable with respect to its second and third arguments for almost all $t \in (0,T)$ with linear and continuous derivative.
The derivative $-f'_u(.,u,\theta_\epsilon)$ moreover satisfies semi-coercivity in the sense

$$\forall u,v \in V, \forall^{a.e.} t \in (0,T) : \langle -f'_u(t,u,\theta_\epsilon)v, v\rangle_{V^*,V} \ge a_0^\theta |v|_V^2 - a_1^\theta(t)|v|_V - a_2^\theta(t)\|v\|_H^2$$

for all θ_ϵ in some neighborhood of θ in $\mathcal{X}$, with some $a_0^\theta > 0, a_1^\theta \in L^2(0,T), a_2^\theta \in L^1(0,T)$ and some seminorm $|.|_V$ satisfying $\forall v \in V : \|v\|_V \le c_{|.|}(|v|_V + \|v\|_H)$ for some $c_{|.|} > 0$.

(R3) f is locally Lipschitz continuous in the sense

$$\begin{aligned}
&\forall M \ge 0, \exists L(M) \ge 0, \forall^{a.e.} t \in (0,T) : \\
&\qquad \|f(t,v_1,\theta_1) - f(t,v_2,\theta_2)\|_{V^*} \le L(M)(\|v_1 - v_2\|_V + \|\theta_1 - \theta_2\|_{\mathcal{X}}), \\
&\qquad\qquad \forall v_i \in V, \theta_i \in \mathcal{X} : \|v_i\|_V, \|\theta_i\|_{\mathcal{X}} \le M, i = 1,2.
\end{aligned}$$

Moreover, let also u_0 be Gâteaux differentiable. Then F, as defined by (2.30), is Gâteaux differentiable on $\mathcal{X}$ and its derivative is given by

$$\begin{aligned}&F'(\theta) : \mathcal{X} \to \mathcal{Y}\\&(F'(\theta)\xi)(t) = g'_u(t, S(\theta)(t), \theta)u^\xi(t) + g'_\theta(t, S(\theta)(t), \theta)\xi,\end{aligned} \tag{2.32}$$

where $u^\xi = S(\theta)'\xi$ solves the sensitivity equation

$$\begin{cases}\dot{u}^\xi(t) = f'_u(t, S(\theta)(t), \theta)u^\xi(t) + f'_\theta(t, S(\theta)(t), \theta)\xi \quad t \in (0, T)\\ u^\xi(0) = u'_0(\theta)\xi.\end{cases} \tag{2.33}$$

Before proving the result, we notice some facts.
Firstly, Assumption 2.4.1 just guarantees that S is a well-defined map from $\mathcal{X}$ to $W^{1,2,2}(0, T; V, V^*)$ [100, Theorems 8.27, 8.31]. In order to ensure that S maps to $\tilde{\mathcal{U}} = W^{1,2,2}(0, T; V, V^*) \cap L^\infty(0, T; V)$, condition (R1) of Proposition 2.4.2 is additionally required. With condition (R1), we strengthen $S(\theta)$ to lie in $L^\infty(0, T; V)$. The proof is basically based on the regularity result in Roubíček's book [100] with extending the operator to be time-dependent (see Section 2.8.1). Secondly, due to the formulation, differentiability of the forward operator in the reduced setting, in principle, is a question of differentiablity of S (and g).

Proof of Proposition 2.4.2. We show Gâteaux differentiability of S. Fixing θ and without loss of generality, we consider ξ lying in a unit ball and $\epsilon \in (0, 1]$.

First, (R1) allows us to apply the regularity result (A.53) to obtain

$$\begin{aligned}&\|S(\theta + \epsilon\xi)\|_{L^\infty(0,T;V)}\\&\quad\le N^\theta\Bigg(2c_2^\theta\|\gamma\|_{L^2(0,T)} + 2\tilde{c}_2^\theta\|\tilde{\gamma}\|_{L^1(0,T)}^{\frac{1}{2}} + 2(c_2^\theta + \tilde{c}_2^\theta)\sqrt{\frac{c_1^\theta}{c_0^\theta}}\|u_0(\theta + \epsilon\xi)\|_H\\&\qquad\qquad + \|f_3(., \theta + \epsilon\xi)\|_{L^2(0,T;H)} + \sqrt{|\varphi(0, u_0(\theta + \epsilon\xi), \theta + \epsilon\xi)|}\Bigg)\\&\quad\le N^\theta\Bigg(2c_2^\theta\|\gamma\|_{L^2(0,T)} + 2\tilde{c}_2^\theta\|\tilde{\gamma}\|_{L^1(0,T)}^{\frac{1}{2}} + 2C_{V\to H}(c_2^\theta + \tilde{c}_2^\theta)\sqrt{\frac{c_1^\theta}{c_0^\theta}}\|u_0(\theta)\|_V\\&\qquad\qquad + \|f_3(., \theta)\|_{L^2(0,T;H)} + \sqrt{|\varphi(0, u_0(\theta), \theta)|} + 1\Bigg)\\&\quad:= M^\theta\end{aligned} \tag{2.34}$$

for any $\epsilon \in [0, \bar{\epsilon}]$, where the constant N^θ depends only on $c_0^\theta, c_1^\theta, c_2^\theta, \tilde{c}_2^\theta, T$. Here we make use of continuity of the embedding $V \hookrightarrow H$ through the constant $C_{V\to H}$.

Let denote $v_\epsilon := \frac{1}{\epsilon}\left(S(\theta+\epsilon\xi) - S(\theta)\right)$. The function v_ϵ solves

$$\begin{aligned}
\dot{v}_\epsilon(t) &= \frac{1}{\epsilon}\left(f(t, S(\theta+\epsilon\xi)(t), \theta+\epsilon\xi) - f(t, S(\theta)(t), \theta)\right) \\
&= \int_0^1 \left(f'_u(t, S(\theta)(t) + \lambda\epsilon v_\epsilon(t), \theta+\lambda\epsilon\xi)v_\epsilon(t) + f'_\theta(t, S(\theta)(t) + \lambda\epsilon v_\epsilon(t), \theta+\lambda\epsilon\xi)\xi\right) d\lambda \\
&=: A_\epsilon(t)v_\epsilon(t) + B_\epsilon(t)\xi \\
v_\epsilon(0) &= \frac{1}{\epsilon}(u_0(\theta+\epsilon\xi) - u_0(\theta)),
\end{aligned}$$

where by local Lipschitz continuity of f and (2.34) we have, for all most all $t \in (0, T)$,

$$\begin{aligned}
\|B_\epsilon(t)\|_{\mathcal{X}\to V^*} &\le L(\|S(\theta)\|_{L^\infty(0,T;V)} + \|S(\theta+\epsilon\xi)\|_{L^\infty(0,T;V)} + \|\theta\|_{\mathcal{X}} + 1) \\
&\le L(2M^\theta + \|\theta\|_{\mathcal{X}} + 1) \\
\|A_\epsilon(t)\|_{V\to V^*} &\le L(2M^\theta + \|\theta\|_{\mathcal{X}} + 1)
\end{aligned}$$

for any $\epsilon \in (0, \bar{\epsilon}]$.
The fact that $-A_\epsilon$ is linear bounded and semi-coercive by (R2) enables the following estimate (Chapter 1, Theorem 1.3.5, Observation 1.3.8)

$$\begin{aligned}
\|v_\epsilon\|_{W^{1,2,2}(0,T;V,V^*)} &\le C^\theta\left(\|v_\epsilon(0)\|_H + \|B_\epsilon(.)\xi\|_{L^2(0,T;V^*)}\right) \\
&\le C^\theta\left(C_{V\to H}\|u'_0(\theta)\|_{\mathcal{X}\to V} + 1 + \sqrt{T}L(2M^\theta + \|\theta\|_{\mathcal{X}} + 1)\right) \quad (2.35)
\end{aligned}$$

for any $\epsilon \in (0, \bar{\epsilon}]$.

Also by boundedness and semi-coercivity of $-f_u$, the solution to the sensitivity equation u^ξ uniquely exists according to the existence theory for linear parabolic PDEs (see Chapter 1, Observation 1.3.8).

Let $\tilde{v}_\epsilon := \frac{1}{\epsilon}\left(S(\theta+\epsilon\xi) - S(\theta) - \epsilon u^\xi\right)$, then $\tilde{v}_\epsilon$ solves

$$\begin{aligned}
\dot{\tilde{v}}_\epsilon(t) &= f'_u(t, S(\theta)(t), \theta)\tilde{v}_\epsilon + \frac{1}{\epsilon}\Big(- \epsilon f'_u(t, S(\theta)(t), \theta)v_\epsilon(t) - \epsilon f'_\theta(t, S(\theta)(t), \theta)\xi \\
&\qquad + f(t, S(\theta)(t) + \epsilon v_\epsilon(t), \theta+\epsilon\xi) - f(t, S(\theta)(t), \theta)\Big) \\
&=: f'_u(t, S(\theta)(t), \theta)\tilde{v}_\epsilon + b_\epsilon(t) \\
\tilde{v}_\epsilon(0) &= \frac{1}{\epsilon}(u_0(\theta+\epsilon\xi) - u_0(\theta) - \epsilon u'_0(\theta)\xi),
\end{aligned}$$

as a result

$$\|\tilde{v}_\epsilon\|_{W^{1,2,2}(0,T;V,V^*)} \le C^\theta\left(\|\tilde{v}_\epsilon(0)\|_H + \|r_\epsilon\|_{L^2(0,T)}\right). \quad (2.36)$$

The fact that v_ϵ is bounded for any $\epsilon \in (0, \bar{\epsilon}]$ allows us to proceed analogously to the proof of Proposition 2.3.1 to get eventually

$$\|\tilde{v}_\epsilon\|_{W^{1,2,2}(0,T;V,V^*)} \to 0 \quad \text{as} \quad \epsilon \to 0,$$

which proves Gâteaux differentiability of S.

Linearity of $S'(\theta)$ is obvious from (2.33). Its boundedness follows similarly to (2.36) with u^ϵ in place of $\tilde{v}_\epsilon$ and $f'_\theta(t, S(\theta)(t), \theta)\xi$ in place of b_ϵ. Continuity of S is the result from unique existence theory for nonlinear PDE (Chapter 1, Theorem 1.3.5) and Caratheodory condition on f.

Using the chain rule (Chapter 1, Remark 1.4.7), we obtain the derivative of F as in (2.32) and the proof is complete. □

Remark 2.4.3. If f, g and u_0 are Fréchet differentiable then so is F.
This assertion follows from the observations:

- $\|S(\theta + \epsilon\xi)\|_{L^\infty(0,T;V)}$ is uniformly bounded for all $\xi \in B(0,1), \epsilon \in [0, \bar{\epsilon}]$,
- $\|v_\epsilon\|_{W^{1,2,2}(0,T;V,V^*)}$ is uniformly bounded with respect to ξ as a consequence of the uniform boundedness of $\|S(\theta + \epsilon\xi)\|_{L^\infty(0,T;V)}$ and $\frac{1}{\epsilon}\big(\|u_0(\theta + \epsilon\xi) - u_0(\theta)\|_V\big)$,
- Uniform convergence of $\tilde{v}_\epsilon(0)$ to the zero function by Fréchet differentiability of u_0,
- Remark 2.3.2.

Concluding differentiability of F, we now derive the adjoint for $F'(\theta)$.

Proposition 2.4.4. *The Hilbert space adjoint of $F'(\theta)$ is given by*

$$F'(\theta)^\star : \mathcal{Y} \to \mathcal{X}$$

$$F'(\theta)^\star z = \int_0^T g'_\theta(t, S(\theta)(t), \theta)^* z(t) + f'_\theta(t, S(\theta)(t), \theta)^* p^z(t) dt + u'_0(\theta)^* p^z(0), \quad (2.37)$$

where p^z solves

$$\begin{cases} -\dot{p^z}(t) = f'_u(t, S(\theta)(t), \theta)^* p^z(t) + g'_u(t, S(\theta)(t), \theta)^* z(t) \quad t \in (0,T) \\ p^z(T) = 0. \end{cases} \quad (2.38)$$

Proof. [58, Proposition 2.7]. □

Remark 2.4.5. For the Kaczmarz approach and discrete measurement, we refer to [58, Section 2.4, Remark 2.8] respectively.

Remark 2.4.6. With Propositions 2.3.1 and 2.4.2 as well as Remarks 2.3.2 and 2.4.3, we obtain Gâteaux and Fréchet differentiability of the forward operators in both all-at-once and reduced setting. Beyond the use of these derivatives in iterative methods (Landweber here, Gauss-Newton type methods in future work), knowledge of this differentiability yields more information on the topology of the function spaces. By utilizing Fréchet differentiability, other properties of the nonlinear operator can be understood.

2.5 Discussion

Remark 2.5.1. In this work, by introducing the new state spaces $\mathcal{U},\tilde{\mathcal{U}}$ and imposing relevant structural conditions on the nonlinear forward operators, which mainly rely on local Lipschitz continuity, the restrictive growth conditions [58, Conditions (2.2)-(2.5), (2.29)-(2.30)] in the all-at-once and reduced formulations are eliminated.

2.5.1 All-at-once versus reduced version

Comparing the all-at-once and reduced formulations of dynamic inverse problems, we observe the followings:

On function spaces and necessary assumptions

- In the all-at-once formulation, well-definedness of $\mathbb{F}$ directly follows from well-definedness of f and g. On the contrary, the reduced formulation involves the need of well-definedness of the parameter-to-state map $S:\mathcal{X}\to\tilde{\mathcal{U}}$, which requires additional conditions on f and u_0. Therefore, the all-at-once setting gives more flexibility in choosing the function spaces and can deal with more general classes of problems than the reduced setting.

- In the all-at-once setting, the choice of $\mathcal{U}$ depends on $\mathcal{W},\mathcal{X},\mathcal{Y}$. The state space $\mathcal{U}$ must be strong enough for the problem to be well-defined and for the adjoint computation to be feasible.
  In the reduced one, $\tilde{\mathcal{U}}$ just needs to be sufficient for the observation $\mathcal{C}:\tilde{\mathcal{U}}\to\mathcal{Y}$ to be well-defined. In the usual case where $\mathcal{X}=L^2(\Omega),\mathcal{Y}=L^2(0,T;L^2(\Omega)),\mathcal{C}=\mathrm{id}$, although the true state space is much stronger, $\tilde{\mathcal{U}}$ is just required to be $L^2(0,T;L^2(\Omega))$, which might lead to a simplification of the adjoint (c.f, Chapter 5).

- In the general case, when $f:(0,T)\times V\times\mathcal{X}\to W^*$ with the Hilbert space W being possibly different from V, the local Lipschitz condition can be written as

$$
\begin{aligned}
&\forall M\geq 0,\exists L(M)\geq 0,\forall^{a.e.}t\in(0,T): \\
&\qquad\|f(t,v_1,\theta_1)-f(t,v_2,\theta_2)\|_{W^*}\leq L(M)(\|v_1-v_2\|_V+\|\theta_1-\theta_2\|_{\mathcal{X}}),\\
&\qquad\qquad\forall v_i\in V,\theta_i\in\mathcal{X}:\|v_i\|_V,\|\theta_i\|_{\mathcal{X}}\leq M, i=1,2
\end{aligned}
\tag{2.39}
$$

  and is applicable for both settings. With this condition, all the proofs proven in the all-at-once setting are unchanged, while in the reduced setting, the proof for well-definedness of S on the new function spaces might be altered.

On Landweber stepsize

- The stepsizes in the two approaches are not correlated since they are constrained by boundedness of the derivative of the forward operators (Assumption 2.6.2), which are formulated differently

$$\begin{aligned}\|F'(\theta)\xi\|_{\mathcal{Y}} &= \|g'_u(\theta)S'(\theta)\xi + g'_\theta(\theta)\xi\|_{\mathcal{Y}}\\ &= \left\| g'_u(\theta)\left[\frac{d}{dt} - f'_u(\theta);\delta_0\right]^{-1}[f'_\theta(\theta)\xi; u'_0(\theta)\xi] + g'_\theta(\theta)\xi\right\|_{\mathcal{Y}}\\ \|\mathbb{F}'(u,\theta)(v,\xi)\|_{\mathcal{W}\times V\times\mathcal{Y}} &= \big(\, \|\dot{v} - f'_u(u,\theta)v + \|f'_\theta(u,\theta)\xi\|^2_{\mathcal{W}}\\ &\qquad + \|\delta_0(v) + u'_0(\theta)\xi\|^2_V + \|g'_u(u,\theta)v + g'_\theta(u,\theta)\xi\|^2_{\mathcal{Y}}\big)^{1/2}.\end{aligned}$$

  In case of source identification problem with $\mathcal{C} = \mathrm{id}$ and g, u_0 being independent of θ, we have

$$\begin{aligned}\|F'(\theta)\xi\|_{\mathcal{Y}} &= \left\|\left[\frac{d}{dt} - f'_u(\theta);\delta_0\right]^{-1}[\xi;0]\right\|_{\mathcal{Y}} \le C_{\tilde{\mathcal{U}}\hookrightarrow\mathcal{Y}}\left\|\left[\frac{d}{dt} - f'_u(\theta);\delta_0\right]^{-1}[\xi;0]\right\|_{\tilde{\mathcal{U}}}\\ \|\mathbb{F}'(u,\theta)(v,\xi)\|_{\mathcal{W}\times V\times\mathcal{Y}} &= \big(\, \|\dot{v} - f'_u(u,\theta)v + \xi\|^2_{\mathcal{W}} + \|\delta_0(v)\|^2_V + \|v\|^2_{\mathcal{Y}}\big)^{1/2}.\end{aligned}$$

On tangential cone condition

- If having full observation, i.e., $\overline{\mathcal{R}(\mathcal{C})} = \mathcal{Y}$, the tangential cone condition of the reduced setting could be computed via the one of the all-at-once setting (c.f., Section 2.8.3 or a detailed explanation in Chapter 3).

On feasibility

- Due to well-definedness of S, usually monotonicity of F must be taken into account (c.f. Section 2.7), which turns out to be a restriction on the sign of the nonlinearity in the reduced version. On the contrary, the all-at-once one allows arbitrary signs.

- Also by this reason, we mostly have $\mathcal{D}(F) \subset \mathcal{D}(\mathbb{F}_2)$ (where $\mathbb{F}_2$ is the second component of $\mathbb{F}$). Obvious examples are shown in Chapter 3 and Chapter 5.

- Unlike the reduced one, the all-at-once formulation approximates also the state u. Avoiding solving nonlinear equations is the key advantage of the all-at-once approach, which facilitates very much the practical implementation.

- The all-at-once formulation naturally carries over to the wave equation (or also fractional sub- or superdiffusion) context by just replacing the first time derivative by a second (or fractional) time derivative. The reduced version, however, requires additional conditions for well-definedness of the parameter-to-state map.

2.5.2 Time-dependent parameter identification

The parameter considered in the previous sections is time-independent. However, the theory developed in this thesis, in principle, allows the case of time-dependent θ by introducing a time-dependent parameter space, for instance, $\mathcal{X} = L^2(0,T;X), \mathcal{X} = W^{1,2,2}(0,T;X,X^*) \times X_0, \mathcal{X} = H^1(0,T;X) \times X_0$.

Relying on this function space, the local Lipschitz condition holding for all-at-once and reduced setting reads as follows

$$\begin{aligned} &\forall M \geq 0, \exists L(M) \geq 0, \forall^{a.e.} t \in (0,T): \\ &\qquad \|f(t,v_1,\theta_1) - f(t,v_2,\theta_2)\|_{V^*} \leq L(M)(\|v_1 - v_2\|_V + \|\theta_1 - \theta_2\|_X), \\ &\qquad\qquad \forall v_i \in V, \theta_i \in X : \|v_i\|_V, \|\theta_i\|_X \leq M, i = 1,2. \end{aligned} \tag{2.40}$$

The proof for differentiability in both settings needs adapting to the new parameter function space. When working with concrete examples, if time point evaluation $\theta(t)$ is needed, the feasible choices are $\mathcal{X} = H^1(0,T;X) \times X_0$ since $H^1(0,T) \hookrightarrow C(0,T)$ or $\mathcal{X} = W^{1,2,2}(0,T;X,X^*) \times X_0$ since $W^{1,2,2}(0,T;X,X^*) \hookrightarrow C(0,T;H)$ with $X \hookrightarrow H \hookrightarrow X^*$. If the parameter plays the role of a source term, the time point evaluation will not be required.

In case $\mathcal{X} = L^2(0,T;X)$ with Hilbert space X and u_0 independent of θ (to avoid problems from existence of $\theta(0,.)$ for θ only in $L^2(0,T;X)$), the adjoint $\mathbb{F}'(u,\theta)^\star$ in the all-at-once setting is derived with the changes as follows

$$f'_\theta(.,u,\theta)^\star w = f'_\theta(.,u,\theta)^* I w, \qquad g'_\theta(.,u,\theta)^\star z = g'_\theta(.,u,\theta)^* z \tag{2.41}$$

with

$$f'_\theta(.,u,\theta)^*(t) : V^{**} = V \to X^* = X, \qquad g'_\theta(.,u,\theta)^*(t) : Y^* = Y \to X^* = X.$$

And for the the reduced setting, one has

$$F'(\theta)^\star z = g'_\theta(.,S(\theta),\theta)^* z + f'_\theta(.,S(\theta),\theta)^* p^z, \tag{2.42}$$

where p^z solves (2.38).

The Kaczmarz approach relying on the idea of time segmenting does not directly carry over to this case and will be a subject for future research.

2.6 Algorithm and convergence

Loping Landweber-Kaczmarz

In this part, we make explicit the steps required in both settings in case of continuous observation. Starting from an initial guess (u_0, θ_0), the Landweber-Kaczmarz iterations run

All-at-once version

S.1: Set argument to adjoint equations

$$
\begin{aligned}
w_k(t) &= (\dot{u}_k(t) - f(t, u_k(t), \theta_k))\,|_{(\tau_j,\tau_{j+1})}\\
h_k &= u_k(0) - u_0(\theta_k)\\
z_k(t) &= \big(g(t, u_k(t), \theta_k) - y^\delta(t)\big)\,|_{(\tau_j,\tau_{j+1})}
\end{aligned}
$$

so that

$$(w_k, h_k, z_k) = \mathbb{F}_{j(k)}(u_k, \theta_k) - \mathbb{Y}^\delta_{j(k)}$$

S.2: Evaluate adjoint states

$$
\begin{aligned}
&\tilde{u}^w_k, f'_\theta(., u_k, \theta_k)^\star w_k &&\longleftarrow && w_k\\
&u^h_k, u'_k(0)^\star h_k &&\longleftarrow && h_k\\
&u^z_k, g'_\theta(., u_k, \theta_k)^\star z_k &&\longleftarrow && z_k
\end{aligned}
$$

S.3: Update (u, θ) by

$$
\begin{aligned}
(u_{k+1}, \theta_{k+1}) &= (u_k, \theta_k)\\
&- \mu_k \mathbb{F}'_{j(k)}(u_k, \theta_k)^\star (\mathbb{F}_{j(k)}(u_k, \theta_k) - \mathbb{Y}^\delta_{j(k)})
\end{aligned}
$$

Reduced version

S.1: Solve nonlinear state equation

$$
\begin{cases}
\dot{u}_k(t) = f(t, u_k(t), \theta_k) & t \in (0, T)\\
u_k(0) = u_0(\theta_k)
\end{cases}
$$

S.2: Set argument to adjoint equation

$$z_k(t) = \big(g(t, u_k(t), \theta_k) - y^\delta(t)\big)\,|_{(\tau_j,\tau_{j+1})}$$

so that

$$z_k = F_{j(k)}(u_k, \theta_k) - y^\delta_{j(k)}$$

S.3: Evaluate the adjoint state

$$p^z_k \quad \longleftarrow \quad z_k$$

S.4: Update θ by

$$
\begin{aligned}
&\theta_{k+1}\\
&= \theta_k - \mu_k F'_{j(k)}(\theta_k)^\star (F_{j(k)}(\theta_k) - y^\delta_{j(k)})
\end{aligned}
$$

with $j(k) := (k \mod n)$ for n sub-intervals. The stopping rule for the Landweber-Kaczmarz method is chosen according to the discrepancy principle stated in Chapter 1, Definition 1.1.3.

Remark 2.6.1. We have some observations on the algorithm

- For each iteration, the all-at-once algorithm works only with linear models in all steps, while the reduced algorithm requires one step solving a nonlinear equation to evaluate the parameter-to-state map.
- Together with the fact that the adjoint states in the all-at-once setting can be analytically represented (see Remarks 2.3.6 and 2.3.7), one step of the all-at-once algorithm is expected to run much faster than one of the reduced algorithm.
- The residual in the all-at-once case comprises both the errors generated from θ and u in the model and in the observations, while in the reduced case, the exact $u = S(\theta)$ is supposed to be computed. Being inserted into the discrepancy principle, the stopping index k_* in the reduced algorithm is therefore possibly smaller than the one in the all-at-once case.

- For the Kaczmarz approach, the all-at-once setting restricts both model and observation operators to the subinterval $[\tau_j, \tau_{j+1}]$. the reduced setting, however, applies the time restriction only for the observation operator, the model needs to be solved on the full time line to construct the parameter-to-state map.
- One can also incorporate the Kaczmarz strategy into the discrete observation case for both settings.

Convergence analysis

We now provide convergence results under certain conditions. These conditions are derived in the context of iterative regularization methods for nonlinear ill-posed problems [62], which are recalled in the Chapter 1, Assumptions (1.8)-(1.9).

Assumption 2.6.2.

- Tangential cone condition in the all-at-once version

$$\begin{aligned}
&\|f(.,\tilde{u},\tilde{\theta}) - f(.,u,\theta) - f'_u(.,u,\theta)(\tilde{u}-u) - f'_\theta(.,u,\theta)(\tilde{\theta}-\theta)\|_{\mathcal{W}} \\
&\qquad + \|u_0(\tilde{\theta}) - u_0(\theta) - u_0'(\theta)(\tilde{\theta}-\theta)\|_V \\
&\qquad + \|g(.,\tilde{u},\tilde{\theta}) - g(.,u,\theta) - g'_u(.,u,\theta)(\tilde{u}-u) - g'_\theta(.,u,\theta)(\tilde{\theta}-\theta)\|_{\mathcal{Y}} \\
&\leq c_{tc}\big(\|\dot{\tilde{u}} - \dot{u} - f(.,\tilde{u},\tilde{\theta}) \\
&\qquad + f(.,u,\theta)\|_{\mathcal{W}} + \|u_0(\tilde{\theta}) - u_0(\theta)\|_V + \|g(.,\tilde{u},\tilde{\theta}) - g(.,u,\theta)\|_{\mathcal{Y}}\big)
\end{aligned} \tag{2.43}$$

$\forall (u,\theta), (\tilde{u},\tilde{\theta}) \in \mathcal{B}_\rho((u_0,\theta_0))$,

and in the reduced version

$$\begin{aligned}
&\|g(.,S(\tilde{\theta}),\tilde{\theta}) - g(.,S(\theta),\theta) - g'_u(.,S(\theta),\theta)v - g'_\theta(.,S(\theta),\theta)(\tilde{\theta}-\theta)\|_{\mathcal{Y}} \\
&\leq \tilde{c}_{tc}\|g(.,S(\tilde{\theta}),\tilde{\theta}) - g(.,S(\theta),\theta)\|_{\mathcal{Y}}
\end{aligned} \tag{2.44}$$

$\forall \theta, \tilde{\theta} \in \mathcal{B}_\rho(\theta_0)$

for some $c_{tc}, \tilde{c}_{tc} < \frac{1}{2}$, where v solves

$$\begin{cases} \dot{v}(t) = f'_u(t,S(\theta)(t),\theta)v(t) + f'_\theta(t,S(\theta)(t),\theta)(\tilde{\theta}-\theta) & t \in (0,T) \\ v(0) = u_0'(\theta)(\tilde{\theta}-\theta) \end{cases}. \tag{2.45}$$

- The constant in the discrepancy principle is sufficiently large, i.e.,
$\tau > 2\dfrac{1+c_{tc}}{1-2c_{tc}}$.

- The stepsize parameter satisfies $\mu_k \in \left(0, \dfrac{1}{\|\mathbb{F}'(u_k,\theta_k)\|^2}\right]$.

Since our methods are considered in Hilbert spaces, we can employ existing convergence results for Landweber-Kaczmarz method collected in Chapter 1, Section 1.1.3.

2.7 Application to inverse source problems

In this section, the first part is dedicated to examining the conditions proposed in the abstract theory for a class of problems. This work will expose the maximum nonlinearity allowed in our setting, which indicates the improvement comparing to the current result [58]. The section is subsequently continued by some numerical experiments running on both continuous and discrete observations.

Inverse source problems

Let us consider the semilinear diffusion system

$$\dot{u} = \Delta u - \Phi(u) + \theta \qquad (t,x) \in (0,T) \times \Omega \tag{2.46}$$

$$u_{|\partial\Omega} = 0 \qquad t \in (0,T) \tag{2.47}$$

$$u(0) = u_0 \qquad x \in \Omega \tag{2.48}$$

$$y = Cu \qquad (t,x) \in (0,T) \times \Omega, \tag{2.49}$$

where $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain.
We investigate this problem in the function spaces

$$\mathcal{X} = L^2(\Omega), \qquad V = H_0^1(\Omega), \qquad H = L^2(\Omega), \qquad Y = L^2(\Omega)$$

with linear observation (i.e., $\mathcal{C}$ is a linear operator). For the reduced setting, we assume that the nonlinear operator Φ is monotone and $\Phi(0) = 0$. By this way, one typical example could be given, for instance, $\Phi(u) = |u|^{\gamma-1}u, \gamma \geq 1$.

We now verify the imposed conditions in the all-at-once and reduced versions. To begin, we decompose the model operator into

$$-f = -\Delta u + \Phi(u) - \theta := f_1 + f_3.$$

It is obvious that $f_3 = -\theta \in L^2(0,T;H)$, and g is Lipschitz continuous with the Lipschitz constant $L(M) = \|\mathcal{C}\|_{V\to Y}$ and satisfies the boundedness condition $g(t,0,0) = 0 \in Y$.

The next part focuses on analyzing the properties of the model operator f.

(R1) *Regularity conditions for* f_1

The regularity condition (R1) holds by the following argument

$$\begin{aligned}
f_1(v) &= -\Delta v + \Phi(v) = \phi'(v) \\
\phi(v) &= \int_\Omega \frac{1}{2}|\nabla v|^2 + \frac{1}{\gamma+1}\Phi(v)v dx \\
&\geq \frac{1}{2}\|v\|_V^2 \text{ and continuous.} \\
\phi''(v)[w,w] &= \int_\Omega |\nabla w|^2 + \Phi'(v)wwdx \geq 0, \quad \forall v,w \in V
\end{aligned}$$

concludes convexity of ϕ. Here, we invoke monotonicity and differentiability of Φ.
Analogously, $-f'_u(.,u,\theta_\epsilon) = -\Delta + \Phi'(u)$ is semi-coercive in the sence of (R2).

(R3) *Local Lipschitz continuity of f*

First, we observe that

$$\begin{aligned}
\|f(t,v_1,\theta_1) - f(t,v_2,\theta_2)\|_{V^*} &= \|\Phi(v_2) - \Phi(v_1) - \Delta(v_2 - v_1) - (\theta_2 - \theta_1)\|_{H^{-1}} \\
&= \sup_{\|\nabla w\|_{L^2(\Omega)} \le 1} \int_\Omega (\Phi(v_2) - \Phi(v_1))w + \nabla(v_2 - v_1)\nabla w - (\theta_2 - \theta_1) w dx \\
&\le \sup_{\|\nabla w\|_{L^2(\Omega)} \le 1} \int_\Omega (\Phi(v_2) - \Phi(v_1)) w dx + \|v_1 - v_2\|_V + \|\theta_1 - \theta_2\|_{\mathcal{X}} . c_{PF},
\end{aligned}$$

where c_{PF} is the constant in the Poincaré-Friedrichs inequality: $\|v\|_{L^2(\Omega)} \le c_{PF}\|\nabla v\|_{L^2(\Omega)}, \forall v \in H_0^1(\Omega)$. Developing the first term on the right hand side by applying Hölder's inequality, we have

$$\begin{aligned}
&\sup_{\|\nabla w\|_{L^2(\Omega)} \le 1} \int_\Omega (\Phi(v_2) - \Phi(v_1)) w dx \\
&= \sup_{\|\nabla w\|_{L^2(\Omega)} \le 1} \left(\int_\Omega w^{\bar{p}} dx \right)^{\frac{1}{\bar{p}}} \left(\int_\Omega (\Phi(v_2) - \Phi(v_1))^{\frac{\bar{p}}{\bar{p}-1}} dx \right)^{\frac{\bar{p}-1}{\bar{p}}} \\
&\le c_{H^1 \to L^{\bar{p}}} \sqrt{1 + c_{PF}^2} \left[\int_\Omega (v_1 - v_2)^{\bar{p}} dx \right]^{\frac{1}{\bar{p}}} . \left[\int_\Omega \left(\int_0^1 \Phi'(v_1 + \lambda(v_2 - v_1)) d\lambda \right)^{\frac{\bar{p}}{\bar{p}-2}} dx \right]^{\frac{\bar{p}-2}{\bar{p}}} \\
&\le \gamma c_\gamma c_{H^1 \to L^{\bar{p}}}^2 (1 + c_{PF}^2) \|v_1 - v_2\|_V \left[\int_0^1 \int_\Omega (\lambda v_2 + (1-\lambda) v_1)^{(\gamma-1)\frac{\bar{p}}{\bar{p}-2}} dx d\lambda \right]^{\frac{\bar{p}-2}{\bar{p}}} \\
&\le 2^{\gamma-1} \gamma c_\gamma c_{H^1 \to L^{\bar{p}}}^2 (1 + c_{PF}^2) \|v_1 - v_2\|_V \left(\|v_1^{\gamma-1}\|_{L^{\frac{\bar{p}}{\bar{p}-2}}(\Omega)} + \|v_2^{\gamma-1}\|_{L^{\frac{\bar{p}}{\bar{p}-2}}(\Omega)} \right),
\end{aligned}$$

provided additionally that Φ' is the Nemytskii operator (that as always, we use the same notation) induced by the real function Φ' with $|\Phi'(v)| \le c_\gamma |v|^{\gamma-1}$. Altogether, we arrive at

$$\|f(t,v_1,\theta_1) - f(t,v_2,\theta_2)\|_{V^*} \le L(\|v_1\|_V^{\gamma-1} + \|v_2\|_V^{\gamma-1})(\|v_1 - v_2\|_V + \|\theta_1 - \theta_2\|_{\mathcal{X}})$$

if

$$(\gamma - 1)\frac{\bar{p}}{\bar{p}-2} \le \bar{p} \quad \Leftrightarrow \quad \gamma \le \bar{p} - 1,$$

where $c_{H^1\to L^{\bar p}}$ is the constant and $\bar p$ is the maximum power such that the embedding $H^1(\Omega) \hookrightarrow L^{\bar p}(\Omega)$ is continuous

$$\bar p \begin{cases} = \infty & \text{if } d = 1 \\ < \infty & \text{if } d = 2 \\ = \frac{2d}{d-2} & \text{if } d \geq 3 \end{cases} \quad \Leftrightarrow \quad \gamma \begin{cases} = \infty & \text{if } d = 1 \\ < \infty & \text{if } d = 2 \\ \leq 5 & \text{if } d = 3 \\ \leq \frac{d+2}{d-2} & \text{if } d \geq 4. \end{cases} \tag{2.50}$$

Remark 2.7.1. Condition (2.50) on γ reveals that in two dimensional space, our proposed setting works out well with all finite powers of the nonlinearity and in one dimensional space, the accepted power is unconstrained. Noticeably, in the practical case, i.e., three dimensions, the largest power we attain is up to 5 which enhances the limit in the nonlinearity of the current work and enables us to include important applications that had been ruled out in [58] due to the growth conditions there.

The remaining task is to examine well-definedness of the parameter-to-state map $S : \mathcal{X} \to W^{1,2,2}(0,T;V,V^*)$, which will be respectively presented in Sections 2.8.2 and 2.8.3. For verifying the tangential cone condition, we additionally impose the growth condition $|\Phi''(v)| \leq c_\gamma |v|^{\gamma-2}$.

Remark 2.7.2. At this point, we can briefly discuss the uniqueness question. Since θ is time-independent, if having observation at any single time instance $t \in (0,T)$ and observation on all of Ω, we are able to determine θ uniquely by

$$\theta = \left(\dot u - \Delta u + \Phi(u)\right)(t). \tag{2.51}$$

The point evaluation at t only works for sufficiently smooth observation, e.g., $u \in C^1(I; L^2(\Omega)) \cap C(I; H^2(\Omega))$ for some neighborhood I of t, so that $\theta \in L^2(\Omega)$. In principle, this also induces a reconstruction scheme, namely, after filtering the given noisy data, applying formula (2.51). However, in contrast to the scheme we propose here, this would not apply to the practically relevant case of partial (e.g. boundary) and time discrete observations.
Alternatively, we can approach this issue by using Theorem 2.4 [62] (or Chapter 1, Theorem 1.1.11), which concludes uniqueness of the solution from the tangential cone condition as in Section 2.6 and the null space condition

$$\mathcal{N}(F'(\theta^\dagger)) \subset \mathcal{N}(F'(\theta)) \quad \text{for all } \theta \in \mathcal{B}_\rho(\theta^\dagger),$$

where $\mathcal{B}_\rho(\theta^\dagger)$ is a closed ball of radius ρ, center $\theta^\dagger$. Since the tangential cone condition has been verified, it remains to examine the null space of $F'(\theta^\dagger)$. This could be done by linearizing the equation then utilizing some results from Isakov's book [56].
The null space of $F'(\theta^\dagger)$ consists all θ such that the solution v of

$$\begin{aligned} &\dot v - \Delta v + \Phi'(u^\dagger)v = \theta && (t,x) \in (0,T) \times \Omega \\ &v_{|\partial\Omega} = 0 && t \in (0,T) \\ &v(0) = 0 && x \in \Omega \end{aligned}$$

leads to vanishing observations.
Theorem 9.1.2 [56] states that the solution to this inverse source problem is unique if a final observation is available and θ has compact support. As a result, the observation operator just needs to convey the information of u, e.g., at the final time $g(u) = u(T)$. In case of discrete measurement and non-compact support θ , uniqueness is still attainable by using Theorem 9.2.6 [56], where one need to additionally measure one intermediate time data, i.e., $u(t_i, x)$ for some $t_i \in (0, T)$. In this situation, also observation of Neumann data on an arbitrary part of the boundary is demanded.

Numerical results

In the following numerical experiments, we select the nonlinear term $\Phi(u) = u^3$ motivated by the superconductivity example. We assume to observe u fully in time and space, i.e., $\mathcal{C}u = u$ on $(0, T) \times \Omega$ and that at initial time $u(0) = u_0 = 0$. The method in use is loping Landweber-Kaczmarz.

The parameters for implementation are as follows: the time line (0,0.1) (101 time steps) is segmented into 5 time subintervals, the space domain is $\Omega = (0, 1)$ (101 grid points) and the system is perturbed by 5% data noise.

Figure 2.1 displays the results of reconstructed parameter and state comparing to the exact ones. Apparently, two settings yield very similar results, except at $t = 0$ where the reduced setting approximates the exact initial state better. This is explained by the fact that the model equations (2.4)-(2.5) in the reduced setting is fully preserved to construct the parameter-to-state map, while in the all-at-once (AAO) setting, u_0 only appears in the forward operator with the index zero. However, the reconstructed parameters in both settings are definitely comparable.

In Figure 2.2, the left and the middle figures show the scalar w_k in the discrepancy principle in Chapter 1, Definition 1.7 for each iteration (horizontal axis). Five operators represent five time subintervals correspondingly. The right figure sums all w_k over five subintervals and plots both settings together. The all-at-once setting stops the iterations after a factor of 1.5 times those in the reduced setting. This means the all-at-once setting requires much more iterations than the reduced one to obtain an accepted error level. Nevertheless, it runs much faster than the reduced case, in particular the cpu times are: 2989s (Reduced method), 1175s (AAO method). The reasons for this effect have been discussed in Remark 2.6.1.

To demonstrate the discrete observation case, we ran tests at several numbers of discrete observation time points. The parameters for implementation stay the same as in the continuous observation case. Here we also use the loping Landweber-Kaczmarz method.

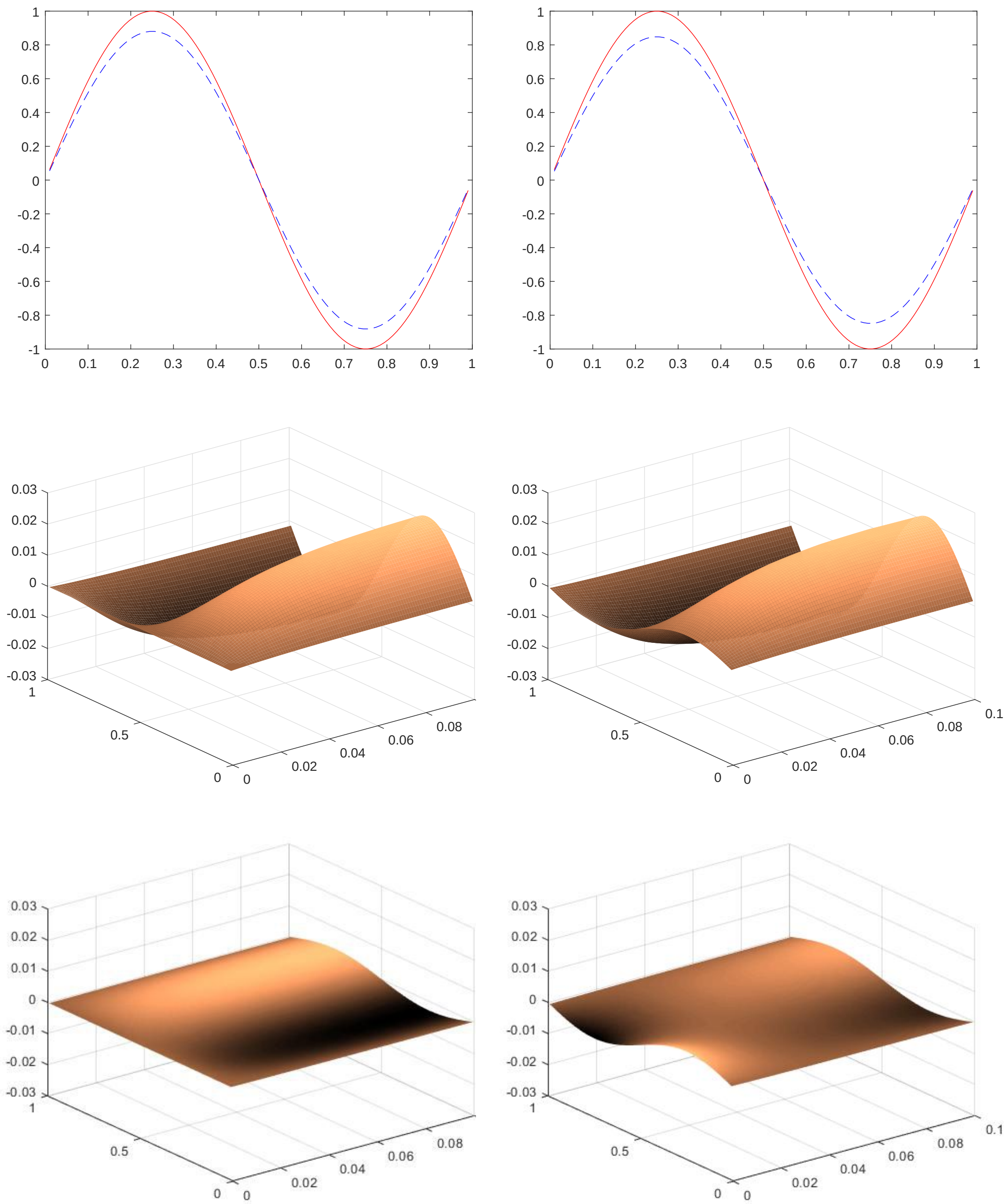

Figure 2.1: Top: exact parameter $\theta^\dagger$ (solid) and reconstruction (dashed). Middle: reconstructed state u. Bottom: difference $u - u^\dagger$. Left: Reduced, right: AAO setting.

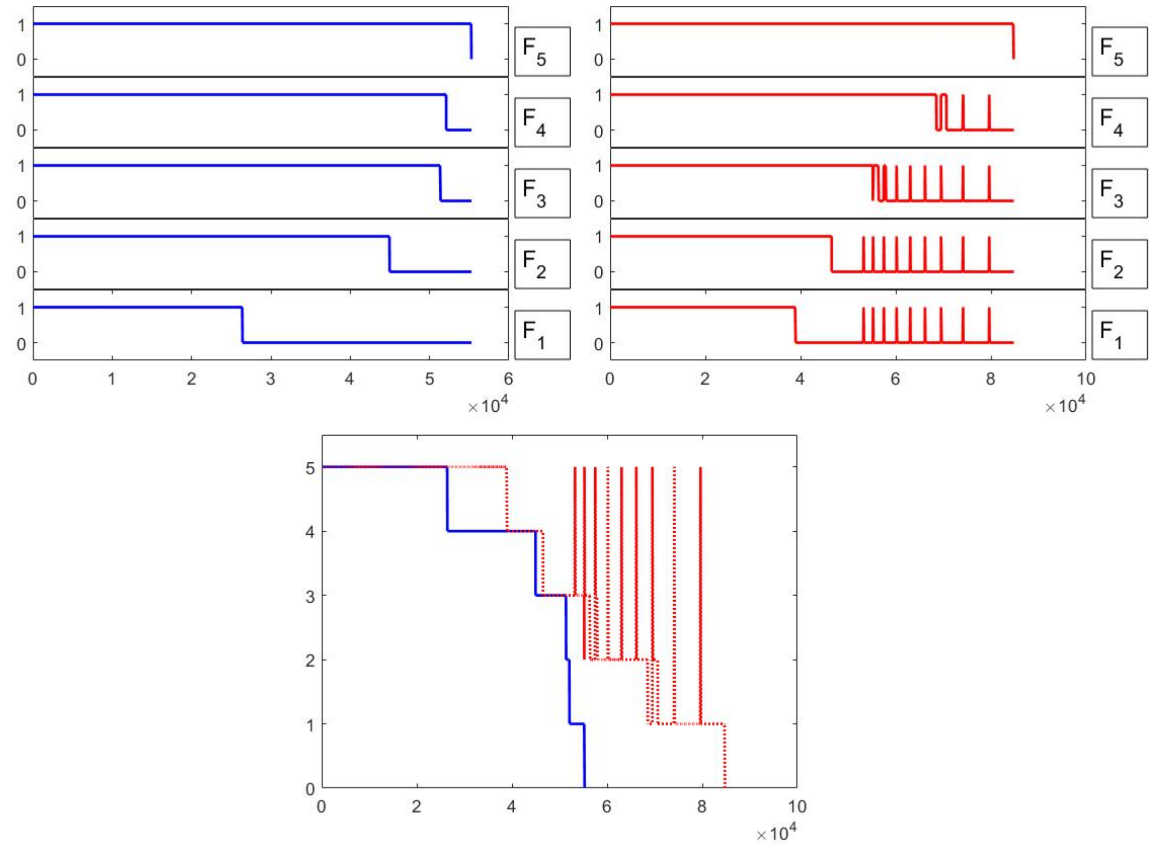

Figure 2.2: On/off iterations on each sub-interval per cycle for Reduced (top left), AAO (top right) methods. Bottom: Number of inner steps per cycle for reduced (solid blue), AAO (dashed red) settings.

Table 2.1: Numerical experiment with different numbers of discrete observation points (np) at 5% noise. Observation points are uniformly distributed on $(0, T]$.

np	Reduced				AAO			
	#iter	#updates	cpu(s)	$\frac{\lVert\theta^{\delta}_{k_*}-\theta^{\dagger}\rVert}{\lVert\theta^{\dagger}\rVert}$	#iter	#updates	cpu(s)	$\frac{\lVert\theta^{\delta}_{k_*}-\theta^{\dagger}\rVert}{\lVert\theta^{\dagger}\rVert}$
3	6191	3854	38	0.113	9722	6215	17	0.158
11	6214	4506	39	0.113	9921	7142	18	0.140
21	6215	4539	39	0.113	10793	7253	19	0.137
51	6278	4556	42	0.112	11831	7303	20	0.135
101	9896	4626	62	0.108	16865	7389	28	0.132

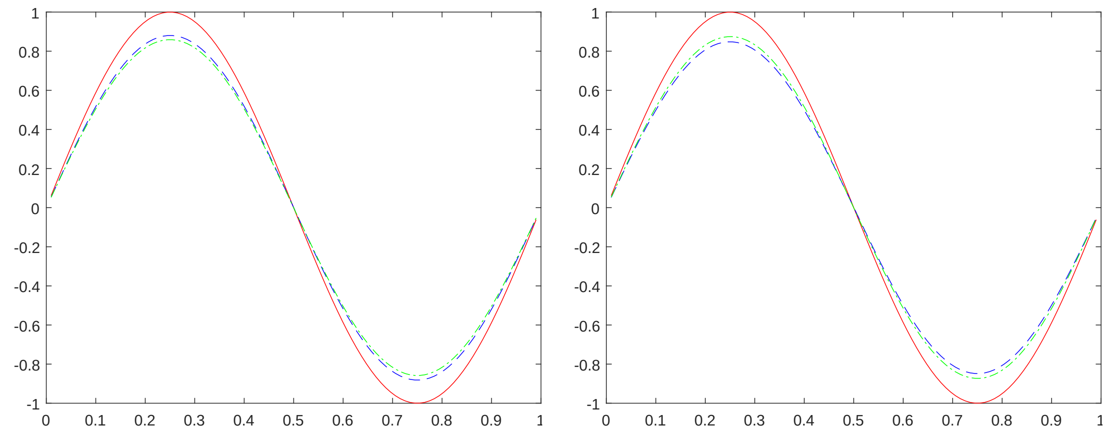

Figure 2.3: Exact parameter $\theta^\dagger$ (solid red), reconstruction by Landweber-Kaczmarz (dashed blue) and reconstruction by Landweber (dashed dotted green). Left: Reduced, right: AAO setting.

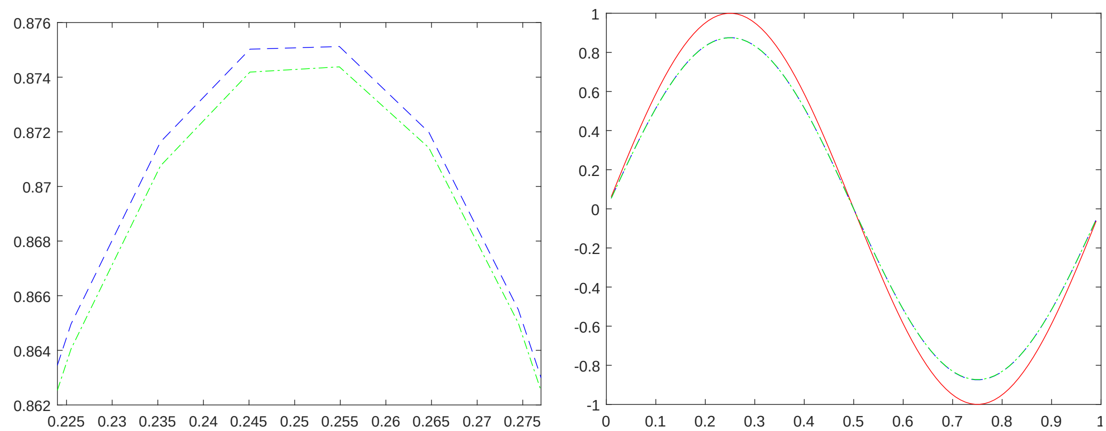

Figure 2.4: Exact parameter $\theta^\dagger$ (solid red), reconstruction by Landweber-Kaczmarz (dashed blue) and reconstruction by Landweber (dashed dotted green). Left: zoom of right, right: AAO setting.

Table 2.1 compares the reduced version with the all-at-once one at different numbers of observation time points np, where each of the time points corresponds to a sub equation, i.e., $np = n$ in the Kaczmarz method (formula (1.3) in Chapter 1). Those numbers vary largely from 3 to the maximum discretization time step 101. Despite the largely varying np, the errors in both settings are quite stable. It gets plausible when looking at the "#updates" columns reporting the "real" iterations, which are the iterations making the update, i.e., iterations with $w_k = 1$. This reveals that significantly increasing the number of discrete observation time points does not bring any improvement. We reason this phenomenon by the above argument (cf. Remark 2.7.2) of uniqueness of the solution according to which, one additional intermediate time data is sufficient to uniquely recover θ.

We turn to a comparison between the Landweber method on the full time horizon and the proposed Landweber-Kaczmarz method. Figure 2.3 shows that in the reduced setting the reconstruction by the Landweber-Kaczmarz method is more accurate, but the performance seems to be reversed in the all-at-once setting. The reason for this is that, in the Landweber-Kaczmarz method the initial data appears only in F_0, while the Landweber method, as working with a fixed operator F on the whole time line $[0,T]$, employs the initial condition always. This leads to the next test in the all-at-once setting displayed in Figure 2.4. In this test, we design the collection of operators $\{F_j\}_{j=0\ldots n-1}$ for Landweber-Kaczmarz method in a slightly different way with the formulation in Remark 2.3.5, namely, we include the initial data in all $F_j, j = 0\ldots n-1$. Figure 2.4 indicates an enhancement in the computed output from the Landweber-Kaczmarz method, and it now performs slightly better than the Landweber method. Besides, we see the all-at-once result in Figure 2.4 is more similar to the reduced result than that in the original Landweber-Kaczmarz method (Figure 2.1). This confirms the advantage of the presence of initial data in the all-at-once formulation of the Landweber-Kaczmarz method in this example.

From this, we observe the followings

- Landweber-Kaczmarz performs slightly better than Landweber in both settings.
- The all-at-once result is more comparable to the reduced one if the initial data is present in all F_j of the all-at-once formulation.
- Relying on the idea of time line segmenting, there would be flexible strategies to generate the collection of F_j for Landweber-Kaczmarz method, e.g.:
    - Method suggested in Remark 2.3.5
    - Method suggested in Remark 2.3.5 with F_j defined on $\{0\} \cup [\tau_j, \tau_{j+1}], j = 0\ldots n-1$
    - Evolution of time subinterval, i.e., F_j is defined on $[0, \tau_{j+1}], j = 0\ldots n-1$

  and those strategies affect the runtime accordingly.

In all tests, we used the finite difference method to compute the exact state. More precisely, an implicit Euler scheme was employed for time discretization and a central difference quotient was used for space discretization. The numerical integration ran with the trapezoid rule. Gaussian noise was generated to be added to the exact data y.

We also point out that in case of having sufficiently smooth data $y^\delta = u^\delta \in L^2(0,T;H_0^1(\Omega))$, we are able to recover the parameter directly from

$$\theta^\delta = \left(\dot{u}^\delta - \Delta u^\delta + {u^\delta}^3\right)(t)$$

if we compute the terms $\dot{u}^\delta$ and Δu^δ, e.g., by filtering (cf. Remark 2.7.2). Getting the output from those in $L^2(0,T;L^2(\Omega))$ and the fact $(u^\delta)^3 \in L^2(0,T;L^2(\Omega))$ since

$H_0^1(\Omega) \hookrightarrow L^6(\Omega)$, it yields the output $\theta^\delta \in L^2(\Omega)$. Nevertheless, if applying the method developed in the thesis, one can deal with much higher nonlinearity, namely, $\Phi(u) = u^5$ as claimed in Remark 2.8.2. Moreover, as apposed to this direct inversion, our approach also works for partial (e.g. boundary) or time discrete observations.

The source identification problem (2.46)-(2.49) examined here can be extended into time-dependent $\theta \in L^2(0, T; X)$ and implemented in the all-at-once as well as the reduced setting using the Landweber method as discussed in Remark 2.5.2.

Conclusion and outlook

In this study, we consider a general evolution system over a finite time line and investigate parameter identification in it by using Landweber-Kaczmarz regularization. We formulate this problem in two different settings: an all-at-once version and a reduced version. In each version, both cases of full and discrete observations are taken into account. The main ingredients for the regularization method are: differentiability and adjoint of the derivative of the forward operators. Differentiability was proved by mainly basing on the choice of appropriate function spaces and a local Lipschitz continuity condition. Segmenting the time line into several subintervals gives the idea to the application of a Kaczmarz method. A loping strategy is incorporated into the method forming the loping Landweber-Kaczmarz iteration. The shown example proves that the method is efficient in the practically relevant situation of high nonlinearity.

Several questions arise for future research:

We plan to extend the theory to time-dependent parameters. For this purpose, we need to build an appropriate function space for θ which, for instance, could allow the local Lipschitz continuity condition. In addition, the assumptions for well-posedness of the parameter-to-state map S need to be carefully considered.

Concerning the model, we intend to also study second order in time equations modeling wave phenomena. Rewriting them as first order in time system by introducing another state $\tilde{u} = \dot{u}$, in principle, allows us to use the present formulation. However, an appropriate function space setting for wave type equations requires different tools for showing, e.g., well-definedness of the parameter-to-state map.

In our test problem (2.46)-(2.49), we consider full space observations in order to establish the tangential cone condition. Practically, relevant partial or boundary observations are yet to be tested numerically.

Regarding numerical implementation of other iterative regularization methods, the difficulty of Newton type methods, which are supposed to give rapid convergence, is the requirement of solving a linear system per iteration step, while this is avoided in the Landweber-Kaczmarz method by the direct use of the Hilbert space adjoints. In the context of this thesis, only numerical experiments for Landweber and Landweber-Kaczmarz methods are provided. Numerical implementation and computational tests for other iterative methods will be a subject of future work.

Considering the all-at-once formulation on a infinite time horizon, by setting $\dot{\theta} = 0$ and $\mathbf{U} := (\theta, u)$, the problem can be written as a dynamical system $\dot{\mathbf{U}}(t) = (0, f)(t) =: \mathbf{F}(., \mathbf{U})(t), t > 0, \mathbf{U}(0) = (\theta^*, u_0)$, where the exact parameter θ^*, being a time constant function, is supposed to be estimated simultaneously to the time evolution of the system and the data collecting process $y(t) = g(., u, \theta)(t) =: \mathbf{G}(\mathbf{U})(t)$. This appears to be a link to online parameter identification methods (see, e.g., [77, 78]). The relation between the proposed all-at-once formulation and online parameter identification for time-independent parameters as well as their analysis (possibly by means of a Lopatinskii condition) will be subject of future research.

2.8 Auxiliary results

2.8.1 Regularity in non-autonomous case

We refer the reader to Regularity Theorem 8.16 in [100], which we are using with exactly the same notations. All the equations referred to "((.))" indicate the ones in the book [100].

Remark 2.8.1. Observations on the Theorem 8.16 in [100]

- This proof still holds for the case A_2 is time-dependent. The condition ((8.59d)) on A_2 could stay fixed or be weakened to $\|A_2(t,v)\|_H \leq C(\gamma(t) + \|v\|_V^{q/2}), \gamma \in L^2(0,T)$, then

$$\|u(t)\|_V \leq N\left(2C\|\gamma\|_{L^2(0,T)} + 2C\sqrt{\frac{c_1}{c_0}}\|u_0\|_H + \|f\|_{L^2(0,T;H)} + \sqrt{|\Phi(u_0)|}\right). \tag{A.52}$$

- We can slightly relax the constraint on c_1 by applying Cauchy's inequality with ϵ for the first estimate in the original proof. In this way, we get $2c_1T < 1/2$ which can be traded off by the scaling $\frac{1}{\epsilon}$ on the right hand side of ((8.63)).
  If $\Phi : V \to \mathbb{R}^+$, the assumption on the smallness of c_1 can be omitted.

Proof. The first expression in the book shows that A_1 does not depend on t. With the hope of generalizing to time-dependent case, our strategy is as follows.
First we set

$$\begin{aligned}
&A_1(t,v) = \Phi'_v(t,v), \quad \Phi : [0,T) \times V \to \mathbb{R} \\
&\Phi'_v \text{ induces a Nemytskii operator, namely, } \Psi'_u \\
&\text{i.e., } \Psi'_u(t,u)(t) = \Phi'_v(t,u(t)),
\end{aligned}$$

thus

$$\begin{aligned}\left\langle A_1(t,u)(t), \frac{d}{dt}u(t)\right\rangle &= \left\langle \Phi'_v(t,u(t)), \frac{d}{dt}u(t)\right\rangle \\ &= \left\langle \Psi'_u(t,u)(t), \frac{d}{dt}u(t)\right\rangle \\ &= \Psi'_t(t,u)(t) - \Phi'_t(t,u(t)).\end{aligned}$$

Looking at the first estimate in the book, we can think of treating $\Psi'_t(t,u)(t)$ as $\Phi'_t(t,u)(t)$ on the left hand side and leaving $\Phi'_t(t,u(t))$ to the right hand side. Choosing

$$\begin{aligned}&\Phi(t,v) \geq c_0\|v\|_V^q - c_1\|v\|_H^2, &&\forall t \in (0,T)\\ &\|\Phi'_t(t,v)\|_H \leq \tilde{C}(\tilde{\gamma}(t) + \|v\|_V^q), &&\tilde{\gamma} \in L^1(0,T)\end{aligned}$$

lets us estimate analogously to ((8.62)) and obtain

$$\begin{aligned}\|u(t)\|_V \leq N\Big(&2C\|\gamma\|_{L^2(0,T)} + 2\tilde{C}\|\tilde{\gamma}\|^{\frac{1}{2}}_{L^1(0,T)} \\ &+ 2(C+\tilde{C})\sqrt{\frac{c_1}{c_0}}\|u_0\|_H + \|f\|_{L^2(0,T;H)} + \sqrt{|\Phi(0,u_0)|}\Big)\end{aligned} \tag{A.53}$$

for all $t \in (0,T)$. This completes the proof. □

2.8.2 Well-definedness of parameter-to-state map

(S1) *Pseudomonotone*

We first recall, $-f(t,.,\theta)$ is called pseudomonotone iff $-f(t,.,\theta)$ is bounded and

$$\left.\begin{aligned}\liminf_{k\to\infty}\langle f(t,u_k,\theta), u_k - u\rangle \geq 0\\ u_k \rightharpoonup u\end{aligned}\right\} \Rightarrow \begin{cases}\langle f(t,u,\theta), u-v\rangle \geq \limsup_{k\to\infty}\langle f(t,u_k,\theta), u_k - v\rangle\\ \forall v \in V.\end{cases} \tag{A.54}$$

We argue this as follows. First, from the left hand side of (A.54) we refer that

$$\begin{aligned}
&\liminf_{k\to\infty} \langle f(t,u_k,\theta) - f(t,u,\theta), u_k - u\rangle \geq \liminf_{k\to\infty} \langle -f(t,u,\theta), u_k - u\rangle = 0\\
&\liminf_{k\to\infty} \langle \Delta(u_k - u) - (\Phi(u_k) - \Phi(u)), u_k - u\rangle \geq 0\\
&\liminf_{k\to\infty} \langle \Delta(u_k - u), u_k - u\rangle \geq \liminf_{k\to\infty} \langle \Phi(u_k) - \Phi(u), u_k - u\rangle \geq 0\\
&\liminf_{k\to\infty} - \|\nabla(u_k - u)\|^2_{L^2(\Omega)} \geq 0\\
&\limsup_{k\to\infty} \|\nabla(u_k - u)\|^2_{L^2(\Omega)} \leq 0\\
&\lim_{k\to\infty} \|\nabla(u_k - u)\|^2_{L^2(\Omega)} = 0\\
&\lim_{k\to\infty} \langle \Delta(u_k - u), u_k - u\rangle = 0,
\end{aligned} \tag{A.55}$$

where monotonicity of Φ is invoked in the above argument.
We then make an observation on the right hand side of (A.54)

$$\begin{aligned}
&\limsup_{k\to\infty} \langle f(t,u_k,\theta), u_k - v\rangle\\
&\qquad = \limsup_{k\to\infty} \{\langle f(t,u_k,\theta), u_k - u\rangle + \langle f(t,u_k,\theta), u - v\rangle\}\\
&\qquad = \limsup_{k\to\infty} \{\langle f(t,u_k,\theta) - f(t,u,\theta), u_k - u\rangle + \langle f(t,u,\theta), u_k - u\rangle\\
&\qquad\qquad\qquad + \langle f(t,u_k,\theta), u - v\rangle\}\\
&\qquad = \limsup_{k\to\infty} \langle \Delta(u_k - u) - (\Phi(u_k) - \Phi(u)), u_k - u\rangle\\
&\qquad\qquad\qquad + \limsup_{k\to\infty} \langle \Delta u_k - \Phi(u_k) + \theta, u - v\rangle\\
&\qquad \leq \limsup_{k\to\infty} \langle \Delta u_k - \Phi(u) + \theta, u - v\rangle ,
\end{aligned}$$

which follows in order from (A.55), monotonicity of Φ, the fact that u_k converges weakly to u and weakly lower semicontinuity of $\langle \Phi(.), u - v\rangle$.
Together with the last evaluation

$$\begin{aligned}
\limsup_{k\to\infty} \langle \Delta u_k, u - v\rangle &= \limsup_{k\to\infty} \langle u_k, \Delta(u - v)\rangle\\
&= \langle u, \Delta(u - v)\rangle\\
&= \langle \Delta u, u - v\rangle ,
\end{aligned}$$

it proves

$$\begin{aligned}
\limsup_{k\to\infty} \langle f(t,u_k,\theta), u_k - v\rangle &\leq \langle \Delta u - \Phi(u) + \theta, u - v\rangle\\
&= \langle f(t,u,\theta), u - v\rangle
\end{aligned}$$

for any $v \in V$. This concludes the validity of (A.54).

To accomplish the monotonicity, the boundedness of $-f(t,.,\theta)$ is needed. We claim this by

$$
\begin{aligned}
&\|-f(t,v,\theta)\|_{V^*}\\
&\quad= \sup_{\|\nabla w\|_{L^2(\Omega)}\le 1} \int_\Omega (-\Delta v + \Phi(v) - \theta) w dx\\
&\quad\le c_{PF}\|\theta\|_{\mathcal{X}} + \|v\|_V + \sup_{\|\nabla w\|_{L^2(\Omega)}\le 1} \left(\int_\Omega w^{\bar p} dx\right)^{\frac{1}{\bar p}} \left(\int_\Omega \Phi(v)^{\frac{\bar p}{\bar p-1}} dx\right)^{\frac{\bar p-1}{\bar p}}\\
&\quad\le c_{PF}\|\theta\|_{\mathcal{X}} + \|v\|_V + c_{H^1\to L^{\bar p}}\sqrt{1+c_{PF}^2}\|v\|^\gamma_{L^{\gamma\frac{\bar p}{\bar p-1}}(\Omega)}\\
&\quad\le c_{PF}\|\theta\|_{\mathcal{X}} + \|v\|_V + c^{1+\gamma}_{H^1\to L^{\bar p}}(1+c_{PF}^2)^{\frac{1+\gamma}{2}}\|v\|_V^\gamma \qquad \forall v \in V,
\end{aligned}
$$

provided that $\gamma\frac{\bar p}{\bar p-1} \le \bar p$ or $\gamma \le \bar p - 1$, which is indeed obtained by (2.50).

(S2) *Semi-coercive*

Since Φ is monotone,

$$
\begin{aligned}
\langle -f(t,v,\theta), v\rangle_{V^*,V} &= \int_\Omega (-\Delta v + \Phi(v) - \theta) v dx\\
&\ge \|\nabla v\|^2_{L^2(\Omega)} - \int_\Omega \theta v dx\\
&\ge \|v\|_V^2 - \|\theta\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)}\\
&\ge \|v\|_V^2 - c_{PF}\|\theta\|_{\mathcal{X}}\|v\|_V,
\end{aligned}
$$

which implies $c_0^\theta = 1 > 0, c_1^\theta = c_{PF}\|\theta\|_{\mathcal{X}} \in L^2(0,T)$ and $c_2^\theta = 0$.

(S3) *Uniqueness*

By again, the monotonicity of Φ,

$$
\begin{aligned}
\langle f(t,u,\theta) - f(t,v,\theta), u-v\rangle_{V^*,V} &= \int_\Omega \left(\Delta(u-v) - (\Phi(u)-\Phi(v))\right)(u-v) dx\\
&\le \int_\Omega \Delta(u-v)(u-v)dx\\
&= -\|\nabla(u-v)\|^2_{L^2(\Omega)}\\
&\le \frac{-1}{c_{PF}^2}\|u-v\|^2_{L^2(\Omega)},
\end{aligned}
$$

which implies $\rho^\theta = \frac{-1}{c_{PF}^2} \in L^1(0,T)$.
This property is used to prove uniqueness of $S(\theta)$ with given θ.

2.8.3 Tangential cone condition

All-at-once setting

Since the observation is linear in our example, the condition (2.43) is fulfilled provided that, for every $u, \tilde{u} \in \mathcal{B}_{2\rho(u_0)}$,

$$\|f(.,\tilde{u},\tilde{\theta}) - f(.,u,\theta) - f'_u(.,u,\theta)(\tilde{u}-u) - f'_\theta(.,u,\theta)(\tilde{\theta}-\theta)\|_{\mathcal{W}} \leq c_{tc}\|\tilde{u}-u\|_{\mathcal{Y}} \quad \text{(A.56)}$$

or, for every $t \in (0,T)$,

$$\|\Phi(\tilde{u}) - \Phi(u) - \Phi'(u)(\tilde{u}-u)\|_{V^*} \leq c_{tc}\|\tilde{u}-u\|_Y. \quad \text{(A.57)}$$

Developing the left hand side (LHS) of (A.57) and assuming that Φ is differentiable up to order two with $|\Phi''(v)| \leq \tilde{c}_\gamma |v|^{\gamma-2}$, we have

$$\begin{aligned}
\text{LHS} &= \left\| \int_0^1 \int_0^1 \Phi''(u+\sigma\lambda(\tilde{u}-u))(\tilde{u}-u)^2 d\lambda d\sigma \right\|_{V^*} \\
&\leq \sup_{0\leq\sigma,\lambda\leq 1} C \left(\int_\Omega \left(\Phi''(u+\sigma\lambda(\tilde{u}-u))(\tilde{u}-u)^2\right)^{\frac{\bar{p}}{\bar{p}-1}} dx \right)^{\frac{\bar{p}-1}{\bar{p}}} \\
&\leq C\|\tilde{u}-u\|_{L^2(\Omega)} \sup_{0\leq\sigma,\lambda\leq 1} \left(\int_\Omega \left(|u+\sigma\lambda(\tilde{u}-u)|^{\gamma-2}(\tilde{u}-u)\right)^{\frac{2\bar{p}}{\bar{p}-2}} dx \right)^{\frac{\bar{p}-2}{2\bar{p}}} \\
&\leq C\|\tilde{u}-u\|_{L^2(\Omega)}\|\tilde{u}-u\|_V \left(\|\tilde{u}^{\gamma-2}\|_{L^{\frac{2\bar{p}}{\bar{p}-4}}} + \|u^{\gamma-2}\|_{L^{\frac{2\bar{p}}{\bar{p}-4}}} \right) \\
&\leq C\|\tilde{u}-u\|_{L^2(\Omega)}\|\tilde{u}-u\|_{\mathcal{U}} \left(\|\tilde{u}\|_{\mathcal{U}} + \|u\|_{\mathcal{U}}\right) \\
&\leq c_{tc}\|\tilde{u}-u\|_Y, \qquad \text{(A.58)}
\end{aligned}$$

where the generic constant C may take different values whenever it appears. The tangential cone coefficient c_{tc}, which depends only on $c_{H^1\to L^{\bar{p}}}, c_{PF}, \gamma$ and T, is sufficiently small if u is sufficiently close to $\tilde{u}$ and $\gamma \leq \frac{\bar{p}}{2}$.

Reduced setting

We need to verify that, for all $\theta, \tilde{\theta} \in \mathcal{B}_{2\tilde{\rho}(\theta_0)}$,

$$\|S(\tilde{\theta}) - S(\theta) - v\|_{\mathcal{Y}} \leq \tilde{c}_{tc}\|S(\tilde{\theta}) - S(\theta)\|_{\mathcal{Y}}, \quad \text{(A.59)}$$

where v solves (2.45). Letting $\xi = \tilde{\theta} - \theta$ then $v = u^\xi$ solves the sensitivity equation (2.33), and by denoting $\tilde{u} := S(\tilde{\theta}), u := S(\theta)$, (A.59) becomes

$$\|\tilde{u} - u - u^\xi\|_{\mathcal{Y}} \leq \tilde{c}_{tc}\|\tilde{u}-u\|_{\mathcal{Y}}.$$

Denoting $\tilde{u} - u - u^\xi =: \tilde{v}_1 = \tilde{v}_{\epsilon=1}$, with $\tilde{v}_\epsilon$ as in (2.36), and using the fact that $\tilde{v}_1(0) = 0$, we deduce

$$\begin{aligned}
\|\tilde{v}_1\|_{L^2(0,T;L^2(\Omega))} &\leq C\|\tilde{v}_1\|_{W^{1,2,2}(0,T;V,V^*)} \leq CC^\theta\|r_1\|_{L^2(0,T)} \\
&\leq CC^\theta c_{tc}\|\tilde{u} - u\|_{L^2(0,T;L^2(\Omega))} \\
&=: \tilde{c}_{tc}\|\tilde{u} - u\|_{L^2(0,T;L^2(\Omega))},
\end{aligned}$$

where $r_1 = r_{\epsilon=1}$, and its $L^2(0,T)$-norm is the left hand side of (A.56).

Remark 2.8.2. In three dimensions, $\gamma = 5$ is achievable with $\mathcal{Y} = L^2(0,T;V)$ by calculating in the same routine to (A.58) for $Y = V$; however the realistic data space is $\mathcal{Y} = L^2(0,T;L^2(\Omega))$.

Chapter 3

Tangential cone condition

The tangential condition was introduced in [49] (recalled in Chapter 1, Assumption (1.8)) as a sufficient condition for convergence of the Landweber iteration for solving ill-posed problems. This condition ensures nonlinearity of the forward operator fits together with the data misfit. In this work, we present a series of time dependent benchmark inverse problems for which we can verify this condition.

The content of this chapter is organized as follows: Section 3.1 gives an introduction to the tangential cone condition and the abstract function space setting for the parabolic problems. Section 3.2 provides results for the all-at-once setting, that are also made use of in the subsequent Section 3.3 for the reduced setting. The proofs of the propositions in Section 3.2 and the notation can be found in auxiliary results.

3.1 Introduction

We consider the problem of recovering a parameter θ in the evolution equation

$$\dot{u}(t) = f(t, \theta, u(t)) \quad t \in (0, T) \tag{3.1}$$

$$u(0) = u_0, \tag{3.2}$$

where for each $t \in (0, T)$ we consider $u(t)$ as a function on a bounded $C^{1,1}$ domain $\Omega \subset \mathbb{R}^d$. In (3.1), $\dot{u}$ denotes the first order time derivative of u and f is a nonlinear function. We here focus on the setting of θ not being time dependent. Problems with state equations of the form $\dot{u}(t) = f(t, \theta(t), u(t))$ could be treated analogously but this would lead to different requirements on the underlying function spaces. These model equations are equipped with additional data obtained from continuous observations over time

$$y(t) = \mathcal{C}(t, u(t)), \tag{3.3}$$

with an observation operator $\mathcal{C}$, which will be assumed to be linear and independent of θ; in particular, in most of what follows $\mathcal{C}$ is the continuous embedding $V \hookrightarrow Y$, with V and Y introduced below.

While formulating the requirements and results first of all in this general framework, we will also apply it to a number of examples as follows.

Identification of a potential

We study the problem of identifying the space-dependent parameter c from observation of the state u in $\Omega \times (0,T)$ in

$$\dot{u} - \Delta u + cu = \varphi \qquad (t,x) \in (0,T) \times \Omega \tag{3.4}$$

$$u_{|\partial\Omega} = 0 \qquad t \in (0,T) \tag{3.5}$$

$$u(0) = u_0 \qquad x \in \Omega, \tag{3.6}$$

where $\varphi \in L^2(0,T;H^{-1}(\Omega))$ and $u_0 \in L^2(\Omega)$ are known. Here, $-\Delta$ could be replaced by any linear elliptic differential operator with smooth coefficients.
With this equation, known, among others, as diffusive Malthus equation [87], one can model the evolution of a population u with diffusion and with exponential growth as time progresses. The latter phenomenon is quantified by the growth rate c, which in this particular case, depends only on the environment.

Identification of a diffusion coefficient

We further consider the problem of recovering the space-dependent parameter a from measurements of u in $\Omega \times (0,T)$ governed by the diffusion equation

$$\dot{u} - \nabla \cdot \Big(a\nabla u\Big) = \varphi \qquad (t,x) \in (0,T) \times \Omega \tag{3.7}$$

$$u_{|\partial\Omega} = 0 \qquad t \in (0,T) \tag{3.8}$$

$$u(0) = u_0 \qquad x \in \Omega, \tag{3.9}$$

where $\varphi \in L^2(0,T;H^{-1}(\Omega))$ and $u_0 \in L^2(\Omega)$ are known. This is, for instance, a simple model of groundwater flow, whose temporal evolution is driven by the divergence of the flux $-a\nabla u$ and the source term φ. The coefficient a represents the diffusivity of the sediment and u is the piezometric head [47].
Banks and Kunisch [7, Chapter I.2] discussed the more general model: $\dot{u}+\nabla\cdot\Big(-a\nabla u + bu\Big) + cu$ describing the sediment formation in lakes and deep seas, in particular, the mixture of organisms near the sediment-water interface.

An inverse source problem with a quadratic first order nonlinearity

Here we are interested in the problem of identifying the space-dependent source term θ from observation of the state u in $\Omega \times (0,T)$

$$\dot{u} - \Delta u - |\nabla u|^2 = \theta \qquad (t,x) \in (0,T) \times \Omega \tag{3.10}$$
$$u_{|\partial\Omega} = 0 \qquad t \in (0,T) \tag{3.11}$$
$$u(0) = u_0 \qquad x \in \Omega\,. \tag{3.12}$$

This quasilinear parabolic equation with a quadratic nonlinearity in ∇u arises, e.g., in stochastic optimal control theory [33, Chapter 3.8].

An inverse source problem with a cubic zero order nonlinearity

The following nonlinear reaction-diffusion equation involves determining the space-dependent source term θ from observation of the state u in $\Omega \times (0,T)$ in a semiliear parabolic equation

$$\dot{u} - \Delta u + \Phi(u) = \varphi - \theta \qquad (t,x) \in (0,T) \times \Omega \tag{3.13}$$
$$u_{|\partial\Omega} = 0 \qquad t \in (0,T) \tag{3.14}$$
$$u(0) = u_0 \qquad x \in \Omega, \tag{3.15}$$

where the possibly space and time dependent source term $\varphi \in L^2(0,T;H^{-1}(\Omega))$ and the initial data $u_0 \in H_0^1(\Omega)$ are known.

For applications of PDEs with cubic nonlinearity, we refer to Chapter 2, Section 2.1.

In part of the analysis we will also consider an additional gradient nonlinearity $\Psi(\nabla u)$ in the PDE, cf. (3.44) below.

Coming back to the general setting (3.1)–(3.3) we will make the following assumptions, where all the considered examples fit into. The operators defining the model and observation equations above are supposed to map between the function spaces

$$f : (0,T) \times \mathcal{X} \times V \to W^* \tag{3.16}$$
$$\mathcal{C} : (0,T) \times V \to Y, \tag{3.17}$$

where $\mathcal{X}, Y, W, V \subseteq Y$ are Banach spaces. More precisely, $\mathcal{X}$ is the parameter space, Y the data space, W^* the space in which the equations is supposed to hold and V the state space. The latter three are first of all the spaces for the respective values at fixed time instances and will also be assigned a version for time-dependent functions, denoted by calligraphic letters. So $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{W}^*$ will denote the parameter, data and equation spaces, respectively, and $\mathcal{U}$ or $\tilde{\mathcal{U}}$ (to distinguish between the different versions in the reduced and all-at-once setting below) the state space. The initial condition $u_0 \in H$, where H is a Banach space as well, will in most of what follows be supposed to be independent of the coefficient θ here. Dependence of the initial data and also of the observation operator on θ can be relevant in some applications but leads to further technicalities, thus for clarity of exposition we shift consideration of these dependencies to future work.

For fixed θ, we assume that the Caratheodory mappings f and $\mathcal{C}$ as defined above induce Nemytskii operators (see Chapter 1, Definitions 1.3.1,1.3.2, and we will use the same notation f and $\mathcal{C}$) on the function space

$$\mathcal{U} = L^2(0,T;V) \cap H^1(0,T;W^*) \text{ or } \tilde{\mathcal{U}} = L^\infty(0,T;V) \cap H^1(0,T;W^*)\,,$$

cf. (3.34) for the all-at-once setting and (3.48) for the reduced setting respectively, in which the state u will be contained, and map into the image space $\mathcal{W}^*$ and observation space $\mathcal{Y}$, respectively, where

$$\mathcal{W}^* = L^2(0,T;W^*), \qquad \mathcal{Y} = L^2(0,T;Y). \tag{3.18}$$

Moreover, $\tilde{\mathcal{U}}$ or $\mathcal{U}$, respectively, will be assumed to continuously embed into $C(0,T;H)$ in order to make sense out of (3.2).

We will consider formulation of the inverse problem, on the one hand, in a classical way, as a nonlinear operator equation

$$F(\theta) = y \tag{3.19}$$

with a forward operator F mapping between Banach spaces $\mathcal{X}$ and $\mathcal{Y}$, and alternatively, on the other hand, as a system of model and observation equation

$$\mathcal{A}(\theta, u) \;=\; 0; \tag{3.20}$$
$$\mathcal{C}(u) \;=\; y. \tag{3.21}$$

Here,

$$\begin{aligned} &\mathcal{A} : \mathcal{X} \times \mathcal{U} \to \mathcal{W}^* \times H, \quad (\theta, u) \mapsto \mathcal{A}(\theta, u) = (\dot{u} - f(\theta, u), u(0) - u_0) \\ &\mathcal{C} : \mathcal{U} \to \mathcal{Y} \end{aligned} \tag{3.22}$$

are the model and observation operators, so that with the parameter-to-state map $S : \mathcal{X} \to \mathcal{U}$ defined by

$$\mathcal{A}(\theta, S(\theta)) = (0,0) \tag{3.23}$$

and

$$F = \mathcal{C} \circ S, \tag{3.24}$$

(3.19) is equivalent to the all-at-once formulation (3.20), (3.21). Defining

$$\mathbb{F} : \mathcal{X} \times \mathcal{U} \to \mathcal{W}^* \times H \times \mathcal{Y}$$

by

$$\mathbb{F}(\theta, u) = (\mathcal{A}(\theta, u), \mathcal{C}(u)),$$

and setting $\mathbb{Y} = (0, 0, y)$, we can rewrite (3.20), (3.21) analogously to (3.19), as

$$\mathbb{F}(\theta, u) = \mathbb{Y}\,. \tag{3.25}$$

All-at-once approaches have been studied for PDE constrained optimization in, e.g., [79, 80, 86, 111, 94, 105, 107] and more recently, for ill-posed inverse problems in, e.g., [18, 19, 44, 57, 59, 111], particularly for time dependent models in [58, 93]. The fact that we are actually using different state spaces $\mathcal{U}$, $\tilde{\mathcal{U}}$ in these two settings is on one hand due to the requirements arising from the need for well-definedness and differentiability of the parameter-to-state map in the reduced setting. On the other hand, while these constraints are not present in the all-at-once setting and a quite general choice of the state space is possible there, whenever a Hilbert space setting is required — e.g., for reasons of easier implementation — this does not only apply to the parameter and data space but also to the state and equation spaces in the all-at-once setting, whereas in a reduced setting these spaces are “hidden” inside the forward operator.

Convergence proofs of iterative regularization methods for solving (3.19) (and likewise (3.25)), such as the Landweber iteration [49, 61] or the iteratively regularized Gauss-Newton method [4, 61, 66], require structural assumptions on the nonlinear forward operator F, such as the tangential cone condition [101] (or Chapter 1, Assumption (1.8))

$$\|F(\theta) - F(\tilde{\theta}) - F'(\theta)(\theta - \tilde{\theta})\|_{\mathcal{Y}} \leq c_{tc}\|F(\theta) - F(\tilde{\theta})\|_{\mathcal{Y}} \quad \forall \theta, \tilde{\theta} \in B_\rho^{\mathcal{X}}(\theta^0)\,. \tag{3.26}$$

for some sufficiently small constant c_{tc}. Here $F'(\theta)$ does not necessarily need to be the Fréchet or Gâteaux derivative of F, but it is just required to be some linear operator that is uniformly bounded in a neighborhood of the initial guess θ_0, i.e., $F'(\theta) \in L(\mathcal{X}, \mathcal{Y})$ such that (Chapter 1, Assumption (1.9))

$$\|F'(\theta)\|_{L(\mathcal{X},\mathcal{Y})} \leq C_F \quad \forall \theta \in B_\rho^{\mathcal{X}}(\theta^0) \tag{3.27}$$

for some $C_F > 0$.

The conditions (3.26) and (3.27) enforce certain local convexity conditions of the residual $\theta \mapsto \|F(\theta) - y\|^2$, cf.[68]. In this sense, the conditions are structurally similar to conditions used in the analysis of Tikhonov regularization, such as those in [23]. The tangential cone condition eventually guarantees convergence to the solution of (3.19) by a gradient descent method for the residual (and also for the Tikhonov functional). Therefore, it ensures that the iterates are not trapped in local minima.

The key contribution of this chapter is to establish (3.26), (3.27) in the reduced setting (3.19) as well as its counterpart in the all-at-once setting (3.25) for the above examples (as well as somewhat more general classes of examples) of parameter identification in initial boundary value problems for parabolic PDEs represented by (3.1), (3.2). In the reduced setting, this also involves the proof of well-definedness and differentiability of the parameter-to-state map S, whereas in the all-at-once setting, this is not needed, thus leaving more freedom in the choice of function spaces. Correspondingly, the examples classes considered in Section 3.2 will be more general than those in Section 3.3.

Some non-trivial static benchmark problems, where the tangential condition has been verified, can be found e.g., in [29, 55, 85].

We mention in passing that in view of existing convergence analysis for such iterative regularization methods for (3.19) or (3.25) in rather general Banach spaces, we will formulate our results in general Lebesgue and Sobolev spaces. Still, we particularly strive for a full Hilbert space setting as preimage and image spaces X and Y, since derivation and implementation of adjoints is much easier then, and also the use of general Banach spaces often introduces additional nonlinearity or nonsmoothness. Moreover, we point out that while in the reduced setting, we will focus on examples of parabolic problems in order to employ a common framework for establishing well-definedness of the parameter-to-state map, the all-at-once version of the tangential condition trivially carries over to the wave equation (or also fractional sub- or superdiffusion) context by just replacing the first time derivative by a second (or fractional) one.

3.2 All-at-once setting

The tangential cone condition and boundedness of the derivative in the all-at-once setting $\mathbb{F}(\theta, u) = \mathbb{Y}$ (3.25) with

$$\mathbb{F} : \mathcal{X} \times \mathcal{U} \to \mathcal{W}^* \times H \times \mathcal{Y}\,, \quad \mathbb{F}(\theta, u) = \begin{pmatrix} \dot{u} - f(\theta, u) \\ u(0) - u_0 \\ \mathcal{C}(u) \end{pmatrix} \tag{3.28}$$

and the norms

$$\|(\theta, u)\|_{\mathcal{X}\times\mathcal{U}} := \left(\|\theta\|^2_{\mathcal{X}} + \|u\|^2_{\mathcal{X}}\right)^{1/2}\,,$$
$$\|(w, h, y)\|_{\mathcal{W}^*\times H\times\mathcal{Y}} := \left(\|w\|^2_{\mathcal{W}^*} + \|h\|^2_H + \|y\|^2_{\mathcal{Y}}\right)^{1/2}$$

on the product spaces read as

$$
\begin{aligned}
&\|f(\theta,u)-f(\tilde{\theta},\tilde{u})-f'_\theta(\theta,u)(\theta-\tilde{\theta})-f'_u(\theta,u)(u-\tilde{u})\|_{\mathcal{W}^*}\\
&\leq c_{tcc}^{AAO}\Big(\|\dot{\tilde{u}}-\dot{u}+f(\theta,u)-f(\tilde{\theta},\tilde{u})\|^2_{\mathcal{W}^*}+\|u(0)-\tilde{u}(0)\|_H^2+\|\mathcal{C}(u-\tilde{u})\|^2_{\mathcal{Y}}\Big)^{1/2},\\
&\quad\forall(\theta,u),(\tilde{\theta},\tilde{u})\in B_\rho^{\mathcal{X}\times\mathcal{U}}(\theta^0,u^0)\,,
\end{aligned}
\tag{3.29}
$$

and

$$
\begin{aligned}
&\Big(\|\dot{v}-f'_\theta(\theta,u)\chi-f'_u(\theta,u)v\|^2_{\mathcal{W}^*}+\|v(0)\|^2_H+\|\mathcal{C}v\|^2_{\mathcal{Y}}\Big)^{1/2}\leq C_{\mathbb{F}}\left(\|\chi\|^2_{\mathcal{X}}+\|v\|^2_{\mathcal{U}}\right)^{1/2},\\
&\quad\forall(\theta,u)\in B_\rho^{\mathcal{X}\times\mathcal{U}}(\theta^0,u^0)\,,\ \chi\in\mathcal{X}\,,\ v\in\mathcal{U}
\end{aligned}
\tag{3.30}
$$

where we have assumed linearity of $\mathcal{C}$.

Since the right hand side terms $\|u(0)-\tilde{u}(0)\|_H$ and $\|f(\theta,u)-f(\tilde{\theta},\tilde{u})\|_{\mathcal{W}^*}$ in (3.29) are usually too weak to help for verification of this condition, we will just skip it in the following and consider

$$
\begin{aligned}
&\|f(\theta,u)-f(\tilde{\theta},\tilde{u})-f'_\theta(\theta,u)(\theta-\tilde{\theta})-f'_u(\theta,u)(u-\tilde{u})\|_{\mathcal{W}^*}\\
&\leq c_{tcc}^{AAO}\|\mathcal{C}(u-\tilde{u})\|_{\mathcal{Y}}\,,\qquad\qquad \forall(\theta,u),(\tilde{\theta},\tilde{u})\in B_\rho^{\mathcal{X}\times\mathcal{U}}(\theta^0,u^0),
\end{aligned}
\tag{3.31}
$$

which under these conditions is obviously sufficient for (3.29). Moreover, in order for the remaining right hand side term to be sufficiently strong in order to be able to dominate the left hand side, we will need to have full observations in the sense that

$$
\overline{\mathcal{R}(\mathcal{C}(t))}=Y\,. \tag{3.32}
$$

In the next section, it will be shown that under certain stability conditions on the generalized ODE in (3.1), together with (3.32), the version (3.31) of the all-at-once tangential cone condition is sufficient for its reduced counterpart (3.26).

Likewise, we will further consider the sufficient condition for boundedness of the derivative

$$
\begin{aligned}
&\|f'_\theta(\theta,u)\|_{L(\mathcal{X},\mathcal{W}^*)}\leq C_{\mathbb{F},1}\,, &&\|f'_u(\theta,u)\|_{L(\mathcal{U},\mathcal{W}^*)}\leq C_{\mathbb{F},2}\,,\\
&\|\partial_t\|_{L(\mathcal{U},\mathcal{W}^*)}\leq C_{\mathbb{F},0}\,, &&\|\mathcal{C}\|_{L(\mathcal{U},\mathcal{Y})}\leq C_{\mathbb{F},3}\\
&\quad\forall(\theta,u)\in B_\rho^{\mathcal{X}\times\mathcal{U}}(\theta^0,u^0)\,.
\end{aligned}
\tag{3.33}
$$

The function space setting considered here will be

$$
\begin{aligned}
&\mathcal{U}=\{u\in L^2(0,T;V):\dot{u}\in L^2(0,T;W^*)\}\hookrightarrow C(0,T;H)\,,\\
&\mathcal{W}=L^2(0,T;W)\,,\qquad \mathcal{Y}=L^2(0,T;Y)\,,
\end{aligned}
\tag{3.34}
$$

so that the third bound in (3.33) is automatically satisfied with $C_{\mathbb{F},0}=1$. We focus on Lebesgue and Sobolev spaces[1]

$$
V=W^{s,m}(\Omega)\,,\quad W=W^{t,n}(\Omega)\,,\quad Y=L^q(\Omega)\,, \tag{3.35}
$$

[1] In place of V, its intersection with $H_0^1(\Omega)$ might be considered in order to take into account homogeneous Dirichlet boundary conditions. For the estimates themselves, this does not change anything.

with $s, t \in [0, \infty)$, $m, n \in [1, \infty]$, $q \in [1, \hat{q}]$, and $\hat{q}$ the maximal index such that V continuously embeds into $L^{\hat{q}}(\Omega)$, i.e. such that

$$s - \frac{d}{m} \succeq -\frac{d}{\hat{q}}, \tag{3.36}$$

so that with $\mathcal{C}$ defined by the embedding operator $\mathcal{U} \to \mathcal{Y}$, the last bound in (3.33) is automatically satisfied [2]. For the notation $\succeq$, we refer to right below.

The parameter space $\mathcal{X}$ may be very general at the beginning of Section 3.2.1 and in Section 3.2.2. We will only specify it in the particular examples of Section 3.2.1.

Notation

- For $a, b \in \mathbb{R}$, the notation $a \succeq b$ means: $a \geq b$ with strict inequality if $b = 0$.
- For normed spaces A, B, the notation $A \hookrightarrow B$ means: A is continuously embedded in B.
- For a normed space A, an element $a \in A$ and $\rho > 0$, we denote by $\mathcal{B}_\rho^A(a)$ the closed ball of radius ρ around a in A.
- For vectors $\vec{a}, \vec{b} \in \mathbb{R}^n$, $\vec{a} \cdot \vec{b}$ denotes the Euclidean inner product. Likewise, $\nabla \cdot \vec{v}$ denotes the divergence of the vector field $\vec{v}$.
- C denotes a generic constant that may take different values whenever it appears.
- For $p \in [1, \infty]$, we denote by $p^* = \frac{p}{p-1}$ the dual index.

We will now verify the conditions (3.31), (3.33) for some (classes of) examples.

3.2.1 Bilinear problems

Many coefficient identification problems in linear PDEs, such as the identification of a potential or of a diffusion coefficient, as mentioned above, can be treated in a general bilinear context.

Consider an evolution driven by a bilinear operator, i.e.,

$$f(\theta, u)(t) = L(t)u(t) + ((B\theta)(t))u(t) - g(t), \tag{3.37}$$

[2] One could possibly think of also extending to more general Lebesgue spaces instead of L^2 with respect to time. As long as the summability index is the same for $\mathcal{W}$ and $\mathcal{Y}$ this would not change anything in Section 3.2.1. As soon as the summability indices differ, one has to think of continuity of the embedding $\mathcal{U} = L^{r_1}(0, T; V) \cap W^{1,r_2}(0, T; W^*) \hookrightarrow \mathcal{Y} = L^{r_3}(0, T; Y)$ as a whole, possibly taking advantage of some interpolation between $L^{r_1}(0, T; V)$ and $W^{1,r_2}(0, T; W^*)$. This could become very technical but might pay off in specific applications.

where for almost all $t \in (0,T)$, and all $\theta \in \mathcal{X}$, $v \in V$ we have $L(t)$, $(B\theta)(t) \in \mathcal{L}(V,W^*)$, $\theta \mapsto (B\theta)(t)v \in \mathcal{L}(\mathcal{X},W^*)$, and $g(t) \in W^*$, with

$$\sup_{t\in[0,T]} \|L(t)\|_{\mathcal{L}(V,W^*)} \leq C_L \,, \quad \sup_{t\in[0,T]} \|(B\theta)(t)\|_{\mathcal{L}(V,W^*)} \leq C_B \|\theta\|_{\mathcal{X}} \tag{3.38}$$

so that the first and second bounds in (3.33) are satisfied, due to the estimates

$$\begin{aligned}
\|f'_\theta(\theta,u)\chi\|_{\mathcal{W}^*} &= \left(\int_0^T \|((B\chi)(t))u(t)\|^2_{W^*}\right)^{1/2} \leq C_B\|\chi\|_{\mathcal{X}} \left(\int_0^T \|u(t)\|^2_V\right)^{1/2} \\
\|f'_u(\theta,u)v\|_{\mathcal{W}^*} &= \left(\int_0^T \|L(t)v(t) + ((B\theta)(t))v(t)\|^2_{W^*}\right)^{1/2} \\
&\leq (C_L + C_B\|\theta\|_{\mathcal{X}}) \left(\int_0^T \|v(t)\|^2_V\right)^{1/2} .
\end{aligned}$$

For the left hand side in (3.31), we have

$$\Big(f(\theta,u) - f(\tilde\theta,\tilde u) - f'_u(\theta,u)(u-\tilde u) - f'_\theta(\theta,u)(\theta-\tilde\theta)\Big)(t) = -((B(\theta-\tilde\theta)(t))(u-\tilde u)(t)\,,$$

and (3.31) is satisfied if and only if (3.32) and

$$\|(B(\theta-\tilde\theta))(u-\tilde u)\|_{\mathcal{W}^*} \leq c^{AAO}_{tcc}\|\mathcal{C}(u-\tilde u)\|_{\mathcal{Y}}\,, \quad \forall(\theta,u),(\tilde\theta,\tilde u) \in B^{\mathcal{X}\times\mathcal{U}}_\rho(\theta^0,u^0)$$

hold. A sufficient condition for this to hold is

$$\begin{aligned}
&\|(B(\theta-\tilde\theta))(t)(v-\tilde v)\|_{W^*} \leq c^{AAO}_{tcc}\|\mathcal{C}(t)(v-\tilde v)\|_Y\,, \\
&\forall(\theta,v),(\tilde\theta,\tilde v) \in B^{\mathcal{X}\times V}_\rho(\theta^0,u^0(t))\,, \quad t\in(0,T).
\end{aligned} \tag{3.39}$$

The proofs of the propositions for the following examples can be found in the Section 3.4, the auxiliary results. Likewise, the conditions on the summability and smoothness indices s,t,p,q,m,n of the used spaces, (B.107), (B.109), (B.112), (B.113), (B.115), (B.119), (B.120), (B.121), (B.122), (B.123) as appearing in the formulation of the propositions, are derived there.

Identification of a potential c

Problem (3.4)-(3.6) can be cast into the form (3.37) by setting $\theta = c$ and

$$L(t) = \Delta\,, \quad (Bc)(t)v = -cv, \tag{3.40}$$

(i.e., $(Bc)(t)$ is a multiplication operator with the multiplier c). We set

$$\mathcal{X} = L^p(\Omega)\,. \tag{3.41}$$

Proposition 3.2.1. *For $\mathcal{U}$, $\mathcal{W}$, $\mathcal{Y}$ according to (3.34) with (3.35) (B.107), $-\Delta \in \mathcal{L}(V, W^*)$, the operator $\mathbb{F}$ defined by (3.28), (3.37), (3.40), $\mathcal{C} = id : \mathcal{U} \to \mathcal{Y}$ satisfies the tangential cone condition (3.31) with a uniformly bounded operator $\mathbb{F}'(c)$, i.e., the family of linear operators $(\mathbb{F}'(c))_{c\in M}$ is uniformly bounded in the operator norm, for c in a bounded subset M of $\mathcal{X}$.*

Remark 3.2.2. A full Hilbert space setting can be achieved by setting $p = q = m = n = 2$ and choosing $s \geq 0$, $t > \frac{d}{2}$.

Identification of a diffusion coefficient a

The a problem (3.7)-(3.9) is defined by setting

$$L(t) \equiv 0\,, \quad (Ba)(t)v = \nabla \cdot (a\nabla v) \tag{3.42}$$

so that

$$\begin{aligned}
&\|(B(\hat{a})(t))\hat{v}\|_{W^*} \\
&= \sup_{w\in W\,,\ \|w\|_W\leq 1} \int_\Omega \hat{a}\nabla\hat{v}\cdot\nabla w\,dx = \sup_{w\in W\,,\ \|w\|_W\leq 1} \int_\Omega \hat{v}\Big(\nabla\hat{a}\cdot\nabla w + \hat{a}\Delta w\Big)\,dx \\
&\leq \|\hat{v}\|_{L^q}\Big(\|\nabla\hat{a}\|_{L^p} \sup_{w\in W\,,\ \|w\|_W\leq 1} \|\nabla w\|_{L^{\frac{p^* q}{q-p^*}}} + \|\hat{a}\|_{L^r} \sup_{w\in W\,,\ \|w\|_W\leq 1} \|\Delta w\|_{L^{\frac{r^* q}{q-r^*}}}\Big)\,.
\end{aligned}$$

Note that since $Y = L^q(\Omega)$ we had to move all derivatives away from $\hat{v}$ by means of integration by parts, which forces us to use spaces of differentiability order at least two in W and at least one in $\mathcal{X}$. Thus, we here consider

$$\mathcal{X} = W^{1,p}(\Omega)\,. \tag{3.43}$$

Proposition 3.2.3. *For $\mathcal{U}$, $\mathcal{W}$, $\mathcal{Y}$ according to (3.34) with (3.35), (B.109), the operator $\mathbb{F}$ defined by (3.28), (3.37), (3.42), $\mathcal{C} = id : \mathcal{U} \to \mathcal{Y}$ satisfies the tangential cone condition (3.31) with a uniformly bounded operator $\mathbb{F}'(a)$.*

Remark 3.2.4. A full Hilbert space setting $p = q = m = n = 2$ requires to choose
$s \geq 0$ and $t \begin{cases} \geq 2 \text{ if } d = 1 \\ > 2 \text{ if } d = 2 \\ > 1 + \frac{d}{2} \text{ if } d \geq 3 \end{cases}$.

3.2.2 Nonlinear inverse source problems

Consider nonlinear evolutions that are linear with respect to the parameter θ, i.e.,

$$f(\theta, u)(t) = L(t)u(t) + \Phi(u(t)) + \Psi(\nabla u(t)) - B(t)\theta, \tag{3.44}$$

where for almost all $t \in (0,T)$, $L(t) \in \mathcal{L}(V, W^*)$, $B(t) \in \mathcal{L}(\mathcal{X}, W^*)$ and $\Phi, \Psi \in C^2(\mathbb{R})$ satisfy the Hölder continuity and growth conditions

$$|\Phi'(\lambda) - \Phi'(\tilde{\lambda})| \leq C_{\Phi''}(1 + |\lambda|^\gamma + |\tilde{\lambda}|^\gamma)|\tilde{\lambda} - \lambda|^\kappa \tag{3.45}$$

for all $\tilde{\lambda}, \lambda \in \mathbb{R}$

$$|\Psi'(\lambda) - \Psi'(\tilde{\lambda})| \leq C_{\Psi''}(1 + |\lambda|^{\hat{\gamma}} + |\tilde{\lambda}|^{\hat{\gamma}})|\tilde{\lambda} - \lambda|^{\hat{\kappa}} \tag{3.46}$$

for all $\tilde{\lambda}, \lambda \in \mathbb{R}^d$, where $\gamma, \hat{\gamma}, \kappa, \hat{\kappa} \geq 0$. We will show that the exponents $\gamma, \hat{\gamma}$ may actually be arbitrary as long as the smoothness s, t of V and W is chosen appropriately.

Proposition 3.2.5. *The operator $\mathbb{F}$ defined by (3.28), (3.37), (3.42), $\mathcal{C} = id : \mathcal{U} \to \mathcal{Y}$ in either of the four following cases*

(a) (3.45) *and Ψ affinely linear and $\mathcal{U}$, $\mathcal{W}$, $\mathcal{Y}$ as in* (3.34) *with* (3.35), (B.112), (B.113)*;*

(b) (3.45), (3.46) *and $\mathcal{U}$, $\mathcal{W}$, $\mathcal{Y}$ as in* (3.34) *with* (3.35), (B.112), (B.113), (B.115), (B.119)*;*

(c) (3.45) *and Ψ affinely linear, $\mathcal{W}$, $\mathcal{Y}$ as in* (3.34)*, $\mathcal{U}$ as in* (B.120) *with* (3.35), (3.36), (B.122)*;*

(d) (3.45), (3.46)*, $\mathcal{W}$, $\mathcal{Y}$ as in* (3.34)*, $\mathcal{U}$ as in* (B.121) *with* (3.35), (3.36), (B.122), (B.123)*;*

satisfies the tangential cone condition (3.31) *with a uniformly bounded operator $\mathbb{F}'(\theta)$.*

Remark 3.2.6. A Hilbert space setting $p = q = m = n = 2$ is, therefore, possible for arbitrary $\gamma, \kappa, \hat{\gamma}$, provided t and s are chosen sufficiently large, cf. (B.112), (B.113) in case $C_{\Psi''} = 0$, and additionally (B.115), (B.119) otherwise.

3.3 Reduced setting

In this section, we formulate the system (3.1)-(3.3) by one operator mapping from the parameter space to the observation space. To this end, we introduce the parameter-to-state map

$$S : \mathcal{D} \subseteq \mathcal{X} \to \tilde{\mathcal{U}}, \qquad \text{where} \qquad u = S(\theta) \text{ solves } (3.1) - (3.2)$$

then, with $\mathcal{D}(F) = \mathcal{D}$ the forward operator for the reduced setting can be expressed as

$$F : \mathcal{D}(F) \subseteq \mathcal{X} \to \mathcal{Y}, \qquad \theta \mapsto \mathcal{C}(S(\theta)), \tag{3.47}$$

and the inverse problem of recovering θ from y can be written as

$$F(\theta) = y.$$

Here, differently from the state space $\mathcal{U}$ in the all-at-once setting, cf., (3.34), we use a non Hilbert state space

$$\tilde{\mathcal{U}} = \{u \in L^\infty(0,T;V) : \dot{u} \in L^2(0,T;W^*)\} \tag{3.48}$$

as this appears to be more appropriate for applying parabolic theory.

We now establish a framework for verifying the tangential cone condition as well as boundedness of the derivative in this general setting.

For this purpose, we make the following assumptions.

Assumption 3.3.1.

(R1) *Local Lipschitz continuity of f*

$$\begin{aligned}\forall M \geq 0, &\exists L(M) \geq 0, \forall^{a.e.} t \in (0,T) :\\ &\|f(t,\theta_1,v_1) - f(t,\theta_2,v_2)\|_{W^*} \leq L(M)(\|v_1 - v_2\|_V + \|\theta_1 - \theta_2\|_{\mathcal{X}}),\\ &\qquad \forall v_i \in V, \theta_i \in \mathcal{X} : \|v_i\|_V, \|\theta_i\|_{\mathcal{X}} \leq M, i = 1,2.\end{aligned}$$

(R2) *Well-definedness of the parameter-to-state map*

$$S : \mathcal{D}(F) \subseteq \mathcal{X} \to \tilde{\mathcal{U}}$$

with $\tilde{\mathcal{U}}$ as in (3.48) as well as its boundedness in the sense that there exists $C_S > 0$ such that for all $\theta \in \mathcal{B}_\rho^{\mathcal{X}}(\theta^0)$ the estimate

$$\|S(\theta)\|_{L^\infty(0,T;V)} \leq C_S$$

holds.

(R3) *Continuous dependence on data of the solution to the linearized problem with zero initial data*, i.e., there exists a constant C_{lin} such that for all $\theta \in \mathcal{B}_\rho^{\mathcal{X}}(\theta^0)$, $b \in \mathcal{W}^*$, and any z solving

$$\dot{z}(t) = f'_u(\theta, S(\theta))(t)z(t) + b(t) \quad t \in (0,T) \tag{3.49}$$
$$z(0) = 0, \tag{3.50}$$

the estimate

$$\|z\|_{\mathcal{Y}} \leq C_{lin}\|b\|_{\mathcal{W}^*} \tag{3.51}$$

holds.

(R4) *Tangential cone condition of the all-at-once setting* (3.31)

$$\exists \rho > 0, \forall (\theta, u), (\tilde{\theta}, \tilde{u}) \in \mathcal{B}_\rho^{\mathcal{X},\mathcal{U}}(\theta^0, u^0):$$
$$\|f(\tilde{\theta}, \tilde{u}) - f(\theta, u) - f'_u(\theta, u)(\tilde{u} - u) - f'_\theta(\theta, u)(\tilde{\theta} - \theta)\|_{\mathcal{W}^*} \leq c_{tcc}^{AAO}\|\mathcal{C}\tilde{u} - \mathcal{C}u\|_{\mathcal{Y}}.$$

The main result of this section is as follows.

Theorem 3.3.2. *Suppose Assumption 3.3.1 holds and $\mathcal{C}$ is the embedding $V \hookrightarrow Y$. Then there exists a constant $\rho > 0$ such that for all $\theta, \tilde{\theta} \in \mathcal{B}_\rho^{\mathcal{X}}(\theta_0) \subset \mathcal{D}(F)$,*

i) *$F'(\theta)$ is uniformly bounded*

$$\|F'(\theta)\|_{L(\mathcal{X},\mathcal{Y})} \leq M \tag{3.52}$$

for some constant M, and

ii) *the tangential cone condition is satisfied*

$$\|F(\tilde{\theta}) - F(\theta) - F'(\theta)(\tilde{\theta} - \theta)\|_{\mathcal{Y}} \leq c_{tcc}^{Red}\|F(\tilde{\theta}) - F(\theta)\|_{\mathcal{Y}} \tag{3.53}$$

for some small constant c_{tcc}^{Red}.

This is a consequence of the following two propositions, in which we combine the all-at-once versions of the tangential cone and boundedness conditons, respectively, with the assumed stability of S and its linearization.

Proposition 3.3.3. *Given $\mathcal{C}$ is the embedding $V \hookrightarrow Y$ and u_0 is independent of θ, the tangential cone condition in the reduced setting (3.53) follows from the one in the all-at-once setting (R4) if the linearized forward operator is boundedly invertible as in (R3) and S is well defined according to (R2).*

Proof. We begin by observing that the functions

$$\begin{aligned}
v &:= S(\theta) - S(\tilde{\theta}) \\
w &:= S'(\theta)h \\
z &:= S(\theta) - S(\tilde{\theta}) - S'(\theta)(\theta - \tilde{\theta})
\end{aligned}$$

solve the corresponding equations

$$\dot{v}(t) = f(\theta, S(\theta))(t) - f(\tilde{\theta}, S(\tilde{\theta}))(t) \qquad t \in (0,T), \quad v(0) = 0 \tag{3.54}$$
$$\dot{w}(t) = f'_u(\theta, S(\theta))w(t) + f'_\theta(\theta, S(\theta))h(t) \qquad t \in (0,T), \quad w(0) = 0 \tag{3.55}$$
$$\begin{aligned}
\dot{z}(t) &= f'_u(\theta, S(\theta))z(t) \\
&\quad + \big(- f'_u(\theta, S(\theta))v(t) - f'_\theta(\theta, S(\theta))(\theta - \tilde{\theta})(t) \\
&\quad + f(\theta, S(\theta))(t) - f(\tilde{\theta}, S(\tilde{\theta}))(t)\big) \\
&=: f'_u(\theta, S(\theta))z(t) + r(t) \qquad t \in (0,T), \quad z(0) = 0.
\end{aligned} \tag{3.56}$$

Hence, we end up with the following estimate, using the assumed bounded invertibility of the linearized problem (3.56) and the fact that $\mathcal{C}$ is the embedding $V \hookrightarrow Y$,

$$\begin{aligned}\|F(\theta) - F(\tilde{\theta}) - F'(\theta)(\theta - \tilde{\theta})\|_{\mathcal{Y}} &= \|S(\theta) - S(\tilde{\theta}) - S'(\theta)(\theta - \tilde{\theta})\|_{\mathcal{Y}} \\ &\leq C_{lin}\|r\|_{\mathcal{W}^*} \\ &\leq C_{lin}c_{tcc}^{AAO}\|F(\theta) - F(\tilde{\theta})\|_{\mathcal{Y}}, \end{aligned} \tag{3.57}$$

where $\|r\|_{\mathcal{W}^*}$ and c_{tcc}^{AAO} are respectively the left hand side and the constant in the all-at-once tangential cone estimate, applied to $u = S(\theta)$ and $\tilde{u} = S(\tilde{\theta})$. □

Remark 3.3.4. The inverse problem (3.19) with (3.22), (3.23), (3.24) can be written as a composition of the linear observation operator $\mathcal{C}$ and the nonlinear parameter-to-state map S. Such problems have been considered and analyzed in [52], but as opposed to that the inversion of our observation operator is ill-posed so the theory of [52] does not apply here.

Note that in (3.57), c_{tcc}^{AAO} must be sufficiently small such that the tangential cone constant in the reduced setting $c_{tcc}^{Red} := C_{lin}c_{tcc}^{AAO}$ fulfills the smallness condition required in convergence proofs as well. Moreover, we wish to emphasize that for the proof of Proposition 3.3.3, the constant C_{lin} does not need to be uniform but could as well depend on θ. Also the uniform boundedness condition on S from (R2) is not yet needed here.

Under further assumptions on the defining functions f, we also get existence and uniform boundedness of the linear operator $F'(\theta)$ as follows.

Proposition 3.3.5. *Let S be well defined and bounded according to (R2), and let (R1), (R3) be satisfied.*
Then $F'(\theta)$ is Gâteaux differentiable and its derivative given by

$$F'(\theta) : \mathcal{X} \to \mathcal{Y}, \qquad \text{where} \qquad F'(\theta)h = w \text{ solves } (3.55) \tag{3.58}$$

is uniformly bounded in $\mathcal{B}_\rho^{\mathcal{X}}(\theta_0)$.

Proof. For differentiablity of F relying on conditions (R1)-(R3), we refer to [93, Proposition 4.2]. Moreover, again using (R1)-(R3), for any $\theta \in \mathcal{B}_\rho^{\mathcal{X}}(\theta_0)$ we get

$$\begin{aligned}\|F'(\theta)h\|_{\mathcal{Y}} = \|S'(\theta)h\|_{\mathcal{Y}} &\leq C_{lin}\|f'_\theta(\theta, S(\theta))h\|_{L^2(0,T;W^*)} \\ &\leq C_{lin}\sqrt{T}\|f'_\theta(\theta, S(\theta))\|_{\mathcal{X}\to W^*}\|h\|_{\mathcal{X}} \\ &\leq C_{lin}\sqrt{T}L(M)\|h\|_{\mathcal{X}}\end{aligned}$$

for $M = C_S + \|\theta_0\|_{\mathcal{X}} + \rho$, where $L(M)$ is the Lipschitz constant in (R1) and C_{lin} is as in (R3). Above, we employ boundedness of S by C_S as assumed in (R2).
This proves uniform boundedness of $F'(\theta)$. □

We now discuss Assumption 3.3.1 in more detail.

Remark 3.3.6. For the case $V = W$.
We rely on the setting of a Gelfand triple $V \subseteq H \subseteq V^*$ for the general framework of nonlinear evolution equations. By this, (R2) can be fulfilled under the conditions suggested in Roubíček [100, Theorems 8.27, 8.31] (or Chapter 1, Assumption 1.3.5):

For every $\theta \in \mathcal{D}(F)$,

(S1) for almost $t \in (0,T)$, the mapping $-f(t,\theta,\cdot)$ is pseudomonotone, i.e., $-f(t,\theta,\cdot)$ is bounded and

$$\left.\begin{aligned}\liminf_{k\to\infty} \langle f(t,\theta,u_k), u_k - u\rangle \geq 0\\ u_k \rightharpoonup u\end{aligned}\right\} \Rightarrow \begin{cases}\forall v \in V : \langle f(t,\theta,u), u - v\rangle\\ \geq \limsup\limits_{k\to\infty} \langle f(t,\theta,u_k), u_k - v\rangle\end{cases}.$$

(S2) $-f(\cdot,\theta,\cdot)$ is semi-coercive, i.e.,

$$\forall v \in V, \forall^{a.e.} t \in (0,T) : \langle -f(t,\theta,v), v\rangle_{V^*,V} \geq C_0^\theta |v|_V^2 - C_1^\theta(t)|v|_V - C_2^\theta(t)\|v\|_H^2$$

for some $C_0^\theta > 0, C_1^\theta \in L^2(0,T), C_2^\theta \in L^1(0,T)$ and some seminorm $|.|_V$ satisfying
$\forall v \in V : \|v\|_V \leq c_{|.|}(|v|_V + \|v\|_H)$ for some $c_{|.|} > 0$.

(S3) f satisfies the growth condition

$$\exists \gamma^\theta \in L^2(0,T), \hbar^\theta : \mathbb{R} \to \mathbb{R} \text{ increasing} : \|f(t,\theta,v)\|_{V^*} \leq \hbar^\theta(\|v\|_H)(\gamma^\theta(t) + \|v\|_V)$$

and a condition for uniqueness of the solution, e.g.,

$$\forall u, v \in V, \forall^{a.e.} t \in (0,T) : \langle f(t,\theta,u) - f(t,\theta,v), u - v\rangle_{V^*,V} \leq \rho^\theta(t)\|u - v\|_H^2$$

for some $\rho^\theta \in L^1(0,T)$

and further conditions for $S(\theta) \in L^\infty(0,T;V)$, e.g, [100, Theorem 8.16, 8.18].

In case of linear $f(t,\theta,\cdot)$, S.1.-S.3. boil down to boundedness and semi-coercivity S.2. of $-f(\cdot,\theta,\cdot)$ according to [100, Theorem 8.27, 8.31, 8.28]. Alternatively, one can observe that linear boundedness implies the growth condition in S.3. with $\gamma^\theta = 0, \hbar^\theta = \|f(\cdot,\theta)\|$, and S.2. implies the rest of S.3. with $\rho^\theta = C_2^\theta$ if $C_1^\theta \leq 0$ as $C_0^\theta > 0$. The pseudomonotonicity assumption S.1., which guarantees week convergence of $f(\cdot,\theta,u_k)$ to $f(\cdot,\theta,u)$ when the approximation solution sequence u_k converges weakly to u, can be replaced by weak continuity of $f(\cdot,\theta,\cdot)$ which holds in this linear bounded case.

Treating the linearized problem (3.49)-(3.50) as an independent problem, we can impose on $f'_u(\theta, S(\theta))$ the boundedness and semi-coercivity properties, and (R3) therefore follows from [100, Theorems 8.27, 8.31].

Remark 3.3.7. For general spaces V, W.
Some examples even in case $V \neq W$ allow using the results quoted in Remark 3.3.6 with an appropriately chosen Gelfand triple, see, e.g., Section 3.3.1 below.

When dealing with linear and quasilinear parabolic problems, detailed discussions for unique exsistence of the solution are exposed in the books, e.g., of Evans [32], Ladyzhenskaya et al. [81], Pao [95]. If constructing the solution to the initial value problem through the semigroup approach, one can find several results, e.g., from Evans [32], Pazy [96] combined with the elliptic results from Ladyzhenskaya et al. [82].

Addressing (R3), a possible strategy is using the following dual argument.
Suppose W is reflexive and z is a solution to the problem (3.49)-(3.50), then by the Hahn-Banach Theorem

$$\begin{aligned}
\|z\|_{L^2(0,T;V)} &= \sup_{\|\phi\|_{L^2(0,T;V^*)}\leq 1} \int_0^T \langle z, \phi\rangle_{V,V^*} dt \\
&= \sup_{\|\phi\|_{L^2(0,T;V^*)}\leq 1} \int_0^T \langle z, -\dot{p} - f_u'(\theta, S(\theta))^* p\rangle_{V,V^*} dt \\
&= \sup_{\|\phi\|_{L^2(0,T;V^*)}\leq 1} \int_0^T \langle \dot{z} - f_u'(\theta, S(\theta))z, p\rangle_{W^*,W} dt \\
&= \sup_{\|\phi\|_{L^2(0,T;V^*)}\leq 1} \int_0^T \langle b, p\rangle_{W^*,W} dt \\
&\leq \sup_{\|\phi\|_{L^2(0,T;V^*)}\leq 1} \|b\|_{L^2(0,T;W^*)} \|p\|_{L^2(0,T;W)},
\end{aligned}$$

where

$$f_u'(\theta, S(\theta))(t) : V \to W^*, \qquad f_u'(\theta, S(\theta))(t)^* : W^{**} = W \to V^*,$$

and p solves the adjoint equation

$$-\dot{p}(t) = f_u'(\theta, S(\theta))^* p(t) + \phi(t) \quad t \in (0,T) \tag{3.59}$$
$$p(T) = 0. \tag{3.60}$$

If in the adjoint problem, the estimate

$$\|p\|_{L^2(0,T;W)} \leq \tilde{C}_{lin} \|\phi\|_{L^2(0,T;V^*)} \tag{3.61}$$

holds for some uniform constant $\tilde{C}_{lin}$, then we obtain

$$\|z\|_{\mathcal{Y}} \leq \|\mathcal{C}\|_{V\to Y} \|z\|_{L^2(0,T;V)} \leq \|\mathcal{C}\|_{V\to Y} \tilde{C}_{lin} \|b\|_{\mathcal{W}^*}. \tag{3.62}$$

Thus (R3) is fulfilled.

So we can replace (R3) by

(R3-dual) *Continuous dependence on data of the solution to the adjoint linearized problem* associated with zero final condition, i.e., there exists a constant $\widetilde{C}_{lin}$ such that for all $\theta \in \mathcal{B}_\rho^{\mathcal{X}}(\theta^0)$, $\phi \in L^2(0,T,V^*)$, and any p solving (3.59)-(3.60), the estimate (3.61) holds.

In the following sections, we examine the specific examples introduced in the introduction in the relevant function space setting

$$\mathcal{X} = L^p(\Omega) \quad \text{or} \quad \mathcal{X} = W^{1,p}(\Omega) \qquad p \in [1,\infty] \tag{3.63}$$
$$Y = L^q(\Omega) \qquad q \in [1,\bar{q}] \tag{3.64}$$
$$\tilde{\mathcal{U}} = \{u \in L^\infty(0,T;V) : \dot{u} \in L^2(0,T;W^*)\}, \tag{3.65}$$

where V, W will be chosen subject to the particular example, where $\hat{q}$ is the maximum power allowing $V \hookrightarrow L^{\hat{q}}(\Omega)$, and $\bar{q} \leq \hat{q}$ is the maximum power such that (3.51) in (R3) holds.

3.3.1 Identification of a potential

We investigate this problem in the function spaces

$$\mathcal{D}(F) = \mathcal{X} = L^p(\Omega), \qquad Y = L^q(\Omega), \qquad V = L^2(\Omega), \qquad W = H^2(\Omega) \cap H_0^1(\Omega).$$

Now we verify the conditions proposed in Assumption 3.3.1.

(R1) *Local Lipschitz continuity of f:*

Applying Hölder's inequality, we have

$$\begin{aligned}\|f(\tilde{c},\tilde{u}) - f(c,u)\|_{W^*} = \|\tilde{c}\tilde{u} - cu\|_{W^*} &= \sup_{\|w\|_W \leq 1} \int_\Omega (\tilde{c}\tilde{u} - cu) w dx \\ &\leq \sup_{\|w\|_W \leq 1} \|w\|_W C_{W\to L^{\bar{p}}} \left(\int_\Omega |\tilde{c}(\tilde{u} - u) + (\tilde{c} - c)u|^{\bar{p}^*} dx \right)^{\frac{1}{\bar{p}^*}} \\ &\leq C_{W\to L^{\bar{p}}} (\|\tilde{c}\|_{L^p} \|\tilde{u} - u\|_{L^r} + \|\tilde{c} - c\|_{L^p} \|u\|_{L^r}) \\ &\leq L(M)(\|\tilde{u} - u\|_V + \|\tilde{c} - c\|_{\mathcal{X}})\end{aligned}$$

with the dual index $\bar{p}^* = \frac{\bar{p}}{\bar{p}-1}$ and $r = \frac{\bar{p}p}{\bar{p}p - p - \bar{p}}$, $L(M) = C_{W\to L^{\bar{p}}} C_{V\to L^r}(\|u\|_V + \|\tilde{u}\|_V + \|c\|_{\mathcal{X}} + \|\tilde{c}\|_{\mathcal{X}}) + 1$. Above, we invoke the continuous embbedings through the constants $C_{W\to L^{\bar{p}}}, C_{V\to L^r}$, where $\bar{p}$ denotes the maximum power allowing $W \subseteq L^{\bar{p}}$. Thus, we are supposing

$$p \geq \max\left\{ \frac{2\bar{p}}{\bar{p}-2}, \frac{\bar{p}}{\bar{p}-1} \right\} = \frac{2\bar{p}}{\bar{p}-2} \text{ and } 2 - \frac{d}{2} \succeq -\frac{d}{\bar{p}} \tag{3.66}$$

in order to guarantee $V = L^2(\Omega) \hookrightarrow L^r(\Omega)$ and $W = H^2(\Omega) \cap H_0^1(\Omega) \hookrightarrow L^{\bar{p}}(\Omega)$

(R2) *Well-definedness and boundedness of the parameter-to-state map:*

Verifying boundedness and semi-coercivity conditions with the Gelfand triple $H_0^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$ (while remaining with $V = L^2(\Omega)$ in the definition of the space $\tilde{\mathcal{U}}$) shows that, for $u_0 \in L^2(\Omega), \varphi \in L^2(0,T;H^{-1}(\Omega))$ the initial value problem (3.4)-(3.6) admits a unique solution $u \in W(0,T) := \{u \in L^2(0,T;H_0^1(\Omega)) : \dot{u} \in L^2(0,T;H^{-1}(\Omega)\} \subset \{u \in L^\infty(0,T;L^2(\Omega)) : \dot{u} \in L^2(0,T;H^{-2}(\Omega))\} = \tilde{\mathcal{U}}$.

Indeed, semi-coercivity is deduced as follows. For

$$p \geq 2, \quad d \leq 3, \tag{3.67}$$

we see

$$\begin{aligned}
\int_\Omega cu^2 dx &\leq \|c\|_{L^2(\Omega)} \left(\int_\Omega u^4 dx\right)^{\frac{1}{2}} \leq \|c\|_{L^2(\Omega)} \left(\int_\Omega u^2 dx\right)^{\frac{1}{4}} \left(\int_\Omega u^6 dx\right)^{\frac{1}{4}} \\
&\leq \|c\|_{L^2(\Omega} \|u\|_{L^2(\Omega)}^{\frac{1}{2}} \|u\|_{L^6(\Omega)}^{\frac{3}{2}} \\
&\leq C_{H_0^1 \to L^6} \|c\|_{L^2(\Omega} \|u\|_{L^2(\Omega)}^{\frac{1}{2}} \|u\|_{H_0^1(\Omega)}^{\frac{3}{2}} \\
&\leq C_{H_0^1 \to L^6} \|c\|_{L^2(\Omega)} \left(\frac{1}{4\epsilon} \|u\|_{L^2(\Omega)} \|u\|_{H_0^1(\Omega)} + \epsilon \|u\|_{H_0^1(\Omega)}^2\right) \\
&\leq C_{H_0^1 \to L^6} \|c\|_{L^2(\Omega)} \left(\frac{1}{16\epsilon\epsilon_1} \|u\|_{L^2(\Omega)}^2 + \frac{\epsilon_1}{4\epsilon} \|u\|_{H_0^1(\Omega)}^2 + \epsilon \|u\|_{H_0^1(\Omega)}^2\right),
\end{aligned} \tag{3.68}$$

which yields semi-coercivity

$$\begin{aligned}
&\langle -f(t,u,c), u\rangle_{H^{-1},H_0^1} = \int_\Omega (-\Delta u + cu) u dx \\
&\geq \left(1 - C_{H_0^1 \to L^6} \|c\|_{L^2(\Omega)} \left(\frac{\epsilon_1}{4\epsilon} + \epsilon\right)\right) \|u\|_{H_0^1(\Omega)}^2 - \frac{C_{H_0^1 \to L^6}}{16\epsilon\epsilon_1} \|c\|_{L^2(\Omega)} \|u\|_{L^2(\Omega)}^2, \\
&=: C_0^c \|u\|_{H_0^1(\Omega)}^2 + C_1^c \|u\|_{L^2(\Omega)}^2,
\end{aligned}$$

where the constant C_0^c is positive if choosing $\epsilon_1 < \epsilon$ and ϵ, ϵ_1 sufficiently small. Boundedness of f can be concluded from

$$\begin{aligned}
\|-f(t,c,u)\|_{H^{-1}(\Omega)} &= \sup_{\|v\|_{H_0^1} \leq 1} \int_\Omega (-\Delta u + cu) v dx \\
&\leq \sup_{\|v\|_{H_0^1} \leq 1} \left(\|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)} + C_{H_0^1 \to L^6} C_{H_0^1 \to L^3} \|c\|_{L^2(\Omega)} \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)}\right) \\
&\leq C \|c\|_{L^2(\Omega)} \|u\|_{H^1(\Omega)}.
\end{aligned}$$

Moreover, by the triangle inequality: $\|c\|_{L^2(\Omega)} \leq \|c^0\|_{L^2(\Omega)} + \|c - c^0\|_{L^2(\Omega)} \leq \|c^0\|_{L^2(\Omega)} + \rho$, semi-coercivity of f is satisfied with the constants C_0, C_1 now depending only on the point c^0. This hence gives us uniform boundedness of S on the ball $\mathcal{B}_\rho^{\mathcal{X}}(c^0)$.

(R3) *Continuous dependence on data of the solution to the linearized problem with zero initial data:*

We use the duality argument mentioned in Remark 3.3.7. To do so, we need to prove existence of the adjoint state $p \in L^2(0,T;W)$ and the associated estimate (R3-dual).

Initially, by the transformation $v = e^{-\lambda t}p$ and putting $\tau = T - t$, the adjoint problem (3.59)-(3.60) is equivalent to

$$\dot{v}(t) - \Delta v(t) + (\lambda + c)v(t) = e^{-\lambda t}\phi(t) \qquad t \in (0,T) \tag{3.69}$$
$$v(0) = 0. \tag{3.70}$$

We note that this problem with $c = \hat{c} \in L^\infty(\Omega), \lambda + \hat{c} > -C_{PF}$, the constant in the Poincaré-Friedrichs inequality, $\phi \in L^2(0,T;L^2(\Omega)), \partial\Omega \in C^2$, admits a unique solution in $L^2(0,T;H^2(\Omega)\cap H_0^1(\Omega))$ [32, Section 7.1.3, Theorem 5] [3] and the operator $\frac{d}{dt} - \Delta + (\lambda + \hat{c}) : L^2(0,T;H^2(\Omega) \cap H_0^1(\Omega)) \to L^2(0,T;L^2(\Omega)) \times H^1(\Omega), p \mapsto (\phi, p_0)$ is boundedly invertible.

Suppose u solves (3.69)-(3.70), by the identity

$$\begin{aligned}
&\dot{u} - \Delta u + (\lambda + c)u = e^{-\lambda t}\phi \quad \Leftrightarrow \quad \dot{u} - \Delta u + (\lambda + \hat{c})u = e^{-\lambda t}\phi + (\hat{c} - c)u \\
&u = \left(\frac{d}{dt} - \Delta + (\lambda + \hat{c})\right)^{-1}\left[e^{-\lambda t}\phi + (\hat{c} - c)u\right] \\
&\quad =: Tu,
\end{aligned}$$

we observe that $T : L^2(0,T;H^2(\Omega) \cap H_0^1(\Omega)) \to L^2(0,T;H^2(\Omega) \cap H_0^1(\Omega))$ is a contraction

$$\begin{aligned}
&\|T(u - v)\|_{L^2(0,T;H^2\cap H_0^1)} \\
&\leq \left\|\left(\frac{d}{dt} - \Delta + (\lambda + \hat{c})\right)^{-1}\right\|_{L^2(0,T;L^2(\Omega))\to L^2(0,T;H^2\cap H_0^1)} \|(\hat{c} - c)(u - v)\|_{L^2(0,T;L^2(\Omega))} \\
&\leq C^{\hat{c}}\|\hat{c} - c\|_{L^p}\|u - v\|_{L^2(0,T;L^{\frac{2p}{p-2}}(\Omega))} \\
&\leq C\epsilon\|u - v\|_{L^2(0,T;H^2\cap H_0^1)},
\end{aligned} \tag{3.71}$$

where $C\epsilon < 1$ if we assume $\hat{c} = c^0 \in L^\infty(\Omega)$ and ρ is sufficiently small. In some case, smallness of ρ can be omitted (discussed at the end of (R3)). Estimate (3.71) holds provided

$$W = H^2(\Omega) \cap H_0^1(\Omega) \hookrightarrow L^{\frac{2p}{p-2}}(\Omega) \text{ i.e.,} \quad p \geq \frac{2\bar{p}}{\bar{p} - 2}. \tag{3.72}$$

[3] where smoothness of the domain can be slightly relaxed to $C^{1,1}$ as assumed here, see, e.g., [40]

Thus, for $\phi \in L^2(0,T;L^2(\Omega))$ there exists a unique solution $v \in L^2(0,T;H^2(\Omega)\cap H_0^1(\Omega))$ to the problem (3.69)-(3.70), which implies $p = e^{\lambda t}v \in L^2(0,T;H^2(\Omega)\cap H_0^1(\Omega))$ is the solution to the adjoint problem (3.59)-(3.60).

Observing that p solves

$$\begin{aligned}&\dot{p}(t) - \Delta p(t) + \hat{c}p(t) = (\hat{c}-c)p(t) + \phi(t) \qquad t \in (0,T)\\ &p(0) = 0,\end{aligned}$$

employing again [32, Section 7.1.3 , Theorem 5] and smallness of ρ yields

$$\begin{aligned}\|p\|_{L^2(0,T;W)} &\leq C(\|(\hat{c}-c)p\|_{L^2(0,T;L^2(\Omega))} + \|\phi\|_{L^2(0,T;L^2(\Omega))})\\ &\leq C(2\rho\|p\|_{L^2(0,T;H^2\cap H_0^1)} + \|\phi\|_{L^2(0,T;L^2(\Omega))})\\ &\leq C\|\phi\|_{L^2(0,T;V^*)}\end{aligned} \tag{3.73}$$

with some constant C independent of $\theta \in \mathcal{B}_\rho^{\mathcal{X}}(c^0)$. This yields (R3-dual) with $\bar{q} = 2$.

If $d = 1, p \geq 2$ or $d = 2, p > 2$ or $d = 3, p \geq \frac{12}{5}$, the smallness condition on ρ can be omitted. Indeed, for $d = 3, p \geq \frac{12}{5}$ testing the adjoint equation by $-\Delta p$ yields

$$\begin{aligned}&\int_\Omega -\dot{p}\Delta p + (\Delta p)^2 dx = \int_\Omega (cp-\phi)\Delta p dx\\ &\frac{1}{2}\frac{d}{dt}\|\nabla p\|^2_{L^2(\Omega)} + \|\Delta p\|^2_{L^2(\Omega)} \leq \frac{1}{2}\|\Delta p\|^2_{L^2(\Omega)} + \|\phi\|^2_{L^2(\Omega)} + \|cp\|^2_{L^2(\Omega)}\\ &\frac{1}{2}\frac{d}{dt}\|\nabla p\|^2_{L^2(\Omega)} + \frac{1}{2}\|\Delta p\|^2_{L^2(\Omega)} \leq \|\phi\|^2_{L^2(\Omega)} + \|c\|^2_{L^p(\Omega)}\left(\int_\Omega p^{\frac{p}{p-2}+\frac{p}{p-2}}dx\right)^{\frac{p-2}{p}}\\ &\qquad\leq \|\phi\|^2_{L^2(\Omega)} + \|c\|^2_{L^p(\Omega)}\|p\|_{L^6(\Omega)}\|p\|_{L^\infty(\Omega)}|\Omega|^{\frac{5p-12}{6p}}\\ &\qquad\leq \|\phi\|^2_{L^2(\Omega)} + (\|c^0\|^2_{\mathcal{X}} + \rho^2)|\Omega|^{\frac{5p-12}{6p}}\left(\frac{C^2_{H_0^1\to L^6}}{4\epsilon}\|\nabla p\|^2_{L^2(\Omega)}\right.\\ &\qquad\qquad\left. + \epsilon C^2_{H^2\cap H_0^1\to L^\infty}\left(\|\Delta p\|^2_{L^2(\Omega)} + \|\nabla p\|^2_{L^2(\Omega)}\right)\right).\end{aligned} \tag{3.74}$$

where in the last estimate we apply Young's inequality. Choosing ϵ sufficiently small allows us to subtract the term involving $\|\Delta p\|^2_{L^2(\Omega)}$ on the right hand side from the one on the left hand side and get a positive coefficient in front. Here, the choice of ϵ depends only on the constants $c^0, \rho, \Omega, C_{H^2\cap H_0^1\to L^\infty}$.

It is also obvious that, if $d < 3$, in the second line of the above calculation, we can directly estimate as follow

$$\begin{aligned}&d = 1, p \geq 2 : \|cp\|^2_{L^2(\Omega)} \leq \|c\|^2_{L^2(\Omega)}\|p\|^2_{L^\infty(\Omega)} \leq C^2_{H_0^1\to L^\infty}\|c\|^2_{L^2(\Omega)}\|\nabla p\|^2_{L^2(\Omega)}\\ &d = 2, p > 2 : \|cp\|^2_{L^2(\Omega)} \leq \|c\|^2_{L^p(\Omega)}\|p\|^2_{L^{\frac{2p}{p-2}}(\Omega)} \leq C^2_{H_0^1\to L^{\frac{2p}{p-2}}}\|c\|^2_{L^p(\Omega)}\|\nabla p\|^2_{L^2(\Omega)}.\end{aligned} \tag{3.75}$$

Employing firstly Gronwall-Bellman inequality with initial data $\nabla p(0) = 0$, then taking the integral on $[0,T]$, we obtain

$$\|p\|_{L^\infty(0,T;H^1(\Omega))} + \|\Delta p\|_{L^2(0,T;L^2(\Omega))} \leq C\|\phi\|_{L^2(0,T;L^2(\Omega))} \tag{3.76}$$

with the constant C depending only on c^0, ρ. This estimate is valid for all $c \in \mathcal{B}_\rho^{\mathcal{X}}(c^0)$. Since the adjoint problem has the same form as the original problem, applying (3.76) in (3.71) we can relax $\hat{c}$, by means of without fixing $\hat{c} = c^0$ but choosing it sufficiently close to c since $\overline{L^\infty(\Omega)} = L^p(\Omega), |\Omega| < \infty$ to have $C^{\hat{c}}\epsilon \leq C\epsilon$ arbitrarily small with constant C as in (3.76). Therefore the constraint on smallness of ρ can be omitted in these cases.

(R4) *All-at-once tangential cone condition:*

According to (3.36), (B.107) with $s = 0$, $t = 2$, $m = n = 2$, this follows if

$$\frac{p}{p-1} \leq q \leq \hat{q} \geq 2 \text{ and } 2 - \frac{d}{2} \succeq -\frac{d(p-1)}{p} + \frac{d}{q}.$$

Corollary 3.3.8. *Assume $u_0 \in L^2(\Omega), \varphi \in L^2(0,T;H^{-1}(\Omega))$, and*

$$\begin{aligned}\mathcal{D}(F) = \mathcal{X} = L^p(\Omega), \quad Y = L^q(\Omega), \quad V = L^2(\Omega), \quad W = H^2(\Omega) \cap H_0^1(\Omega)\\ p \geq 2, \quad q \in \left[\underline{q}, 2\right], \quad d \leq 3\end{aligned} \tag{3.77}$$

with $\underline{q} = \max\left\{\frac{p}{p-1}\,,\ \min_{q\in[1,\infty]}\left\{2 - \frac{d}{2} \succeq -\frac{d(pq-p-q)}{pq}\right\}\right\}$.

Then F defined by $F(c) = u$ solving (3.4)-(3.6) satisfies the tangential cone condition (3.53) with a uniformly bounded operator $F'(c)$ defined by (3.58), see also [49] for the static case.

Remark 3.3.9. This allows a full Hilbert space setting of X and Y by choosing $p = q = 2$ as long as $d \leq 3$.

3.3.2 Identification of a diffusion coefficient

We pose this problem in the function spaces

$$\mathcal{X} = W^{1,p}(\Omega), \quad Y = L^q(\Omega), \quad V = L^2(\Omega), \quad W = H^2(\Omega) \cap H_0^1(\Omega) \qquad p > d \tag{3.78}$$

so that $\mathcal{X} \hookrightarrow L^\infty(\Omega)$ and define the domain of F by

$$\mathcal{D}(F) = \{a \in \mathcal{X} : a \geq \underline{a} > 0 \text{ a.e. on } \Omega\}. \tag{3.79}$$

Now we examine the conditions (R1)-(R3).

(R1) *Local Lipschitz continuity of f:*

$$
\begin{aligned}
\| -\nabla\cdot\Big(\tilde a\nabla\tilde u\Big) + \nabla\cdot\Big(a\nabla u\Big)\|_{W^*}&\\
&= \sup_{\|w\|_W\le 1}\int_\Omega (\tilde a\nabla\tilde u - a\nabla u)\nabla w dx\\
&= \sup_{\|w\|_W\le 1}\int_\Omega (a\nabla(\tilde u-u) + (\tilde a - a)\nabla\tilde u)\nabla w dx\\
&= \sup_{\|w\|_W\le 1}\int_\Omega (\tilde u-u)(\nabla a\nabla w + a\Delta w) + \tilde u(\nabla(\tilde a-a)\nabla w + (\tilde a-a)\Delta w)dx\\
&\le \sup_{\|w\|_W\le 1}\int_\Omega (\|\tilde u-u\|_{L^2}\|\nabla a\|_{L^p} + \|\tilde u\|_{L^2}\|\nabla(a-a)\|_{L^p})\|\nabla w\|_{L^{\frac{2p}{p-2}}}\\
&\qquad\qquad + (\|\tilde u-u\|_{L^2}\|a\|_{L^\infty} + \|\tilde u\|_{L^2}\|a-a\|_{L^\infty})\|\Delta w\|_{L^2}dx\\
&\le L(M)(\|\tilde u-u\|_V + \|\tilde a - a\|_{\mathcal X})
\end{aligned}
$$

with $M = \Big(C_{W\to W^{1,\frac{2p}{p-2}}} + C_{\mathcal X\to L^\infty}\Big)(\|u\|_V + \|\tilde u\|_V + \|c\|_{\mathcal X} + \|\tilde c\|_{\mathcal X})$, subject to the constraint

$$
W = H^2(\Omega)\cap H^1_0(\Omega)\hookrightarrow L^{\frac{2p}{p-2}}(\Omega) \text{ i.e.,}\quad p\ge\frac{2\bar p}{\bar p-2}. \tag{3.80}
$$

(R2) *Well-definedness and boundedness of the parameter-to-state map:*

A straightforward verification of boundedness and coercivity gives unique existence of the solution $u\in W(0,T)\subset\tilde{\mathcal U}$ for $a\in\mathcal D(F), \varphi\in L^2(0,T;H^{-1}(\Omega)), u_0\in L^2(\Omega)$.

Similarly to the c-problem, the fact that the semi-coercivity property of f holds

$$
\langle -f(t,a,u),u\rangle_{H^{-1},H^1_0} = \int_\Omega -\nabla\cdot(a\nabla u)udx \ge \underline{a}\|u\|_{H^1_0(\Omega)}
$$

with the coefficient $\underline{a}$ being independent of a shows uniform boundedness of S.

(R3) *Continuous dependence on data of the solution to the linearized problem with zero initial data:*

We employ the result in [32, Section 7.1.3, Theorem 5] with noting that the actual smoothness condition needed for the coefficient is that, a is differentiable a.e on Ω and $a\in W^{1,\infty}(\Omega)$ rather than $a\in C^1(\Omega)$. From the observation $a\in\mathcal D(F) = W^{1,p}(\Omega), p>d$ is differentiable a.e and the fact that $W^{1,\infty}(\Omega)$ is dense in $W^{1,p}(\Omega)$, it enables us to imitate the contraction scenario and the dual argument as in the c-problem.

Taking u, v solving (3.7)-(3.9), we see

$$T : L^2(0,T;H^2(\Omega)\cap H^1_0(\Omega)) \to L^2(0,T;H^2(\Omega)\cap H^1_0(\Omega))$$
$$T = \left(\frac{d}{dt} - \nabla\cdot\Big(\hat{a}\nabla\Big)\right)^{-1}\nabla\cdot\Big((a-\hat{a})\nabla\Big)$$

is a contraction

$$\begin{aligned}
&\|T(u-v)\|_{L^2(0,T;H^2\cap H^1_0)}\\
&\leq \left\|\left(\frac{d}{dt} - \nabla\cdot\Big(\hat{a}\nabla\Big)\right)^{-1}\right\|_{L^2(0,T;L^2(\Omega))\to L^2(0,T;H^2\cap H^1_0)} \|\nabla\cdot\Big((a-\hat{a})\nabla(u-v)\Big)\|_{L^2(0,T;L^2(\Omega))}\\
&\leq C^{\hat{a}}\|\hat{a}-a\|_{\mathcal{X}}\|u-v\|_{L^2(0,T;H^2\cap H^1_0)}\\
&\leq C\epsilon\|u-v\|_{L^2(0,T;H^2\cap H^1_0)},
\end{aligned} \tag{3.81}$$

where $C\epsilon < 1$ if we assume $\hat{a} = a^0 \in W^{1,\infty}(\Omega)$ and ρ is sufficiently small. If the index p is large enough, smallness of ρ can be omitted (discussed at the end of (R3)). Therefore, given $\phi \in L^2(0,T;L^2(\Omega))$, the adjoint state $p \in L^2(0,T;H^2\cap H^1_0)$ uniquely exists.

We also have the estimate

$$\begin{aligned}
\|p\|_{L^2(0,T;W)} &\leq C\|\nabla\cdot\Big((a-\hat{a})\nabla p\Big)\|_{L^2(0,T;L^2(\Omega))} + \|\phi\|_{L^2(0,T;L^2(\Omega))})\\
&\leq C(2\rho\|p\|_{L^2(0,T;H^2\cap H^1_0)} + \|\phi\|_{L^2(0,T;L^2(\Omega))})\\
&\leq C\|\phi\|_{L^2(0,T;V^*)},
\end{aligned}$$

which proves continuous dependence of p on $\phi \in L^2(0,T;V^*)$, consequently, continuous dependence of the solution $z \in L^2(0,T;V)$ on the data $b \in L^2(0,T;W^*)$ in (3.49)-(3.50). Here smallness of ρ is assumed.
If $p \geq 4$, smallness of ρ is not required. To verify this, we test the adjoint equation by $-\Delta p$

$$\begin{aligned}
&\int_\Omega -\dot{p}\Delta p + a(\Delta p)^2 dx = \int_\Omega (-\nabla a\nabla p - \phi)\Delta p dx\\
&\frac{1}{2}\frac{d}{dt}\|\nabla p\|^2_{L^2(\Omega)} + \underline{a}\|\Delta p\|^2_{L^2(\Omega)} \leq \frac{\underline{a}}{2}\|\Delta p\|^2_{L^2(\Omega)} + \frac{1}{\underline{a}}\|\phi\|^2_{L^2(\Omega)} + \frac{1}{\underline{a}}\|\nabla a\nabla p\|^2_{L^2(\Omega)}\\
&\frac{1}{2}\frac{d}{dt}\|\nabla p\|^2_{L^2(\Omega)} + \frac{\underline{a}}{2}\|\Delta p\|^2_{L^2(\Omega)} \leq \frac{1}{\underline{a}}\|\phi\|^2_{L^2(\Omega)} + \frac{1}{\underline{a}}\|\nabla a\nabla p\|^2_{L^2(\Omega)},
\end{aligned} \tag{3.82}$$

where the last term on the right hand side can be estimated as in (3.68) of the c-problem with $(\nabla a)^2$ in place of c, ∇p in place of u and the assumption

$\mathcal{X} \hookrightarrow W^{1,4}(\Omega)$

$$\begin{aligned}
&\frac{1}{\underline{a}}\|\nabla a \nabla p\|^2_{L^2(\Omega)} \\
&\leq \frac{C_{H_0^1\to L^6}}{\underline{a}}\|\nabla a\|^2_{L^4(\Omega)}\left(\frac{1}{16\epsilon\epsilon_1}\|\nabla p\|^2_{L^2(\Omega)} + \left(\frac{\epsilon_1}{4\epsilon}+\epsilon\right)\|\nabla p\|^2_{H_0^1(\Omega)}\right) \\
&\leq \frac{2C_{H_0^1\to L^6}}{\underline{a}}(\|a^0\|^2_{\mathcal{X}}+\rho^2)\left(\frac{1}{16\epsilon\epsilon_1}\|\nabla p\|^2_{L^2(\Omega)} + \left(\frac{\epsilon_1}{4\epsilon}+\epsilon\right)\|\Delta p\|^2_{L^2(\Omega)}\right). \qquad (3.83)
\end{aligned}$$

Choosing $\epsilon_1 < \epsilon$, and ϵ_1, ϵ sufficiently small such that we can move the term involving $\|\Delta p\|^2_{L^2(\Omega)}$ from the right hand side to the left hand side of (3.82). Note that, this choice of ϵ_1, ϵ is just subject to a^0 and ρ.
Proceeding similarly to the c-problem, meaning applying Gronwall-Bellman inequality then taking the integral on $[0,T]$, we obtain

$$\|p\|_{L^\infty(0,T;H^1(\Omega))} + \|\Delta p\|_{L^2(0,T;L^2(\Omega))} \leq C\|\phi\|^2_{L^2(0,T;L^2(\Omega))} \qquad (3.84)$$

with a constant C depending only on a^0, ρ.

Observing the similarity in the form of the adjoint problem and the original problem, invoking the uniform bound (3.84) w.r.t parameter a and the fact $\overline{W^{1,\infty}(\Omega)} = W^{1,p}(\Omega)$ one can eliminate the need of smallness of ρ.

(R4) *All-at-once tangential cone condition:*

According to (3.36), (B.109) with $s=0$, $t=2$, $m=n=2$, we require

$$\frac{p}{p-1} \leq q \leq \hat{q} \geq 2 \text{ and } 1-\frac{d}{2} \succeq -\frac{d(p-1)}{p} + \frac{d}{q} \text{ and } -\frac{d}{2} \geq -d + \frac{d}{p} - 1\,.$$

Corollary 3.3.10. *Assume $u_0 \in L^2(\Omega), \varphi \in L^2(0,T;H^{-1}(\Omega))$, and*

$$\begin{aligned}
\mathcal{X} = W^{1,p}(\Omega), \qquad Y = L^q(\Omega), \qquad & V = L^2(\Omega), \qquad W = H^2(\Omega)\cap H_0^1(\Omega) \\
& p \geq 2, \quad q \in [\underline{q}, 2]\,, \quad d < p,
\end{aligned} \qquad (3.85)$$

where $\underline{q} = \max\left\{\frac{p}{p-1}\,,\ \min_{q\in[1,\infty]}\left\{1-\frac{d}{2} \succeq -\frac{d(p-1)}{p} + \frac{d}{q} \wedge -\frac{d}{2} \geq -d + \frac{d}{p} - 1\right\}\right\}.$

Then F defined by $F(a) = u$ solving (3.7)-(3.9) satisfies the tangential cone condition (3.53) with a uniformly bounded operator $F'(a)$ defined by (3.58).

Remark 3.3.11. This yields the possibility of a full Hilbert space setting $p=q=2$ of X and Y in case $d=1$, see also [47] and, for the static case, [49].

3.3.3 Inverse source problem with quadratic nonlinearity

By the transformation $U := e^u$, the initial-value problem (3.10)–(3.12) can be converted into an inverse potential problem as considered in Section 3.3.1

$$\dot{U} - \Delta U + \theta U = 0 \qquad (t,x) \in (0,T) \times \Omega \tag{3.86}$$
$$U_{|\partial\Omega} = 1 \qquad t \in (0,T) \tag{3.87}$$
$$U(0) = U_0 \qquad x \in \Omega \tag{3.88}$$

with $U_0 = e^{u_0}$. Thus, in principle it is covered by the analysis from the previous section, as long as additionally positivity of U can be established. So the purpose of this section is to investigate whether we can allow different function spaces X, Y by directly considering (3.10)–(3.12) instead of (3.86)–(3.88).

We show that f verifies the hypothesis proposed for the tangential cone condition in the reduced setting on the function spaces

$$\mathcal{X} = L^p(\Omega), \qquad Y = L^q(\Omega), \qquad V = W = H^2(\Omega) \cap H_0^1(\Omega). \tag{3.89}$$

(R1) *Local Lipschitz continuity of f:*

$$\begin{aligned}
\| - |\nabla\tilde{u}|^2 + |\nabla u|^2 - \tilde{\theta} + \theta\|_{W^*} &= \sup_{\|w\|_W \le 1} \int_\Omega \left(\nabla(u-\tilde{u}) \cdot \nabla(u+\tilde{u}) - \tilde{\theta} + \theta\right) w dx \\
&\le C_{W\to L^{\bar{p}}} \left(\|(\nabla(u-\tilde{u}) \cdot \nabla(u+\tilde{u})\|_{L^{\frac{\bar{p}}{\bar{p}-1}}} + \|\theta - \tilde{\theta}\|_{L^{\frac{\bar{p}}{\bar{p}-1}}} \right) \\
&\le C_{W\to L^{\bar{p}}} \left(\|\nabla(u-\tilde{u})\|_{L^{\frac{2\bar{p}}{\bar{p}-1}}} \|\nabla(u+\tilde{u})\|_{L^{\frac{2\bar{p}}{\bar{p}-1}}} + \|\theta - \tilde{\theta}\|_{L^{\frac{\bar{p}}{\bar{p}-1}}} \right) \\
&\le C_{W\to L^{\bar{p}}} \left(C^2_{V\to W^{1,\frac{2\bar{p}}{\bar{p}-1}}} \|u-\tilde{u}\|_V \|u+\tilde{u}\|_V + C_{\mathcal{X}\to L^{\frac{\bar{p}}{\bar{p}-1}}} \|\theta - \tilde{\theta}\|_{\mathcal{X}} \right).
\end{aligned}$$

We can chose $L(M) = C_{W\to L^{\bar{p}}} \left(C^2_{V\to W^{1,\frac{2\bar{p}}{\bar{p}-1}}} (\|u\|_V + \|\tilde{u}\|_V) + C_{\mathcal{X}\to L^{\frac{\bar{p}}{\bar{p}-1}}} \right) + 1$, under the conditions

$$\begin{aligned}
&V = H^2(\Omega) \cap H_0^1(\Omega) \hookrightarrow W^{1,\frac{2\bar{p}}{\bar{p}-1}}(\Omega) \text{ i.e.,} \quad 1 - \frac{d}{2} \ge -\frac{d(\bar{p}-1)}{2\bar{p}} \\
&\mathcal{X} = L^p(\Omega) \hookrightarrow L^{\frac{\bar{p}}{\bar{p}-1}}(\Omega) \text{ i.e.,} \quad p \ge \frac{\bar{p}}{\bar{p}-1}.
\end{aligned} \tag{3.90}$$

(R2) *Well-definedness and boundedness of parameter-to-state map:*

We argue unique existence of the solution to (3.10)–(3.12) via the transformed problem (3.86)–(3.88) for $U = e^u$.

To begin, by a similar argument to (3.71) with the elliptic operator $A = -\Delta + \theta, \theta \in L^p(\Omega)$ in place of the parabolic operator, we show that the corresponding elliptic problem admits a unique solution in $H^2(\Omega) \cap H_0^1(\Omega)$ if the index p

satisfies (3.72). Employing next the semigroup theory in [32, Section 7.4.3, Theorem 5] or [96, Chapter 7, Corollary 2.6] with assuming that $U_0 \in D(A) = H^2(\Omega) \cap H_0^1(\Omega)$ implies unique existence of a solution $U \in C^1(0,T;H^2(\Omega))$ to (3.86)–(3.88).

Let $U, \hat{U}$ respectively solve (3.86)–(3.88) associated with the coefficients $\theta \in \mathcal{X}, \hat{\theta} \in L^\infty(\Omega)$ with the same boundary and initial data, then $v = U - \hat{U}$ solves

$$\begin{aligned}
&\dot{v}(t) - \Delta v(t) + \hat{\theta} v(t) = (\hat{\theta} - \theta)U(t) \qquad t \in (0,T)\\
&v(0) = 0.
\end{aligned}$$

Owing to the regularity from [32, Section 7.1.3, Theorem 5] and estimating similarly to (3.71), we obtain

$$\begin{aligned}
\|U - \hat{U}\|_{L^\infty(0,T;H^2(\Omega))} &\leq C^{\hat{\theta}} \|(\hat{\theta} - \theta)U\|_{H^1(0,T;L^2(\Omega))}\\
&\leq C\|\hat{\theta} - \theta\|_{\mathcal{X}} \|U\|_{H^1(0,T;H^2(\Omega))}
\end{aligned} \tag{3.91}$$

with positive $\hat{U}$ since $\hat{\theta} \in L^\infty(\Omega)$ and the constant C depending only on θ^0, ρ. Here we assume $\hat{\theta} = \theta^0 \in L^\infty(\Omega)$ and ρ is sufficiently small such that the right hand side is sufficiently small. Then $U \in L^\infty(0,T;H^2(\Omega)) \subseteq L^\infty((0,T) \times \Omega)$ is close to $\hat{U}$ and therefore positive as well. This assertion is valid if $0 < U_0 = e^{u_0} \in H^2(\Omega) \cap H_0^1(\Omega)$, $0 < U|_{\delta\Omega}$, which is chosen as $U|_{\delta\Omega} = 1$ in this case (such that $\log(U|_{\delta\Omega}) = 0$) and

$$\begin{aligned}
&H^2(\Omega) \hookrightarrow L^{\frac{2p}{p-2}}(\Omega) \text{ i.e.,} \quad p \geq \frac{2\bar{p}}{\bar{p}-2}\\
&V = H^2(\Omega) \hookrightarrow L^\infty(\Omega) \text{ i.e.,} \quad d \leq 3.
\end{aligned} \tag{3.92}$$

This leads to unique existence of the solution $u := \log(U)$ to the problem (3.10)–(3.12), moreover $0 < \underline{c} \leq U \in C^1(0,T;H^2(\Omega))$ allows $u = \log(U) \in C^1(0,T;H^2(\Omega) \cap H_0^1(\Omega))$.
If $d = 1, p \geq 2$, no assumption on smallness of ρ is required since

$$\|U - \hat{U}\|_{L^\infty(0,T;H^1(\Omega))} \leq C^\theta \|(\hat{\theta} - \theta)\hat{U}\|_{L^2(0,T;L^2(\Omega))} \leq C\|\hat{\theta} - \theta\|_{\mathcal{X}} \|\hat{U}\|_{L^2(0,T;H^2(\Omega))} \tag{3.93}$$

due to the estimates (3.74)–(3.76) in Section 3.3.1. Here the constant C depends only on θ^0, ρ as claimed in (3.76). This and the fact $\overline{L^\infty(\Omega)} = L^p(\Omega)$ allow us to chose $\hat{\theta} \in L^\infty(\Omega)$ being sufficiently close to $\theta \in L^p(\Omega)$ to make the right hand side of (3.93) arbitrarily small without the need of smallness of ρ.
We have observed that, with the same positive boundary and initial data, the solution $U = U(\theta)$ to (3.86)–(3.88) is bounded away from zero for all $\theta \in \mathcal{B}_\rho^{\mathcal{X}}(\theta^0)$.

Besides, $S : \theta \mapsto U$ is a bounded operator as proven in (R2) of Section 3.3.1. Consequently, $u = \log(U)$ with $\Delta u = -\frac{|\nabla U|^2}{U^2} + \frac{\Delta U}{U}$ is uniformly bounded in $L^2(0,T; H^2(\Omega) \cap H_0^1(\Omega))$ for all $\theta \in \mathcal{B}_\rho^{\mathcal{X}}(\theta^0)$, thus $S : \theta \mapsto u$ is a bounded operator on $\mathcal{B}_\rho^{\mathcal{X}}(\theta^0)$.

Moreover, we can derive a uniform bound for U in $H^1(0,T; H^2(\Omega))$ with respect to θ. From

$$(\dot{U} - \dot{\hat{U}}) - \Delta(U - \hat{U}) + (\theta - \hat{\theta})(U - \hat{U}) = -\hat{\theta}(U - \hat{U}) - (\theta - \hat{\theta})\hat{U},$$

by taking the time derivative of both sides then test them with $-\Delta(\dot{U} - \dot{\hat{U}})$ we have

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\|\nabla(\dot{U} - \dot{\hat{U}})\|^2_{L^2(\Omega)} &+ \|\Delta(\dot{U} - \dot{\hat{U}})\|^2_{L^2(\Omega)} \\
&\le C_{H^2 \hookrightarrow L^\infty}\|\theta - \hat{\theta}\|_{L^2(\Omega)}\|\Delta(\dot{U} - \dot{\hat{U}})\|^2_{L^2(\Omega)} \\
&\quad + \|\hat{\theta}\|_{L^\infty(\Omega)}\|\dot{U} - \dot{\hat{U}}\|_{L^2(\Omega)}\|\Delta(\dot{U} - \dot{\hat{U}})\|_{L^2(\Omega)} \\
&\quad + C_{H^2 \hookrightarrow L^\infty}\|\theta - \hat{\theta}\|_{L^2(\Omega)}\|\Delta\dot{\hat{U}}\|_{L^2(\Omega)}\|\Delta(\dot{U} - \dot{\hat{U}})\|_{L^2(\Omega))} \\
\frac{1}{2}\frac{d}{dt}\|\nabla(\dot{U} - \dot{\hat{U}})\|^2_{L^2(\Omega)} &+ (1 - \rho C_{H^2 \hookrightarrow L^\infty} - \epsilon)\|\Delta(\dot{U} - \dot{\hat{U}})\|^2_{L^2(\Omega)} \\
&\le \frac{1}{2\epsilon}\left(\|\hat{\theta}\|^2_{L^\infty(\Omega)}\|\dot{U} - \dot{\hat{U}}\|^2_{L^2(\Omega)} + C^2_{H^2 \hookrightarrow L^\infty}\rho^2\|\Delta\dot{\hat{U}}\|^2_{L^2(\Omega)}\right),
\end{aligned}$$

where $\|\Delta\dot{\hat{U}}\|_{L^2(\Omega)}$ is attained by estimating with the same technique for (3.86)–(3.88) with the coefficient $\hat{\theta} \in L^\infty(\Omega)$. Since ϵ is arbitrarily small, if ρ is sufficiently small and the following condition holds

$$\mathcal{X} = L^p(\Omega) \hookrightarrow L^2(\Omega) \text{ i.e.,} \quad p \ge 2, \tag{3.94}$$

applying Gronwall's inequality then integrating on $[0,T]$ yields

$$\|U - \hat{U}\|_{H^1(0,T;H^2(\Omega))} \le C\|\hat{\theta} - \theta\|_{\mathcal{X}}\|\hat{U}\|_{H^1(0,T;H^2(\Omega))} \tag{3.95}$$

for fixed $\hat{U} = S(\hat{\theta}) = S(\theta^0)$. So, $S(\mathcal{B}_\rho^{\mathcal{X}}(\theta^0))$ is bounded in $H^1(0,T; H^2(\Omega))$ and its diameter can be controlled by ρ. In case $d = 1$, smallness of ρ can be omitted if one uses the estimate (3.93).

(R3) *Continuity of the inverse of the linearized model:*

Now we consider the linearized problem

$$\dot{z}(t) - \Delta z(t) + 2\nabla u(t) \cdot \nabla z(t) = r(t) \qquad t \in (0,T) \tag{3.96}$$

$$z(0) = 0, \tag{3.97}$$

whose adjoint problem after transforming $t = T - \tau$ is

$$\dot{p}(t) - \Delta p(t) - 2\nabla \cdot (\nabla u(t) p(t)) = \phi(t) \qquad t \in (0,T) \tag{3.98}$$
$$p(0) = 0. \tag{3.99}$$

Since $u \in C^1(0,T; H^2(\Omega) \cap H_0^1(\Omega))$ as proven in (R2), this equation with the coefficients $m := -2\nabla u \in C^1(0,T; H^1(\Omega)), n := -2\Delta u \in C^1(0,T; L^2(\Omega))$ is feasible to attain the estimate (R3) by the contraction argument.

Indeed, let us take p solving (3.98)–(3.99), then

$$\begin{aligned}
&\dot{p} - \Delta p + \hat{m} \cdot \nabla p + \hat{n} p = \phi + (\hat{m} - m) \cdot \nabla p + (\hat{n} - n) p \\
&p = \left(\frac{d}{dt} - \Delta + \hat{m} \cdot \nabla + \hat{n} \right)^{-1} [\phi + (\hat{m} - m) \cdot \nabla p + (\hat{n} - n) p] \\
&\quad =: Tp
\end{aligned}$$

with some $\hat{m} \in L^\infty((0,T) \times \Omega)$ and some $\hat{n} \in L^\infty((0,T) \times \Omega)$ approximating m and n. Then for $d \leq 3$, $T : L^2(0,T; H^2(\Omega) \cap H_0^1(\Omega)) \to L^2(0,T; H^2(\Omega) \cap H_0^1(\Omega))$ is a contraction

$$\begin{aligned}
&\|T(p-q)\|_{L^2(0,T;H^2 \cap H_0^1)} \\
&\leq \left\| \left(\frac{d}{dt} - \Delta + \hat{m} \cdot \nabla + \hat{n} \right)^{-1} \right\|_{L^2(0,T;L^2(\Omega)) \to L^2(0,T;H^2 \cap H_0^1)} \\
&\qquad . \left(\|(\hat{m} - m) \cdot \nabla (p-q)\|_{L^2(0,T;L^2(\Omega))} + \|(\hat{n} - n)(p-q)\|_{L^2(0,T;L^2(\Omega))} \right) \\
&\leq C^{\hat{\theta}} \Big(\|\hat{m} - m\|_{L^\infty(0,T;H^1(\Omega))} \|\nabla (p-q)\|_{L^2(0,T;H^1(\Omega))} \\
&\qquad\qquad + \|\hat{n} - n\|_{L^\infty(0,T;L^2(\Omega))} \|p-q\|_{L^2(0,T;L^\infty(\Omega))} \Big) \\
&\leq C\epsilon \|p-q\|_{L^2(0,T;H^2 \cap H_0^1)},
\end{aligned} \tag{3.100}$$

where $H_0^1(\Omega) \hookrightarrow L^6(\Omega), H^2(\Omega) \cap H_0^1(\Omega) \hookrightarrow L^\infty(\Omega)$ for $d \leq 3$. Above, we apply from [32, Section 7.1.3 , Theorem 5] the continuity of $\left(\frac{d}{dt} - \Delta + \hat{m} \cdot \nabla + \hat{n}\right)^{-1}$ with noting that, although the theorem is stated for time-independent coefficients, the proof reveals it is still applicable for $\hat{m} = \hat{m}(t,x), \hat{n} = \hat{n}(t,x)$ being bounded in time and space.
The above constant $C^{\hat{\theta}}$, which depends on $\hat{m} \in \nabla \cdot S(\mathcal{B}_\rho^\mathcal{X}(\theta^0)) \cap L^\infty(0,T; L^\infty(\Omega))$, $\hat{n} \in \Delta S(\mathcal{B}_\rho^\mathcal{X}(\theta^0)) \cap L^\infty(0,T; L^\infty(\Omega))$ can be bounded by some constant C depending only on $S(\theta^0)$ and the diameter of $S(\mathcal{B}_\rho^\mathcal{X}(\theta^0))$ similarly to Sections 3.3.1 and 3.3.2 if choosing $\hat{\theta} = \theta^0$. In order to make $C\epsilon$ less than one, we require $\|\hat{m} - m\|_{L^\infty(0,T;H^1(\Omega))}$ and $\|\hat{n} - n\|_{L^\infty(0,T;L^2(\Omega))}$ to be sufficiently small. Those conditions turn out to be uniform boundedness of $\|\hat{U} - U\|_{L^\infty(0,T;H^2(\Omega))}$ (or the diameter of $S(\mathcal{B}_\rho^\mathcal{X}(\theta^0))$, which can be seen as smallness of ρ as in

(3.95) since $H^1(0,T) \hookrightarrow L^\infty(0,T)$. From that, existence of the dual state $p \in L^2(0,T;H^2(\Omega)\cap H_0^1(\Omega))$ for given $\phi \in L^2(0,T;L^2(\Omega))$ is shown.

Then (R3-dual) follows without adding further constraints on p

$$\begin{aligned}&\|p\|_{L^2(0,T;H^2(\Omega))}\\ &\quad\le C(\|(\hat{m}-m)\cdot\nabla p\|_{L^2(0,T;L^2(\Omega))}+\|(\hat{n}-n)p\|_{L^2(0,T;L^2(\Omega))}+\|\phi\|_{L^2(0,T;L^2(\Omega))})\\ &\quad\le C\|\phi\|_{L^2(0,T;L^2(\Omega))}\end{aligned}$$

with constant C depending only on some fixed $\hat{m},\hat{n}$ and the assumption on smallness of ρ. Here with the L^2-norm on the right hand side, the maximum q is limited by $\bar{q}=2$.

Observing that the problem (3.98)–(3.99) has the form of the a-problem written in (3.82), with $\underline{a}=1, \nabla a=-2\nabla u(t)\in L^6(\Omega)$ and the additional term in the last line of the right hand side, namely,

$$\begin{aligned}\frac{1}{\underline{a}}\|np\|^2_{L^2(\Omega)}&=\|\Delta up\|^2_{L^2(\Omega)}\le\|\Delta u\|^2_{L^2(\Omega)}\|p\|^2_{L^\infty(\Omega)}\\ &\le C^2_{H_0^1\to L^\infty}\|\Delta u\|^2_{L^2(\Omega)}\|\nabla p\|^2_{L^2(\Omega)}\end{aligned}\tag{3.101}$$

if the dimension $d=1$.
The solution $u=S(\theta)$ also lies in some ball in $C^1(0,T;H^2(\Omega)\cap H_0^1(\Omega))$ for all $\theta\in\mathcal{B}_\rho^{\mathcal{X}}(\theta^0)$, as in (R2) we have shown boundedness of the operator S.
It allows us to evaluate analogously to (3.82)–(3.83) with taking into account the additional term (3.101) to eventually get

$$\|\Delta p\|_{L^2(0,T;L^2(\Omega))}\le C\|\phi\|^2_{L^2(0,T;L^2(\Omega))}$$

with the constant C depending only on θ^0,ρ. Hence, if $d=1$, ρ is not required to be small.

(R4) *All-at-once tangential cone condition:*

According to (3.36), (B.123) with $s=t=2$, $m=n=2$, $\hat{\gamma}=0,\rho=2$ this follows if

$$\begin{aligned}&2-\frac{d}{2}\succeq 1-\frac{d}{q^*}+\frac{d}{R}\text{ and}\\ &1\le\frac{R}{q^*}\text{ and }q\le\hat{q}\text{ and }2-\frac{d}{2}\succeq\max\left\{-\frac{d}{\hat{q}},\ 1-\frac{d}{R}\right\},\end{aligned}$$

where the latter conditions come from the requirements $V=H^2(\Omega)\cap H_0^1(\Omega)\hookrightarrow W^{1,R}(\Omega)$.

Corollary 3.3.12. *Assume* $u_0 \in V$ *and*

$$\begin{aligned} \mathcal{D}(F) = \mathcal{X} = L^p(\Omega), \qquad Y = L^q(\Omega), \qquad & V = W = H^2(\Omega) \cap H_0^1(\Omega) \\ & p \geq 2, \quad q \in \left[\underline{q}, 2\right], \quad d \leq 3 \end{aligned} \tag{3.102}$$

with $\underline{q} = \min_q \left\{ 2 - \frac{d}{2} \succeq 1 - d + \frac{d}{q} + \frac{d}{\check{p}} \wedge q \geq 1 + \frac{1}{\check{p}-1} \right\}$.

Then F *defined by* $F(\theta) = u$ *solving* (3.10)-(3.12) *satisfies the tangential cone condition (3.53) with a uniformly bounded operator* $F'(\theta)$ *defined by (3.58).*

Remark 3.3.13. To achieve a Hilbert space setting for X and Y, one can choose $p = q = 2$ if $d \leq 3$, see also [93].

3.3.4 Inverse source problem with cubic nonlinearity

We investigate this problem in the function spaces

$$\mathcal{X} = L^p(\Omega), \qquad Y = L^q(\Omega), \qquad V = W = H_0^1(\Omega).$$

In the following we examine the conditions required for deriving the tangential cone condition and boundedness of the derivative of the forward operator.

(R1) *Local Lipschitz continuity of* f*:*

$$\begin{aligned} &\|\tilde{u}^3 - u^3 + \tilde{\theta} - \theta\|_{W^*} = \sup_{\|w\|_W \leq 1} \int_\Omega (\tilde{u} - u)(\tilde{u}^2 + \tilde{u}u + u^2)w + (\tilde{\theta} - \theta) w dx \\ &\leq C_{W\to L^{\bar{p}}} \left(\|(\tilde{u} - u)(\tilde{u}^2 + \tilde{u}u + u^2)\|_{L^{\frac{\bar{p}}{\bar{p}-1}}} + \|\tilde{\theta} - \theta\|_{L^{\frac{\bar{p}}{\bar{p}-1}}} \right) \\ &\leq C_{W\to L^{\bar{p}}} \left(2\|\tilde{u} - u\|_{L^{\bar{p}}} (\|\tilde{u}\|^2_{L^{\frac{2\bar{p}}{\bar{p}-2}}} + \|u\|^2_{L^{\frac{2\bar{p}}{\bar{p}-2}}}) + \|\tilde{\theta} - \theta\|_{L^{\frac{\bar{p}}{\bar{p}-1}}} \right) \\ &\leq C_{W\to L^{\bar{p}}} \left(2C_{V\to L^{\bar{p}}} C^2_{V\to L^{\frac{2\bar{p}}{\bar{p}-2}}} \|\tilde{u} - u\|_V (\|\tilde{u}\|_V^2 + \|u\|_V^2) + \|\tilde{\theta} - \theta\|_{\mathcal{X}} C_{\mathcal{X}\to L^{\frac{\bar{p}}{\bar{p}-1}}} \right). \end{aligned}$$

We chose $L(M) = C_{W\to L^{\bar{p}}} \left(2C_{V\to L^{\bar{p}}} C^2_{V\to L^{\frac{2\bar{p}}{\bar{p}-2}}} (\|\tilde{u}\|_V^2 + \|u\|_V^2) + C_{\mathcal{X}\to L^{\frac{\bar{p}}{\bar{p}-1}}} \right) + 1$, subject to the conditions

$$\begin{aligned} & V = W = H_0^1(\Omega) \hookrightarrow L^{\bar{p}}(\Omega) \text{ i.e.,} \quad 1 - \frac{d}{2} \succeq -\frac{d}{\bar{p}} \\ & V = H_0^1(\Omega) \hookrightarrow L^{\frac{2\bar{p}}{\bar{p}-2}}(\Omega) \text{ i.e.,} \quad d \leq 4 \\ & \mathcal{X} = L^p(\Omega) \hookrightarrow L^{\frac{\bar{p}}{\bar{p}-1}}(\Omega) \text{ i.e.,} \quad p \geq \frac{\bar{p}}{\bar{p}-1}. \end{aligned} \tag{3.103}$$

(R2) *Well-definedness and boundedness of the parameter-to-state map:*

Verifying the conditions S.1.-S.3. with the Gelfand triple $H_0^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$ shows that the problem (3.13)-(3.15) admits a unique solution in the space $W(0,T)$. Subsequently, [100, Theorem 8.16] strengthens the solution to belong to $L^\infty(0,T;V)$. To validate this regularity result, the following additional assumptions are made

$$\mathcal{X} = L^p(\Omega) \hookrightarrow L^2(\Omega) \text{ i.e.,} \quad p \geq 2, \tag{3.104}$$

the initial data $u_0 \in V$ and the known source term $\varphi \in L^2(0,T;L^2(\Omega))$.

From [93, Proposition 4.2, Section 6.1], we have

$$\begin{aligned}\|S(\theta)\|_{L^\infty(0,T;V)} &\leq N\left(\|\theta+\varphi\|_{L^2(0,T;L^2(\Omega))} + \sqrt{\int_\Omega \frac{1}{2}|\nabla u_0|^2 + \frac{1}{4}u_0^4 dx}\right)\\ &\leq N\left(\sqrt{T}(\|\theta_0\|_{L^2(\Omega)}+\rho) + \|\varphi\|_{L^2(0,T;L^2(\Omega))} + \sqrt{\int_\Omega \frac{1}{2}|\nabla u_0|^2 + \frac{1}{4}u_0^4 dx}\right)\end{aligned}$$

for some N depending only on $c_0^\theta = c_0 = \frac{1}{2}$. This thus implies uniform boundedness of S on $\mathcal{B}_\rho^{\mathcal{X}}(\theta_0)$.

(R3) *Continuous dependence on data of the solution to the linearized problem with zero initial data:*

For this purpose, semi-coercivity of the linearized forward operator is obvious

$$\begin{aligned}\langle -f'_u(t,\theta,v), v\rangle_{V^*,V} &= \int_\Omega (-\Delta v + 3u^2 v)v dx\\ &\geq \|\nabla v\|^2_{L^2(\Omega)} = \|v\|_V^2.\end{aligned}$$

(R4) *All-at-once tangential cone condition:*

According to (3.36), (B.122), with $s=t=1$, $m=n=2$, $\gamma=\kappa=1, r=\hat{q}=\bar{p}$ this follows if

$$2 \leq \frac{\bar{p}}{q^*} \text{ and } 1-\frac{d}{2} \succeq -\frac{d}{q^*} + \frac{2d}{\bar{p}} \text{ and } q \leq \bar{p} \text{ and } 1-\frac{d}{2} \succeq -\frac{d}{\bar{p}},$$

where the latter condition comes from the requirement $V = H_0^1(\Omega) \hookrightarrow L^{\bar{p}}(\Omega)$.

Corollary 3.3.14. *Assume $u_0 \in H_0^1(\Omega), \varphi \in L^2(0,T;L^2(\Omega))$, and*

$$\begin{aligned}\mathcal{D}(F) = \mathcal{X} = L^p(\Omega), \qquad & Y = L^q(\Omega), \qquad V = W = H_0^1(\Omega)\\ & p \geq 2, \quad q \in \left[\underline{q}, \bar{q}\right], \quad d \leq 4,\end{aligned} \tag{3.105}$$

where $\underline{q} = \min_q \left\{1 - \frac{d}{2} \succeq -d + \frac{d}{q} + \frac{2d}{\bar{p}} \wedge q \geq 1 + \frac{2}{\bar{p}-2}\right\}$ *with*

$$d = 1 \text{ and } \bar{q} = \infty, \qquad d = 2 \text{ and } \bar{q} < \infty, \qquad d \geq 3 \text{ and } \bar{q} = \frac{2d}{d-2}. \tag{3.106}$$

Then F *defined by* $F(\theta) = u$ *solving* (3.13)-(3.15) *satisfies the tangential cone condition (3.53) with a uniformly bounded operator* $F'(\theta)$ *defined by (3.58).*

Remark 3.3.15. Here X and Y can be chosen as Hilbert spaces with $p = q = 2$ and $d \leq 3$.

Conclusion and outlook

The key contribution of this chapter is establishing two main ingredients for the iterative regularization methods: the tangential cone condition and locally uniform boundedness of the derivative of the forward operator F. We establish these structural assumptions on the nonlinear operator F for some class of parabolic model problems in the reduced setting as well as its counterpart in the all-at-once setting.

The following questions are considered to be open problems:

In this study, one crucial assumption enabling justification of the tangential cone condition is the full observation. Instead of this, when considering measurement of u only on the boundary or a part of Ω (which means $\mathcal{C}$ is the trace or the restriction), one might need to investigate alternative nonlinearity conditions, e.g., the adjoint range invariance (which is sufficient for the tangential cone condition) or range invariance [62, Chapter 4.3] to ensure convergence of the iterative methods.

It is well-known for regularization theory that convergence rates can be obtained under an additional smoothness property of the true solution, known as the "source condition" [31]: $x^\dagger - x^0 = \phi(F'(x^\dagger)^* F'(x^\dagger))v$ for some $v \in \mathcal{N}(F'(x^\dagger))^\perp$ and some index function ϕ, i.e., $\phi : [0,\infty) \to [0,\infty)$ is continunous increasing and $\phi(0) = 0$. A recent concept of source condition named "variational source condition" [54, 53, 34] has been introduced in the form

$$x \in \mathcal{D}(F): \quad \frac{\beta}{2}\|x - x^\dagger\|_{\mathcal{X}}^2 \leq \frac{1}{2}\|x\|_{\mathcal{X}}^2 - \frac{1}{2}\|x^\dagger\|_{\mathcal{X}}^2 + \Phi\left(\|F(x) - F(x^\dagger)\|_{\mathcal{Y}}^2\right)$$

with some $\beta \in (0,1]$ and concave index function Φ. Such a variational souce condition does not only yield the classical source condition but also incorporates the nonlinear structure of F. So far, this variational source conditions has been verified only for a few cases. Establishing the variational source condition for further practical examples would be desirable.

3.4 Auxiliary results

Proof of Proposition 3.2.1

On (3.41) for some $p \in [1,\infty]$, we can estimate by applying Hölder's inequality, once with exponent p and once with exponent $\frac{q}{p^*}$ (where $p^* = \frac{p}{p-1}$ is the dual index)

$$\|(B\hat{c})(t)\hat{v}\|_{W^*} = \sup_{w\in W\,,\ \|w\|_W\leq 1} \int_\Omega \hat{c}\,\hat{v}\,w\,dx \leq \|\hat{c}\|_{L^p}\|\hat{v}\|_{L^q} \sup_{w\in W\,,\ \|w\|_W\leq 1} \|w\|_{L^{\frac{p^*q}{q-p^*}}}\,,$$

where we need to impose $q \geq p^*$ and in case of equality formally set $\frac{p^*q}{q-p^*} = \infty$. In order to guarantee continuity of the embedding $W \hookrightarrow L^{\frac{p^*q}{q-p^*}}(\Omega)$ as needed here, we therefore, together with (3.36), require the conditions

$$s - \frac{d}{m} \succeq -\frac{d}{\hat{q}} \text{ and } \hat{q} \geq q \geq p^* \text{ and } t - \frac{d}{n} \succeq -\frac{d(q-p^*)}{p^*q}\,. \tag{B.107}$$

□

Proof of Proposition 3.2.3

With $\mathcal{X}$ as in (3.43), in order to guarantee the required boundedness of the embeddings

$$\mathcal{X} \hookrightarrow L^r(\Omega)\,, \quad W \hookrightarrow W^{1,\frac{p^*q}{q-p^*}}(\Omega)\,,\ W \hookrightarrow W^{2,\frac{r^*q}{q-r^*}}(\Omega)\,,$$
for some $r \in [1,\infty]$ such that $r^* \leq q$

we impose, additionally to (3.36), the conditions

$$(a)\ \hat{q} \geq q \geq \max\{p^*, r^*\} \text{ and} \qquad (b)\ t - 1 - \frac{d}{n} \succeq -\frac{d(q-p^*)}{p^*q} \text{ and}$$
$$(c)\ t - 2 - \frac{d}{n} \succeq -\frac{d(q-r^*)}{r^*q} \text{ and} \qquad (d)\ 1 - \frac{d}{p} \succeq -\frac{d}{r}$$

for some $r \in [1,\infty]$. To eliminate r, observe that the requirement (c), i.e., $t - 2 - \frac{d}{n} \succeq -\frac{d}{r^*} + \frac{d}{q}$ gets weakest when r^* is chosen minimal, which, subject to requirement (d) is

$$r \begin{cases} = \infty \text{ if } p > d \\ < \infty \text{ if } p = d \\ = \frac{dp}{d-p} \text{ if } p < d \end{cases}, \text{ i.e.,} \quad r^* \begin{cases} = 1 \text{ if } p > d \\ > 1 \text{ if } p = d \\ = \frac{dp}{dp-d+p} \text{ if } p < d \end{cases}. \tag{B.108}$$

Inserting this into (c) and taking into account (3.36), we end up with the following requirements on s, t, p, q, m, n (using the fact that $q \geq p^*$ implies $q \geq \frac{dp}{dp-d+p}$):

$$s - \frac{d}{m} \succeq -\frac{d}{\hat{q}} \text{ and } \hat{q} \geq q \geq p^* \text{ and}$$
$$t - 1 - \frac{d}{n} \succeq -\frac{d(q-p^*)}{p^* q} \text{ and } t - 2 - \frac{d}{n} \begin{cases} \succeq -d + \frac{d}{q} \text{ if } p > d \\ > -d + \frac{d}{q} \text{ and } q > 1 \text{ if } p = d \\ \succeq -\frac{dp-d+p}{p} + \frac{d}{q} \text{ if } p < d \end{cases} . \tag{B.109}$$

□

Proof of Proposition 3.2.5

Here we have

$$\Big(f(\theta,u) - f(\tilde{\theta},\tilde{u}) - f'_u(\theta,u)(u-\tilde{u}) - f'_\theta(\theta,u)(\theta-\tilde{\theta})\Big)(t)$$
$$= \int_0^1 \Big(\Phi'((u(t) + \sigma(\tilde{u}(t) - u(t)) - \Phi'(u(t))\Big)\, d\sigma\, (\tilde{u}(t) - u(t))$$
$$+ \int_0^1 \Big(\Psi'(\nabla u(t) + \sigma(\nabla\tilde{u}(t) - \nabla u(t)) - \Psi'(\nabla u(t))\Big)\, d\sigma\, \nabla(\tilde{u}(t) - u(t))\,.$$

This shows that the only condition which has to be taken into account when choosing the space $\mathcal{X}$ is that $B(t) \in \mathcal{L}(\mathcal{X}, W^*)$. Again we assume $\mathcal{C}(t)$ to be the embedding operator $V \hookrightarrow Y$.

As opposed to Section 3.2.1, where we could do the estimates pointwise in time, we will now also have to to use Hölder estimates with respect to time. To this end, we dispose over the following continuous embeddings

$$\mathcal{U} \hookrightarrow L^2(0,T; W^{s,m}(\Omega))$$
$$\mathcal{U} \hookrightarrow L^\infty(0,T; H^{\tilde{s}}(\Omega)) \text{ provided } W^{s-\tilde{s},m}(\Omega) \hookrightarrow W^{t+\tilde{s},n}(\Omega)\,,$$

where the first holds just by definition of $\mathcal{U}$ and the second follows from [100, Lemma 7.3][4] with $\widetilde{W} = W^{t+\tilde{s},n}(\Omega)$, using the fact that

$$u \in L^2(0,T; W^{s,m}(\Omega)) \cap H^1(0,T; (W^{t,n}(\Omega))^*)$$
$$\Leftrightarrow\ D^{\tilde{s}} u \in L^2(0,T; W^{s-\tilde{s},m}(\Omega)) \cap H^1(0,T; (W^{t+\tilde{s},n}(\Omega))^*),$$

where $D^{\tilde{s}} v = \sum_{|\alpha| \leq \tilde{s}} D^\alpha v$.

[4] $L^2(0,T;\widetilde{W}) \cap H^1(0,T;\widetilde{W}^*) \hookrightarrow L^\infty(0,T;L^2(\Omega))$

We first consider the case of an affinely linear (or just vanishing) function Ψ, which still comprises, e.g., models with linear drift and diffusion, so that $C_{\Psi''}$ can be set to zero. We can then estimate

$$\|f(\theta,u)-f(\tilde\theta,\tilde u)-f_u'(\theta,u)(u-\tilde u)-f_\theta'(\theta,u)(\theta-\tilde\theta)\|_{L^2(0,T;W^*)}$$
$$\leq C_{\Phi''}\left(\int_0^T\Big(\sup_{w\in W\,,\ \|w\|_W\leq 1}\int_\Omega(1+|u(t)|^\gamma+|\tilde u(t)|^\gamma)\,|\tilde u(t)-u(t)|^{1+\kappa}\,w\,dx\Big)^2\,dt\right)^{1/2},$$

where, using Hölder's inequality three times ($P=q$, $P=\frac{r}{q^*(\gamma+\kappa)}$, $P=\frac{\gamma+\kappa}{\gamma}$) and continuity of the embedding $H^{\tilde s}(\Omega)\hookrightarrow L^r(\Omega)$ provided $\tilde s-\frac{d}{2}\succeq-\frac{d}{r}$

$$\begin{aligned}
&\left(\int_0^T\Big(\sup_{w\in W\,,\ \|w\|_W\leq 1}\int_\Omega|u(t)|^\gamma\,|\tilde u(t)-u(t)|^{1+\kappa}\,w\,dx\Big)^2\,dt\right)^{1/2}\\
&\leq\|\tilde u-u\|_{L^2(0,T;L^q(\Omega))}\sup_{w\in W\,,\ \|w\|_W\leq 1}\Big\||u|^\gamma|\tilde u-u|^\kappa w\Big\|_{L^\infty(0,T;L^{q^*}(\Omega))}\\
&\leq\|\tilde u-u\|_{\mathcal{Y}}\Big\|\Big(|u|^\gamma|\tilde u-u|^\kappa\Big)^{\frac{1}{\gamma+\kappa}}\Big\|^{\gamma+\kappa}_{L^\infty(0,T;L^r(\Omega))}\sup_{w\in W\,,\ \|w\|_W\leq 1}\|w\|_{L^{\frac{rq^*}{r-q^*(\gamma+\kappa)}}(\Omega)}\\
&\leq\|\tilde u-u\|_{\mathcal{Y}}\|u\|^\gamma_{L^\infty(0,T;L^r(\Omega))}\|\tilde u-u\|^\kappa_{L^\infty(0,T;L^r(\Omega))}\sup_{w\in W\,,\ \|w\|_W\leq 1}\|w\|_{L^{\frac{rq^*}{r-q^*(\gamma+\kappa)}}(\Omega)}\\
&\leq(C^\Omega_{H^{\tilde s}\to L^r})^{\gamma+\kappa}\|u\|^\gamma_{L^\infty(0,T;H^{\tilde s}(\Omega))}\|\tilde u-u\|^\kappa_{L^\infty(0,T;H^{\tilde s}(\Omega))}\\
&\qquad\|\tilde u-u\|_{\mathcal{Y}}\sup_{w\in W\,,\ \|w\|_W\leq 1}\|w\|_{L^{\frac{rq^*}{r-q^*(\gamma+\kappa)}}(\Omega)}
\end{aligned}\tag{B.110}$$

(and likewise for the term containing $|\tilde u(t)|^\gamma$) for some $r\in[1,\infty]$ with $\frac{r}{q^*}\geq\gamma+\kappa$. In order to get finiteness of the $L^\infty(0,T;H^{\tilde s}(\Omega))$ norms appearing here by means of [100, Lemma 7.3], we assume the embedding $W^{s-\tilde s,m}(\Omega)\hookrightarrow W^{t+\tilde s,n}(\Omega)$ to be continuous, which leads to the condition

$$s-\tilde s-\frac{d}{m}\succeq t+\tilde s-\frac{d}{n}\text{ and }s-\tilde s\geq t+\tilde s\,.$$

Moreover, in order to guarantee continuity of the embedding $W\hookrightarrow L^{\frac{rq^*}{r-q^*(\gamma+\kappa)}}(\Omega)$ and for the above Hölder estimate to make sense we impose

$$\gamma+\kappa\leq\frac{r}{q^*}\text{ and }t-\frac{d}{n}\succeq-\frac{d(r-q^*(\gamma+\kappa))}{rq^*}$$

for some $r\in[1,\infty]$. Summarizing, we have the following conditions

$$\begin{aligned}
&\tilde s-\frac{d}{2}\succeq-\frac{d}{r}\text{ and }s-\tilde s-\frac{d}{m}\succeq t+\tilde s-\frac{d}{n}\text{ and }s-\tilde s\geq t+\tilde s\text{ and}\\
&\gamma+\kappa\leq\frac{r}{q^*}\text{ and }t-\frac{d}{n}\succeq-\frac{d(r-q^*(\gamma+\kappa))}{rq^*}=-\frac{d}{q^*}+\frac{d(\gamma+\kappa)}{r},
\end{aligned}\tag{B.111}$$

which imply

$$s \geq \frac{d}{m} + d - \frac{d}{q^*} + d\frac{\gamma+\kappa-2}{r}\,.$$

This lower bound on s gets weakest for maximal r, if $\gamma+\kappa > 2$ and for minimal r if $\gamma+\kappa < 2$. We therefore make the following case distinction.
If $\gamma+\kappa > 2$ or $\gamma+\kappa = 2$ and $q = 1$ we set $r = \infty$, which leads to $\tilde{s} > \frac{d}{2}$, hence, according to (B.111), we can choose

$$\begin{aligned}
&\text{case } \gamma+\kappa > 2 \text{ or } (\gamma+\kappa = 2 \text{ and } q = 1):\\
&t > \frac{d}{n} - \frac{d}{q^*}\,, \quad q \leq \hat{q}\,,\\
&s > \max\left\{t + d + \max\left\{0, \frac{d}{m} - \frac{d}{n}\right\}, \frac{d}{m} - \frac{d}{\hat{q}}\right\}.
\end{aligned} \tag{B.112}$$

If $\gamma+\kappa < 2$ or $\gamma+\kappa = 2$ and $q > 1$ we set $r = \max\{1, q^*(\gamma+\kappa)\} < \infty$, $\tilde{s} := \max\{0, \frac{d}{2} - \frac{d}{r}\}$ and, according to (B.111), can therefore choose

$$\begin{aligned}
&\text{case } \gamma+\kappa < 2 \text{ or } (\gamma+\kappa = 2 \text{ and } q > 1):\\
&t > \frac{d}{n} + \min\left\{0, -\frac{d}{q^*} + d(\gamma+\kappa)\right\}, \quad q \leq \hat{q}\,,\\
&s > \max\left\{t + \max\left\{0, d - \frac{2d}{\max\{1, q^*(\gamma+\kappa)\}}\right\}, \frac{d}{m} - \frac{d}{\hat{q}}\right\}.
\end{aligned} \tag{B.113}$$

Now we consider the situation of nonvanishing gradient nonlinearites $C_{\Psi''} > 0$ where we additionally need to estimate terms of the form

$$\left(\int_0^T \Big(\sup_{w\in W\,,\ \|w\|_W\leq 1} \int_\Omega |\nabla u(t)|^{\hat{\gamma}}\, |\nabla\tilde{u}(t) - \nabla u(t)|^{1+\hat{\kappa}}\, w\, dx\Big)^2\, dt\right)^{1/2},$$

which, in order to end up with an estimate in terms of $\|\tilde{u}-u\|_{L^2(0,T;L^q(\Omega))}$ requires us to move the gradient by means of integration by parts. Assuming for simplicity that $\hat{\kappa} = 1$ we get

$$\begin{aligned}
&\left(\int_0^T \Big(\sup_{w\in W\,,\ \|w\|_W\leq 1} \int_\Omega |\nabla u(t)|^{\hat{\gamma}}\, |\nabla\tilde{u}(t) - \nabla u(t)|^2\, w\, dx\Big)^2\, dt\right)^{1/2}\\
&= \left(\int_0^T \Big(\sup_{w\in W\,,\ \|w\|_W\leq 1} \int_\Omega (\tilde{u}(t) - u(t))\, g^w(t)\, dx\Big)^2\, dt\right)^{1/2}\\
&\leq \|\tilde{u} - u\|_{L^2(0,T;L^q(\Omega))} \sup_{w\in W\,,\ \|w\|_W\leq 1} \|g^w\|_{L^\infty(0,T;L^{q^*}(\Omega))},
\end{aligned}$$

where

$$\begin{aligned} g^w(t) &= \nabla\cdot\Big(|\nabla u(t)|^{\hat\gamma}\,\nabla(\tilde u(t)-u(t))\,w\Big)\\ &= \hat\gamma|\nabla u(t)|^{\hat\gamma-2}(\nabla^2 u(t)\nabla u(t))\cdot\nabla(\tilde u(t)-u(t))\,w\\ &\quad+|\nabla u(t)|^{\hat\gamma}\,\Delta(\tilde u(t)-u(t))\,w+|\nabla u(t)|^{\hat\gamma}\,\nabla(\tilde u(t)-u(t))\cdot\nabla w\\ &\quad=: g_1(t)+g_2(t)+g_3(t), \end{aligned}$$

where ∇^2 denotes the Hessian. For the last term we proceed analogously to above (basically replacing u by ∇u and w by ∇w) to obtain

$$\begin{aligned} &\|g_3\|_{L^\infty(0,T;L^{q^*}(\Omega))} = \|\,|\nabla u(t)|^{\hat\gamma}\,\nabla(\tilde u(t)-u(t))\cdot\nabla w\|_{L^\infty(0,T;L^{q^*}(\Omega))}\\ &\le \|\nabla u\|^{\hat\gamma}_{L^\infty(0,T;L^R(\Omega))}\|\nabla(\tilde u-u)\|_{L^\infty(0,T;L^R(\Omega))}\sup_{w\in W,\ \|w\|_W\le 1}\|\nabla w\|_{L^{\frac{Rq^*}{R-q^*(\hat\gamma+1)}}(\Omega)} \end{aligned} \tag{B.114}$$

and use [100, Lemma 7.3] with $\nabla u\in L^2(0,T;W^{s-1,m}(\Omega))\cap H^1(0,T;(W^{t+1,n}(\Omega))^*)$, which under the conditions

$$\begin{aligned} &t-\frac{d}{n}\succeq 1-\frac{d(R-q^*(\hat\gamma+1))}{Rq^*},\\ &s-1-\tilde s-\frac{d}{m}\succeq t+1+\tilde s-\frac{d}{n},\ \ s-1-\tilde s\ge t+1+\tilde s,\ \ \tilde s-\frac{d}{2}\succeq-\frac{d}{R} \end{aligned} \tag{B.115}$$

yields $\nabla u\in L^\infty(0,T;H^{\tilde s}(\Omega))\subseteq L^\infty(0,T;L^R(\Omega))$ and $W\hookrightarrow W^{1,\frac{Rq^*}{R-q^*(\hat\gamma+1)}}(\Omega)$.

The other two terms can be bounded by

$$|g_1(t)+g_2(t)|\le\Big(\hat\gamma|\nabla^2 u(t)|\,|\nabla u(t)|^{\hat\gamma-1}\,|\nabla(\tilde u(t)-u(t))|+|\nabla^2(\tilde u(t)-u(t))|\,|\nabla u(t)|^{\hat\gamma}\Big)\,|w|$$

(note that here $|\cdot|$ denotes the Frobenius norm of a matrix) so that it suffices to find an estimate on expressions of the form

$$\||\nabla^2 z|\,|\nabla v|^{\hat\gamma-1}\,|\nabla y|\,|w|\|_{L^\infty(0,T;L^2(\Omega))}$$

for $z,v,y\in\mathcal U$, $w\in W$. To this end, we will again employ [100, Lemma 7.3], making use of the fact that for any $\varrho,R\in[1,\infty)$, due to Hölder's inequality with $P=\frac{\varrho}{2}$ and with $P=\frac{R(\varrho-2)}{2\varrho\hat\gamma}$, the estimate

$$\begin{aligned} &\|\,|\nabla^2 z|\,|\nabla v|^{\hat\gamma-1}|\nabla y|\,|w|\|_{L^2(\Omega)}\\ &\le\|\,|\nabla^2 z|\,\|_{L^\varrho(\Omega)}\|\Big(|\nabla v|^{\hat\gamma-1}|\nabla y|\Big)^{\frac{1}{\hat\gamma}}\|^{\hat\gamma}_{L^R(\Omega)}\|w\|_{L^{\frac{2R\varrho}{R(\varrho-2)-2\varrho\hat\gamma}}(\Omega)}\\ &\le C^\Omega_{H^{\hat s}\to L^\varrho}(C^\Omega_{H^{\check s}\to L^\varrho})^{\hat\gamma}\|\,|\nabla^2 z|\|_{H^{\hat s}(\Omega)}\|\Big(|\nabla v|^{\hat\gamma-1}|\nabla y|\Big)^{\frac{1}{\hat\gamma}}\|^{\hat\gamma}_{H^{\check s}(\Omega)}\|w\|_{L^{\frac{2R\varrho}{R(\varrho-2)-2\varrho\hat\gamma}}(\Omega)} \end{aligned} \tag{B.116}$$

holds. To make sense of these Hölder estimates and to guarantee continuity of the embedding $W \hookrightarrow L^{\frac{2R\varrho}{R(\varrho-2)-2\varrho\hat{\gamma}}}(\Omega)$ we impose

$$\varrho \geq 2 \text{ and } R \geq \frac{2\varrho\hat{\gamma}}{\varrho-2} \text{ and } t - \frac{d}{n} \succeq -\frac{d(R(\varrho-2)-2\varrho\hat{\gamma})}{2R\varrho} = -\frac{d}{2} + \frac{d}{\varrho} + \frac{d\hat{\gamma}}{R} \tag{B.117}$$

Taking into account the fact that here $\nabla^2 z$ contains second and $\left(|\nabla v|^{\hat{\gamma}-1}|\nabla y|\right)^{\frac{1}{\hat{\gamma}}}$ first derivatives of elements of $\mathcal{U}$, we therefore aim at continuity of the embeddings

$$\begin{aligned} &L^2(0,T;W^{s-2,m}(\Omega)) \cap H^1(0,T;W^{t+2,n}(\Omega)) \hookrightarrow L^\infty(0,T;H^{\hat{s}}(\Omega)) \hookrightarrow L^\infty(0,T;L^\varrho(\Omega)) \\ &L^2(0,T;W^{s-1,m}(\Omega)) \cap H^1(0,T;W^{t+1,n}(\Omega)) \hookrightarrow L^\infty(0,T;H^{\check{s}}(\Omega)) \hookrightarrow L^\infty(0,T;L^R(\Omega)), \end{aligned}$$

which can be achieved by means of [100, Lemma 7.3] under the conditions

$$\begin{aligned} &s-2-\hat{s}-\frac{d}{m} \succeq t+2+\hat{s}-\frac{d}{n} \text{ and } s-2-\hat{s} \geq t+2+\hat{s} \text{ and } \hat{s}-\frac{d}{2} \succeq -\frac{d}{\varrho} \\ &s-1-\check{s}-\frac{d}{m} \succeq t+1+\check{s}-\frac{d}{n} \text{ and } s-1-\check{s} \geq t+1+\check{s} \text{ and } \check{s}-\frac{d}{2} \succeq -\frac{d}{R}. \end{aligned} \tag{B.118}$$

For instance, we may set $\varrho = 2$, $R = \infty$ to obtain, inserting into (B.115), (B.117), (B.118), that $\hat{s} \geq 0$, $\check{s} > \frac{d}{2}$ hence

$$\begin{aligned} &t > \frac{d}{n},\ t - \frac{d}{n} \succeq 1 - \frac{d}{q^*},\ s \succeq t + 2 + \max\{2,d\} + \frac{d}{m} - \frac{d}{n},\ s \geq t + 2 + \max\{2,d\} \\ &s - \frac{d}{m} \succeq -\frac{d}{\hat{q}},\ q \leq \hat{q}. \end{aligned} \tag{B.119}$$

In order to avoid the use of too high values of s and t, we can alternatively skip the use of [100, Lemma 7.3] and instead set

$$\mathcal{U} = \{u \in L^\infty(0,T;L^r(\Omega)) \cap L^2(0,T;V) : \dot{u} \in L^2(0,T;W^*)\} \tag{B.120}$$

in case $C_{\Psi''} = 0$, or

$$\mathcal{U} = \{u \in L^\infty(0,T;L^r(\Omega) \cap W^{1,R}(\Omega) \cap W^{2,\varrho}(\Omega)) \cap L^2(0,T;V) : \dot{u} \in L^2(0,T;W^*)\} \tag{B.121}$$

otherwise. This can also be embedded in a Hilbert space setting by replacing $L^\infty(0,T)$ with $H^\sigma(0,T)$ for some $\sigma > \frac{1}{2}$. Going back to estimate (B.110) in case $C_{\Phi''} = 0$ we end up with the conditions

$$\gamma + \kappa \leq \frac{r}{q^*} \text{ and } t - \frac{d}{n} \succeq -\frac{d}{q^*} + \frac{d(\gamma+\kappa)}{r}, \tag{B.122}$$

cf. (B.111), and in case $C_{\Psi''} > 0$, considering estimates (B.114), (B.116) otherwise, we require

$$\begin{aligned} & t - \frac{d}{n} \succeq \max\left\{1 - \frac{d}{q^*} + \frac{d(\hat{\gamma}+1)}{R}, -\frac{d}{2} + \frac{d}{\varrho} + \frac{d\hat{\gamma}}{R}\right\} \text{ and} \\ & \hat{\gamma} + 1 \leq \frac{R}{q^*} \text{ and } \varrho \geq 2 \text{ and } \hat{\gamma} \leq \frac{R(\varrho-2)}{2\varrho}, \end{aligned} \tag{B.123}$$

cf. (B.115) (B.117), and in both cases we additionally need to impose (3.36). □

Part II

MAGNETIC PARTICLE IMAGING

Chapter 4

How Magnetic Particle Imaging works

In 2005, a completely new quantitative imaging method called *Magnetic Particle Imaging* (MPI) has been invented. MPI detects the nonlinear magnetization behavior of magnetic nanoparticles to an oscillating magnetic field for the purpose of determining the particle concentration. This technique promises high-resolution, high-sensitivity and real-time imaging.

In this chapter, we introduce some basic concepts of how MPI works. Essentially, the folowing information is summarized from the book by Knopp and Buzug [73].

4.1 MPI: a novel imaging technique

Magnetic particle imaging (MPI) is a dynamic imaging modality for medical applications that has first been introduced in 2005 by B. Gleich and J. Weizenecker [39]. Magnetic nanoparticles, consisting of a magnetic iron oxide core and a nonmagnetic coating, are inserted into the body to serve as a tracer. The key idea is to measure the nonlinear response of the nanoparticles to a temporally changing external magnetic field in order to draw conclusions on the spatial concentration of the particles inside the body. Since the particles are distributed along the bloodstream of a patient, the particle concentration yields information on the blood flow and is thus suitable for cardiovascular diagnosis or cancer detection [73, 74]. An overview of MPI basics is given in [73]. Since MPI requires the nanoparticles as a tracer, it mostly yields quantitative information on their distribution, but does not image the morphology of the body, such as the tissue density. The latter can be visualized using computerized tomography (CT) [90] or magnetic resonance imaging (MRI) [51]. These do not require a tracer, but involve ionizing radiation in the case of CT or, in the case of MRI, a strong magnetic field and a potentially high acquisition time. Other tracer-based methods are, e.g., single photon emission computerized tomography (SPECT) and positron emission tomography (PET) [35, 91, 106], which both involve radioactive radiation. The magnetic nanoparticles that are used in MPI, on the other hand, are not harmful for organisms.

Table 4.1: Quantitative comparison of different imaging modalities.

	CT	MRI	PET	SPECT	MPI
spatial resolution	0.5 mm	1 mm	4 mm	10 mm	< 1 mm
acquisition time	1 s	1 s - 1 h	1 min	1 min	< 0.1 s
sensitivity	low	low	high	high	high
harmfulness	X-ray	heating	radiation	radiation	heating

MPI physics

MPI uses a magnetic *gradient field*, known as a *selection field*, to saturate all particles outside a *field-free point* (FFP). The FFP is then rapidly traveled across a scanned volume by an oscillating *excitation* or *drive field*, that causes a change in magnetization of the particles. This temporal change in magnetization induces a voltage in the receive coil, which can be assigned to the instantaneous FFP location, generating an image of the *particle concentration*. These terminologies will be explained further in the next sections.

The MPI process could be summarized as follow:

- Magnetic nanoparticles consisting of a magnetic iron oxide core, whose diameter is in the range 1-100 nm and a nonmagnetic coating, are injected into the body to serve as a tracer.
- A magnetic *gradient field* is applied to saturate all particles outside a *field-free point* (FFP), the point where the applied field is zero.
- The FFP is then rapidly driven along a given trajectory across the scanned volume by an oscillating *excitation field*, causing a change in *magnetization* of the particles. [73].
- According to Faraday's law of induction, this temporal change in magnetization induces a voltage in the receive coil. This signal can be assigned to the instantaneous FFP location, which implies an image of the particle concentration.

4.2 Magnetic particles

Particle material

The goal of MPI is to identify the spatial distribution of magnetic material injected into the blood stream of a human body. One suitable magnetic material for MPI is iron oxide, which is usually available in the form of iron oxide based nanoparticles. Such particles consist of a magnetic core with a diameter in the range 1-100 nm, which causes its magnetization behavior. The magnetic core is covered by a magnetically neutral coating in order to prevent the impact among particles (*superparamagnetic*).

Particle concentration

Due to the small particle size in the nanometer range, it is impractical to capture the precise position of a particular particle. Instead, MPI aims at producing an image of the spatial *particle concentration*, which is defined by the number of particles per volume

$$c(x) = \frac{\text{number of particles}}{\text{volume}}.$$

Particle magnetization

The magnetic behavior of a single particle is described by its *magnetic moment vector*. When applying an external magnetic field, the particles start to align with the applied field causing the change in *particle magnetization*, i.e., the sum of magnetic moments as in Figure 4.1.

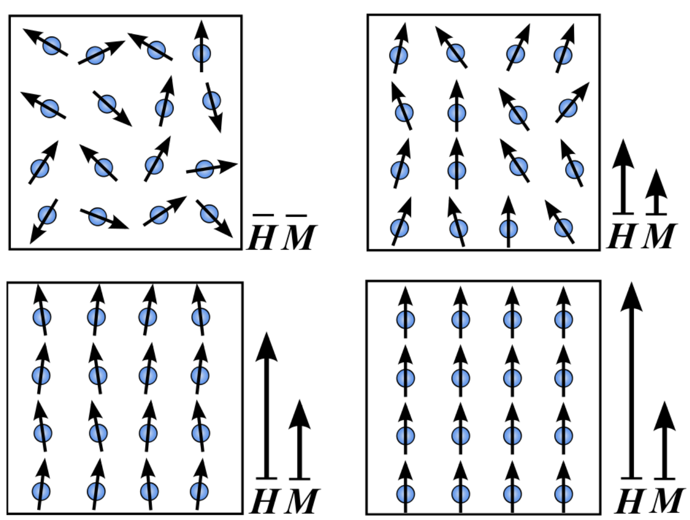

Figure 4.1: Magnetic behavior of superparamagnetic nanoparticles **M** in an applied magnetic field **H**.

The relation between the external magnetic field and the magnetization is not linear but exhibits a nonlinear behavior. When increasing the applied field, the magnetization initially shows a sharp increase, after that it goes flat into *saturation*, where the magnetization hardly changes anymore (see Figure 4.2).

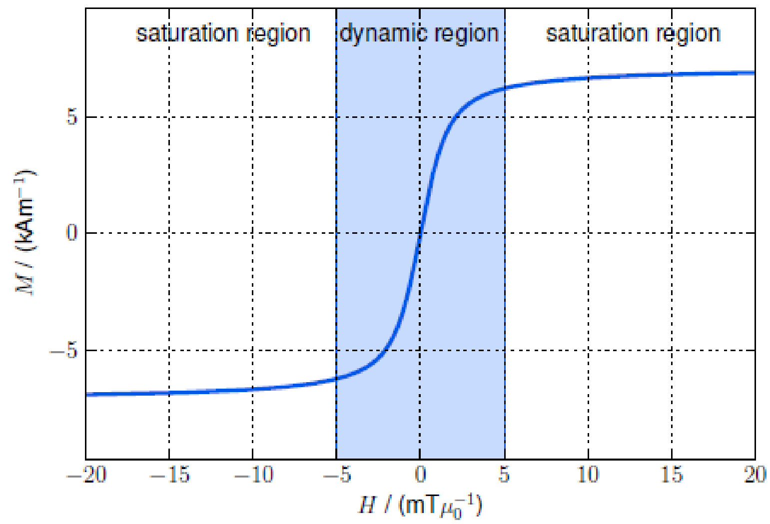

Figure 4.2: The relation between the external magnetic field and the magnetization of the particles of 30 nm diameter.

Relaxation effect

In reality, when applying a time varying magnetic field, the magnetic moment vector does not immediate follow the field **H**. It needs a certain time to reach the direction of **H**, in particular, the change in magnetization will be slightly later than the change in the applied field. This delay is called the *relaxation time* τ.

In general, there are two ways of how a particle changes its direction according to an applied magnetic field. Either the particle itself performs a physical rotation, which is named *Brownian rotation*; or only the magnetic moment in a fixed particle rotates, which is named *Neél relaxation* (Figure 4.3).

The relaxation effect heavily depends on the frequency f^H of the applied magnetic field. If $f^H > 1/\tau$, the magnetization can not follow the change of the applied field. This means, MPI technique must be performed with a relevant applied frequency.

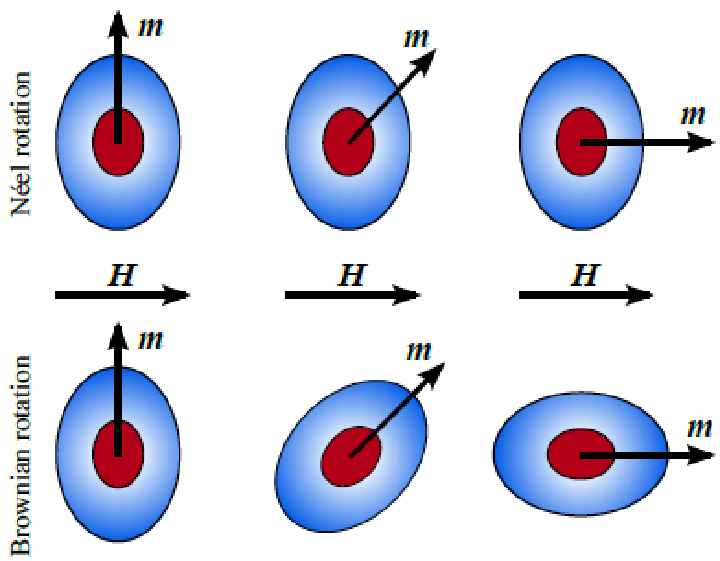

Figure 4.3: Comparison of Néel and Brownian rotation.

4.3 Signal generation

Induction principle

According to Faraday's law of induction, whenever the *magnetic flux density* $\mathbf{B}$ changes temporally, it induces an *electric field* $\mathbf{E}$ (Figure 4.4)

$$\nabla \times \mathbf{E} = -\frac{d\mathbf{B}}{dt}, \qquad \mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}),$$

where μ_0 is the *magnetic permeability* in vacuum. And the voltage recorded in the receive coil is the integral of the electric field strength along the conductor

$$v(t) = \int_{\partial S} E(\mathbf{l}) \cdot d\mathbf{l} = -\frac{d}{dt} \int_S \mathbf{B}(\mathbf{r}, t) \cdot d\mathbf{A}.$$

Detection of particle magnetization

As the magnetic flux density includes both applied excitation field $\mathbf{H}$ and particle magnetization $\mathbf{M}$, one needs to distinguish the induced excitation signal v^H and the induced particle signal v^P, which is the desirable signal

$$v^P(t) = -\frac{d}{dt}\mu_0 \int_S \mathbf{M}(\mathbf{r}, t) \cdot d\mathbf{A}. \tag{4.1}$$

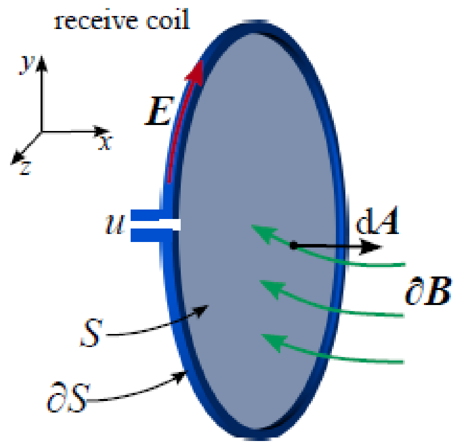

Figure 4.4: Voltage induced in a receive coil.

Indeed, it is achievable since v^P and v^H are different on the frequency domain, more precisely, v^P has higher frequency components than v^H (see Figure 4.5). This frequency discrimination is the result of the nonlinear magnetization with respect to the applied field captured in Figure 4.2. This motivates a method to eliminate the unwanted part v^H by applying a bandpass filter.

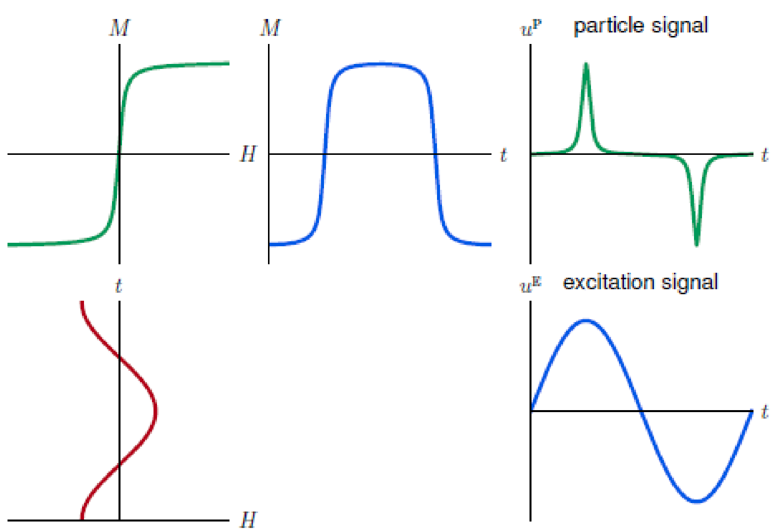

Figure 4.5: Signal generation in MPI: The particles are excited by a sinusoidal magnetic field causing a magnetization progression, which resembles a rectangular function. The induced particle signal contains two sharp peaks and can be distinguished from the sinusoidal excitation signal.

4.4 Selection field and drive field

The unique feature of MPI is the oscillating applied magnetic field, which causes the particles passing through the *field free point* to flip back and forth resulting the temporal change in their magnetization. This magnetization change is the key factor inducing the measurable voltage in the receive coil.

The applied magnetic field $\mathbf{H}$ (the *total field*) usually consists of two fields: *a selection* or *gradient field* and *a drive* or *excitation* $\mathbf{H}^D$ as in Figure 4.6.

The selection field is designed to be space-inhomogeneous and contains one special location named field free point (FFP), which is simply characterized by the point where the field magnitude is zero. Outside of the vicinity of the FFP, the field strength is quickly increasing in a linear fashion. The name "gradient field" indicates the gradient of the field being constant according to its linearity.

The drive field $\mathbf{H}^D$ is realized as constant in space but varies in time. This time-dependent field is employed to drive the FFP to every position in the scanned object as time progresses. The temporal change of this field is translated to a movement of the FFP, when it is superimposed with the selection field. Rapidly moving the FFP across the sample in an particular trajectory excites the particle in the FFP to be magnetized, thus induces the particle signal.

The most prominent *FFP trajectories* are Cartesian and Lissajous (Figure 4.7).

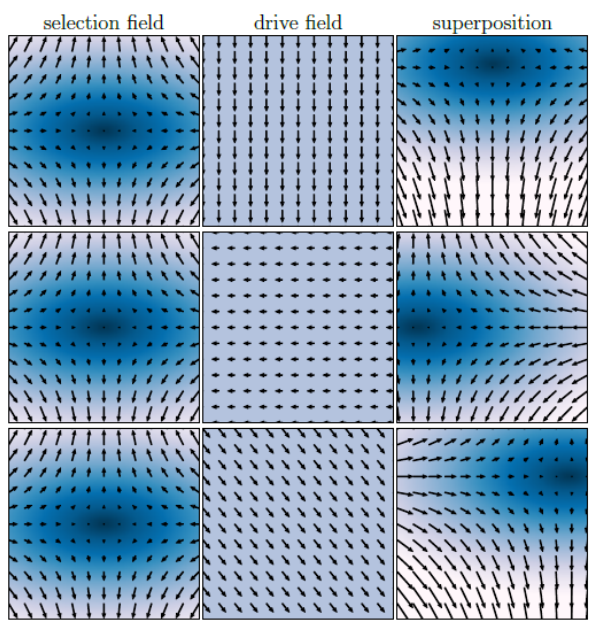

Figure 4.6: 2D translation of the FFP achieved by the superposition of a drive field with a selection field.

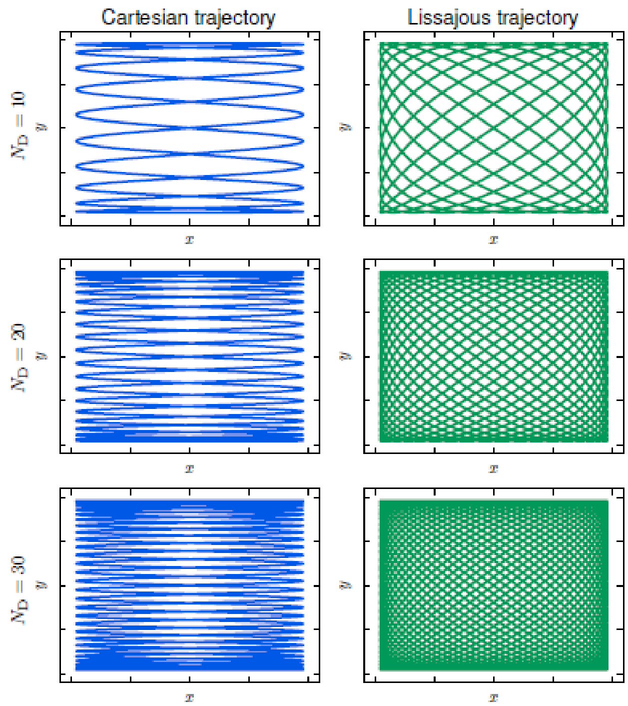

Figure 4.7: Cartesian and Lissajous trajectories for different density parameters N_D.

4.5 From data to images

The reconstruction problem in MPI is determining the particle distribution given the measured voltages in (4.1)

$$\begin{aligned} v(t) &= -\frac{d}{dt}\mu_0 \int_S \mathbf{M} \cdot d\mathbf{A} = -\frac{d}{dt}\mu_0 \int_S c\,\mathbf{m} \cdot d\mathbf{A} \\ &=: \int_\Omega c(\mathbf{x}) s(x,t)\, d\mathbf{x}. \end{aligned} \tag{4.2}$$

Here, $\mathbf{m}$ is magnetization of the material defined from the particle magnetization $\mathbf{M}$ via the relation $\mathbf{M} = c.\mathbf{m}$, and s is the *system function* encoding $\mathbf{m}$ and involves physical constants.

In order to reconstruct the particle distribution, it is obligatory to known the system function s, which appears as the kernel of the Fredholm integral equation of the first kind formulated by (4.2). Obviously, the system function depends on $\frac{d}{dt}\mathbf{m}$; hence the material magnetization $\mathbf{m}$ is required to be specified.

In MPI, the problem of finding the correct integral kernel/system function remains unsolved. This is the reason for our research in the next two chapters.

Chapter 5

MPI in the context of inverse problem

A decent mathematical model demands for a data-driven computation of the system function which does not only describe the measurement geometry but also encodes the interaction of the particles with the external magnetic field. The physical model of this interaction is given by the Landau-Lifshitz-Gilbert (LLG) equation. The determination of the system function can be seen as an inverse problem of its own, which can be interpreted as a calibration problem for MPI. In this chapter, the calibration problem is formulated as an inverse parameter identification problem for the LLG equation.

We begin with a detailed introduction to the modelling in MPI. Section 5.2 describes the full forward problem and presents the initial boundary value problem for the LLG equation that we use to describe the magnetization evolution. In Section 5.3, we formulate the inverse problem of calibration both in the all-at-once and in the reduced setting to obtain the final operator equation that is analyzed in the subsequent section. First, in Section 5.4.1, we present an analysis for the all-at-once setting. The inverse problem in the reduced setting is then addressed in Section 5.4.2. Finally, we conclude our findings and give an outlook on further research.

5.1 Introduction

At this point, there have been promising preclinical studies on the performance of MPI showing that this imaging modality has a great potential for medical diagnosis since it is highly sensitive with a good spatial and temporal resolution, and the data acquisition is very fast [74]. However, particularly in view of an application to image the human body, there remain some obstacles. One obstacle is the time-consuming calibration process. In this work, we assume that the concentration of the nanoparticles inside the body remains static throughout both the calibration process and the actual image acquisition. Mathematically, the forward problem of MPI then can essentially be formulated as an integral equation of the first kind for the particle concentration (or distribution) c,

$$u(t) = \int_{\Omega} c(x) s(x,t) \,\mathrm{d}x,$$

where the integration kernel s is called the *system function*. The system function encodes some geometrical aspects of the MPI scanner, such as the coil sensitivities of the receive coils in which the particle signal u is measured, but mostly it is determined by the particle behavior in response to the applied external magnetic field.

The actual inverse problem in MPI is to reconstruct the concentration c under the knowledge of the system function s from the measured data u. To this end, the system function has to be determined prior to the scanning procedure. This is usually done by evaluating a series of full scans of the field of view, where in each scan, a delta sample is placed in a different pixel until the entire field of view is covered [73]. Another option is a model-based approach for s (see [72] for an example), which basically involves a model for the particle magnetization. Since this model often depends on unknown parameters, the model-based determination of the system function itself can again be formulated as an inverse problem. This article now addresses this latter type of inverse problem, i.e., the identification of the system function for a known set of concentrations from calibration measurements. More precisely, our goal is to find a decent model for the time-derivative of the particle magnetization $\mathbf{m}$, which is proportional to s.

So far, in model-based approaches for the system function, the particle magnetization $\mathbf{m}$ is not modeled directly. Instead, one describes the mean magnetization $\overline{\mathbf{m}}$ of the particles via the Langevin function, i.e., the response of the particles is modeled on the mesoscopic scale [73, 71]. This approach is based on the assumption that the particles are in thermodynamic equilibrium and respond directly to the external field. For this reason, the mean magnetization is assumed to be a function of the external field, such that the mean magnetization is always aligned with the external field. The momentum of the mean magnetization is calculated via the Langevin function. This model, however, neglects some properties of the particle behavior. In particular, the magnetic moments of the particles do not align instantly with the external field [25].

In this work, we thus address an approach from micromagnetics, which models the time-dependent behavior of the magnetic material inside the particles' cores on the micro scale and allows for taking into account various additional physical properties such as particle-particle interaction. For an overview, see for example [76]. Since the core material is iron oxide, which is a ferrimagnetic material that shows a similar behavior as ferromagnets [26, 27], we use the *Landau-Lifshitz-Gilbert (LLG) equation*

$$\frac{\partial}{\partial t}\mathbf{m} = -\widetilde{\alpha}_1 \mathbf{m} \times (\mathbf{m} \times \mathbf{H}_{\text{eff}}) + \widetilde{\alpha}_2 \mathbf{m} \times \mathbf{H}_{\text{eff}},$$

see also [37, 83], for the evolution of the magnetization $\mathbf{m}$ of the core material. The field $\mathbf{H}_{\text{eff}}$ incorporates the external magnetic field together with other relevant physical effects. According to the LLG equation, the magnetization $\mathbf{m}$ performs a damped precession around the field vector of the external field, which leads to a relaxation effect. The LLG equation has been widely applied to describe the time evolution in micromagnetics [16, 43, 30].

In contrast to the imaging problem of MPI, the inverse problem of determining the magnetization $\mathbf{m}$ along with the constants $\widetilde{\alpha}_1, \widetilde{\alpha}_2$ turns out to be a nonlinear inverse problem, which is typical for parameter identification problems for partial differential equations, for example electrical impedance tomography [14], terahertz tomography [113], ultrasound imaging [24] and other applications from imaging and nondestructive testing [70].
We use the *all-at-once* as well as the *reduced* formulation of this inverse problem in a Hilbert space setting, see also [57, 58, 93], and analyze both cases including well-definedness of the forward mapping, continuity, Fréchet differentiability and calculate the adjoint mappings for the Fréchet derivatives. By consequence, iterative methods such as the Landweber method [49, 84], also in combination with Kaczmarz' method [45, 46], Newton methods (see, e.g., [99]), or subspace techniques [112] can be applied for the numerical solution. An overview of suitable regularization techniques is given in [63, 67].

Notation

The differential operators $-\Delta$ and ∇ are applied component wise to a vector field. In particular, this means that by $\nabla \mathbf{u}$ we denote the transpose of the Jacobian of $\mathbf{u}$. Moreover, $\langle \mathbf{a}, \mathbf{b} \rangle$ or $\mathbf{a} \cdot \mathbf{b}$ denotes the Euclidean inner product between two vectors; $A\colon B$ the Frobenius inner product between two matrices.

5.2 Underlying physical model for MPI

The basic physical principle that is exploited in MPI is Faraday's law of induction, which states that whenever the magnetic flux density $\mathbf{B}$ through a coil changes in time, this change induces an electric current in the coil. This current, or rather the respective voltage, can be measured. In MPI, the magnetic flux density $\mathbf{B}$ consists of the external applied magnetic field $\mathbf{H}_{\text{ext}}$ and the particle magnetization $\mathbf{M}^{\text{P}}$, i.e.,

$$\mathbf{B} = \mu_0 \left(\mathbf{H}_{\text{ext}} + \mathbf{M}^{\text{P}} \right),$$

where μ_0 is the magnetic permeability in vacuum. The particle magnetization $\mathbf{M}^{\text{P}}(x,t)$ in $x \in \Omega \subseteq \mathbb{R}^3$ depends linearly on the concentration $c(x)$ of magnetic material, which corresponds to the particle concentration, in $x \in \Omega$ and on the magnetization $\mathbf{m}(x,t)$ of the magnetic material. We thus have

$$\mathbf{M}^{\text{P}}(x,t) = c(x)\mathbf{m}(x,t),$$

where $|\mathbf{m}| = m_{\mathrm{S}} > 0$, i.e., the vector $\mathbf{m}$ has the fixed length m_{S} that depends on the magnetic core material inside the particles. At this point it is important to remark that we use a slightly different approach to separate the particle concentration, which carries the spatial information on the particles, from the magnetization behavior of the magnetic material and the measuring process. In our approach, the concentration is a dimensionless quantity, whereas in most models, it is defined as the number of particles per unit volume (see, e.g. [73], also Section 4.2 in Chapter 4).

A detailed derivation of the forward model in MPI, based on the equilibrium model for the magnetization, can be found in [73]. The steps that are related to the measuring process can be adapted to our approach. For the reader's convenience, we want to give a short overview and introduce the parameters related to the scanner setup.

If the receive coil is a simple conductor loop, which encloses a surface S, the voltage that is induced can be expressed by

$$u(t) = -\frac{\mathrm{d}}{\mathrm{d}t}\int_S \mathbf{B}(x,t)\cdot \,\mathrm{d}\mathbf{A} = -\mu_0 \frac{\mathrm{d}}{\mathrm{d}t}\int_S \left(\mathbf{H}_{\mathrm{ext}} + \mathbf{M}^{\mathrm{P}}\right)\cdot \,\mathrm{d}\mathbf{A}. \tag{5.1}$$

The signal that is recorded in the receive coil thus originates from temporal changes of the external magnetic field $\mathbf{H}$ as well as of the particle magnetization $\mathbf{M}^{\mathrm{P}}$,

$$\begin{aligned} u(t) &= -\mu_0 \left(\int_\Omega \mathbf{p}^{\mathrm{R}}(x)\cdot \frac{\partial}{\partial t}\mathbf{H}_{\mathrm{ext}}(x,t)\,\mathrm{d}x + \int_\Omega \mathbf{p}^{\mathrm{R}}(x)\cdot \frac{\partial}{\partial t}\mathbf{M}^{\mathrm{P}}(x,t)\,\mathrm{d}x \right) && (5.2)\\ &=: u^{\mathrm{E}}(t) + u^{\mathrm{P}}(t) && (5.3) \end{aligned}$$

For the signal that is caused by the change in the particle magnetization we obtain

$$\begin{aligned} u^{\mathrm{P}}(t) &= -\mu_0 \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \mathbf{p}^{\mathrm{R}}(x)\cdot \mathbf{M}^{\mathrm{P}}(x,t)\,\mathrm{d}x \\ &= -\mu_0 \int_\Omega \mathbf{p}^{\mathrm{R}}(x)\cdot \frac{\partial}{\partial t}\mathbf{M}^{\mathrm{P}}(x,t)\,\mathrm{d}x \\ &= -\mu_0 \int_\Omega c(x)\mathbf{p}^{\mathrm{R}}(x)\cdot \frac{\partial}{\partial t}\mathbf{m}(x,t)\,\mathrm{d}x \\ &= -\mu_0 \int_\Omega c(x)s(x,t)\,\mathrm{d}x. \end{aligned}$$

The function

$$s(x,t) := \mathbf{p}^{\mathrm{R}}(x)\cdot \frac{\partial}{\partial t}\mathbf{m}(x,t) = \left\langle \mathbf{p}^{\mathrm{R}}(x), \frac{\partial}{\partial t}\mathbf{m}(x,t)\right\rangle_{\mathbb{R}^3} \tag{5.4}$$

is called the *system function* and can be interpreted as a potential to induce a signal in the receive coil. The function $\mathbf{p}^{\mathrm{R}}$ is called the coil sensitivity and is determined by the architecture of the respective receive coil. For our purposes, we assume that $\mathbf{p}^{\mathrm{R}}$ is known. The measured signal that originates from the magnetic particles can thus essentially be calculated via an integral equation of the first kind with a time-dependent integration kernel s.

The particle magnetization, however, changes in time in response to changes of the external field. It is thus an important objective to encode the interplay of the external field and the particles in a sufficiently accurate physical model. The magnetization of the magnetic particles that are used in MPI can be considered on different scales. The following characterization from ferromagnetism has been taken from [76]:

- On the *atomic level*, one can describe the behavior of a magnetic material as a spin system and take into account stochastic effects that arise, for example, from Brownian motion.
- In the *microscopic scale*, continuum physics is applied to work with deterministic equations describing the magnetization of the magnetic material.
- In the *mesoscopic scale*, we can describe the magnetization behavior via a mean magnetization, which is an average particle magnetic moment.
- Finally, on a *macroscopic scale*, all aspects that arise from the microstructure are neglected and the magnetization is described by phenomenological constitutive laws.

In this work, we intend to use a model from micromagnetism allowing us to work with a deterministic equation to describe the magnetization of the magnetic material. The core material of the nanoparticles consists of iron-oxide or magnetite, which is a ferrimagnetic material. The magnetization curve of ferrimagnetic materials is similar to the curve that is observed for ferromagnets but with a lower saturation magnetization (see, e.g., [26, 27]). This approach has also been suggested in [98]. The evolution of the magnetization in time is described by the *Landau-Lifshitz-Gilbert (LLG) equation*

$$\mathbf{m}_t := \frac{\partial}{\partial t}\mathbf{m} = -\widetilde{\alpha}_1 \mathbf{m} \times (\mathbf{m} \times \mathbf{H}_{\mathrm{eff}}) + \widetilde{\alpha}_2 \mathbf{m} \times \mathbf{H}_{\mathrm{eff}}, \tag{5.5}$$

see [37, 76] and the therein cited literature. The coefficients

$$\widetilde{\alpha}_1 := \frac{\gamma \alpha_{\mathrm{D}}}{m_{\mathrm{S}}(1+\alpha_{\mathrm{D}}^2)} > 0, \quad \widetilde{\alpha}_2 := \frac{\gamma}{(1+\alpha_{\mathrm{D}}^2)} > 0$$

are material parameters that contain the gyromagnetic constant γ, the saturation magnetization m_{S} of the core material and a damping parameter α_{D}. The vector field $\mathbf{H}_{\mathrm{eff}}$ is called the *effective magnetic field*. It is defined as the negative gradient

$-D\mathcal{E}(\mathbf{m})$ of the *Landau energy* $\mathcal{E}(\mathbf{m})$ of a ferromagnet, see, e.g., [76]. Taking into account only the interaction with the external magnetic field $\mathbf{H}$ and particle-particle interactions, this energy is given by

$$\mathcal{E}_A(\mathbf{m}) = A \int_\Omega |\nabla \mathbf{m}|^2 \, \mathrm{d}x - \mu_0 m_\mathrm{S} \int_\Omega \langle \mathbf{H}, \mathbf{m} \rangle_{\mathbb{R}^3} \, \mathrm{d}x,$$

where $A \geq 0$ is a scalar parameter (the exchange stiffness constant [37]). We thus have

$$\mathbf{H}_\mathrm{eff} = 2A\Delta\mathbf{m} + \mu_0 m_\mathrm{S} \mathbf{H}_\mathrm{ext}. \tag{5.6}$$

Together with Neumann boundary conditions and a suitable initial condition, our model for the magnetization thus reads

$$\begin{aligned} \mathbf{m}_t &= -\alpha_1 \mathbf{m} \times (\mathbf{m} \times (\Delta\mathbf{m} + \mathbf{h}_\mathrm{ext})) + \alpha_2 \mathbf{m} \times (\Delta\mathbf{m} + \mathbf{h}_\mathrm{ext}) && \text{in } [0,T] \times \Omega && (5.7)\\ 0 &= \partial_\nu \mathbf{m} && \text{on } [0,T] \times \partial\Omega && (5.8)\\ \mathbf{m}_0 &= \mathbf{m}(t=0) \; |\mathbf{m}_0| = m_\mathrm{S} && \text{in } \Omega, && (5.9) \end{aligned}$$

where $\mathbf{h}_\mathrm{ext} = \frac{\mu_0 m_\mathrm{S}}{2A} \mathbf{H}_\mathrm{ext}$ and $\alpha_1 := 2A\widetilde{\alpha}_1, \alpha_2 := 2A\widetilde{\alpha}_2 > 0$. The initial value $\mathbf{m}_0 = \mathbf{m}(t=0)$ corresponds to the magnetization of the magnetic material in the beginning of the measurement. To obtain a reasonable value for $\mathbf{m}_0$, we take into account that the external magnetic field is switched on before the measuring process starts, i.e., $\mathbf{m}_0$ is the state of the magnetization that is acquired when the external field is static. This allows us to precompute $\mathbf{m}_0$ as the solution of the stationary problem

$$\alpha_1 \mathbf{m}_0 \times (\mathbf{m}_0 \times (\Delta\mathbf{m}_0 + \mathbf{h}_\mathrm{ext}(t=0))) = \alpha_2 \mathbf{m}_0 \times (\Delta\mathbf{m}_0 + \mathbf{h}_\mathrm{ext}(t=0)) \tag{5.10}$$

with Neumann boundary conditions.

Remark 5.2.1. In the stationary case, damping does not play a role, and if we additionally neglect particle-particle interactions, we obtain the approximative equation

$$\hat{\mathbf{m}}_0 \times (\hat{\mathbf{m}}_0 \times \mathbf{h}_\mathrm{ext}(t=0)) = 0$$

with an approximation $\hat{\mathbf{m}}_0$ to $\hat{\mathbf{m}}$, since $\alpha_2 \approx 0$ and $\mathbf{H}_\mathrm{eff} \approx \mu_0 m_\mathrm{S} \mathbf{H}_\mathrm{ext}$. The above equation yields $\hat{\mathbf{m}}_0 \parallel \mathbf{h}_\mathrm{ext}(t=0)$. Together with $|\hat{\mathbf{m}}_0| = m_\mathrm{S}$ this yields

$$\hat{\mathbf{m}}_0 = m_\mathrm{S} \frac{\mathbf{h}_\mathrm{ext}(t=0)}{|\mathbf{h}_\mathrm{ext}(t=0)|}.$$

This represents a good approximation to $\mathbf{m}_0$ where $\mathbf{h}_\mathrm{ext}$ is strong at the time point $t = 0$:

$$\mathbf{m}_0 \approx \hat{\mathbf{m}}_0 = m_\mathrm{S} \frac{\mathbf{h}_\mathrm{ext}(t=0)}{|\mathbf{h}_\mathrm{ext}(t=0)|}.$$

5.2.1 Observation operator

Faraday's law states that a temporally changing magnetic field induces an electric current in a conductor loop or coil, which yields the relation (5.1). By consequence, not only the change in the particle magnetization contributes to the induced current but also the dynamic external magnetic field $\mathbf{H}_{\text{ext}}$. Since we need the particle signal for the determination of the particle magnetization, we need to separate the particle signal from the excitation signal due to the external field. This is realized by processing the signal in a suitable way using filters.
MPI scanners usually use multiple receive coils to measure the induced particle signal at different positions in the scanner. We assume that we have $L \in \mathbb{N}$ receive coils with coil sensitivities $\mathbf{p}_\ell^{\mathrm{R}}$, $\ell = 1, ..., L$, and the measured signal is given by

$$\widetilde{v}_\ell(t) = -\mu_0 \int_0^T \widetilde{a}_\ell(t-\tau) \int_\Omega c(x)\mathbf{p}_\ell^{\mathrm{R}}(x) \cdot \frac{\partial}{\partial \tau}\mathbf{m}(x,\tau)\,\mathrm{d}x\,\mathrm{d}\tau, \tag{5.11}$$

where T is the repetition time of the acquisition process, i.e., the time that is needed for one full scan of the object, and $a_\ell : [0,T] \to \mathbb{R}$ is the transfer function with periodic continuation $\widetilde{a}_\ell : \mathbb{R} \to \mathbb{R}$. The transfer function serves as a filter to separate particle and excitation signal, i.e., it is chosen such that

$$\widetilde{v}_\ell^{\mathrm{E}}(t) := \left(\widetilde{a}_\ell * u_\ell^{\mathrm{E}}\right)(t) = -\mu_0 \int_0^T \widetilde{a}_\ell(t-\tau) \int_\Omega \mathbf{p}_\ell^{\mathrm{R}}(x) \cdot \frac{\partial}{\partial t}\mathbf{H}_{\text{ext}}(x,t)\,\mathrm{d}x\,\mathrm{d}t \approx 0.$$

In practice, $\widetilde{a}_\ell$ is often a band pass filter. For a more detailed discussion of the transfer function, see also [73]. In this work, the transfer function is known analytically.

We define

$$\mathbf{K}_\ell(t,\tau,x) := -\mu_0 \widetilde{a}_\ell(t-\tau)c(x)\mathbf{p}_\ell^{\mathrm{R}}(x),$$

such that the measured particle signals are given by

$$v_\ell(t) = \int_0^T \int_\Omega \mathbf{K}_\ell(t,\tau,x) \cdot \frac{\partial}{\partial \tau}\mathbf{m}(x,\tau)\,\mathrm{d}\tau\,\mathrm{d}x, \tag{5.12}$$

where $\mathbf{m}$ fulfills (5.7), (5.8), (5.9).

To determine $\mathbf{m}$ in $\Omega \times (0,T)$, we use the data $v_{k\ell}(t)$, $k = 1, ..., K$, $\ell = 1, ..., L$, from the scans that we obtain for different particle concentrations c_k, $k = 1, ..., K$, $K \in \mathbb{N}$. The forward operator thus reads

$$v_{k\ell}(t) = \int_0^T \int_\Omega \mathbf{K}_{k\ell}(t,\tau,x) \cdot \frac{\partial}{\partial \tau}\mathbf{m}(x,\tau)\,\mathrm{d}x\,\mathrm{d}\tau, \quad \mathbf{K}_{k\ell}(t,\tau,x) := -\mu_0 \widetilde{a}_\ell(t-\tau)c_k(x)\mathbf{p}_\ell^{\mathrm{R}}(x). \tag{5.13}$$

5.2.2 Landau-Lifshitz-Gilbert equation

In this section, we derive additional formulations of (5.7) - (5.9) that are suitable for the analysis. The approach is motivated by [76], where only particle-particle interactions are taken into account.
First of all, we observe that multiplying (5.7) with $\mathbf{m}$ on both sides yields

$$\frac{1}{2} \cdot \frac{\mathrm{d}}{\mathrm{d}t} |\mathbf{m}(x,t)|^2 = \mathbf{m}(x,t) \cdot \mathbf{m}_t(x,t) = 0, \tag{5.14}$$

which shows that the absolute value of $\mathbf{m}$ does not change in time. Since $|\mathbf{m}_0| = m_{\mathrm{S}}$, we have $\mathbf{m}(x,t) \in m_{\mathrm{S}} \cdot \mathcal{S}^2$, where $\mathcal{S}^2 := \{\mathbf{v} \in \mathbb{R}^3 \, : \, |\mathbf{v}| = 1\}$ is the unit sphere in $\mathbb{R}^3$. As a consequence, we have $0 = \nabla |\mathbf{m}|^2 = 2 \nabla \mathbf{m} \cdot \mathbf{m}$ in Ω, so that by taking the divergence, we get

$$\langle \mathbf{m}, \Delta \mathbf{m} \rangle = - \langle \nabla \mathbf{m}, \nabla \mathbf{m} \rangle. \tag{5.15}$$

Now we make use of the identity

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \langle \mathbf{a}, \mathbf{c} \rangle \mathbf{b} - \langle \mathbf{a}, \mathbf{b} \rangle \mathbf{c}$$

for $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ to derive

$$\mathbf{m} \times (\mathbf{m} \times \Delta \mathbf{m}) = \langle \mathbf{m}, \Delta \mathbf{m} \rangle \mathbf{m} - |\mathbf{m}|^2 \Delta \mathbf{m} = -|\nabla \mathbf{m}|^2 \mathbf{m} - m_{\mathrm{S}}^2 \Delta \mathbf{m}, \tag{5.16}$$

$$\mathbf{m} \times (\mathbf{m} \times \mathbf{h}_{\mathrm{ext}}) = \langle \mathbf{m}, \mathbf{h}_{\mathrm{ext}} \rangle \mathbf{m} - |\mathbf{m}|^2 \mathbf{h}_{\mathrm{ext}} = \langle \mathbf{m}, \mathbf{h}_{\mathrm{ext}} \rangle \mathbf{m} - m_{\mathrm{S}}^2 \mathbf{h}_{\mathrm{ext}}. \tag{5.17}$$

Using (5.15) together with (5.16), (5.17) and $|\mathbf{m}| = m_{\mathrm{S}}$, we obtain from (5.7) - (5.9)

$$\begin{aligned} \mathbf{m}_t - \alpha_1 \, m_{\mathrm{S}}^2 \, \Delta \mathbf{m} = \alpha_1 |\nabla \mathbf{m}|^2 \mathbf{m} + \alpha_2 \mathbf{m} \times \Delta \mathbf{m} - \alpha_1 \langle \mathbf{m}, \mathbf{h}_{\mathrm{ext}} \rangle \mathbf{m} & \\ + \alpha_1 \, m_{\mathrm{S}}^2 \, \mathbf{h}_{\mathrm{ext}} + \alpha_2 \mathbf{m} \times \mathbf{h}_{\mathrm{ext}} \qquad & \text{in } [0,T] \times \Omega \qquad (5.18) \\ 0 = \partial_\nu \mathbf{m} \qquad & \text{on } [0,T] \times \partial\Omega \quad (5.19) \\ \mathbf{m}_0 = \mathbf{m}(t=0), \; |\mathbf{m}_0| = m_{\mathrm{S}} \qquad & \text{in } \Omega. \qquad (5.20) \end{aligned}$$

Taking the cross product of $\mathbf{m}$ with (5.18) and multiplying with $-\hat{\alpha}_2$, where $\hat{\alpha}_1 = \frac{\alpha_1}{m_{\mathrm{S}}^2 \alpha_1^2 + \alpha_2^2}$, $\hat{\alpha}_2 = \frac{\alpha_2}{m_{\mathrm{S}}^2 \alpha_1^2 + \alpha_2^2}$, by (5.16), (5.17) and cancellation of the first and third term on the right hand side, we get

$$\begin{aligned} & - \hat{\alpha}_2 \mathbf{m} \times \mathbf{m}_t + \alpha_1 \hat{\alpha}_2 m_{\mathrm{S}}^2 \, \mathbf{m} \times \Delta \mathbf{m} \\ & = \frac{\alpha_2^2}{m_{\mathrm{S}}^2 \alpha_1^2 + \alpha_2^2} \left(|\nabla \mathbf{m}|^2 \mathbf{m} + m_{\mathrm{S}}^2 \Delta \mathbf{m} \right) \\ & \qquad \alpha_1 \hat{\alpha}_2 m_{\mathrm{S}}^2 \, \mathbf{m} \times \mathbf{h}_{\mathrm{ext}} + \frac{\alpha_2^2}{m_{\mathrm{S}}^2 \alpha_1^2 + \alpha_2^2} \left(m_{\mathrm{S}}^2 \mathbf{h}_{\mathrm{ext}} - \langle \mathbf{m}, \mathbf{h}_{\mathrm{ext}} \rangle \mathbf{m} \right) , \end{aligned}$$

where the second term on the left hand side can be expressed via (5.18) as

$$\begin{aligned} & \alpha_1 \hat{\alpha}_2 \mathbf{m} \times \Delta \mathbf{m} \\ & = \hat{\alpha}_1 \mathbf{m}_t + \frac{\alpha_1^2}{m_{\mathrm{S}}^2 \alpha_1^2 + \alpha_2^2} \left(-m_{\mathrm{S}}^2 \Delta \mathbf{m} - |\nabla \mathbf{m}|^2 \mathbf{m} + \langle \mathbf{m}, \mathbf{h}_{\mathrm{ext}} \rangle \mathbf{m} - m_{\mathrm{S}}^2 \mathbf{h}_{\mathrm{ext}} \right) - \alpha_1 \hat{\alpha}_2 \mathbf{m} \times \mathbf{h}_{\mathrm{ext}}. \end{aligned}$$

This yields the alternative formulation

$$\begin{aligned}
\hat{\alpha}_1 m_{\mathrm{S}}^2 \mathbf{m}_t - \hat{\alpha}_2 \mathbf{m} \times \mathbf{m}_t - m_{\mathrm{S}}^2 \Delta \mathbf{m} &= |\nabla \mathbf{m}|^2 \mathbf{m} + m_{\mathrm{S}}^2 \mathbf{h}_{\mathrm{ext}} - \langle \mathbf{m}, \mathbf{h}_{\mathrm{ext}} \rangle \mathbf{m} \\
& \qquad \text{in } [0,T] \times \Omega, && (5.21) \\
0 &= \partial_\nu \mathbf{m} \qquad \text{on } [0,T] \times \partial\Omega, && (5.22) \\
\mathbf{m}_0 = \mathbf{m}(t=0),\ |\mathbf{m}_0| &= m_{\mathrm{S}} \qquad \text{in } \Omega. && (5.23)
\end{aligned}$$

5.3 Inverse problem for calibration process

Apart from the obvious inverse problem of determining the concentration c of magnetic particles inside a body from the measurements v_ℓ, $\ell = 1, ..., L$, MPI gives rise to a range of further parameter identification problems of entirely different nature. In this work, we are not addressing the imaging process itself, but consider an inverse problem that is essential for the calibration process. Here, calibration refers to determining the system function s_ℓ, which serves as an integral kernel in the imaging process. The system function includes all system parameters of the tomograph and encodes the physical behaviour of the magnetic material in the cores of the magnetic particles inside a temporally changing external magnetic field. Experiments show that a simple model for the magnetization based on the assumption that the particles are in their equilibrium state at all times is insufficient for the imaging, see, e.g., [72]. A model-based approach with an enhanced physical model has been so far omitted due to the complexity of the involved physics and, the system function is usually measured in a time-consuming calibration process [73, 74].

In this work, we address the inverse problem of calibrating an MPI system for a given set of standard calibration concentrations c_k, $k = 1, ..., K$, for which we measure the corresponding signals and obtain the data $v_{k\ell}(t)$, $k = 1, ..., K$, $\ell = 1, ..., L$. Here we assume that the coil sensitivity $\mathbf{p}_\ell^{\mathrm{R}}$ as well as the transfer function $\widetilde{a}_\ell$ are known.

This, together with the fact that $\mathbf{m}$ is supposed to satisfy the LLG equation (5.21)–(5.23), is used to determine the system function (5.4). Actually, since $\mathbf{p}^{\mathrm{R}}$ is known, the inverse problem under consideration here consists of reconstructing $\mathbf{m}$ from (5.13), (5.21)–(5.23). As the initial boundary value problem (5.21)–(5.23) has a unique solution $\mathbf{m}$ for given $\hat{\alpha}_1$, $\hat{\alpha}_2$, it actually suffices to determine these two parameters. This is the point of view that we take when using a classical reduced formulation of the calibration problem

$$F(\hat{\alpha}) = y \tag{5.24}$$

with the data $y_{k\ell} = v_{k\ell}$ and the forward operator

$$F : \mathcal{D}(F)(\subseteq \mathcal{X}) \to \mathcal{Y}, \qquad \hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2) \mapsto \mathcal{K} \frac{\partial}{\partial t} S(\hat{\alpha}) \tag{5.25}$$

containing the parameter-to-state map

$$S : \mathcal{X} \to \tilde{\mathcal{U}} \tag{5.26}$$

that maps the parameters $\hat{\alpha}$ into the solution $\mathbf{m} := S(\hat{\alpha})$ of the LLG initial boundary value problem (5.21)–(5.23). The linear operator $\mathcal{K}$ is the integral operator defined by the kernels $\mathbf{K}_{k\ell}$ $k = 1, ..., K$, $\ell = 1, ..., L$, i.e.,

$$\mathcal{K}_{k\ell}\mathbf{u} = \int_0^T \int_\Omega \mathbf{K}_{k\ell}(t, \tau, \mathbf{x}) \cdot \mathbf{u}(\mathbf{x}, \tau)\, \mathrm{d}\tau\, \mathrm{d}\mathbf{x}\,. \tag{5.27}$$

Here, the preimage and image spaces are defined by

$$\mathcal{X} = \mathbb{R}^2, \qquad \mathcal{Y} = L^2(0,T)^{KL}, \tag{5.28}$$

and the state space $\tilde{\mathcal{U}}$ will be chosen appropriately below, see Section 5.4.2.

Alternatively, we also consider the all-at-once formulation of the inverse problem as a simultaneous system

$$\mathbb{F}(\mathbf{m}, \hat{\alpha}) = \mathbf{y} := (0, y)^T \tag{5.29}$$

for the state $\mathbf{m}$ and the parameters $\hat{\alpha}$, with the forward operator

$$\mathbb{F}(\mathbf{m}, \hat{\alpha}) = \begin{pmatrix} \mathbb{F}_0(\mathbf{m}, \hat{\alpha}) \\ \Big(\mathbb{F}_{k\ell}(\mathbf{m}, \hat{\alpha})\Big)_{k=1,...,K\,,\ \ell=1,...,L} \end{pmatrix}$$

where

$$\mathbb{F}_0(\mathbf{m}, \hat{\alpha}_1, \hat{\alpha}_2) =: \hat{\alpha}_1\mathbf{m}_t - \Delta\mathbf{m} - \hat{\alpha}_2\mathbf{m} \times \mathbf{m}_t - |\nabla\mathbf{m}|^2\mathbf{m} - \mathbf{h}_{\text{ext}} + (\mathbf{m} \cdot \mathbf{h}_{\text{ext}})\mathbf{m}$$

and

$$\mathbb{F}_{k\ell}(\mathbf{m}, \hat{\alpha}_1, \hat{\alpha}_2) = \mathcal{K}_{k,\ell}\mathbf{m}_t$$

with $\mathcal{K}_{k,\ell}$ as in (5.27). Here $\mathbb{F}$ maps between $\mathcal{U} \times \mathcal{X}$ and $\mathcal{W} \times \mathcal{Y}$ with $\mathcal{X}$, $\mathcal{Y}$ as in (5.28), and $\mathcal{U}$, $\mathcal{W}$ appropriately chosen function spaces, see Section 5.4.1.

Iterative methods for solving inverse problems usually require the linearization $F'(\hat{\alpha})$ of the forward operator F and its adjoint $F'(\hat{\alpha})^*$ (and likewise for $\mathbb{F}$) in the given Hilbert space setting.

For example, consider Landwebers' iteration cf., e.g., [84, 49] defined by a gradient decent method for the least squares functional $\|F(\hat{\alpha}) - y\|^2_{\mathcal{Y}}$ as

$$\hat{\alpha}_{n+1} = \hat{\alpha}_n - \mu_n F'(\hat{\alpha}_n)^*(F(\hat{\alpha}_n) - y)$$

with an appropriately chosen step size μ_n. Alternatively, one can split the forward operator into a system by considering it row wise $F_k(\hat{\alpha}) = y_k$ with $F_k = (F_{kl})_{\ell=1...L}$ or column wise $F_\ell(\hat{\alpha}) = y_\ell$ with $F_\ell = (F_{kl})_{k=1,...,K}$, or even element wise $F_{kl}(\hat{\alpha}) = y_{kl}$, and cyclically iterating over these equations with gradient descent steps in a Kaczmarz version of the Landweber iteration cf., e.g., [46, 45]. The same can be done with the respective all-at-once versions [57]. These methods extend to Banach spaces as well by using duality mappings, cf., e.g., [104], however, for the sake of simplicity of exposition and implementation, we will concentrate on a Hilbert space setting here; in particular, all adjoints will be Hilbert space adjoints.

5.4 Derivatives and adjoints

Motivated by their need in iterative reconstruction methods, we now derive and rigorously justify derivatives of the forward operators as well as their adjoints, both in an all-at-once and in a reduced setting.

Notation wise there will be a slight move from the somewhat physics oriented paradigm of the previous section to a more mathematically conventional one: For instance, the subscript "ext" in the external magnetic field will be skipped and boldface notation of x will be not be continued here. Moreover, to avoid confusion with the dual pairing, we will use the dot notation for the Euclidean inner product.

5.4.1 All-at-once formulation

We split the magnetization additively into its given initial value $\mathbf{m}_0$ and the unknown rest $\hat{\mathbf{m}}$, so that the forward operator reads

$$\mathbb{F}(\hat{\mathbf{m}}, \hat{\alpha}_1, \hat{\alpha}_2) = \begin{pmatrix} \mathbb{F}_0(\hat{\mathbf{m}}, \hat{\alpha}_1, \hat{\alpha}_2) \\ \\ \Big(\mathbb{F}_{k\ell}(\hat{\mathbf{m}}, \hat{\alpha}_1, \hat{\alpha}_2)\Big)_{k=1,\dots,K\,,\;\ell=1,\dots,L} \end{pmatrix}$$
$$:= \begin{pmatrix} \hat{\alpha}_1\hat{\mathbf{m}}_t - \Delta_N(\mathbf{m}_0+\hat{\mathbf{m}}) - \hat{\alpha}_2(\mathbf{m}_0+\hat{\mathbf{m}}) \times \hat{\mathbf{m}}_t \\ \qquad -|\nabla(\mathbf{m}_0+\hat{\mathbf{m}})|^2(\mathbf{m}_0+\hat{\mathbf{m}}) - \mathbf{h} + ((\mathbf{m}_0+\hat{\mathbf{m}})\cdot\mathbf{h})(\mathbf{m}_0+\hat{\mathbf{m}}) \\ \\ \Big(\int_0^T\int_\Omega \mathbf{K}_{k\ell}(t,\tau,x)\cdot\mathbf{m}_t(x,\tau)\,dx\,d\tau\Big)_{k=1,\dots,K\,,\;\ell=1,\dots,L} \end{pmatrix}$$

for given $\mathbf{h} \in L^2(0,T;L^p(\Omega;\mathbb{R}^3))$, $p \geq 2$, where $\Delta_N : H^1(\Omega) \to H^1(\Omega)^*$ and, using the same notation, $\Delta_N : H^2_N(\Omega) \to L^2(\Omega)(\subseteq H^1(\Omega)^*)$ with $H^2_N(\Omega) = \{u \in H^2(\Omega) : \partial_\nu u = 0 \text{ on } \partial\Omega\}$ [1] is equipped with homogeneous Neumann boundary conditions, i.e, it is defined by

$$\langle -\Delta_N u, v\rangle_{H^1(\Omega)^*,H^1(\Omega)} = (\nabla u, \nabla v)_{L^2(\Omega)} \quad \forall u,v \in H^1(\Omega)$$

and thus satisfies

$$(-\Delta_N u, v)_{L^2(\Omega)} = \int_\Omega \nabla u\cdot\nabla v\,dx \quad \forall u \in H^2_N(\Omega)\,,\; v \in H^1(\Omega)\,. \tag{5.30}$$

The forward operator is supposed to act between Hilbert spaces

$$\mathbb{F} : \mathcal{U}\times\mathbb{R}^2 \to \mathcal{W}\times L^2(0,T)^{KL}$$

with the linear space

$$\begin{aligned} \mathcal{U} &= \{\mathbf{u} \in L^2(0,T;H^2_N(\Omega;\mathbb{R}^3)) \cap H^1(0,T;L^2(\Omega;\mathbb{R}^3)) \,:\, \mathbf{u}(0) = 0\} \\ &\subseteq C(0,T;H^1(\Omega)) \cap H^s(0,T;H^{2-2s}(\Omega)) \end{aligned} \tag{5.31}$$

[1]Note that as opposed to $H^1(\Omega)$ functions, $H^2(\Omega)$ functions do have a Neumann boundary trace

for $s \in [0,1]$, where the latter embedding is continuous by, e.g, [100, Lemma 7.3] or Chapter 1, Theorem 1.2.8 , applied to $\frac{\partial u_i}{\partial x_j}$, and interpolation, as well as

$$\mathcal{W} = H^1(0,T;H^1(\Omega;\mathbb{R}^3))^* \text{ or, in case } p>2,\ \mathcal{W} = H^1(0,T;L^2(\Omega;\mathbb{R}^3))^*\,. \tag{5.32}$$

We equip $\mathcal{U}$ with the inner product

$$(\mathbf{u}_1,\mathbf{u}_2)_{\mathcal{U}} := \int_0^T \int_\Omega \Big((-\Delta_N \mathbf{u}_1)\cdot(-\Delta_N \mathbf{u}_2) + \mathbf{u}_{1t}\cdot \mathbf{u}_{2t}\Big)\,dx\,dt + \int_\Omega \nabla \mathbf{u}_1(T) \colon \nabla \mathbf{u}_2(T)\,dx\,,$$

which, in spite of the nontrivial nullspace of the Neumann Laplacian $-\Delta_N$, defines a norm equivalent to the usual norm on $L^2(0,T;H^2(\Omega;\mathbb{R}^3))\cap H^1(0,T;L^2(\Omega;\mathbb{R}^3))$ due to the estimates

$$\begin{aligned}
\|\mathbf{u}\|^2_{L^2(0,T;L^2(\Omega))} &= -\int_0^T \int_\Omega \int_0^t \mathbf{u}(s)\,ds\,\mathbf{u}_t(t)\,dx\,dt + \int_\Omega \int_0^t \mathbf{u}(s)\,ds\,\mathbf{u}(T)\,dx\\
&\le \Big(T\|\mathbf{u}_t\|_{L^2(0,T;L^2(\Omega))} + \sqrt{T}\|\mathbf{u}(T)\|_{L^2(\Omega)}\Big)\|\mathbf{u}\|_{L^2(0,T;L^2(\Omega))}\\
\|\mathbf{u}(T)\|_{L^2(\Omega)} &= \left\|\int_0^T \mathbf{u}_t(t)\,dt\right\|_{L^2(\Omega)} \le \sqrt{T}\|\mathbf{u}_t\|_{L^2(0,T;L^2(\Omega))}\,.
\end{aligned}$$

This, together with the definition of the Neumann Laplacian (5.30) and the use of solutions $\mathbf{z}$, $\mathbf{v}$ to the auxiliary problems

$$\left\{\begin{array}{rcl} \mathbf{z}_t - \Delta\mathbf{z} &=& \mathbf{v} \text{ in } (0,T)\times\Omega\\ \partial_\nu \mathbf{z} &=& 0 \text{ on } (0,T)\times\partial\Omega\\ \mathbf{z}(0) &=& 0 \text{ in } \Omega \end{array}\right. ,\quad \left\{\begin{array}{rcl} -\mathbf{v}_t - \Delta\mathbf{v} &=& \mathbf{f} \text{ in } (0,T)\times\Omega\\ \partial_\nu \mathbf{v} &=& 0 \text{ on } (0,T)\times\partial\Omega\\ \mathbf{v}(T) &=& \mathbf{g} \text{ in } \Omega \end{array}\right. \tag{5.33}$$

allows us to derive the identity

$$\begin{aligned}
(\mathbf{u},\mathbf{z})_{\mathcal{U}} &= \int_0^T \int_\Omega \Big(\nabla\mathbf{u}\colon \nabla(-\Delta_N\mathbf{z}) - \mathbf{u}\cdot\mathbf{z}_{tt}\Big)dx\,dt + \int_\Omega \mathbf{u}(T)\cdot\Big(\mathbf{z}_t(T) - \Delta_N\mathbf{z}(T)\Big)\,dx\\
&= \int_0^T \int_\Omega \Big(\nabla\mathbf{u}\colon \nabla(\mathbf{v}-\mathbf{z}_t) - \mathbf{u}\cdot(\mathbf{v}_t + \Delta_N\mathbf{z}_t)\Big)dx\,dt + \int_\Omega \mathbf{u}(T)\cdot\mathbf{v}(T)\,dx\\
&= \int_0^T \int_\Omega \mathbf{u}\cdot\Big(-\Delta_N\mathbf{v} - \mathbf{v}_t\Big)dx\,dt + \int_\Omega \mathbf{u}(T)\cdot\mathbf{v}(T)\,dx\\
&= \int_0^T \int_\Omega \mathbf{u}\cdot\mathbf{f}\,dx\,dt + \int_\Omega \mathbf{u}(T)\cdot\mathbf{g}\,dx\,,
\end{aligned} \tag{5.34}$$

which will be needed later on for deriving the adjoint.

On $\mathcal{W} = H^1(0,T;H^1(\Omega;\mathbb{R}^3))^*$ we use the inner product

$$\begin{aligned}
(\mathbf{w}_1,\mathbf{w}_2)_{\mathcal{W}} := \int_0^T \int_\Omega \Big(& I_1[\nabla(-\Delta_N + \mathrm{id})^{-1}\mathbf{w}_1](t) \colon I_1[\nabla(-\Delta_N + \mathrm{id})^{-1}\mathbf{w}_2](t)\\
&+ I_1[(-\Delta_N + \mathrm{id})^{-1}\mathbf{w}_1](t)\cdot I_1[(-\Delta_N + \mathrm{id})^{-1}\mathbf{w}_2](t)\,dx\,dt
\end{aligned}$$

with the isomorphism $-\Delta_N + \mathrm{id} : H^1(\Omega) \to (H^1(\Omega))^*$ and the time integral operators

$$I_1[w](t) := \int_0^t w(s)\,ds - \frac{1}{T}\int_0^T (T-s)w(s)\,ds\,,$$
$$I_2[w](t) := -\int_0^t (t-s)w(s)\,ds + \frac{t}{T}\int_0^T (T-s)w(s)\,ds\,,$$

so that $I_2[w]_t(t) = -I_1[w](t)$, $I_1[w]_t(t) = -I_2[w]_{tt}(t) = w(t)$ and $I_2[w](0) = I_2[w](T) = 0$, hence

$$\int_0^T I_1[w_1](t)\, I_1[w_2](t)\,dt = \int_0^T I_2[w_1](t)\, w_2(t)\,dt,$$

so that in case $\mathbf{w}_2 \in L^2(0,T;L^2(\Omega;\mathbb{R}^3))$,

$$\begin{aligned}(\mathbf{w}_1,\mathbf{w}_2)_{\mathcal{W}} &= \int_0^T \int_\Omega \Big(I_2[\nabla(-\Delta_N+\mathrm{id})^{-1}\mathbf{w}_1](t) \colon [\nabla(-\Delta_N+\mathrm{id})^{-1}\mathbf{w}_2](t) \\ &\qquad + I_2[(-\Delta_N+\mathrm{id})^{-1}\mathbf{w}_1](t)\cdot[(-\Delta_N+\mathrm{id})^{-1}\mathbf{w}_2](t)\,dx\,dt \\ &= \int_0^T \int_\Omega I_2[(-\Delta_N+\mathrm{id})^{-1}\mathbf{w}_1](t)\cdot \mathbf{w}_2(t)\,dx\,dt\,. \end{aligned} \tag{5.35}$$

In case $p > 2$ in the assumption on $\mathbf{h}$, we can set $\mathcal{W} = H^1(0,T;L^2(\Omega;\mathbb{R}^3))^*$ and use the simpler inner product

$$(\mathbf{w}_1,\mathbf{w}_2)_{\mathcal{W}} := \int_0^T \int_\Omega I_1[\mathbf{w}_1](t)\cdot I_1[\mathbf{w}_2](t)\,dx\,dt\,,$$

which in case $\mathbf{w}_2 \in L^2(0,T;L^2(\Omega;\mathbb{R}^3))$ satisfies

$$(\mathbf{w}_1,\mathbf{w}_2)_{\mathcal{W}} = \int_0^T \int_\Omega I_2[\mathbf{w}_1](t)\cdot \mathbf{w}_2(t)\,dx\,dt\,.$$

Well-definedness of the forward operator

Indeed it can be verified that $\mathbb{F}$ maps between the function spaces introduced above, cf. (5.31), (5.32). For the linear (with respect to $\hat{\mathbf{m}}$) parts $\hat\alpha_1\hat{\mathbf{m}}_t$, $-\Delta_N\hat{\mathbf{m}}$, and $\int_0^T\int_\Omega \mathbf{K}_{k\ell}(t,\tau,x)\cdot\mathbf{m}_t(x,\tau)\,dx\,d\tau$ of $\mathbb{F}$, this is obvious and for the nonlinear terms $\hat\alpha_2(\mathbf{m}_0+\hat{\mathbf{m}})\times\hat{\mathbf{m}}_t$, $|\nabla(\mathbf{m}_0+\hat{\mathbf{m}})|^2(\mathbf{m}_0+\hat{\mathbf{m}})$, $((\mathbf{m}_0+\hat{\mathbf{m}})\cdot\mathbf{h})(\mathbf{m}_0+\hat{\mathbf{m}})$ we use the following estimates (5.36), (5.37), (5.38), (5.39), (5.40), (5.41), holding for any $\mathbf{u},\mathbf{w},\mathbf{z}\in\mathcal{U}$. For the term $\hat\alpha_2(\mathbf{m}_0+\hat{\mathbf{m}})\times\hat{\mathbf{m}}_t$, we estimate

$$\begin{aligned}&\|\mathbf{u}\times\mathbf{w}_t\|_{H^1(0,T;H^1(\Omega;\mathbb{R}^3))^*} \\ &\quad\le \|\mathbf{u}\times\mathbf{w}_t\|_{L^2(0,T;(H^1(\Omega;\mathbb{R}^3))^*)} \\ &\quad\le C^\Omega_{H^1\to L^3}\|\mathbf{u}\times\mathbf{w}_t\|_{L^2(0,T;L^{3/2}(\Omega;\mathbb{R}^3))} \\ &\quad\le C^\Omega_{H^1\to L^3}\|\mathbf{u}\|_{C(0,T;L^6(\Omega;\mathbb{R}^3))}\|\mathbf{w}_t\|_{L^2(0,T;L^2(\Omega;\mathbb{R}^3))} \\ &\quad\le C^\Omega_{H^1\to L^3}C^\Omega_{H^1\to L^6}\|\mathbf{u}\|_{C(0,T;H^1(\Omega;\mathbb{R}^3))}\|\mathbf{w}_t\|_{L^2(0,T;L^2(\Omega;\mathbb{R}^3))}\,,\end{aligned} \tag{5.36}$$

where we have used duality and continuity of the embeddings $H^1(0,T;H^1(\Omega;\mathbb{R}^3)) \hookrightarrow L^2(0,T;H^1(\Omega;\mathbb{R}^3)) \hookrightarrow L^2(0,T;L^3(\Omega))$ in the first and second estimate, and Hölder's inequality with exponent 4 in the third estimate;
For the term $|\nabla(\mathbf{m}_0+\hat{\mathbf{m}})|^2(\mathbf{m}_0+\hat{\mathbf{m}})$, we use

$$\begin{aligned}
&\|(\nabla\mathbf{u}\colon\nabla\mathbf{w})\mathbf{z}\|_{H^1(0,T;H^1(\Omega;\mathbb{R}^3))^*}\\
&\quad\le C^{(0,T)}_{H^1\to L^\infty}\|(\nabla\mathbf{u}\colon\nabla\mathbf{w})\mathbf{z}\|_{L^1(0,T;(H^1(\Omega;\mathbb{R}^3))^*)}\\
&\quad\le C^{(0,T)}_{H^1\to L^\infty}C^{\Omega}_{H^1\to L^6}\|(\nabla\mathbf{u}\colon\nabla\mathbf{w})\mathbf{z}\|_{L^1(0,T;L^{6/5}(\Omega;\mathbb{R}^3))}\\
&\quad\le C^{(0,T)}_{H^1\to L^\infty}C^{\Omega}_{H^1\to L^6}\\
&\qquad\quad\|\nabla\mathbf{u}\|_{L^2(0,T;L^6(\Omega;\mathbb{R}^3))}\|\nabla\mathbf{w}\|_{L^2(0,T;L^6(\Omega;\mathbb{R}^3))}\|\mathbf{z}\|_{C(0,T;L^2(\Omega;\mathbb{R}^3))}\\
&\quad\le C^{(0,T)}_{H^1\to L^\infty}C^{\Omega}_{H^1\to L^6}\\
&\qquad\quad\|\mathbf{u}\|_{L^2(0,T;H^2(\Omega;\mathbb{R}^3))}\|\mathbf{w}\|_{L^2(0,T;H^2(\Omega;\mathbb{R}^3))}\|\mathbf{z}\|_{C(0,T;H^1(\Omega;\mathbb{R}^3))}\,,
\end{aligned}\tag{5.37}$$

again using duality and the embeddings $H^1(0,T;H^1(\Omega;\mathbb{R}^3)) \hookrightarrow L^\infty(0,T;H^1(\Omega)) \hookrightarrow L^\infty(0,T;L^6(\Omega))$;
For the term $((\mathbf{m}_0+\hat{\mathbf{m}})\cdot\mathbf{h})(\mathbf{m}_0+\hat{\mathbf{m}})$, we estimate

$$\begin{aligned}
&\|(\mathbf{u}\cdot\mathbf{h})\mathbf{z}\|_{H^1(0,T;H^1(\Omega;\mathbb{R}^3))^*}\\
&\quad\le C^{\Omega}_{H^1\to L^6}\|(\mathbf{u}\cdot\mathbf{h})\mathbf{z}\|_{L^2(0,T;L^{6/5}(\Omega;\mathbb{R}^3))}\\
&\quad\le C^{\Omega}_{H^1\to L^6}\|\mathbf{u}\|_{C(0,T;L^6(\Omega;\mathbb{R}^3))}\|\mathbf{z}\|_{C(0,T;L^6(\Omega;\mathbb{R}^3))}\|\mathbf{h}\|_{L^2(0,T;L^2(\Omega;\mathbb{R}^3))}\\
&\quad\le (C^{\Omega}_{H^1\to L^6};\mathbb{R}^3)\|\mathbf{u}\|_{C(0,T;H^1(\Omega;\mathbb{R}^3))}\|\mathbf{z}\|_{C(0,T;H^1(\Omega;\mathbb{R}^3))}\|\mathbf{h}\|_{L^2(0,T;L^2(\Omega;\mathbb{R}^3))}
\end{aligned}\tag{5.38}$$

by duality and the embedding $H^1(0,T;H^1(\Omega;\mathbb{R}^3)) \hookrightarrow L^2(0,T;L^6(\Omega))$, as well as Hölder's inequality.

In case $p>2$, $\mathbb{F}$ maps into the somewhat stronger space $\mathcal{W}=H^1(0,T;L^2(\Omega;\mathbb{R}^3))^*$, due to the estimates

$$\begin{aligned}
&\|\mathbf{u}\times\mathbf{w}_t\|_{H^1(0,T;L^2(\Omega;\mathbb{R}^3))^*}\\
&\quad\le C^{(0,T)}_{H^1\to L^\infty}\|\mathbf{u}\times\mathbf{w}_t\|_{L^1(0,T;L^2(\Omega;\mathbb{R}^3))}\\
&\quad\le C^{(0,T)}_{H^1\to L^\infty}\|\mathbf{u}\|_{L^2(0,T;L^\infty(\Omega;\mathbb{R}^3))}\|\mathbf{w}_t\|_{L^2(0,T;L^2(\Omega;\mathbb{R}^3))}\\
&\quad\le C^{(0,T)}_{H^1\to L^\infty}C^{\Omega}_{H^2\to L^\infty}\|\mathbf{u}\|_{L^2(0,T;H^2(\Omega;\mathbb{R}^3))}\|\mathbf{w}_t\|_{L^2(0,T;L^2(\Omega;\mathbb{R}^3))}\,,
\end{aligned}\tag{5.39}$$

as well as

$$\begin{aligned}
&\|(\nabla\mathbf{u}\colon\nabla\mathbf{w})\mathbf{z}\|_{H^1(0,T;L^2(\Omega;\mathbb{R}^3))^*}\\
&\quad\le C^{(0,T)}_{H^1\to L^\infty}\|(\nabla\mathbf{u}\colon\nabla\mathbf{w})\mathbf{z}\|_{L^1(0,T;L^2(\Omega;\mathbb{R}^3))}\\
&\quad\le C^{(0,T)}_{H^1\to L^\infty}\|\nabla\mathbf{u}\|_{L^2(0,T;L^6(\Omega;\mathbb{R}^3))}\|\nabla\mathbf{w}\|_{L^2(0,T;L^6(\Omega;\mathbb{R}^3))}\|\mathbf{z}\|_{C(0,T;L^6(\Omega;\mathbb{R}^3))}\\
&\quad\le C^{(0,T)}_{H^1\to L^\infty}(C^{\Omega}_{H^1\to L^6};\mathbb{R}^3)\|\mathbf{u}\|_{L^2(0,T;H^2(\Omega;\mathbb{R}^3))}\\
&\qquad\quad\|\mathbf{w}\|_{L^2(0,T;H^2(\Omega;\mathbb{R}^3))}\|\mathbf{z}\|_{C(0,T;H^1(\Omega;\mathbb{R}^3))}\,,
\end{aligned}\tag{5.40}$$

and

$$
\begin{aligned}
&\|(\mathbf{u}\cdot\mathbf{h})\mathbf{z}\|_{H^1(0,T;L^2(\Omega;\mathbb{R}^3))^*}\\
&\quad\le C^{(0,T)}_{H^1\to L^\infty}\|(\mathbf{u}\cdot\mathbf{h})\mathbf{z}\|_{L^1(0,T;L^2(\Omega;\mathbb{R}^3))}\\
&\quad\le C^{(0,T)}_{H^1\to L^\infty}\|\mathbf{u}\|_{L^4(0,T;L^{p^{**}}(\Omega;\mathbb{R}^3))}\|\mathbf{z}\|_{L^4(0,T;L^{p^{**}}(\Omega;\mathbb{R}^3))}\|\mathbf{h}\|_{L^2(0,T;L^p(\Omega;\mathbb{R}^3))}\\
&\quad\le C^{(0,T)}_{H^1\to L^\infty}(C^{(0,T)}_{H^{1/4},L^4})^2(C^{\Omega}_{H^{3/2},L^{p^{**}}})^2\\
&\qquad\qquad\|\mathbf{u}\|_{H^{1/4}(0,T;H^{3/2}(\Omega;\mathbb{R}^3))}\|\mathbf{z}\|_{H^{1/4}(0,T;H^{3/2}(\Omega;\mathbb{R}^3))}\|\mathbf{h}\|_{L^2(0,T;L^p(\Omega;\mathbb{R}^3))}\,,
\end{aligned}
\tag{5.41}
$$

for $p^{**}=\frac{2p}{p-2}<\infty$, which can be bounded by the $\mathcal{U}$ norm of $\mathbf{u}$ and $\mathbf{z}$, using interpolation with $s=\frac{1}{4}$ in (5.31).

Differentiability of the forward operator

Formally, the derivative of $\mathbb{F}$ is given by

$$
\begin{aligned}
&\mathbb{F}'(\hat{\mathbf{m}},\hat\alpha_1,\hat\alpha_2)(\mathbf{u},\beta_1,\beta_2)\\
&=\begin{pmatrix}
\begin{aligned}
&\beta_1\hat{\mathbf{m}}_t-\beta_2(\mathbf{m}_0+\hat{\mathbf{m}})\times\hat{\mathbf{m}}_t\\
&\qquad+\hat\alpha_1\mathbf{u}_t-\Delta_N\mathbf{u}-\hat\alpha_2\mathbf{u}\times\hat{\mathbf{m}}_t-\hat\alpha_2(\mathbf{m}_0+\hat{\mathbf{m}})\times\mathbf{u}_t\\
&\qquad-2(\nabla(\mathbf{m}_0+\hat{\mathbf{m}})\colon\nabla\mathbf{u})(\mathbf{m}_0+\hat{\mathbf{m}})-|\nabla(\mathbf{m}_0+\hat{\mathbf{m}})|^2\mathbf{u}\\
&\qquad+((\mathbf{m}_0+\hat{\mathbf{m}})\cdot\mathbf{h})\mathbf{u}+(\mathbf{u}\cdot\mathbf{h})(\mathbf{m}_0+\hat{\mathbf{m}})
\end{aligned}\\[2ex]
\left(\int_0^T\int_\Omega\mathbf{K}_{k\ell}(t,\tau,x)\cdot\mathbf{u}_t(x,\tau)\,dx\,d\tau\right)_{k=1,\dots,K,\ \ell=1,\dots,L}
\end{pmatrix}\\
&=\begin{pmatrix}
\frac{\partial\mathbb{F}_0}{\partial\hat{\mathbf{m}}}(\hat{\mathbf{m}},\hat\alpha) & \frac{\partial\mathbb{F}_0}{\partial\hat\alpha_1}(\hat{\mathbf{m}},\hat\alpha) & \frac{\partial\mathbb{F}_0}{\partial\hat\alpha_2}(\hat{\mathbf{m}},\hat\alpha)\\
(\frac{\partial\mathbb{F}_{k\ell}}{\partial\hat{\mathbf{m}}}(\hat{\mathbf{m}},\hat\alpha))_{k=1,\dots,K,\ell=1,\dots,L} & 0 & 0
\end{pmatrix}
\begin{pmatrix}\mathbf{u}\\ \beta_1\\ \beta_2\end{pmatrix}
\end{aligned}
$$

where $\frac{\partial\mathbb{F}_0}{\partial\hat{\mathbf{m}}}(\hat{\mathbf{m}},\hat\alpha):\mathcal{U}\to\mathcal{W}$, $\frac{\partial\mathbb{F}_0}{\partial\hat\alpha_1}(\hat{\mathbf{m}},\hat\alpha):\mathbb{R}\to\mathcal{W}$, $\frac{\partial\mathbb{F}_0}{\partial\hat\alpha_2}(\hat{\mathbf{m}},\hat\alpha):\mathbb{R}\to\mathcal{W}$, $(\frac{\partial\mathbb{F}_{k\ell}}{\partial\hat{\mathbf{m}}}(\hat{\mathbf{m}},\hat\alpha))_{k=1,\dots,K,\ell=1,\dots,L}:\mathcal{U}\to L^2(0,T)^{KL}$. Fréchet differentiability follows from the fact that in

$$
\mathbb{F}(\hat{\mathbf{m}}+\mathbf{u},\hat\alpha_1+\beta_1,\hat\alpha_2+\beta_2)-\mathbb{F}(\hat{\mathbf{m}},\hat\alpha_1,\hat\alpha_2)-\mathbb{F}'(\hat{\mathbf{m}},\hat\alpha_1,\hat\alpha_2)(\mathbf{u},\beta_1,\beta_2)
$$

all linear terms cancel out and the nonlinear ones are given by (abbreviating $\mathbf{m}=\mathbf{m}_0+\hat{\mathbf{m}}$)

$$
\begin{aligned}
&(\hat\alpha_1+\beta_1)(\mathbf{m}_t+\mathbf{u}_t)-\hat\alpha_1\mathbf{m}_t-\hat\alpha_1\mathbf{u}_t-\beta_1\mathbf{m}_t=\beta_1\mathbf{u}_t\\
&(\hat\alpha_2+\beta_2)(\mathbf{m}+\mathbf{u})\times(\mathbf{m}_t+\mathbf{u}_t)-\hat\alpha_2\mathbf{m}\times\mathbf{m}_t-\beta_2\mathbf{m}\times\mathbf{m}_t-\hat\alpha_2\mathbf{u}\times\mathbf{m}_t-\hat\alpha_2\mathbf{m}\times\mathbf{u}_t\\
&\qquad=\hat\alpha_2\mathbf{u}\times\mathbf{u}_t+\beta_2\mathbf{m}\times\mathbf{u}_t+\beta_2\mathbf{u}\times\mathbf{m}_t+\beta_2\mathbf{u}\times\mathbf{u}_t\\
&|\nabla\mathbf{m}+\nabla\mathbf{u}|^2(\mathbf{m}+\mathbf{u})-|\nabla\mathbf{m}|^2\mathbf{m}-2(\nabla\mathbf{m}\colon\nabla\mathbf{u})\mathbf{m}-|\nabla\mathbf{m}|^2\mathbf{u}\\
&\qquad=|\nabla\mathbf{u}|^2(\mathbf{m}+\mathbf{u})+2(\nabla\mathbf{m}\colon\nabla\mathbf{u})\mathbf{u}\\
&((\mathbf{m}+\mathbf{u})\cdot\mathbf{h})(\mathbf{m}+\mathbf{u})-(\mathbf{m}\cdot\mathbf{h})\mathbf{m}-(\mathbf{u}\cdot\mathbf{h})\mathbf{m}-(\mathbf{m}\cdot\mathbf{h})\mathbf{u}=(\mathbf{u}\cdot\mathbf{h})\mathbf{u}\,,
\end{aligned}
$$

hence, using again (5.36)–(5.38) can be estimated by some constant times $\|\mathbf{u}\|^2_{\mathcal{U}}+\beta_1^2+\beta_2^2$.

Adjoints

Starting with the adjoint of $\frac{\partial \mathbb{F}_0}{\partial \hat{\mathbf{m}}}(\hat{\mathbf{m}}, \hat{\alpha})$, for any $\mathbf{u} \in \mathcal{U}$, $\mathbf{y} \in L^2(0,T;L^2(\Omega))$, we have, using the definition of $-\Delta_N$, i.e., (5.30),

$$
\begin{aligned}
&\int_0^T \int_\Omega \big(\frac{\partial \mathbb{F}_0}{\partial \hat{\mathbf{m}}}(\hat{\mathbf{m}}, \hat{\alpha})\mathbf{u}\big) \cdot \mathbf{y}\, dx\, dt \\
&= \int_0^T \int_\Omega \Big(\hat{\alpha}_1 \mathbf{u}_t \cdot \mathbf{y} + \nabla \mathbf{u} \colon \nabla \mathbf{y} - \hat{\alpha}_2 (\mathbf{u} \times \hat{\mathbf{m}}_t) \cdot \mathbf{y} - \hat{\alpha}_2 ((\mathbf{m}_0 + \hat{\mathbf{m}}) \times \mathbf{u}_t) \cdot \mathbf{y} \\
&\quad - 2(\nabla(\mathbf{m}_0 + \hat{\mathbf{m}}) \colon \nabla \mathbf{u})\, ((\mathbf{m}_0 + \hat{\mathbf{m}}) \cdot \mathbf{y}) - |\nabla(\mathbf{m}_0 + \hat{\mathbf{m}})|^2 \,(\mathbf{u} \cdot \mathbf{y}) \\
&\quad + ((\mathbf{m}_0 + \hat{\mathbf{m}}) \cdot \mathbf{h})\, (\mathbf{u} \cdot \mathbf{y}) + (\mathbf{u} \cdot \mathbf{h})\, ((\mathbf{m}_0 + \hat{\mathbf{m}}) \cdot \mathbf{y}) \Big)\, dx\, dt \\
&= \int_0^T \int_\Omega \mathbf{u} \cdot \Big(-\hat{\alpha}_1 \mathbf{y}_t + (-\Delta \mathbf{y}) - \hat{\alpha}_2 \hat{\mathbf{m}}_t \times \mathbf{y} + \hat{\alpha}_2 \mathbf{y}_t \times (\mathbf{m}_0 + \hat{\mathbf{m}}) + \hat{\alpha}_2 \mathbf{y} \times \hat{\mathbf{m}}_t \\
&\qquad\qquad - 2((\mathbf{m}_0 + \hat{\mathbf{m}}) \cdot \mathbf{y})\, (-\Delta_N(\mathbf{m}_0 + \hat{\mathbf{m}})) + 2((\nabla(\mathbf{m}_0 + \hat{\mathbf{m}})^T (\nabla \mathbf{y}))\, (\mathbf{m}_0 + \hat{\mathbf{m}}) \\
&\qquad\qquad + 2((\nabla(\mathbf{m}_0 + \hat{\mathbf{m}})^T (\nabla(\mathbf{m}_0 + \hat{\mathbf{m}})))\, \mathbf{y} - |\nabla(\mathbf{m}_0 + \hat{\mathbf{m}})|^2 \mathbf{y} \\
&\qquad\qquad + ((\mathbf{m}_0 + \hat{\mathbf{m}}) \cdot \mathbf{h})\, \mathbf{y} + ((\mathbf{m}_0 + \hat{\mathbf{m}}) \cdot \mathbf{y})\, \mathbf{h} \Big)\, dx\, dt \\
&\quad + \int_\Omega \mathbf{u}(T) \cdot \Big(\hat{\alpha}_1 \mathbf{y}(T) - \hat{\alpha}_2 \mathbf{y}(T) \times (\mathbf{m}_0 + \hat{\mathbf{m}}(T)) \Big)\, dx \\
&=: \int_0^T \int_\Omega \mathbf{u} \cdot \mathbf{f}^{\mathbf{y}}\, dx\, dt + \int_\Omega \mathbf{u}(T) \cdot \mathbf{g}_T^{\mathbf{y}}\, dx\,,
\end{aligned}
\tag{5.42}
$$

where we have integrated by parts with respect to time and used the vector identities

$$
\vec{a} \cdot (\vec{b} \times \vec{c}) = \vec{b} \cdot (\vec{c} \times \vec{a}) = \vec{c} \cdot (\vec{a} \times \vec{b})\,.
$$

Matching the integrals over $\Omega \times (0,T)$ and $\Omega \times \{T\}$, respectively, and taking into account the homogeneous Neumann boundary conditions implied by the definition of $-\Delta_N$, (5.30), as well as the identities (5.34), (5.35), we find that $\frac{\partial \mathbb{F}_0}{\partial \hat{\mathbf{m}}}(\hat{\mathbf{m}}, \hat{\alpha})^* \mathbf{y} =: \mathbf{z}$ is the solution of (5.33) with $\mathbf{f} = \mathbf{f}^{\mathbf{y}}$, $\mathbf{g} = \mathbf{g}_T^{\mathbf{y}}$, where in case $\mathcal{W} = H^1(0,T;H^1(\Omega;\mathbb{R}^3))^*$, $\mathbf{y} = I_2[\widetilde{y}]$, with $\widetilde{y}(t)$ solving

$$
\begin{cases}
-\Delta \widetilde{y}(t) + \widetilde{y}(t) &= \mathbf{w}(t) \text{ in } \Omega \\
\partial_\nu \widetilde{y} &= 0 \text{ on } \partial\Omega
\end{cases}
$$

for each $t \in (0,T)$, or in case $\mathcal{W} = H^1(0,T;L^2(\Omega;\mathbb{R}^3))^*$, just $\mathbf{y} = I_2[\mathbf{w}]$.

With the same $\mathbf{y}$, after pointwise projection onto the mutually orthogonal vectors $\hat{\mathbf{m}}_t(x,t)$ and $(\mathbf{m}_0(x) + \hat{\mathbf{m}}(x,t)) \times \hat{\mathbf{m}}_t(x,t)$ and integration over space and time, we also get the adjoints of $\frac{\partial \mathbb{F}_0}{\partial \hat{\alpha}_1}(\hat{\mathbf{m}}, \hat{\alpha})$, $\frac{\partial \mathbb{F}_0}{\partial \hat{\alpha}_2}(\hat{\mathbf{m}}, \hat{\alpha})$

$$
\frac{\partial \mathbb{F}_0}{\partial \hat{\alpha}_1}(\hat{\mathbf{m}}, \hat{\alpha})^* \mathbf{w} = \int_0^T \int_\Omega \hat{\mathbf{m}}_t \cdot \mathbf{y}\, dx\, dt\,, \quad \frac{\partial \mathbb{F}_0}{\partial \hat{\alpha}_2}(\hat{\mathbf{m}}, \hat{\alpha})^* \mathbf{w} = -\int_0^T \int_\Omega ((\mathbf{m}_0 + \hat{\mathbf{m}}) \times \hat{\mathbf{m}}_t) \cdot \mathbf{y}\, dx\, dt\,.
$$

Finally, the fact that for $\mathbf{u} \in \mathcal{U}$, $y \in L^2(0,T)^{KL}$

$$
\begin{aligned}
&((\frac{\partial \mathbb{F}_{k\ell}}{\partial \hat{\mathbf{m}}}(\hat{\mathbf{m}},\hat{\alpha}))_{k=1,...,K,\ell=1,...,L}\mathbf{u}, y)_{L^2(0,T)^{KL}} \\
&\qquad = \sum_{k=1}^{K}\sum_{\ell=1}^{L}\int_0^T ((\frac{\partial \mathbb{F}_{k\ell}}{\partial \hat{\mathbf{m}}}(\hat{\mathbf{m}},\hat{\alpha}))_{k=1,...,K,\ell=1,...,L}\mathbf{u})_{k\ell}(t) y_{k\ell}(t)\,dt \\
&\qquad = \sum_{k=1}^{K}\sum_{\ell=1}^{L}\int_0^T \int_0^T \int_\Omega \mathbf{K}_{k\ell}(t,\tau,x)\cdot \mathbf{u}_t(x,\tau)\,dx\,d\tau y_{k\ell}(t)\,dt \\
&\qquad = \sum_{k=1}^{K}\sum_{\ell=1}^{L}\int_0^T \Big(-\int_0^T \int_\Omega \frac{\partial}{\partial \tau}\mathbf{K}_{k\ell}(t,\tau,x)\cdot \mathbf{u}(x,\tau)\,dx\,d\tau \\
&\qquad\qquad + \int_\Omega \mathbf{K}_{k\ell}(t,T,x)\cdot \mathbf{u}(x,T)\,dx\Big) y_{k\ell}(t)\,dt\,,
\end{aligned}
\tag{5.43}
$$

where we have integrated by parts with respect to time, implies that due to (5.34), $(\frac{\partial \mathbb{F}_{k\ell}}{\partial \hat{\mathbf{m}}}(\hat{\mathbf{m}},\hat{\alpha}))^*_{k=1,...,K,\ell=1,...,L}\mathbf{w} = \mathbf{z}$ is obtained by solving another auxiliary problem (5.33) with

$$
\mathbf{f}(x,\tau) = -\int_0^T \sum_{k=1}^{K}\sum_{\ell=1}^{L} \frac{\partial}{\partial \tau}\mathbf{K}_{k\ell}(t,\tau,x) y_{k\ell}(t)\,dt,\ \ \mathbf{g}(x) = \int_0^T \sum_{k=1}^{K}\sum_{\ell=1}^{L} \mathbf{K}_{k\ell}(t,T,x) y_{k\ell}(t)\,dt.
\tag{5.44}
$$

Remark 5.4.1. In case of a Landweber-Kaczmarz method iterating cyclically over the equations defined by $\mathbb{F}_0, \mathbb{F}_{k\ell}$, $k = 1,...,K$, $\ell = 1,...,L$, adjoints of derivatives of $\mathbb{F}_0$ remain unchanged while adjoints of $\frac{\partial \mathbb{F}_{k\ell}}{\partial \hat{\mathbf{m}}}(\hat{\mathbf{m}},\hat{\alpha}))_{k=1,...,K,\ell=1,...,L}$ are defined as in (5.43), (5.44) by just skipping the sums over k and ℓ there.

5.4.2 Reduced formulation

We rely on the formulation (5.24) with F defined by (5.25), (5.26), (5.27). Due to the estimate

$$
\|\mathcal{K}_{k\ell}\mathbf{m}_t\|^2_{L^2(0,T)} \le T\|\widetilde{a}_\ell\|^2_{L^2(0,T)}\|c_k\mathbf{p}^R_\ell\|^2_{L^2(\Omega,\mathbb{R}^3)}\|\mathbf{m}\|^2_{H^1(0,T;L^2(\Omega,\mathbb{R}^3))},
$$

if $\widetilde{a}_\ell \in L^2(0,T), c_k\mathbf{p}^R_\ell \in L^2(\Omega,\mathbb{R}^3)$ we can choose the state space in the reduced setting as

$$
\tilde{\mathcal{U}} = H^1(0,T;L^2(\Omega,\mathbb{R}^3)),
\tag{5.45}
$$

which is different from the one in the all-at-once setting.

Adjoint equation

From (5.25) the derivative of the forward operation takes the form

$$F'(\hat{\alpha})\beta = \mathcal{K}\mathbf{u}_t, \tag{5.46}$$

where $\mathbf{u}$ solves the linearized LLG equation

$$\begin{aligned}
&\hat{\alpha}_1\mathbf{u}_t - \hat{\alpha}_2\mathbf{m}\times\mathbf{u}_t - \hat{\alpha}_2\mathbf{u}\times\mathbf{m}_t - \Delta\mathbf{u} - 2(\nabla\mathbf{u} : \nabla\mathbf{m})\mathbf{m} && \\
&\qquad + \mathbf{u}(-|\nabla\mathbf{m}|^2 + (\mathbf{m}\cdot\mathbf{h})) + (\mathbf{u}\cdot\mathbf{h})\mathbf{m} = -\beta_1\mathbf{m}_t + \beta_2\mathbf{m}\times\mathbf{m}_t && \text{in } (0,T)\times\Omega \\
&\partial_\nu\mathbf{u} = 0 && \text{on } (0,T)\times\partial\Omega \\
&\mathbf{u}(0) = 0 && \text{in } \Omega,
\end{aligned}$$

and $\mathbf{m}$ is the solution to (5.21)-(5.23). This equation can be obtained by formally taking directional derivatives (in the direction of $\mathbf{u}$) in all terms of the LLG equation (5.21)–(5.23), or alternatively by subtracting the defining boundary value problems for $S(\mathbf{m}+\epsilon\mathbf{u})$ and $S(\mathbf{m})$, dividing by ϵ and then letting ϵ tend to zero.

The Hilbert space adjoint of $F'(\hat{\alpha})$

$$F'(\hat{\alpha})^* : L^2(0,T)^{KL} \to \mathbb{R}^2$$

satisfies, for each $z \in L^2(0,T)^{KL}$,

$$\begin{aligned}
&(F'(\hat{\alpha})^* z, \beta)_{\mathbb{R}^2} \\
&= (z, F'(\hat{\alpha})\beta)_{L^2(0,T)^{KL}} \\
&= \sum_{k=1}^{K}\sum_{\ell=1}^{L}\int_0^T z_{k\ell}(t)\int_0^T\int_\Omega(-\mu_0)\widetilde{a}_\ell(t-\tau)c_k(x)\mathbf{p}_\ell^R(x)\cdot\mathbf{u}_\tau(\tau,x)dx\,d\tau\,dt \\
&= \sum_{k=1}^{K}\sum_{\ell=1}^{L}\int_0^T z_{k\ell}(t)\Big(-\int_0^T\int_\Omega(-\mu_0)\widetilde{a}_{\ell\,\tau}(t-\tau)c_k(x)\mathbf{p}_\ell^R(x)\cdot\mathbf{u}(\tau,x)\,dx\,d\tau \\
&\qquad\qquad + \int_\Omega(-\mu_0)\widetilde{a}_\ell(t-T)c_k(x)\mathbf{p}_\ell^R(x)\cdot\mathbf{u}(T,x)\,dx\Big)dt \\
&= \int_0^T\int_\Omega\mathbf{u}(\tau,x)\cdot\sum_{k=1}^{K}\sum_{\ell=1}^{L}\int_0^T(-\mu_0)\widetilde{a}_{\ell\,t}(t-\tau)z_{k\ell}(t)\,c_k(x)\mathbf{p}_\ell^R(x)\,dt\,dx\,d\tau \\
&\qquad\qquad + \int_\Omega\mathbf{u}(T,x)\cdot\sum_{k=1}^{K}\sum_{\ell=1}^{L}\int_0^T(-\mu_0)\widetilde{a}_\ell(t)z_{k\ell}(t)\,c_k(x)\mathbf{p}_\ell^R(x)\,dt\,dx \\
&=: (\mathbf{u}, \tilde{K}z)_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))} + (\mathbf{u}(T), \tilde{K}_T z)_{L^2(\Omega,\mathbb{R}^3)}
\end{aligned} \tag{5.47}$$

as the transfer function $\tilde{a}$ is periodic with period T, and the continuous embedding $H(0,T) \hookrightarrow C[0,T]$ allows us to evaluate $\mathbf{u}(t=T)$.

Observing

$$\int_0^T \int_\Omega -\hat{\alpha}_1 \mathbf{q}_t^z \cdot \mathbf{u} \, dx\, dt = \int_0^T \int_\Omega \hat{\alpha}_1 \mathbf{u}_t \cdot \mathbf{q}^z \, dx - \int_\Omega \hat{\alpha}_1 \mathbf{q}^z(T) \cdot \mathbf{u}(T)\, dx \,,$$

$$\begin{aligned}\int_0^T \int_\Omega & -\hat{\alpha}_2 (\mathbf{m} \times \mathbf{q}^z)_t \cdot \mathbf{u} \, dx\, dt \\ &= \int_0^T \int_\Omega -\hat{\alpha}_2 (\mathbf{m} \times \mathbf{u}_t) \cdot \mathbf{q}^z \, dx\, dt - \int_\Omega \hat{\alpha}_2 (\mathbf{m} \times \mathbf{q}^z)(T) \cdot \mathbf{u}(T)\, dx \,,\end{aligned}$$

$$\int_0^T \int_\Omega \hat{\alpha}_2 (\mathbf{q}^z \times \mathbf{m}_t) \cdot \mathbf{u} \, dx\, dt = \int_0^T \int_\Omega -\hat{\alpha}_2 (\mathbf{u} \times \mathbf{m}_t) \cdot \mathbf{q}^z \, dx\, dt \,,$$

$$\int_0^T \int_\Omega -\Delta \mathbf{q}^z \cdot \mathbf{u} \, dx\, dt = \int_0^T \int_\Omega -\mathbf{q}^z \cdot \Delta \mathbf{u} \, dx\, dt - \int_0^T \int_{\partial\Omega} \partial_\nu \mathbf{q}^z \cdot \mathbf{u} \, dx\, dt \,,$$

$$\int_0^T \int_\Omega \mathbf{q}^z (-|\nabla \mathbf{m}|^2 + (\mathbf{m} \cdot \mathbf{h})) \cdot \mathbf{u} \, dx\, dt = \int_0^T \int_\Omega \big(\mathbf{u}(-|\nabla \mathbf{m}|^2 + (\mathbf{m} \cdot \mathbf{h}))\big) \cdot \mathbf{q}^z \, dx\, dt \,,$$

$$\int_0^T \int_\Omega (\mathbf{q}^z \cdot \mathbf{m})\, \mathbf{h} \cdot \mathbf{u} \, dx\, dt = \int_0^T \int_\Omega (\mathbf{u} \cdot \mathbf{h}) \mathbf{m} \cdot \mathbf{q}^z \, dx\, dt \,,$$

$$\begin{aligned}\int_0^T \int_\Omega & 2 (\mathbf{m} \cdot \mathbf{q}^z) \Delta \mathbf{m} \cdot \mathbf{u} \, dx\, dt \\ &= - \int_0^T \int_\Omega 2 (\nabla \mathbf{m} : \nabla \mathbf{u})(\mathbf{m} \cdot \mathbf{q}^z)\, dx\, dt \\ &\qquad + 2 \int_0^T \int_\Omega -\mathbf{u} \cdot ((\nabla \mathbf{m})^\top \nabla \mathbf{m}) \mathbf{q}^z - \mathbf{u} \cdot ((\nabla \mathbf{m})^\top \nabla \mathbf{q}^z) \mathbf{m} \, dx\, dt \,,\end{aligned}$$

we see that, if $\mathbf{q}^z$ solves the adjoint equation

$$\begin{aligned} & -\hat{\alpha}_1 \mathbf{q}_t^z - \hat{\alpha}_2 \mathbf{m} \times \mathbf{q}_t^z - 2\hat{\alpha}_2 \mathbf{m}_t \times \mathbf{q}^z - \Delta \mathbf{q}^z \\ & \quad + 2 \left((\nabla \mathbf{m})^\top \nabla \mathbf{m} \right) \mathbf{q}^z + 2 \left((\nabla \mathbf{m})^\top \nabla \mathbf{q}^z \right) \mathbf{m} \\ & \quad + (-|\nabla \mathbf{m}|^2 + (\mathbf{m} \cdot \mathbf{h})) \mathbf{q}^z + (\mathbf{m} \cdot \mathbf{q}^z)(\mathbf{h} + 2\Delta \mathbf{m}) = \tilde{K} z \qquad \text{in } (0,T) \times \Omega \qquad (5.48) \\ & \partial_\nu \mathbf{q}^z = 0 \qquad \text{on } (0,T) \times \partial\Omega \qquad (5.49) \\ & \hat{\alpha}_1 \mathbf{q}^z(T) + \hat{\alpha}_2 (\mathbf{m} \times \mathbf{q}^z)(T) = \tilde{K}_T z \qquad \text{in } \Omega \qquad (5.50) \end{aligned}$$

then with (5.47), we have

$$\begin{aligned}(F'(\hat{\alpha})^* z, \beta)_{\mathbb{R}^2} &= (\mathbf{u}, \tilde{K} z)_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))} + (\mathbf{u}(T), \tilde{K}_T z)_{L(\Omega,\mathbb{R}^3)} \\ &= \int_0^T \int_\Omega (-\beta_1 \mathbf{m}_t + \beta_2 \mathbf{m} \times \mathbf{m}_t) \cdot \mathbf{q}^z \, dx\, dt \\ &= (\beta_1, \beta_2) \cdot \left(\int_0^T \int_\Omega -\mathbf{m}_t \cdot \mathbf{q}^z \, dx\, dt, \int_0^T \int_\Omega (\mathbf{m} \times \mathbf{m}_t) \cdot \mathbf{q}^z \, dx\, dt \right),\end{aligned}$$

which implies the Hilbert space adjoint $F'(\hat{\alpha})^* : \mathcal{Y} \to \mathbb{R}^2$

$$F'(\hat{\alpha})^* z = \left(\int_0^T \int_\Omega -\mathbf{m}_t \cdot \mathbf{q}^z \, dx \, dt, \int_0^T \int_\Omega (\mathbf{m} \times \mathbf{m}_t) \cdot \mathbf{q}^z \, dx \, dt \right), \tag{5.51}$$

provided that the adjoint state $\mathbf{q}^z$ exists and belongs to sufficiently smooth space (see Subsection 5.4.2 below).

The end condition (5.50) is equivalent to

$$\begin{pmatrix} \hat{\alpha}_1 & -\hat{\alpha}_2 \mathbf{m}_3(T) & \hat{\alpha}_2 \mathbf{m}_2(T) \\ \hat{\alpha}_2 \mathbf{m}_3(T) & \hat{\alpha}_1 & -\hat{\alpha}_2 \mathbf{m}_1(T) \\ -\hat{\alpha}_2 \mathbf{m}_2(T) & \hat{\alpha}_2 \mathbf{m}_1(T) & \hat{\alpha}_1 \end{pmatrix} \mathbf{q}^z(T) =: M_T^{\hat{\alpha}} \mathbf{q}^z(T) = \tilde{K}_T z,$$

where $\mathbf{m}_i(T), i = 1, 2, 3,$ denotes the i-th component of $\mathbf{m}(T)$. The matrix $M_T^{\hat{\alpha}}$ with $\det(M_T^{\hat{\alpha}}) = |\hat{\alpha}_1(\hat{\alpha}_1^2 + \hat{\alpha}_2^2)|$ is invertible if $\hat{\alpha}_1 > 0$, which fits into the condition for existence of the solution to the LLG equation. Hence, we are able to rewrite the adjoint equation in the following form

$$\begin{aligned} & -\hat{\alpha}_1 \mathbf{q}_t^z - \hat{\alpha}_2 \mathbf{m} \times \mathbf{q}_t^z - 2\hat{\alpha}_2 \mathbf{m}_t \times \mathbf{q}^z - \Delta \mathbf{q}^z \\ & \quad + 2\left((\nabla \mathbf{m})^\top \nabla \mathbf{m} \right) \mathbf{q}^z + 2\left((\nabla \mathbf{m})^\top \nabla \mathbf{q}^z \right) \mathbf{m} \\ & \quad + (-|\nabla \mathbf{m}|^2 + (\mathbf{m} \cdot \mathbf{h}))\mathbf{q}^z + (\mathbf{m} \cdot \mathbf{q}^z)(\mathbf{h} + 2\Delta \mathbf{m}) = \tilde{K} z \qquad \text{in } (0,T) \times \Omega && (5.52) \\ & \partial_\nu \mathbf{q}^z = 0 \qquad \text{on } (0,T) \times \partial\Omega && (5.53) \\ & \mathbf{q}^z(T) = (M_T^{\hat{\alpha}})^{-1} \tilde{K}_T z \qquad \text{in } \Omega. && (5.54) \end{aligned}$$

Remark 5.4.2. Formula (5.51) inspires a Kaczmarz scheme relying on restricting the observation operator to time subintervals for every fixed k, ℓ, namely, we segment $(0, T)$ into several subintervals (t^j, t^{j+1}) with the break points $0 = t^0 < \ldots < t^{n-1} = T$

$$F_{k\ell}^j : \mathcal{D}(F)(\subseteq \mathcal{X}) \to \mathcal{Y}^j, \qquad \hat{\alpha} \mapsto y^j := \mathcal{K}_{k\ell} \frac{\partial}{\partial t} S(\hat{\alpha})|_{(t^j, t^{j+1})} \tag{5.55}$$

with

$$\mathcal{Y}^j = L^2(t^j, t^{j+1})^{KL} \qquad j = 0 \ldots n-1, \tag{5.56}$$

hence

$$y_{k\ell}^j(t) = \int_{t^j}^{t^{j+1}} \int_\Omega -\mu_0 \widetilde{a}_\ell(t-\tau) c_k(x) \mathbf{p}_\ell^R(x) \cdot \mathbf{m}_\tau(x,\tau) dx d\tau. \tag{5.57}$$

Here we distinguish superscript j for time subinterval index and subscripts k, ℓ for the index of different receive coils and concentrations.

For $z^j \in \mathcal{Y}^j$,

$$
\begin{aligned}
(\tilde{K}z^j)(x,t) &= \sum_{k=1}^{K}\sum_{\ell=1}^{L} -\mu_0 c_k(x)\mathbf{p}_\ell^R(x)\int_{t^j}^{t^{j+1}} \widetilde{a}_{\ell\,\tau}(\tau-t) z_{k\ell}^j(\tau)\,d\tau \qquad t\in(0,T)\,,\\
(\tilde{K}_T z^j)(x) &= \sum_{k=1}^{K}\sum_{\ell=1}^{L} -\mu_0 c_k(x)\mathbf{p}_\ell^R(x)\int_{t^j}^{t^{j+1}} \widetilde{a}_{\ell}(\tau) z_{k\ell}^j(\tau)\,d\tau
\end{aligned}
$$

yield the same Hilbert space adjoint $F^{j'}(\hat{\alpha})^* : \mathcal{Y}^j \to \mathbb{R}^2$ as in formula (5.51), and the adjoint state $\mathbf{q}^{z^j}$ still needs to be solved on the whole time line $[0,T]$

$$
\begin{aligned}
&-\hat{\alpha}_1\mathbf{q}_t^{z^j} - \hat{\alpha}_2\mathbf{m}\times\mathbf{q}_t^{z^j} - 2\hat{\alpha}_2\mathbf{m}_t\times\mathbf{q}^{z^j} - \Delta\mathbf{q}^{z^j}\\
&\quad + 2\left((\nabla\mathbf{m})^\top\nabla\mathbf{m}\right)\mathbf{q}^{z^j} + 2\left((\nabla\mathbf{m})^\top\nabla\mathbf{q}^{z^j}\right)\mathbf{m}\\
&\quad + (-|\nabla\mathbf{m}|^2 + (\mathbf{m}\cdot\mathbf{h}))\mathbf{q}^{z^j} + (\mathbf{m}\cdot\mathbf{q}^{z^j})(\mathbf{h}+2\Delta\mathbf{m}) = \tilde{K}z^j && \text{in } (0,T)\times\Omega && (5.58)\\
&\partial_\nu\mathbf{q}^{z^j} = 0 && \text{on } (0,T)\times\partial\Omega && (5.59)\\
&\mathbf{q}^{z^j}(T) = (M_T^{\hat{\alpha}})^{-1}\tilde{K}_T z^j && \text{in } \Omega. && (5.60)
\end{aligned}
$$

Besides this, the conventional Kaczmarz method resulting from the collection of observation operators $\mathcal{K}_{k\ell}$ with $k=1\dots K, \ell=1\dots L$ as in (5.13) is always applicable

$$
F_{k\ell} : \mathcal{D}(F)(\subseteq \mathcal{X}) \to \mathcal{Y}_{k\ell}, \qquad \hat{\alpha}\mapsto y_{k\ell} := \mathcal{K}_{k\ell}\frac{\partial}{\partial t}(S(\hat{\alpha})) \tag{5.61}
$$

with

$$
\mathcal{Y}_{k\ell} = L^2(0,T) \qquad k=1\dots K, \ell = 1\dots \tag{5.62}
$$

Thus $F_{k\ell}'(\hat{\alpha})^*$ can be seen as (5.51), where the adjoint state $\mathbf{q}_{k\ell}^z$ solves (5.52)-(5.54) with corresponding data

$$
\begin{aligned}
\tilde{K}_{k\ell}z(x,t) &= -\mu_0 c_k(x)\mathbf{p}_\ell^R(x)\int_0^T \widetilde{a}_{\ell\,\tau}(\tau-t)z(\tau)\,d\tau \qquad t\in(0,T)\,,\\
\tilde{K}_{T\,k\ell}z(x) &= -\mu_0 c_k(x)\mathbf{p}_\ell^R(x)\int_0^T \widetilde{a}_{\ell}(\tau)z(\tau)\,d\tau
\end{aligned}
$$

for each $z\in\mathcal{Y}_{k\ell}$.

Solvability of the adjoint equation

First of all, we derive a bound for $\mathbf{q}^z$. To begin with, we set $\tau = T-t$ to convert (5.52)-(5.54) into an initial boundary value problem then test (5.52) with $\mathbf{q}_t^z$

$$
\int_\Omega \hat{\alpha}_1\mathbf{q}_t^z(t)\cdot\mathbf{q}_t^z(t)\,dx = \hat{\alpha}_1\|\mathbf{q}_t^z(t)\|_{L^2(\Omega,\mathbb{R}^3)}^2\,,
$$

$$\int_\Omega \hat{\alpha}_2(\mathbf{m}(t)\times\mathbf{q}^z_t(t))\cdot\mathbf{q}^z_t(t)\,dx = 0\,,$$
$$\int_\Omega \hat{\alpha}_2(\mathbf{m}_t(t)\times\mathbf{q}^z(t))\cdot\mathbf{q}^z_t(t)\,dx \le |\hat{\alpha}_2|\|\mathbf{m}_t(t)\|_{L^3(\Omega,\mathbb{R}^3)}\|\mathbf{q}^z(t)\|_{L^6(\Omega,\mathbb{R}^3)}\|\mathbf{q}^z_t(t)\|_{L^2(\Omega,\mathbb{R}^3)}\,,$$
$$\int_\Omega -\Delta\mathbf{q}^z(t)\cdot\mathbf{q}^z_t(t)\,dx = \frac{1}{2}\frac{d}{dt}\|\nabla\mathbf{q}^z(t)\|^2_{L^2(\Omega,\mathbb{R}^3)}\,,$$
$$\int_\Omega \big(((\nabla\mathbf{m}(t))^\top\nabla\mathbf{m}(t))\mathbf{q}^z(t)\big)\cdot\mathbf{q}^z_t(t)\,dx$$
$$\le (C^\Omega_{H^1\to L^6})^2\|\nabla\mathbf{m}\|^2_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))}\|\mathbf{q}^z(t)\|_{L^6(\Omega,\mathbb{R}^3)}\|\mathbf{q}^z_t(t)\|_{L^2(\Omega,\mathbb{R}^3)}\,,$$
$$\int_\Omega \big(((\nabla\mathbf{m}(t))^\top\nabla\mathbf{q}^z(t))\mathbf{m}(t)\big)\cdot\mathbf{q}^z_t(t)\,dx$$
$$\le C^\Omega_{H^2\to L^\infty}\|\nabla\mathbf{m}(t)\|_{H^2(\Omega,\mathbb{R}^3)}\|\nabla\mathbf{q}^z(t)\|_{L^2(\Omega,\mathbb{R}^3)}\|\mathbf{q}^z_t(t)\|_{L^2(\Omega,\mathbb{R}^3)}\,,$$
$$\int_\Omega(-|\nabla\mathbf{m}(t)|^2+(\mathbf{m}(t)\cdot\mathbf{h}))\mathbf{q}^z(t)\cdot\mathbf{q}^z_t(t)\,dx$$
$$\le \Big((C^\Omega_{H^1\to L^6})^2\|\nabla\mathbf{m}\|^2_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))}+\|\mathbf{h}(t)\|_{L^3(\Omega,\mathbb{R}^3)}\Big)\,\|\mathbf{q}^z(t)\|_{L^6(\Omega,\mathbb{R}^3)}\|\mathbf{q}^z_t(t)\|_{L^2(\Omega,\mathbb{R}^3)}\,,$$
$$\int_\Omega(\mathbf{m}(t)\cdot\mathbf{q}^z(t))\,\mathbf{h}(t)\cdot\mathbf{q}^z_t(t)\,dx \le \|\mathbf{h}(t)\|_{L^3(\Omega,\mathbb{R}^3)}\|\mathbf{q}^z(t)\|_{L^6(\Omega,\mathbb{R}^3)}\|\mathbf{q}^z_t(t)\|_{L^2(\Omega,\mathbb{R}^3)}\,,$$
$$\int_\Omega(\mathbf{m}(t)\cdot\mathbf{q}^z(t))\Delta\mathbf{m}(t)\cdot\mathbf{q}^z_t(t)\,dx$$
$$\le C^\Omega_{H^1\to L^3}\|\Delta\mathbf{m}(t)\|_{H^1(\Omega,\mathbb{R}^3))}\|\mathbf{q}^z(t)\|_{L^6(\Omega,\mathbb{R}^3)}\|\mathbf{q}^z_t(t)\|_{L^2(\Omega,\mathbb{R}^3)}\,,$$
$$\int_\Omega \tilde{K}z(t)\cdot\mathbf{q}^z_t(t)\,dx \le \|\tilde{K}z(t)\|_{L^2(\Omega,\mathbb{R}^3)}\|\mathbf{q}^z_t(t)\|_{L^2(\Omega,\mathbb{R}^3)}\,.$$

Above, we employ the fact that the solution $\mathbf{m}$ to the LLG equation has $|\mathbf{m}|=1$ and continuity of the embeddings $H^1(\Omega,\mathbb{R}^3)\hookrightarrow L^6(\Omega,\mathbb{R}^3)\hookrightarrow L^3(\Omega,\mathbb{R}^3)$, $H^2(\Omega,\mathbb{R}^3)\hookrightarrow L^\infty(\Omega,\mathbb{R}^3)$ through the constants $C^\Omega_{H^1\to L^6}, C^\Omega_{H^1\to L^3}$ and $C^\Omega_{H^2\to L^\infty}$, respectively. Employing Young's inequality we deduce , for each $t\le T$ and $\epsilon>0$ sufficiently small,

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}\|\nabla\mathbf{q}^z(t)\|^2_{L^2(\Omega,\mathbb{R}^3)}+(\hat{\alpha}_1-\epsilon)\|\mathbf{q}^z_t(t)\|^2_{L^2(\Omega,\mathbb{R}^3)}\\ &\le\bigg[\Big(\|\nabla\mathbf{m}\|^4_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))}+\|\nabla\mathbf{m}(t)\|^2_{H^2(\Omega,\mathbb{R}^3)}+\|\mathbf{m}_t(t)\|^2_{L^3(\Omega,\mathbb{R}^3)}+\|\mathbf{h}(t)\|^2_{L^3(\Omega,\mathbb{R}^3)}\Big)\\ &\qquad .\|\mathbf{q}^z(t)\|^2_{H^1(\Omega,\mathbb{R}^3)}+\|\tilde{K}z(t)\|^2_{L^2(\Omega,\mathbb{R}^3)}\bigg]\frac{C}{4\epsilon}.\end{aligned}\tag{5.63}$$

The generic constant C might take different values whenever it appears.

To have the full H^1-norm on the left hand side of this estimate, we apply the transformation $\tilde{\mathbf{q}^z}(t)=e^t\mathbf{q}^z(t)$, which yields $\tilde{\mathbf{q}^z}_t(t)=e^t(\mathbf{q}^z(t)+\mathbf{q}^z_t(t))$. After testing by $\mathbf{q}^z_t$, the term $\int_\Omega\mathbf{q}^z(t)\cdot\mathbf{q}^z_t(t)\,dx=\frac{1}{2}\frac{d}{dt}\|\mathbf{q}^z(t)\|^2_{L^2(\Omega,\mathbb{R}^3)}$ will contribute to $\frac{1}{2}\frac{d}{dt}\|\nabla\mathbf{q}^z(t)\|^2_{L^2(\Omega,\mathbb{R}^3)}$ forming the full H^1-norm on the left hand side. Alternatively, one can add $\mathbf{q}^z$ to both sides of (5.52) and evaluate the right hand side with $\int_\Omega \mathbf{q}^z(t)\cdot\mathbf{q}_t^z(t)\,dx \le \frac{1}{4\epsilon}\|\mathbf{q}^z(t)\|^2_{H^1(\Omega,\mathbb{R}^3)} + \epsilon\|\mathbf{q}_t^z(t)\|^2_{L^2(\Omega,\mathbb{R}^3)}$.

Integrating over $(0,t)$, we get

$$
\begin{aligned}
&\frac{1}{2}\|\mathbf{q}^z(t)\|^2_{H^1(\Omega,\mathbb{R}^3)} + (\hat{\alpha}_1-\epsilon)\|\mathbf{q}_t^z\|^2_{L^2(0,t;L^2(\Omega,\mathbb{R}^3))}\\
&\le \left[\int_0^t \Big(\|\nabla\mathbf{m}\|^4_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))} + \|\nabla\mathbf{m}(\tau)\|^2_{H^2(\Omega,\mathbb{R}^3)} + \|\mathbf{m}_t(\tau)\|^2_{L^3(\Omega,\mathbb{R}^3)} + \|\mathbf{h}(\tau)\|^2_{L^3(\Omega,\mathbb{R}^3)}\Big)\right.\\
&\qquad\qquad \left. .\|\mathbf{q}^z(\tau)\|^2_{H^1(\Omega,\mathbb{R}^3)}\,d\tau + \|\tilde{K}z\|^2_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))} + \|(M_T^{\hat{\alpha}})^{-1}\tilde{K}_T z\|^2_{H^1(\Omega,\mathbb{R}^3)}\right]\frac{C}{4\epsilon}
\end{aligned}
$$

with the evaluation for the terms $\|\tilde{K}z\|_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))}$ and $\|(M_T^{\hat{\alpha}})^{-1}\tilde{K}_T z\|^2_{H^1(\Omega,\mathbb{R}^3)}$ (not causing any misunderstanding, we omit here the subscripts k,ℓ for indices of concentrations and coil sensitivities)

$$
\begin{aligned}
\|\tilde{K}z(t)\|^2_{L^2(\Omega,\mathbb{R}^3)} &\le C\|c\mathbf{p}^R\|^2_{L^2(\Omega,\mathbb{R}^3)}\|\tilde{a}\|^2_{H^1(0,T)}\|z\|^2_{L^2(0,T)}\\
&\le C^{\tilde{a},c,\mathbf{p}^R}\|z\|^2_{L^2(0,T)}\,,
\end{aligned}
$$

$$
\begin{aligned}
&|(M_T^{\hat{\alpha}})^{-1}\tilde{K}_T z\|^2_{H^1(\Omega,\mathbb{R}^3)}\\
&\le C^{\hat{\alpha}}\|z\|^2_{L^2(0,T)}\|\tilde{a}\|^2_{L^2(0,T)}\\
&\qquad\qquad .\Big(\|c\mathbf{p}^R\|^2_{H^1(\Omega,\mathbb{R}^3)} + \|c\mathbf{p}\mathbf{m}_i(T)\|^2_{H^1(\Omega,\mathbb{R}^3)} + \|c\mathbf{p}^R\mathbf{m}_j(T)\mathbf{m}_k(T)\|^2_{H^1(\Omega,\mathbb{R}^3)}\Big)\\
&\le C^{\hat{\alpha}_0,\rho,\tilde{a}}\|z\|^2_{L^2(0,T)}\Big(\|c\mathbf{p}^R\|^2_{H^1(\Omega,\mathbb{R}^3)} + \|c\mathbf{p}^R\|^2_{L^6(\Omega,\mathbb{R}^3)}\|\nabla\mathbf{m}(T)\|^2_{L^3(\Omega,\mathbb{R}^3)}\Big)\\
&\le C^{\tilde{a}}\|z\|^2_{L^2(0,T)}\Big(\|c\mathbf{p}^R\|^2_{H^1(\Omega,\mathbb{R}^3)} + (C^\Omega_{H^1\to L^6}C^\Omega_{H^1\to L^3})^2\|c\mathbf{p}^R\|^2_{H^1(\Omega,\mathbb{R}^3)}\|\nabla\mathbf{m}\|^2_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))}\Big)\\
&\le C^{\tilde{a},c,\mathbf{p}^R}\|z\|^2_{L^2(0,T)}\|\nabla\mathbf{m}\|^2_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))}
\end{aligned}
$$

with some $i,j,k=1,2,3$. This estimate holds for $c\mathbf{p}^R\in H^1(\Omega,\mathbb{R}^3)$ thus requires some smoothness of the concentration c, while the coil sensitivity $\mathbf{p}^R$ is usually smooth in practice.

Then applying Grönwall's inequality yields

$$
\begin{aligned}
&\|\mathbf{q}^z\|_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))}\\
&\le C\exp\Big(\|\nabla\mathbf{m}\|^2_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))} + \|\nabla\mathbf{m}\|_{L^2(0,T;H^2(\Omega,\mathbb{R}^3))} + \|\mathbf{m}_t\|_{L^2(0,T;L^3(\Omega,\mathbb{R}^3))}\\
&\qquad\qquad + \|\mathbf{h}\|_{L^2(0,T;L^3(\Omega,\mathbb{R}^3))}\Big).\Big(\|\tilde{K}z\|_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))} + \|(M_T^{\hat{\alpha}})^{-1}\tilde{K}_T z\|_{H^1(\Omega,\mathbb{R}^3)}\Big)\\
&\le C^{\tilde{a},c,\mathbf{p}^R}\Big(\|\nabla\mathbf{m}\|_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))\cap L^2(0,T;H^2(\Omega,\mathbb{R}^3))}, \|\mathbf{m}_t\|_{L^2(0,T;L^3(\Omega,\mathbb{R}^3))}, \|\mathbf{h}\|_{L^2(0,T;L^3(\Omega,\mathbb{R}^3))}\Big)\\
&\qquad\qquad .\|z\|_{L^2(0,T)}.
\end{aligned}
$$

Integrating (5.63) on $(0,T)$, we also get

$$
\begin{aligned}
&\|\mathbf{q}^z_t\|_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))}\\
&\leq C^{\tilde{a},c,\mathbf{p}^R}\left(\|\nabla\mathbf{m}\|_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))\cap L^2(0,T;H^2(\Omega,\mathbb{R}^3))}, \|\mathbf{m}_t\|_{L^2(0,T;L^3(\Omega,\mathbb{R}^3))}, \|\mathbf{h}\|_{L^2(0,T;L^3(\Omega,\mathbb{R}^3))}\right)\\
&\qquad .\|z\|_{L^2(0,T)}.
\end{aligned}
$$

Altogether, we obtain

$$
\begin{aligned}
&\|\mathbf{q}^z\|_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))} + \|\mathbf{q}^z_t\|_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))}\\
&\leq C^{\tilde{a},c,\mathbf{p}^R}\left(\|\nabla\mathbf{m}\|_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))\cap L^2(0,T;H^2(\Omega,\mathbb{R}^3))}, \|\mathbf{m}_t\|_{L^2(0,T;L^3(\Omega,\mathbb{R}^3))}, \|\mathbf{h}\|_{L^2(0,T;L^3(\Omega,\mathbb{R}^3))}\right)\\
&\qquad .\|z\|_{L^2(0,T)}.
\end{aligned}
\tag{5.64}
$$

This result applied to the Galerkin approximation implies existence of the solution to the adjoint equation. Uniqueness also follows from (5.64).

Regularity of the solution to the LLG equation

In (5.64), first of all we need the solution to the LLG equation $\mathbf{m}\in L^\infty(0,T;H^2(\Omega,\mathbb{R}^3))$ $\cap L^2(0,T;H^3(\Omega,\mathbb{R}^3))$. This could be referred from the regularity result in [43, Lemma 2.3] for $\mathbf{m}_0\in H^2(\Omega,\mathbb{R}^3)$ with small $\|\nabla\mathbf{m}_0\|_{L^2(\Omega,\mathbb{R}^3)}$. The remaining task is verifying that the estimate still holds in case $\mathbf{h}$ is present, i.e., the right hand side of (5.21) contains the additional the term $\mathrm{Proj}_{\mathbf{m}^\perp}\mathbf{h}$.

Following the line of the proof in [43, Lemma 2.3], we take the second space derivative of $\mathrm{Proj}_{\mathbf{m}^\perp}\mathbf{h}$ then test it by $\Delta\mathbf{m}$

$$
\begin{aligned}
&\int_\Omega \Delta\mathbf{h}(t)\cdot\Delta\mathbf{m}(t)\,dx\\
&\leq\begin{cases}\|\Delta\mathbf{h}(t)\|_{L^2(\Omega,\mathbb{R}^3)}\|\Delta\mathbf{m}(t)\|_{L^2(\Omega,\mathbb{R}^3)} & \text{if } \mathbf{h}\in L^2(0,T;H^2(\Omega,\mathbb{R}^3))\\ \|\nabla\mathbf{h}(t)\|_{L^2(\Omega,\mathbb{R}^3)}\|\nabla^3\mathbf{m}(t)\|_{L^2(\Omega,\mathbb{R}^3)} & \text{if } \mathbf{h}\in L^2(0,T;H^1(\Omega,\mathbb{R}^3)),\ \partial_\nu\mathbf{h}=0 \text{ on } \partial\Omega\end{cases},
\end{aligned}
$$

$$
\begin{aligned}
&\int_\Omega \Delta((\mathbf{m}(t)\cdot\mathbf{h}(t))\mathbf{m}(t))\cdot\Delta\mathbf{m}(t)\,dx\\
&\leq\begin{cases}C\|\mathbf{h}(t)\|_{H^2(\Omega,\mathbb{R}^3)}\left(1+6\|\nabla\mathbf{m}(t)\|_{H^1(\Omega,\mathbb{R}^3)}+2\|\nabla\mathbf{m}(t)\|_{H^2(\Omega,\mathbb{R}^3)}\|\nabla\mathbf{m}\|_{L^\infty(0,T;L^2(\Omega,\mathbb{R}^3))}\right)\\ \qquad .\|\Delta\mathbf{m}(t)\|_{L^2(\Omega,\mathbb{R}^3)} \qquad \text{if } \mathbf{h}\in L^2(0,T;H^2(\Omega,\mathbb{R}^3))\\ C\|\mathbf{h}(t)\|_{H^1(\Omega,\mathbb{R}^3)}\left(1+2\|\nabla\mathbf{m}(t)\|_{L^3(\Omega,\mathbb{R}^3)}\right)\|\nabla^3\mathbf{m}(t)\|_{L^2(\Omega,\mathbb{R}^3)}\\ \qquad \text{if } \mathbf{h}\in L^2(0,T;H^1(\Omega,\mathbb{R}^3)),\ \partial_\nu\mathbf{h}=0 \text{ on } \partial\Omega\end{cases}
\end{aligned}
$$

with C just depending on the constants in the embeddings $H^1(\Omega,\mathbb{R}^3)\hookrightarrow L^6(\Omega,\mathbb{R}^3)\hookrightarrow L^3(\Omega,\mathbb{R}^3)$. Then we can proceed similarly to the proof of [43, Lemma 2.3] with applying Yong's inequality, Gronwall's inequality and time integration to arrive at

$$
\begin{aligned}
&\|\nabla\mathbf{m}\|_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))\cap L^2(0,T;H^2(\Omega,\mathbb{R}^3))}\\
&\qquad\leq\left(\|\nabla\mathbf{m}_0\|_{H^1(\Omega,\mathbb{R}^3)}+\|\mathbf{h}\|\right)C(\|\nabla\mathbf{m}_0\|_{H^1(\Omega,\mathbb{R}^3)},\|\mathbf{h}\|),
\end{aligned}
\tag{5.65}
$$

where $\|\mathbf{h}\|$ is evaluated in $L^2(0,T;H^1(\Omega,\mathbb{R}^3))$ or $L^2(0,T;H^2(\Omega,\mathbb{R}^3))$ as in the two cases mentioned above.

It remains to prove $\mathbf{m}_t \in L^2(0,T;H^1(\Omega,\mathbb{R}^3)) \hookrightarrow L^2(0,T;L^3(\Omega,\mathbb{R}^3))$ to validate (5.64). For this purpose, instead of working with (5.21) we test (5.18) by $-\Delta\mathbf{m}_t$

$$
\begin{aligned}
&\int_\Omega \mathbf{m}_t(t)\cdot(-\Delta\mathbf{m}_t(t))\,dx = \|\nabla\mathbf{m}_t(t)\|^2_{L^2(\Omega,\mathbb{R}^3)}\,,\\
&\int_\Omega -\alpha_1\Delta\mathbf{m}(t)\cdot(-\Delta\mathbf{m}_t(t))\,dx = \frac{\alpha_1}{2}\frac{d}{dt}\|\Delta\mathbf{m}(t)\|^2_{L^2(\Omega,\mathbb{R}^3)}\,,\\
&\int_\Omega -\alpha_1|\nabla\mathbf{m}(t)|^2\mathbf{m}(t)\cdot(-\Delta\mathbf{m}_t(t))\,dx = -\alpha_1\int_\Omega \nabla\left(|\nabla\mathbf{m}(t)|^2\mathbf{m}(t)\right):\nabla\mathbf{m}_t(t)\,dx\\
&\le \alpha_1\Big(2C^\Omega_{H^1\to L^6}C^\Omega_{H^1\to L^3}\|\nabla\mathbf{m}\|_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))}\|\Delta\mathbf{m}(t)\|_{H^1(\Omega,\mathbb{R}^3)}\\
&\qquad\qquad + (C^\Omega_{H^1\to L^6})^3\|\nabla\mathbf{m}\|^3_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))}\Big)\|\nabla\mathbf{m}_t(t)\|_{L^2(\Omega,\mathbb{R}^3)}\,,\\
&\int_\Omega -\alpha_1(\mathbf{h}(t)-(\mathbf{m}(t)\cdot\mathbf{h}(t))\mathbf{m}(t))\cdot(-\Delta\mathbf{m}_t(t))\,dx\\
&= -\alpha_1\int_\Omega \nabla(\mathbf{h}(t)-(\mathbf{m}(t)\cdot\mathbf{h}(t))\mathbf{m}(t)):\nabla\mathbf{m}_t(t)\,dx\\
&\le 2\alpha_1\left(\|\nabla\mathbf{h}(t)\|_{L^2(\Omega,\mathbb{R}^3)} + C^\Omega_{H^1\to L^6}\|\mathbf{h}(t)\|_{L^3(\Omega,\mathbb{R}^3)}\|\nabla\mathbf{m}\|_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))}\right)\|\nabla\mathbf{m}_t(t)\|_{L^2(\Omega,\mathbb{R}^3)}\,,\\
&\int_\Omega -\alpha_2(\mathbf{m}(t)\times\Delta\mathbf{m}(t))\cdot(-\Delta\mathbf{m}_t(t))\,dx = \int_\Omega -\alpha_2\nabla(\mathbf{m}(t)\times\Delta\mathbf{m}(t)):\nabla\mathbf{m}_t(t)\,dx\\
&\le |\alpha_2|\big(C^\Omega_{H^1\to L^6}C^\Omega_{H^1\to L^3}\|\nabla\mathbf{m}\|_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))}\|\Delta\mathbf{m}(t)\|_{H^1(\Omega,\mathbb{R}^3)}\\
&\qquad\qquad + \|\nabla^3\mathbf{m}(t)\|_{L^2(\Omega,\mathbb{R}^3)}\big)\|\nabla\mathbf{m}_t(t)\|_{L^2(\Omega,\mathbb{R}^3)}\,,\\
&\int_\Omega -\alpha_2(\mathbf{m}(t)\times\mathbf{h}(t))\cdot(-\Delta\mathbf{m}_t(t))\,dx = \int_\Omega -\alpha_2\nabla(\mathbf{m}(t)\times\mathbf{h}(t)):(\nabla\mathbf{m}_t(t))\,dx\\
&\le |\alpha_2|\left(C^\Omega_{H^1\to L^6}\|\mathbf{h}(t)\|_{L^3(\Omega,\mathbb{R}^3)}\|\nabla\mathbf{m}\|_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))} + \|\nabla\mathbf{h}(t)\|_{L^2(\Omega,\mathbb{R}^3)}\right)\|\nabla\mathbf{m}_t(t)\|_{L^2(\Omega,\mathbb{R}^3)}.
\end{aligned}
$$

Integrating over $(0,T)$ then employing Hölder's inequality, Young's inequality and (5.65), it follows that

$$
\begin{aligned}
&(1-\epsilon)\|\nabla\mathbf{m}_t\|_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))}\\
&\le \frac{C}{4\epsilon}\Big(\|\nabla\mathbf{m}\|_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))}\|\nabla\mathbf{m}\|_{L^2(0,T;H^2(\Omega,\mathbb{R}^3))} + \|\nabla\mathbf{m}\|^3_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))}\\
&+ \|\nabla\mathbf{m}\|_{L^2(0,T;H^2(\Omega,\mathbb{R}^3))} + \|\mathbf{h}\|_{L^2(0,T;H^1(\Omega,\mathbb{R}^3))}\|\nabla\mathbf{m}\|_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))} + \|\mathbf{h}\|_{L^2(0,T;H^1(\Omega,\mathbb{R}^3))}\Big)\\
&\le \left(\|\nabla\mathbf{m}_0\|_{H^1(\Omega,\mathbb{R}^3)} + \|\mathbf{h}\|\right)C(\|\nabla\mathbf{m}_0\|_{H^1(\Omega,\mathbb{R}^3)},\|\mathbf{h}\|).
\end{aligned}\tag{5.66}
$$

Also $\|\mathbf{m}_t\|_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))} < C\left(\|\nabla\mathbf{m}_0\|_{L^2(\Omega,\mathbb{R}^3)} + \|\mathbf{h}\|_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))}\right)$ according to [76] with taking into account the present of $\mathbf{h}$, we arrive at

$$
\|\mathbf{m}_t\|_{L^2(0,T;H^1(\Omega,\mathbb{R}^3))} \le \left(\|\nabla\mathbf{m}_0\|_{H^1(\Omega,\mathbb{R}^3)} + \|\mathbf{h}\|\right)C(\|\nabla\mathbf{m}_0\|_{H^1(\Omega,\mathbb{R}^3)},\|\mathbf{h}\|),\tag{5.67}
$$

where $\|\mathbf{h}\|$ is evaluated in $L^2(0,T;H^1(\Omega,\mathbb{R}^3))$ or $L^2(0,T;H^2(\Omega,\mathbb{R}^3))$.

In conclusion, the fact that $\mathbf{m} \in L^\infty(0,T;H^2(\Omega,\mathbb{R}^3)) \cap L^2(0,T;H^3(\Omega,\mathbb{R}^3)) \cap H^1(0,T;H^1(\Omega,\mathbb{R}^3))$ for $\mathbf{m}_0 \in H^2(\Omega,\mathbb{R}^3)$ with small $\|\nabla \mathbf{m}_0\|_{L^2(\Omega,\mathbb{R}^3)}$, and $\mathbf{h} \in L^2(0,T;H^1(\Omega,\mathbb{R}^3)), \partial_\nu \mathbf{h} = 0$ on $\partial\Omega$ or $\mathbf{h} \in L^2(0,T;H^2(\Omega,\mathbb{R}^3))$ guarantee unique existence of the adjoint state $\mathbf{q}^z \in L^\infty(0,T;H^1(\Omega,\mathbb{R}^3)) \cap H^1(0,T;L^2(\Omega,\mathbb{R}^3))$. And this regularity of $\mathbf{q}^z$ ensures the adjoint $F'(\hat{\alpha})^*$ in (5.51) to be well-defined.

Remark 5.4.3.

- The LLG equation (5.21)-(5.23) is uniquely solvable for $\hat{\alpha}_1 > 0$ and arbitrary $\hat{\alpha}_2$. Therefore, the regularization problem should be locally solved within the ball $\mathcal{B}_\rho(\hat{\alpha}^0)$ of center $\hat{\alpha}^0$ with $\hat{\alpha}^0_1 > 0$ and radius $\rho < \hat{\alpha}^0_1$.

- [43, Lemma 2.3] requires smallness $\|\nabla \mathbf{m}_0\|_{L^2(\Omega,\mathbb{R}^3)} \le \lambda$, and this smallness depends on $\hat{\alpha}$ through the relation $C^I\left(\lambda^2 + 2\lambda + \frac{\hat{\alpha}_2}{\hat{\alpha}_1}\lambda\right) < 1$ with C^I depending on the constants in the interpolation inequalities.

Altogether, we arrive at

$$\mathcal{D}(F) = \left\{ \hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2) \in \mathcal{B}_\rho(\hat{\alpha}^0) : 0 < \hat{\alpha}^0_1, \rho < \hat{\alpha}^0_1, C^I\left(\lambda^2 + 2\lambda + \frac{\hat{\alpha}_2}{\hat{\alpha}_1}\lambda\right) < 1 \right\}. \tag{5.68}$$

Differentiability of the forward operator

Since the observation operator $\mathcal{K}$ is linear, differentiability of F is just the question of differentiability of S.

Let us rewrite the LLG equation (5.21) in the following form

$$\tilde{g}(\hat{\alpha}, \mathbf{m}) - \Delta\mathbf{m} = \tilde{f}(\mathbf{m})$$

and denote

$$\tilde{\mathbf{v}}^\epsilon := \frac{S(\hat{\alpha} + \epsilon\beta) - S(\hat{\alpha})}{\epsilon} - \mathbf{u} =: \frac{\mathbf{n} - \mathbf{m}}{\epsilon} - \mathbf{u} =: \mathbf{v}^\epsilon - \mathbf{u}.$$

Considering the system of equations

$$\begin{aligned}
&\tilde{g}(\hat{\alpha} + \epsilon\beta, \mathbf{n}) && - \Delta\mathbf{n} = \tilde{f}(\mathbf{n}),\\
&\tilde{g}(\hat{\alpha}, \mathbf{m}) && - \Delta\mathbf{m} = \tilde{f}(\mathbf{m}),\\
&\tilde{g}'_{\mathbf{m}}(\hat{\alpha}, \mathbf{m})\mathbf{u} + \tilde{g}'_{\hat{\alpha}}(\hat{\alpha}, \mathbf{m})\beta && - \Delta\mathbf{u} = \tilde{f}'_{\mathbf{m}}(\mathbf{m})\mathbf{u}
\end{aligned}$$

with the same boundary and initial data for each, we see that $\tilde{\mathbf{v}}^\epsilon$ solves

$$
\begin{aligned}
&\tilde{g}'_{\mathbf{m}}(\hat{\alpha},\mathbf{m})\tilde{\mathbf{v}}^\epsilon - \Delta\tilde{\mathbf{v}}^\epsilon - \tilde{f}'_{\mathbf{m}}(\mathbf{m})\tilde{\mathbf{v}}^\epsilon \\
&\quad = \frac{\tilde{f}(\mathbf{n})-\tilde{f}(\mathbf{m})}{\epsilon} - \tilde{f}'_{\mathbf{m}}(\mathbf{m})\mathbf{v}^\epsilon - \frac{\tilde{g}(\hat{\alpha}+\epsilon\beta,\mathbf{n})-\tilde{g}(\hat{\alpha},\mathbf{m})}{\epsilon} \\
&\qquad + \tilde{g}'_{\mathbf{m}}(\hat{\alpha},\mathbf{m})\mathbf{v}^\epsilon + \tilde{g}'_{\hat{\alpha}}(\hat{\alpha},\mathbf{m})\beta && \text{in } (0,T)\times\Omega && (5.69)\\
&\partial_\nu\tilde{\mathbf{v}}^\epsilon = 0 && \text{on } [0,T]\times\partial\Omega && (5.70)\\
&\tilde{\mathbf{v}}^\epsilon(0) = 0 && \text{in } \Omega, && (5.71)
\end{aligned}
$$

explicitly

$$
\begin{aligned}
&\hat{\alpha}_1\tilde{\mathbf{v}}^\epsilon_t - \hat{\alpha}_2\mathbf{m}\times\tilde{\mathbf{v}}^\epsilon_t - \hat{\alpha}_2\tilde{\mathbf{v}}^\epsilon\times\mathbf{m}_t - \Delta\tilde{\mathbf{v}}^\epsilon \\
&\quad - 2(\nabla\tilde{\mathbf{v}}^\epsilon : \nabla\mathbf{m})\mathbf{m} + \tilde{\mathbf{v}}^\epsilon(-|\nabla\mathbf{m}|^2 + (\mathbf{m}\cdot\mathbf{h})) + (\tilde{\mathbf{v}}^\epsilon\cdot\mathbf{h})\mathbf{m} \\
&= \frac{1}{\epsilon}\left(|\nabla\mathbf{n}|^2\mathbf{n} + \mathrm{Proj}_{\mathbf{n}^\perp}\mathbf{h} - |\nabla\mathbf{m}|^2\mathbf{m} - \mathrm{Proj}_{\mathbf{m}^\perp}\mathbf{h}\right) && && (5.72)\\
&\quad - 2(\nabla\mathbf{v}^\epsilon : \nabla\mathbf{m})\mathbf{m} + \mathbf{v}^\epsilon(-|\nabla\mathbf{m}|^2 + (\mathbf{m}\cdot\mathbf{h})) + (\mathbf{v}^\epsilon\cdot\mathbf{h})\mathbf{m} \\
&\quad - \frac{1}{\epsilon}\left((\hat{\alpha}_1+\epsilon\beta_1)\mathbf{n}_t - (\hat{\alpha}_2+\epsilon\beta_2)\mathbf{n}\times\mathbf{n}_t - \hat{\alpha}_1\mathbf{m}_t + \hat{\alpha}_2\mathbf{m}\times\mathbf{m}_t\right) \\
&\quad + \hat{\alpha}_1\mathbf{v}^\epsilon_{\mathbf{t}} - \hat{\alpha}_2\mathbf{m}\times\mathbf{v}^\epsilon_{\mathbf{t}} - \hat{\alpha}_2\mathbf{v}^\epsilon\times\mathbf{m}_t \\
&\quad + \beta_1\mathbf{m}_t - \beta_2\mathbf{m}\times\mathbf{m}_t && \text{in } (0,T)\times\Omega \\
&\partial_\nu\tilde{\mathbf{v}}^\epsilon = 0 && \text{on } [0,T]\times\partial\Omega && (5.73)\\
&\tilde{\mathbf{v}}^\epsilon(0) = 0 && \text{in } \Omega. && (5.74)
\end{aligned}
$$

Observing the similarity of (5.72)-(5.74) to the adjoint equation (5.52)-(5.54) with $\tilde{\mathbf{v}}^\epsilon$ in place of $\mathbf{q}^z$ and denoting by $\mathbf{b}^\epsilon$ the right hand side of (5.69) or (5.72), one can evaluate $\|\tilde{\mathbf{v}}^\epsilon\|$ using the same technique as in Section 5.4.2. By this way, one achieves, for each $\epsilon\in[0,\bar{\epsilon}]$,

$$
\|\tilde{\mathbf{v}}^\epsilon\|_{L^\infty(0,T;H^1(\Omega,\mathbb{R}^3))\cap H^1(0,T;L^2(\Omega,\mathbb{R}^3))} \le C\|\mathbf{b}^\epsilon\|_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))}
$$

with $\mathbf{b}^\epsilon\in L^2(0,T;L^2(\Omega,\mathbb{R}^3))$ also by analogously estimating and employing $\mathbf{m},\mathbf{n}\in L^\infty(0,T;H^2(\Omega,\mathbb{R}^3))\cap L^2(0,T;H^3(\Omega,\mathbb{R}^3))\cap H^1(0,T;H^1(\Omega,\mathbb{R}^3))$. We note that the constant C here is independent of ϵ.

Next letting $\mathcal{V} := L^\infty(0,T;H^1(\Omega,\mathbb{R}^3)) \cap H^1(0,T;L^2(\Omega,\mathbb{R}^3))$, we have

$$
\begin{aligned}
&\|\mathbf{b}^\epsilon\|_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))}\\
&= \Bigg\| \frac{\tilde{f}(\mathbf{n}) - \tilde{f}(\mathbf{m})}{\epsilon} - \tilde{f}'_{\mathbf{m}}(\mathbf{m})\mathbf{v}^\epsilon - \frac{\tilde{g}(\hat{\alpha}+\epsilon\beta,\mathbf{n}) - \tilde{g}(\hat{\alpha},\mathbf{m})}{\epsilon}\\
&\qquad\qquad + \tilde{g}'_{\mathbf{m}}(\hat{\alpha},\mathbf{m})\mathbf{v}^\epsilon + \tilde{g}'_{\hat{\alpha}}(\hat{\alpha},\mathbf{m})\beta \Bigg\|_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))}\\
&\le \Bigg\| \int_0^1 \Big(\big(\tilde{f}'_{\mathbf{m}}(\mathbf{m}+\lambda\epsilon\mathbf{v}^\epsilon) - \tilde{f}'_{\mathbf{m}}(\mathbf{m})\big)\mathbf{v}^\epsilon - \big(\tilde{g}'_{\mathbf{m}}(\hat{\alpha}+\lambda\epsilon\beta,\mathbf{m}+\lambda\epsilon\mathbf{v}^\epsilon) - \tilde{g}'_{\mathbf{m}}(\hat{\alpha},\mathbf{m})\big)\mathbf{v}^\epsilon\\
&\qquad\qquad - \big(\tilde{g}'_{\hat{\alpha}}(\hat{\alpha}+\lambda\epsilon\beta,\mathbf{m}+\lambda\epsilon\mathbf{v}^\epsilon) - \tilde{g}'_{\hat{\alpha}}(\hat{\alpha},\mathbf{m})\big)\beta \Big)\, d\lambda \Bigg\|_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))}\\
&\le 2 \sup_{\substack{\lambda\in[0,1]\\ \epsilon\in[0,\bar{\epsilon}]}} \Big(\|\tilde{f}'_{\mathbf{m}}(\mathbf{m}+\lambda\epsilon\mathbf{v}^\epsilon)\|_{\mathcal{V}\to L^2(0,T;L^2(\Omega,\mathbb{R}^3))}\|\mathbf{v}^\epsilon\|_{\mathcal{V}}\\
&\qquad\qquad + \|\tilde{g}'_{\mathbf{m}}(\hat{\alpha}+\lambda\epsilon\beta,\mathbf{m}+\lambda\epsilon\mathbf{v}^\epsilon)\|_{\mathcal{V}\to L^2(0,T;L^2(\Omega,\mathbb{R}^3))}\|\mathbf{v}^\epsilon\|_{\mathcal{V}}\\
&\qquad\qquad + \|\tilde{g}'_{\hat{\alpha}}(\hat{\alpha}+\lambda\epsilon\beta,\mathbf{m}+\lambda\epsilon\mathbf{v}^\epsilon)\|_{\mathbb{R}^2\to L^2(0,T;L^2(\Omega,\mathbb{R}^3))}|\beta| \Big).
\end{aligned}
$$

In order to prove uniform boundedness of the derivatives of $\tilde{f},\tilde{g}$ w.r.t λ,ϵ in the above estimate, we again proceed in similar a manner to Section 5.4.2 since the space for $\mathbf{q}^z$ in Section 5.4.2 (c.f. (5.65)) coincides with $\mathcal{V}$ here and by the fact that for $\mathbf{m},\mathbf{n} \in L^\infty(0,T;H^2(\Omega,\mathbb{R}^3)) \cap L^2(0,T;H^3(\Omega,\mathbb{R}^3)) \cap H^1(0,T;H^1(\Omega,\mathbb{R}^3))$

$$
\begin{aligned}
\max\{\|\mathbf{m}\|,\|\mathbf{n}\|\} &\le \max\Big\{\frac{1}{\hat{\alpha}_1}, \frac{1}{\hat{\alpha}_1+\epsilon\beta_1}\Big\} C\left(\|\mathbf{m}_0\|_{H^2(\Omega,\mathbb{R}^3)}, \|\mathbf{h}\|_{L^2(0,T;H^2(\Omega,\mathbb{R}^3))}\right)\\
&\le \frac{C}{\hat{\alpha}_1^0 - \rho}. \qquad (5.75)
\end{aligned}
$$

If $\partial_\nu \mathbf{h} = 0$ on $\partial\Omega$, we just need the $\|.\|_{L^2(0,T;H^1(\Omega,\mathbb{R}^3))}$-norm for $\mathbf{h}$ as claimed in (5.65). This estimate holds for any $\epsilon \in [0,\bar{\epsilon}]$, and the constant C is independent of ϵ.

To accomplish uniform boundedness for $\|\mathbf{b}^\epsilon\|_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))}$, we need to show that $\|\mathbf{v}^\epsilon\|_{\mathcal{V}}$ is also uniformly bounded w.r.t ϵ. It is seen from

$$
\begin{aligned}
\tilde{g}(\hat{\alpha}+\epsilon\beta,\mathbf{n}) - \Delta\mathbf{n} &= \tilde{f}(\mathbf{n}),\\
\tilde{g}(\hat{\alpha},\mathbf{m}) \qquad - \Delta\mathbf{m} &= \tilde{f}(\mathbf{m})
\end{aligned}
$$

that $\mathbf{v}^\epsilon$ solves

$$\begin{aligned}
\int_0^1 \tilde{g}'_{\mathbf{m}}(\hat{\alpha}+\lambda\epsilon\beta, \mathbf{m}+\lambda\epsilon\mathbf{v}^\epsilon)\mathbf{v}^\epsilon + \tilde{g}'_{\hat{\alpha}}(\hat{\alpha}+\lambda\epsilon\beta, \mathbf{m}+\lambda\epsilon\mathbf{v}^\epsilon)\beta \, d\lambda - \Delta\mathbf{v}^\epsilon & \\
= \int_0^1 \tilde{f}'_{\mathbf{m}}(\mathbf{m}+\lambda\epsilon\mathbf{v}^\epsilon)\mathbf{v}^\epsilon \, d\lambda & \quad \text{in } (0,T)\times\Omega \qquad (5.76)\\
\partial_\nu \mathbf{v}^\epsilon = 0 & \quad \text{on } [0,T]\times\partial\Omega \qquad (5.77)\\
\mathbf{v}^\epsilon(0) = 0 & \quad \text{in } \Omega. \qquad (5.78)
\end{aligned}$$

Noting that $\mathbf{M} := \mathbf{m} + \lambda\epsilon\mathbf{v}^\epsilon = \lambda\mathbf{n} + (1-\lambda)\mathbf{m}$ has $\|\mathbf{M}\| \le \frac{C}{\hat{\alpha}_1^0 - \rho}, \forall \lambda \in [0,1]$ with C being independent of ϵ, and $\tilde{g}$ is first order in $\hat{\alpha}$, we can rewrite (5.76) into the following linear equation

$$\tilde{G}(\hat{\alpha}+\lambda\epsilon\beta, \mathbf{M})\mathbf{v}^\epsilon - \Delta\mathbf{v}^\epsilon + \tilde{F}(\mathbf{M})\mathbf{v}^\epsilon = \tilde{B}(\mathbf{M})\beta. \tag{5.79}$$

Following the line of the proof in Section 5.4.2, boundedness of the terms $-\Delta, \tilde{F}(\mathbf{M})$, $\tilde{B}(\mathbf{M})$ are straightforward, while the main term in $\tilde{G}(\hat{\alpha}+\lambda\epsilon\beta, \mathbf{M})$ producing the single square norm of $\mathbf{v}^\epsilon_{\mathbf{t}}$, after being tested by $\mathbf{v}^\epsilon_{\mathbf{t}}$ is

$$\begin{aligned}
\int_0^1 (\hat{\alpha}_1 + \lambda\epsilon\beta_1) \int_\Omega \mathbf{v}^\epsilon_{\mathbf{t}}(t)\cdot\mathbf{v}^\epsilon_{\mathbf{t}}(t)\, dx\, d\lambda &= \|\mathbf{v}^\epsilon_{\mathbf{t}}(t)\|^2_{L^2(\Omega,\mathbb{R}^3)} \left(\hat{\alpha}_1 + \frac{\epsilon\beta_1}{2}\right)\\
&\ge \|\mathbf{v}^\epsilon_{\mathbf{t}}(t)\|^2_{L^2(\Omega,\mathbb{R}^3)}(\hat{\alpha}_1^0 - \rho).
\end{aligned}$$

According to this, one gets, for all $\epsilon \in [0,\bar{\epsilon}]$,

$$\|\mathbf{v}^\epsilon\|_{\mathcal{V}} \le C|\beta| \|\tilde{B}(\mathbf{M})\|_{\mathbb{R}^2 \to L^2(0,T;L(\Omega,\mathbb{R}^3))} \le |\beta| C \tag{5.80}$$

with C depending only on $\mathbf{m}_0, \mathbf{h}, \hat{\alpha}^0, \rho$.

Since $\mathbf{b}^\epsilon \to 0$ pointwise and $\|\mathbf{b}^\epsilon\|_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))} \le C$ for all $\epsilon \in [0,\bar{\epsilon}]$, applying Lebesgue's Dominated Convergence Theorem yields convergence of $\|\mathbf{b}^\epsilon\|_{L^2(0,T;L^2(\Omega,\mathbb{R}^3))}$, thus of $\|\tilde{\mathbf{v}}^\epsilon\|_{\mathcal{V}}$, to zero. Fréchet differentiability of the forward operator in the reduced setting is therefore proved.

Conclusion and outlook

In this chapter, we outlined the mathematical model of MPI which led us to the LLG equation describing the magnetic particle interaction on a microscale level. For calibrating the MPI device it is necessary to compute the system function which mathematically can be interpreted as an inverse parameter identification problem for an initial value problem based on the LLG equation. To this end we deduced a detailed analysis of the forward model, i.e., the operator mapping the coefficients to the solution of the PDE as well as of underlying inverse problem. The inverse problem itself was investigated for an all-at-once and a reduced approach. The analysis includes representations of the adjoint operators and Fréchet derivatives. These results are necessary for a subsequent numerical computation of the system function in a robust manner, what will be subject of future research. Even beyond this, the analysis might be useful for the development of solution methods for any inverse problems that are connected to the LLG equation.

Concerning the model, the following directions are of our interest:

One might include further energy contributions to the effective magnetic field, e.g., the anisotropy, which is modeled as the additive term $(\mathbf{m} \cdot \mathbf{n})\mathbf{n}$ for a given anisotropy direction $\mathbf{n} \in \mathcal{S}^2$. We expect a change in the form of the adjoint equation in the reduced setting, while in the all-at-once one, the procedure to compute the adjoints stays untouched except that the anisotropy term will contribute to $\mathbf{f}^y$ in the expression (5.42).

We plan to extend to the case of time-dependent particle density. In this situation, we replace $c(x)$ by $c(x,t)$ in all formulas, and where the kernel K is differentiated in time (as a result of integration-by-parts), we need to differentiate c as well.

Chapter 6

Numerical reconstruction in MPI

In this chapter, we present a numerical study for the theory introduced in Chapter 5, including approximation of solutions to the LLG equation and reconstruction of the involved physical parameters for given measured MPI data.

The chapter is outlined as follows: Section 6.1 discusses the iterative Landweber and Landweber-Kaczmarz schemes for the reconstruction problem in all-at-once and reduced versions. This leads to an LLG solver relying on the all-at-once approach, which offers an efficient computation scheme for solving the nonlinear LLG equation and will be utilized in the reduced setting. Section 6.2 provides several synthetic examples, also physically meaningful simulations followed by comparisons between the two versions to visualize as well as analyze the proposed algorithm.

6.1 Algorithm

Let us first of all refer to Chapter 1, Section 1.1.3 for the definition of the Landweber and Landweber-Kaczmarz regularization methods. We now make explicit the steps comprised in each of these methods in the all-at-once version as well as its counterpart reduced version. For the all-at-once setting, the algorithm relies on Section 5.4.1, and for the reduced setting, it results from Section 5.4.2, both sections are in Chapter 5.

6.1.1 All-at-once Landweber

Start from $(\hat{\mathbf{m}}, \hat{\alpha}_1, \hat{\alpha}_2)_{j=0} = (\mathbf{m} - \mathbf{m}_0, \hat{\alpha}_1, \hat{\alpha}_2)_{j=0}$, run

S.1. Set argument to adjoint equations

Let $\hat{\mathbf{m}} := \hat{\mathbf{m}}_j, \hat{\alpha}_1 := \hat{\alpha}_{1\,j}, \hat{\alpha}_2 := \hat{\alpha}_{2\,j}$

Compute

$$\begin{aligned}\mathbf{w} = \hat{\alpha}_1 \hat{\mathbf{m}}_t - \Delta_N(\mathbf{m}_0 + \hat{\mathbf{m}}) - \hat{\alpha}_2(\mathbf{m}_0 + \hat{\mathbf{m}}) \times \hat{\mathbf{m}}_t \\ - |\nabla(\mathbf{m}_0 + \hat{\mathbf{m}})|^2(\mathbf{m}_0 + \hat{\mathbf{m}}) - \mathbf{h} + ((\mathbf{m}_0 + \hat{\mathbf{m}}) \cdot \mathbf{h})(\mathbf{m}_0 + \hat{\mathbf{m}})\end{aligned}$$

$$r(t) = \left(\int_0^T \int_\Omega \mathbf{K}_{k\ell}(x,t,\tau) \mathbf{m}_t(x,\tau)\, dx\, d\tau \right)_{k=1,\dots,K,\ell=1,\dots,L} - y^\delta(t).$$

Check: Stopping rule according to discrepancy principle.

S.2. Compute the adjoints

A. Compute $\mathbf{z} = \frac{\partial \mathbb{F}_0}{\partial \hat{\mathbf{m}}}(\hat{\mathbf{m}}, \hat{\alpha})^* \mathbf{w}$

A.1. Input: $\mathbf{w}$
Solve

$$\begin{aligned} -\Delta \widetilde{y}(t) + \widetilde{y}(t) &= \mathbf{w}(t) && \text{in } \Omega \\ \partial_\nu \widetilde{y} &= 0 && \text{on } \partial\Omega. \end{aligned}$$

Output: $\tilde{y}$

A.2. Input: $\tilde{y}$
Compute

$$\mathbf{y}(t) = I_2[\tilde{y}](t) = -\int_0^t (t-s)\tilde{y}(s)\, ds + \frac{t}{T}\int_0^T (T-s)\tilde{y}(s)\, ds.$$

Output: $\mathbf{y}$

A.3. Input: $\mathbf{y}$
Compute

$$\begin{aligned} \mathbf{f}^{\mathbf{y}} = -\hat{\alpha}_1 \mathbf{y}_t + (-\Delta_N \mathbf{y}) - \hat{\alpha}_2 \hat{\mathbf{m}}_t \times \mathbf{y} + \hat{\alpha}_2 \mathbf{y}_t \times (\mathbf{m}_0 + \hat{\mathbf{m}}) + \hat{\alpha}_2 \mathbf{y} \times \hat{\mathbf{m}}_t \\ - 2((\mathbf{m}_0 + \hat{\mathbf{m}}) \cdot \mathbf{y})\, (-\Delta_N(\mathbf{m}_0 + \hat{\mathbf{m}})) \\ + 2((\nabla(\mathbf{m}_0 + \hat{\mathbf{m}})^T (\nabla \mathbf{y}))\, (\mathbf{m}_0 + \hat{\mathbf{m}}) \\ + 2((\nabla(\mathbf{m}_0 + \hat{\mathbf{m}})^T (\nabla(\mathbf{m}_0 + \hat{\mathbf{m}})))\, \mathbf{y} - |\nabla(\mathbf{m}_0 + \hat{\mathbf{m}})|^2 \mathbf{y} \\ + ((\mathbf{m}_0 + \hat{\mathbf{m}}) \cdot \mathbf{h})\, \mathbf{y} + ((\mathbf{m}_0 + \hat{\mathbf{m}}) \cdot \mathbf{y})\, \mathbf{h} \\ \mathbf{g}_T^{\mathbf{y}} = \hat{\alpha}_1 \mathbf{y}(T) - \hat{\alpha}_2 \mathbf{y}(T) \times (\mathbf{m}_0 + \hat{\mathbf{m}}(T)). \end{aligned}$$

Output: $\mathbf{f}^{\mathbf{y}}, \mathbf{g}_T^{\mathbf{y}}$

A.4. Input: $\mathbf{f}^{\mathbf{y}}, \mathbf{g}_T^{\mathbf{y}}$
Solve

$$\begin{aligned} -\mathbf{v}_t - \Delta \mathbf{v} &= \mathbf{f}^{\mathbf{y}} && \text{in } (0,T) \times \Omega \\ \partial_\nu \mathbf{v} &= 0 && \text{on } (0,T) \times \partial\Omega \\ \mathbf{v}(T) &= \mathbf{g}_T^{\mathbf{y}} && \text{in } \Omega. \end{aligned}$$

Output: $\mathbf{v}$

A.5. Input: $\mathbf{v}$

Solve

$$\begin{aligned} \mathbf{z}_t - \Delta \mathbf{z} &= \mathbf{v} && \text{in } (0,T) \times \Omega \\ \partial_\nu \mathbf{z} &= 0 && \text{on } (0,T) \times \partial\Omega \\ \mathbf{z}(0) &= 0 && \text{in } \Omega. \end{aligned}$$

Output: $\mathbf{z}$

B. Compute $\mathbf{s} = \left(\frac{\partial \mathbb{F}_{k\ell}}{\partial \hat{\mathbf{m}}}(\hat{\mathbf{m}}, \hat{\alpha})_{k=1,\dots,K,\ell=1,\dots,L}\right)^* r$

B.1. Input: r from Step S.1.

Compute

$$\mathbf{f}^r(x,\tau) = -\int_0^T \sum_{k=1}^K \sum_{\ell=1}^L \frac{\partial}{\partial \tau} \mathbf{K}_{k\ell}(t,\tau,x) r_{k\ell}(t)\ dt$$
$$\mathbf{g}^r(x) = \int_0^T \sum_{k=1}^K \sum_{\ell=1}^L \mathbf{K}_{k\ell}(t,T,x) r_{k\ell}(t)\ dt$$

with

$$\mathbf{K}_{k\ell}(t,\tau,x) := -\mu_0 \widetilde{a}_\ell(t-\tau) c_k(x) \mathbf{p}_\ell^{\mathrm{R}}(x).$$

Output: $\mathbf{f}^r, \mathbf{g}^r$

B.2. Step A.4. with input: $\mathbf{f}^r, \mathbf{g}^r$

B.3. Step A.5. with output: $\mathbf{s}$

C. Compute $\beta_1 = \frac{\partial \mathbb{F}_0}{\partial \hat{\alpha}_1}(\hat{\mathbf{m}}, \hat{\alpha})^* \mathbf{w}$

C.1. Input: $\mathbf{y}$ from Step A.2.

Compute

$$\beta_1 = \int_0^T \int_\Omega \hat{\mathbf{m}}_t \cdot \mathbf{y}\, dx\, dt$$

Output: β_1

D. Compute $\beta_2 = \frac{\partial \mathbb{F}_0}{\partial \hat{\alpha}_2}(\hat{\mathbf{m}}, \hat{\alpha})^* \mathbf{w}$

D.1. Input: $\mathbf{y}$ from Step A.2.

Compute

$$\beta_2 = -\int_0^T \int_\Omega ((\mathbf{m}_0 + \hat{\mathbf{m}}) \times \hat{\mathbf{m}}_t) \cdot \mathbf{y}\, dx\, dt\,.$$

Output: β_2

S.3. Update $\hat{\mathbf{m}}, \hat{\alpha}_1, \hat{\alpha}_2$

$$\hat{\mathbf{m}}_{j+1} = \hat{\mathbf{m}}_j - \mu(\mathbf{z} + \mathbf{s}) \tag{6.1}$$
$$\hat{\alpha}_{1\,j+1} = \hat{\alpha}_{1\,j} - \mu\beta_1 \tag{6.2}$$
$$\hat{\alpha}_{2\,j+1} = \hat{\alpha}_{2\,j} - \mu\beta_2. \tag{6.3}$$

In the implementation, one can consider each of the vector fields as a three-dimensional matrix, i.e., Figure 6.1. Steps A.1., A.2., A.4., A.5. then operate on each time-space slice, meanwhile Step A.3. needs to be calculated among 3D-matrices.

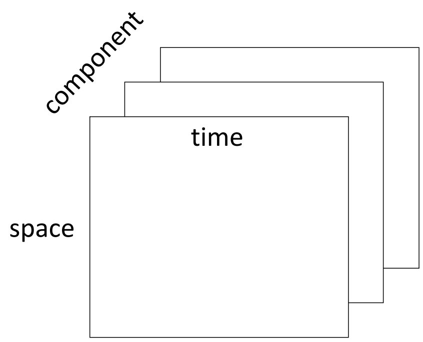

Figure 6.1: Matrix representation for a vector field in the all-at-once setting.

Remark 6.1.1. In case of having given parameters $\hat{\alpha}_1, \hat{\alpha}_2$, one can omit Steps B., C., D. and Steps (6.2)-(6.3) in S.3.to obtain a numerical solution to $\mathbf{m}$. This procedure can be seen as a LLG solver.

As far as a numerical solution of LLG is concerned, numerous investigations have been recently published. Bartels and Prohl in 2008 [8] considered a p-harmonic heat flow with $1 < p < \infty$ and established a numerical implementation strategy based on solving a nonlinear system per time step. These authors integrated a fixed-point method with stopping criterion (to handle the nonlinear system) into a lowest order conforming finite element method. The work of Alouges et. al. in 2008 [2] described a new implicit finite element scheme, which avoids solving nonlinear systems. This θ-scheme introduces the new term $\mathbf{v} := \mathbf{m}_t$ to form a linear equation in $\mathbf{v}$. At each time step n, $\theta \in [0,1]$ involves $\nabla(\mathbf{m}_n + \theta k \mathbf{v}_n)$ in the variation formula (hence implicit), where k is the time step size. After solving $\mathbf{v}_n$ by an implicit finite element scheme, $\mathbf{m}$ at the next time step $(n+1)$ is updated by $\mathbf{m}_{n+1} := \frac{\mathbf{m}_n + k\mathbf{v}_n}{|\mathbf{m}_n + k\mathbf{v}_n|}$. Inspired by the θ-scheme, Baňas et. al. in 2014 [6] studied the more general model including the magnetostriction effect instead of the magnetostatic simplification. The phenomenon is governed by a coupled problem of a LLG equation and a second time-dependent PDE representing the magnetostrictive contribution. The authors later dealt with the full Maxwell-LLG equation in [5] and proposed a fully decoupled scheme. In 2014, Alouges et. al. [3] upgraded the original first order θ-scheme with replacing the tangential update for $\mathbf{m}$ by a higher approximation via $\mathrm{Proj}_{\mathbf{m}^\perp}$ and $\mathbf{m}_{tt}$, creating a new (almost) order two θ-scheme. The algorithm initializes with differentiating with respect to time the LLG equation then ends up with a variation formula linear in $\mathbf{v}$. The approximation in space is still of order one (P^1 Lagrange finite elements), and the convergence of the scheme does not hold for higher order elements.

Independently of these existing works, our algorithm is attractive from a practical point of view since we also require to solve only linear PDEs per iteration step. Our contribution as well as the advantage of the method can be summarized as follows

- Unlike the mentioned novel schemes, that find $\mathbf{m}$ at each time point per iteration step, our scheme solves $\mathbf{m}$ at all time instances per iteration. The loop in our scheme is dedicated to the Landweber(-Kaczmarz) iteration, which improve the whole $\mathbf{m}([0,T]\times\Omega)$ gradually. This shall grant access to space-time adaptive discretization.

- In each step, only three separate and conventional linear PDEs, i.e. A.1., A.4. and A.5. are required to be solved. Our method, therefore, does not need to derive new theory for proving unique existence and convergence of finite element solutions, which are the main results in most of the current literature.

- Also with this reason, the suggested method is favorable in implementation since one can make use of existing standard finite difference or finite element codes.

- We also remark that our method is able to be upgraded to higher order through
  - higher order standard FD/FE for solving the conventional PDEs in A.1., A.4., A.5. (feasible).
  - higher order numerical approximation for the integrals and derivatives in A.2., A.3. (feasible).
  - higher convergence rate of the Landweber. This fact depends on how smooth the exact solution is (source condition) [62, Section 2.3]. And this turns out to be smoothness of $\mathbf{m}_0$ and $\mathbf{h}$, which is feasible if one proceeds similarly to the proof of regularity of $\mathbf{m}$ in the reduced setting (c.f., Chapter 5, formulas (5.65), (5.67)).

- Although not being presented here, our method can be extended to the case when the anisotropy $\pi(\mathbf{m})$ is involved in the effective field, under certain conditions on π. This is due to the fact that after calculating $\mathbb{F}'(\hat{\mathbf{m}},\hat{\alpha}_1,\hat{\alpha}_2)^*\mathbf{w}$, only the term $\pi'(\hat{\mathbf{m}})^*\mathbf{w}$ is added to $\mathbf{f}^{\mathbf{y}}$ in Step A.3., while all other steps remain unchanged.

- As the examples illustrated in this chapter are in simple one-dimensional domain, a finite difference discretization is well adapted. On the other hand, there is no reason preventing the proposed method to be implemented by finite elements to better suit complex geometries.

- Concerning memory requirement, in each of the Landweber iterations if using an implicit Euler time stepping scheme, the proposed method demands a storage for computing the cross product of two matrices of size nx×3 (number of space grids × number of components) or multiplication a matrix of size nx×nx (finite difference matrices or stiffness matrix) with a vector of nx elements. This yields similarity in the memory requirement with the other existing schemes.

However

- As a nature of a nonlinear inverse problem, our scheme is able to solve it just locally, i.e., a sufficiently good initial guess needs to be known.

- Concerning memory requirement, our algorithm demands more memory than the others as the whole $m([0,T]\times\Omega)$ needs to be allocated in RAM for each Landweber iteration.

6.1.2 Reduced Landweber

Start from $\hat{\alpha}_{j=0} = (\hat{\alpha}_1, \hat{\alpha}_2)_{j=0}$, run

S.1. Compute the state $\mathbf{m} := S(\hat{\alpha}_j)$ according to the LLG equation

S.2. Set argument to the adjoint equation
Compute

$$r(t) = \left(\int_0^T \int_\Omega \mathbf{K}_{k\ell}(x,t,\tau)\mathbf{m}_\tau(x,\tau)\,dx\,d\tau\right)_{k=1,\dots,K,\ell=1,\dots,L} - y^\delta(t).$$

Check: Stopping rule according to discrepancy principle.

S.3. Compute the adjoint state $\mathbf{q}^z = F'(\hat{\alpha})^* r$ according to

$$\begin{aligned}
&-\hat{\alpha}_1\mathbf{q}^z_t - \hat{\alpha}_2\mathbf{m}\times\mathbf{q}^z_t - 2\hat{\alpha}_2\mathbf{m}_t\times\mathbf{q}^z - \Delta\mathbf{q}^z \\
&\quad + 2\left((\nabla\mathbf{m})^\top\nabla\mathbf{m}\right)\mathbf{q}^z + 2\left((\nabla\mathbf{m})^\top\nabla\mathbf{q}^z\right)\mathbf{m} \\
&\quad + (-|\nabla\mathbf{m}|^2 + (\mathbf{m}\cdot\mathbf{h}))\mathbf{q}^z + (\mathbf{m}\cdot\mathbf{q}^z)(\mathbf{h}+2\Delta\mathbf{m}) = \tilde{K}r \quad \text{in } (0,T)\times\Omega && (6.4)\\
&\partial_\nu\mathbf{q}^z = 0 \quad \text{on } (0,T)\times\partial\Omega && (6.5)\\
&\hat{\alpha}_1\mathbf{q}^z(T) + \hat{\alpha}_2(\mathbf{m}\times\mathbf{q}^z)(T) = \tilde{K}_T r \quad \text{in } \Omega && (6.6)
\end{aligned}$$

with

$$(\tilde{K}r)(x,t) = \sum_{k=1}^{K}\sum_{\ell=1}^{L} -\mu_0 c_k(x)\mathbf{p}_\ell^R(x)\int_0^T \widetilde{a}_{\ell\,\tau}(\tau - t) r_{k\ell}(\tau)\,d\tau \quad t\in(0,T)$$
$$(\tilde{K}_T r)(x) = \sum_{k=1}^{K}\sum_{\ell=1}^{L} -\mu_0 c_k(x)\mathbf{p}_\ell^R(x)\int_0^T \widetilde{a}_{\ell}(\tau) r_{k\ell}(\tau)\,d\tau.$$

S.4. Update $\hat{\alpha}_1, \hat{\alpha}_2$

$$\hat{\alpha}_{1\,j+1} = \hat{\alpha}_{1\,j} - \mu\int_0^T\int_\Omega -\mathbf{m}_t\cdot\mathbf{q}^z\,dx\,dt \tag{6.7}$$
$$\hat{\alpha}_{2\,j+1} = \hat{\alpha}_{2\,j} - \mu\int_0^T\int_\Omega (\mathbf{m}\times\mathbf{m}_t)\cdot\mathbf{q}^z\,dx\,dt. \tag{6.8}$$

In contrast to the all-at-once setting, which solves the conventional PDEs on time-space slices, the PDE in the adjoint equation (Step S.3.) of the reduced one involves cross product and modulus, it, therefore, interacts between three components. Figure 6.2 is an example for assigning a vector field to a matrix when dealing with implementing the reduced version. One might reshape the matrix in Figure 6.2 into one vector of $nx\times 3\times nt$ elements.

For Step S.2. here, one can employ the LLG solver introduced in Remark 6.1.1. In order to search for $\mathbf{m}$, the program needs an initial guess, which can be chosen as the initial state $\mathbf{m}_0$. Then the computed $\mathbf{m}_j$ of the Landweber iteration j, besides being the input to the next steps S.3.-S.4. in the stream, on the other hand, plays the role of an initial guess for S.2. in the next iteration $(j+1)$.

The reduced setting is supposed to run more slowly than the all-at-once one, if using the same step size, as each of the Landweber iteration calls a LLG solver leading to another inner loop.

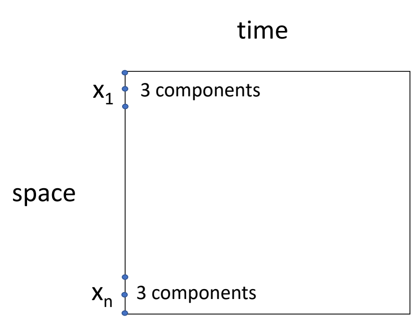

Figure 6.2: Matrix representation for a vector field in the reduced setting.

Besides the Landweber method, we consider also the Landweber-Kaczmarz method, which is especially attractive for problems with large datasets or high dimensional model operator; that our MPI problem with signals from several concentration samples and multiple receive coils in calibration is one typical example. In particular, Kaczmarz method does not need to finish operating the whole system of KL equations, instead it just successively sweeps through each of those equations in each iteration. Motivated by this idea, we propose two independent Kaczmarz schemes, namely, we either segment the time line into several subintervals or use one single data measured from one concentration sample and one receive coil in each iteration. The algorithms for those schemes are detailed in the following remarks.

Remark 6.1.2. Kaczmarz based on time segmenting

Start from $\hat{\alpha}_{j=0} = (\hat{\alpha}_1, \hat{\alpha}_2)_{j=0}$, run

S.1. Compute the state $\mathbf{m} := S(\hat{\alpha}_j)$ according to the LLG equation

S.2. Set argument to adjoint equation
Compute

$$r(t) = \left(\int_{t^j}^{t^{j+1}} \int_\Omega \mathbf{K}_{k\ell}(x,t,\tau)\mathbf{m}_\tau(x,\tau)\,dx\,d\tau \right)_{k=1,\ldots,K,\ell=1,\ldots,L} - y^\delta(t)\chi_{[t^j,t^{j+1}]}(t)$$

with $t^j = \left\lfloor \dfrac{j}{n} \right\rfloor, 0 = t^0 < \ldots < t^{n-1} = T$.

Check: Stopping rule according to discrepancy principle.

S.3. Compute the adjoint state $\mathbf{q}^z = F'(\hat{\alpha})^* r$ according to (6.4)-(6.6) with

$$(\tilde{K}r)(x,t) = \sum_{k=1}^K \sum_{\ell=1}^L -\mu_0 c_k(x)\mathbf{p}_\ell^R(x) \int_{t^j}^{t^{j+1}} \widetilde{a}_{\ell\,\tau}(\tau - t) r_{k\ell}(\tau)\,d\tau \qquad t \in (0,T)$$
$$(\tilde{K}_T r)(x) = \sum_{k=1}^K \sum_{\ell=1}^L -\mu_0 c_k(x)\mathbf{p}_\ell^R(x) \int_{t^j}^{t^{j+1}} \widetilde{a}_\ell(\tau) r_{k\ell}(\tau)\,d\tau.$$

S.4. Update $\hat{\alpha}_1, \hat{\alpha}_2$ as (6.7)-(6.8).

Remark 6.1.3. Kaczmarz based on data selection

Start from $\hat{\alpha}_{j=0} = (\hat{\alpha}_1, \hat{\alpha}_2)_{j=0}$, run

S.1. Compute the state $\mathbf{m} := S(\hat{\alpha}_j)$ according to the LLG equation

S.2. Set argument to adjoint equation
Compute

$$r(t) = \int_0^T \int_\Omega \mathbf{K}_{k\ell}(x,t,\tau)\mathbf{m}_\tau(x,\tau)\,dx\,d\tau - y^\delta_{k\ell}(t)$$

with k, ℓ satisfying $k\ell = \left\lfloor \frac{j}{KL} \right\rfloor$.

Check: Stopping rule according to discrepancy principle.

S.3. Compute the adjoint state $\mathbf{q}^z = F'(\hat{\alpha})^* r$ according to (6.4)-(6.6) with

$$(\tilde{K}r)(x,t) = -\mu_0 c_k(x)\mathbf{p}^R_\ell(x) \int_0^T \widetilde{a}_{\ell\tau}(\tau - t) r(\tau)\,d\tau \qquad t \in (0,T)$$

$$(\tilde{K}_T r)(x) = -\mu_0 c_k(x)\mathbf{p}^R_\ell(x) \int_0^T \widetilde{a}_\ell(\tau) r(\tau)\,d\tau.$$

S.4. Update $\hat{\alpha}_1, \hat{\alpha}_2$ as (6.7)-(6.8).

6.2 Numerical experiments

6.2.1 LLG solver

When using the algorithm in Remark 6.1.1 as an LLG solver, there is no noise involved in the process as the exact data ($y = 0$) is known. Hence, the algorithm should be run with as large number of iterations as possible in order to reach an acceptable accuracy for the reconstructed $\hat{\mathbf{m}}$.

In the following, we shall numerically test this algorithm using the finite difference method for Steps A.1., A.3.-A.5.. In particular, central difference quotients were employed to approximate time and space derivatives. Numerical integration in Step A.2. runs with the trapezoidal rule. For computing the respective L^2-norms, we use also the trapezoidal rule. For time discretization, the interval $[0, 0.2]$ is partitioned into 51 time steps, and for the space domain $[0, 2\pi]$, we impose a discretization of 101 grid points. The method in use is Landweber.

We now analyze the numerical performance of the LLG solver by means of three test cases specified in Table 6.1. Test 1 with the corresponding parameter set is indeed a true solution to the LLG equation. Test 2, however, does not completely yield the original LLG equation since in the LLG model, $\mathbf{m}_0$ is supposed to have constant length over Ω. For the same reason, initial data in Test 3 although fulfilling the requirement of constant length ($|\mathbf{m}_0(x)| = 2$), does not have a homogeneous Neumann boundary. Nevertheless, those test cases still reflect the the all-at-once model with unknown $\hat{\mathbf{m}}$, while $\mathbf{m}_0$ being considered as just an additive term, thus are still recognized as meaningful examples helping to increase diversity of the experiments.

Figures 6.3, 6.5 and 6.7 display the results in time and space of the reconstructed states together with comparisons to the true ones. In Figure 6.4, the initial error in $L^2(0,T,L^2(\Omega,\mathbb{R}^3))$-norm measuring the distance between $\mathbf{m}$ and $\mathbf{m}_{\text{exact}}$ from 40% declines to 0.4% after 3050 iterations. Considering the test cases 2 and 3 presented in Figures 6.6 and 6.8. Starting from 11% and 22% the errors drop to 0.2% and 1.1% after 5350 and 1850 iterations, respectively. In those tests, we create initial guesses for $\hat{\mathbf{m}}$, thus $\mathbf{m}$, by perturbing the exact ones by different amounts to closely inspect the convergence of the method. In practice, one can choose the initial guess for the LLG solver as $\mathbf{m}_0$, which means just $(0,0,0)$ in Test 1 and $(0,0,1)$ in the latter two.

Relying on monotonicity of the residual sequence, we implement an adaptive Landweber step size scheme in order to search for an appropriate one (Figures 6.4, 6.6, 6.8, left). In particular, in each iteration, a residual comparison with the previous step takes place. If the current residual shows a decrease, the current step size μ is accepted, otherwise it is bisected. The iterations are terminated after reaching a level in smallness of the step size, alternatively speaking, the residual is not able to get significantly smaller. One can stop the iterations earlier by checking smallness of step size together with residual tolerance. The runtime reports: 289 seconds, 512 seconds and 205 seconds respectively for 3 tests.

Table 6.1: Test cases and run parameters

Test	1	2	3
$\hat{\alpha}_1$	1	2	1
$\hat{\alpha}_2$	-1	0	0
$\mathbf{h}$	$\frac{2}{5}(0,3,4)$	$-(\cos(x),\cos(x),0)$	(0,0,0)
$\mathbf{m}_{\text{exact}}$	$\frac{1}{5}(0,3,4)$	$(\cos(x),\cos(x),e^t)$	$(\sin(x),\cos(x),e^t)$
$\mathbf{m}_0$	$\frac{1}{5}(0,3,4)$	$(\cos(x),\cos(x),1)$	$(\sin(x),\cos(x),1)$
$\hat{\mathbf{m}}_{exact}$	(0,0,0)	(0,0,e^t-1)	(0,0,e^t-1)
Initial guess $\hat{\mathbf{m}}$	$-5t(1,1,1)$	$-5t\cos(x)(1,1,1)$	$-\frac{\sin(30t)}{5}(1,1,1)$
Step size μ	150	75	300
# iterations	3050	5350	680

Table 6.2: Common physical parameters.

Parameter		Value	Unit
Magnetic permeability	μ_0	$4\pi \times 10^{-7}$	H m^{-1}
Sat. magnetization	Ms	474 000	J m^3T^{-1}
Gyromagnetic ratio	γ	$1.75{\times}10^{11}$	rad s^{-1}
Damping parameter	α_{D}	0.1	
Exchange constant	A	0	J m^{-1}
Field of view	Ω	[-0.006, 0.006]	m
Max observation time	T	$0.03{\times}10^{-3}$	s
External field strength	$\|\mathbf{h}\|$	10^{-5}	T

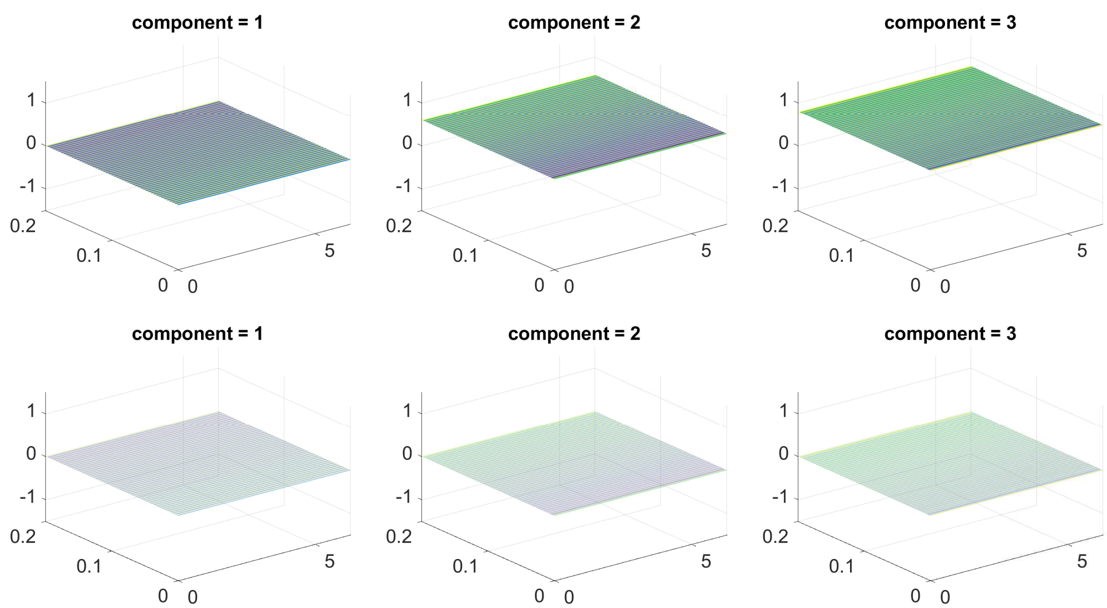

Figure 6.3: Test 1. Top: reconstructed $\mathbf{m}$. Bottom: $\mathbf{m} - \mathbf{m}_{\text{exact}}$. Left to right: each component.

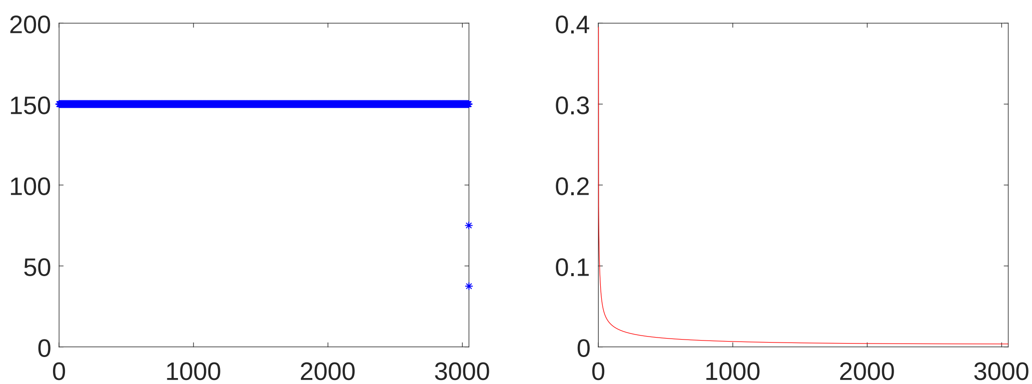

Figure 6.4: Test 1. Left: step size μ. Right: $\frac{\|\mathbf{m}_k - \mathbf{m}_{\text{ex}}\|}{\|\mathbf{m}_{\text{ex}}\|}$

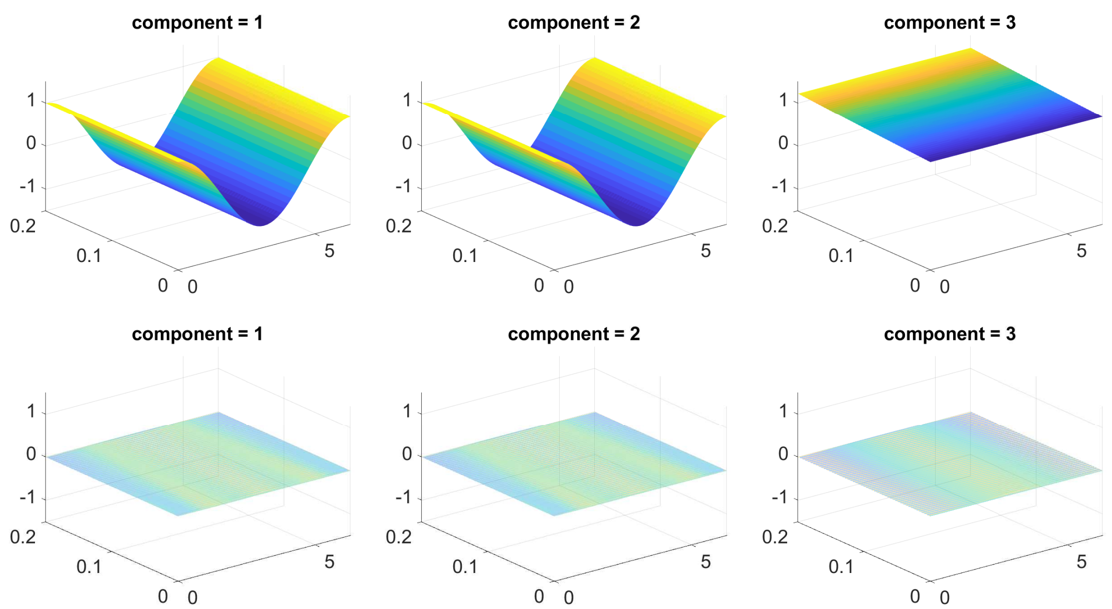

Figure 6.5: Test 2. Top: reconstructed $\mathbf{m}$. Bottom: $\mathbf{m} - \mathbf{m}_{\text{exact}}$. Left to right: each component.

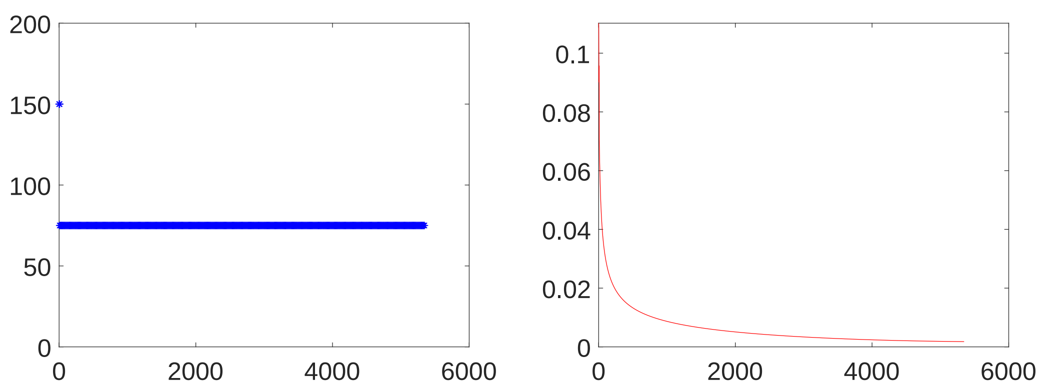

Figure 6.6: Test 2. Left: step size μ. Right: $\frac{\|\mathbf{m}_k - \mathbf{m}_{\text{ex}}\|}{\|\mathbf{m}_{\text{ex}}\|}$

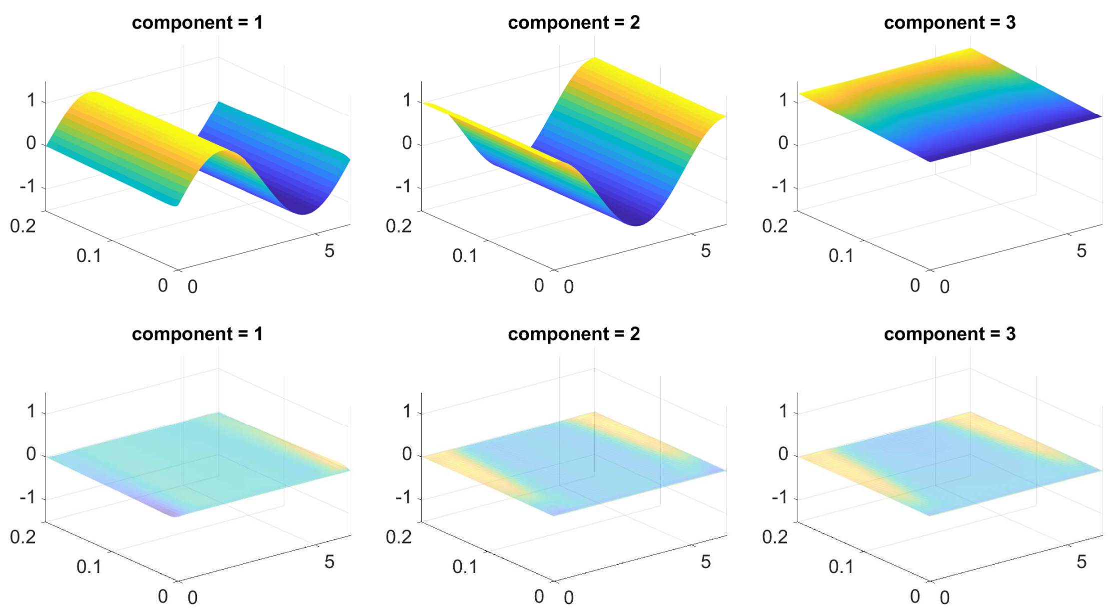

Figure 6.7: Test 3. Top: reconstructed $\mathbf{m}$. Bottom: $\mathbf{m} - \mathbf{m}_{\text{exact}}$. Left to right: each component.

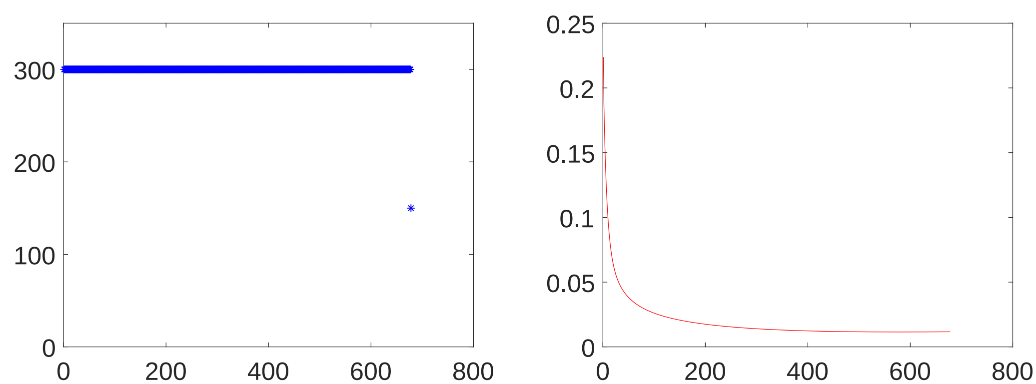

Figure 6.8: Test 3. Left: step size μ. Right: $\frac{\|\mathbf{m}_k - \mathbf{m}_{\text{ex}}\|}{\|\mathbf{m}_{\text{ex}}\|}$

6.2.2 LLG solver with physical parameters

In this section, we perform some simulations with real physical parameters. Table 6.2 gives an overview on common parameters, that can be found, e.g., in [71, 5]. The length of the magnetic moment vector $\mathbf{m}$ is specified by $m_\mathrm{S} = \mathrm{Ms}$, also $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are different by a factor of m_S. For better numerical computing, we shall do the following scaling for the LLG equation.

Let $\overline{\mathbf{m}} := \frac{\mathbf{m}}{m_\mathrm{S}}, \widetilde{\alpha}_1 := m_\mathrm{S}\widetilde{\alpha}_1$, i.e.,

$$\widetilde{\alpha}_1 := \frac{\gamma\alpha_\mathrm{D}}{1+\alpha_\mathrm{D}^2} > 0, \qquad \widetilde{\alpha}_2 := \frac{\gamma}{1+\alpha_\mathrm{D}^2} > 0$$

and let

$$\hat{\alpha}_1 = \frac{\widetilde{\alpha}_1}{\widetilde{\alpha}_1^2+\widetilde{\alpha}_2^2}, \qquad \hat{\alpha}_2 = \frac{\widetilde{\alpha}_2}{\widetilde{\alpha}_1^2+\widetilde{\alpha}_2^2}, \qquad \mathbf{h} = \mu_0 m_\mathrm{S}\mathbf{H}_\mathrm{ext},$$

we obtain the LLG equation for $\overline{\mathbf{m}}$ with $|\overline{\mathbf{m}}| = 1$

$$\begin{aligned} \hat{\alpha}_1\overline{\mathbf{m}}_t - \hat{\alpha}_2\overline{\mathbf{m}}\times\overline{\mathbf{m}}_t - 2Am_\mathrm{S}\Delta\overline{\mathbf{m}} &= 2Am_\mathrm{S}|\nabla\overline{\mathbf{m}}|^2\overline{\mathbf{m}} + \mathbf{h} - (\overline{\mathbf{m}}\cdot\mathbf{h})\overline{\mathbf{m}} && \text{in } [0,T]\times\Omega \\ \partial_\nu\overline{\mathbf{m}} &= 0 && \text{on } [0,T]\times\partial\Omega \\ \overline{\mathbf{m}}(t=0) = \overline{\mathbf{m}}_0,&\ |\overline{\mathbf{m}}_0| = 1 && \text{in } \Omega. \end{aligned}$$

In Figure 6.9, the left columns display three states of the applied magnetic field $\mathbf{h}$, namely, one static field in e_3-direction and two other time-dependent fields. Starting from an initial state (homogeneous in space), the magnetization vector $\overline{\mathbf{m}}$ (right columns) follows the trajectory of $\mathbf{h}$ with a delay known as the relaxation effect. The length of $\overline{\mathbf{m}}$ is not fully preserved, but it is obviously not far from the unit length. Figure 6.10 displays the progression of magnetization $\overline{\mathbf{m}}$ when the static field in e_3-direction is applied to a space-inhomogeneous initial state $\overline{\mathbf{m}}_0$ (top 6 plots) and another $\overline{\mathbf{m}}_0$ distributed randomly in space (bottom 6 plots).

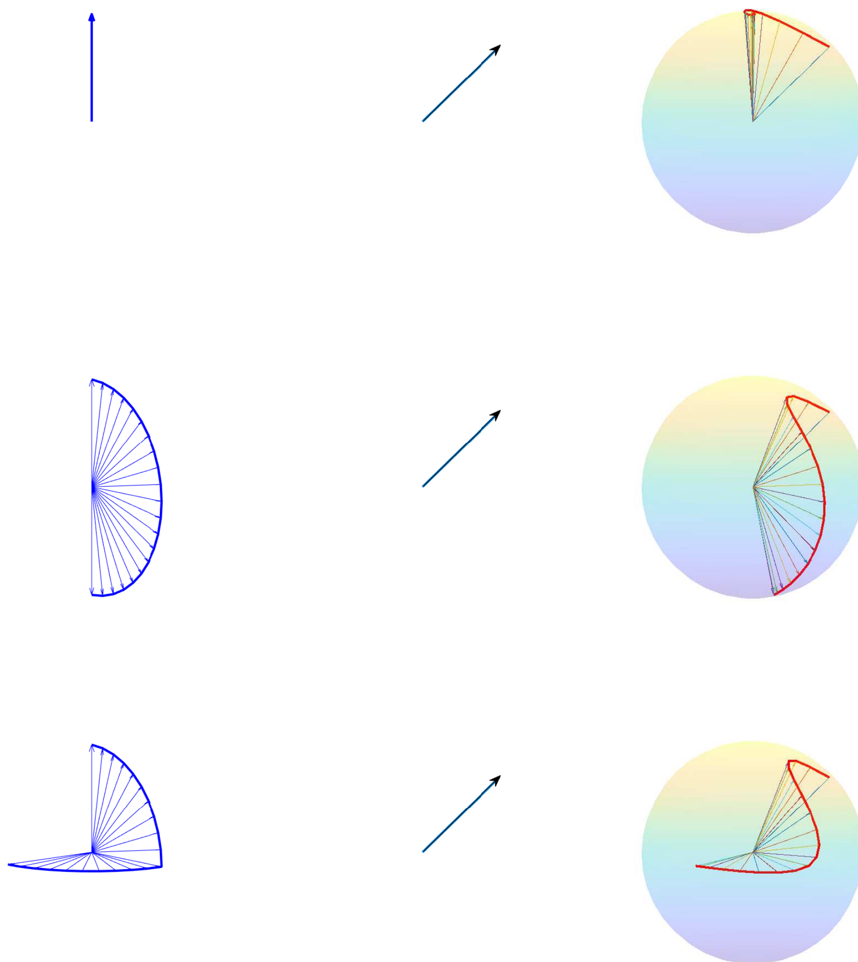

Figure 6.9: Left: applied field $\mathbf{h}$. Middle: initial state $\overline{\mathbf{m}}_0$. Right: trajectory of $\overline{\mathbf{m}}(t)$.

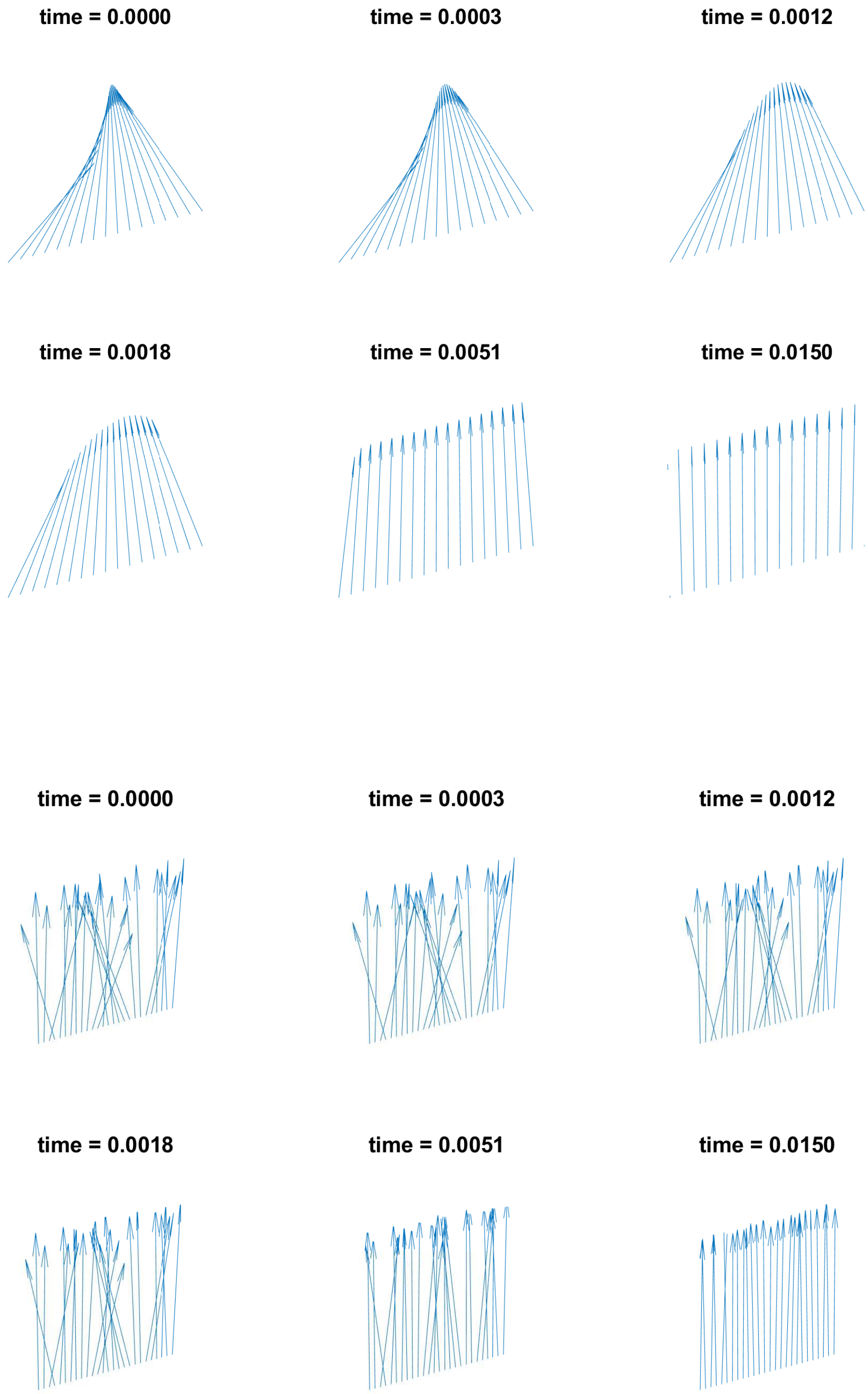

Figure 6.10: Magnetization $\overline{\mathbf{m}}$ at time different time instances.

6.2.3 Reconstruction in all-at-once and reduced settings

In this section, we compare the reconstruction outcome from all-at-once and reduced settings in case of exact measured data. Staying with the implementation method from Section 6.2.1, we carry out the Landweber approximation for Test 2 in Table 6.1. For the measurement process, we define the observation operator via: $\mu_0 = 1, \tilde{a}_l = 1, c_k = 1, \mathbf{p}_\ell^R = (1, 1, 1)$. And the initial guess for the state is $\mathbf{m}_{\text{init}} = \mathbf{m}_0$.

Figures 6.11 and 6.12 respectively present the reconstructed state in the all-at-once setting and in the reduced setting. The all-at-once Landweber ran with $350,000$ iterations and Landweber step size $\mu = 1$, while the reduced one ran with 250 iterations of step size $\mu = 1$. The reconstructed parameters $\hat{\alpha}_1, \hat{\alpha}_2$ are depicted in Figures 6.13, 6.14 (left), where both settings confirm the results at acceptable error level.

In the all-at-once setting, the Landweber approximation is applied to both state $\mathbf{m}$ and parameter $\hat{\alpha}$. In the reduced setting, only the parameter is approximated, while the state is solved exactly with the help of the parameter-to-state map. Hence, the reduced version require more Landweber iterations to reach an acceptable tolerance compared to the reduced one. Despite the lower number of Landweber iterations, the reduced setting, on the other hand, executes another amount of internal loops each time it calls the LLG solver. Figure 6.14 (right) views the number of internal loops in 250 reduced Landweber iterations. The sum of the internal loops is 11095.

With the same step size $\mu = 1$, the runtime reports: 79,000 seconds for 350,000 all-at-once iterations and 90,000 seconds for 250 reduced iterations. The reduced Landweber can be sped up by using a larger step size $\mu = 10$, which is feasible in the this setting; however, not feasible in the all-at-once one. Indeed, the step sizes in the two settings are not correlated since they are chosen subject to the criterion $\mu \in \big(0, \frac{1}{\|F'(x)\|^2}\big], \forall x \in \mathcal{B}_\rho(x^0)$, and according to the problem formulation, the forward operators, thus their derivatives, are different in each setting.

In Figure 6.15, the left plot displays the observation residual $\|\mathcal{K}\mathbf{m}_t - y\|_{L^2(0,T)}$ of each iteration in the reduced version. The middle and right plots display together the observation residual, the LLG residual and the ℓ^2-norm total residual in the all-at-once version. The LLG residual is measured in the $H^1(0, T, H^1(\Omega, \mathbb{R}^3))^*$-norm.

We now examine the reconstruction for Test 3, Table 6.1 in case of noisy data. We perturb the exact measured voltage with 10%, 5% and 3% random noise. Also, instead of initializing the algorithm at $\mathbf{m}_0$, we choose a perturbed version of it, namely, $\mathbf{m}_{\text{init}} = \mathbf{m}_0 - 0.1 \sin(20t)$. The stopping rule is according to the discrepancy principle. Table 6.3 respectively reports the number iterations (#it), the LLG residual (r_{llg}), the observation residual (r_{onb}) and the reconstruction error ($e_{\alpha_1}, e_{\alpha_2}$) in both settings. The LLG residuals in the reduced setting are typically smaller than that in the all-at-once one, since there the states are solved exactly.

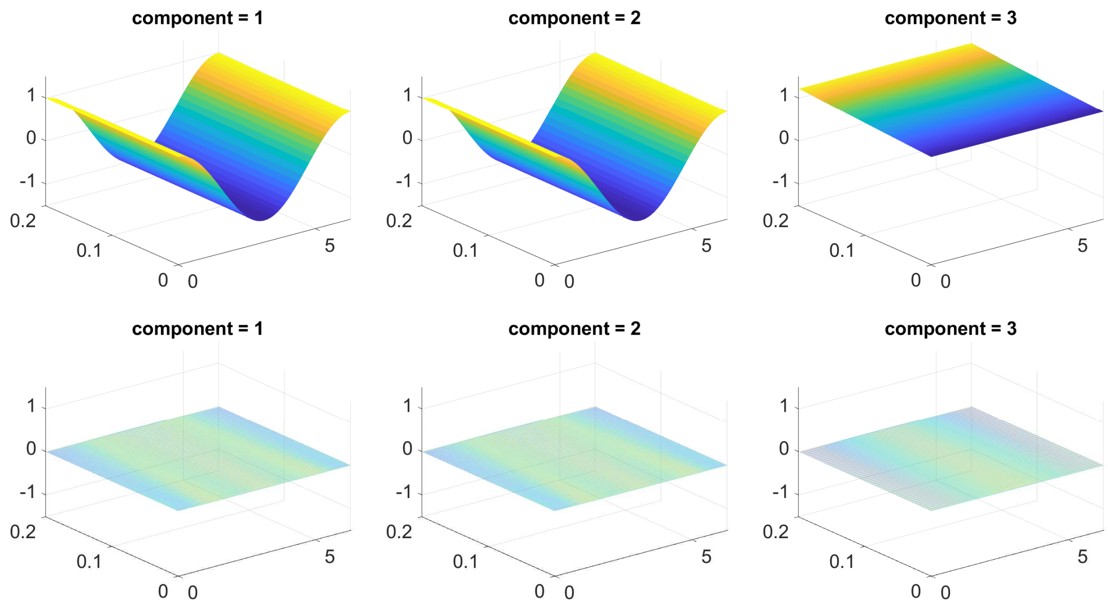

Figure 6.11: All-at-once setting. Top: reconstructed $\mathbf{m}$. Bottom: $\mathbf{m} - \mathbf{m}_{\text{exact}}$. Left to right: each component.

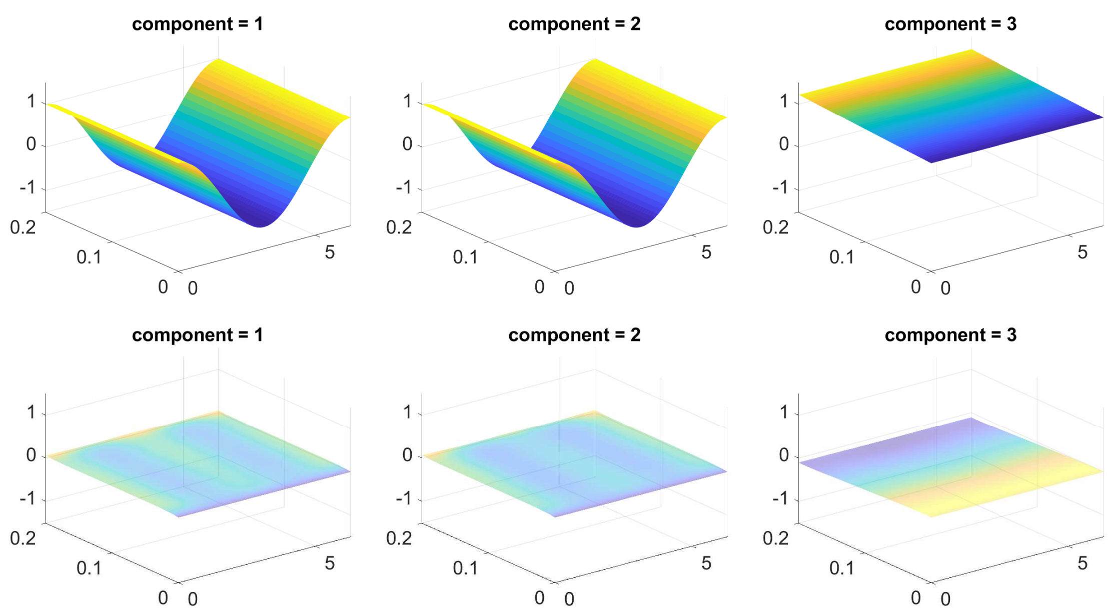

Figure 6.12: Reduced setting. Top: reconstructed $\mathbf{m}$. Bottom: $\mathbf{m} - \mathbf{m}_{\text{exact}}$. Left to right: each component.

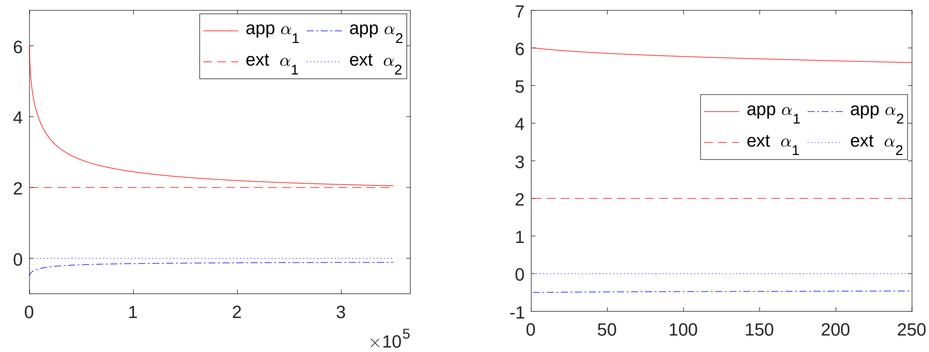

Figure 6.13: All-at-once setting. Left: reconstructed parameter. Right: zoom of first 250 iterations.

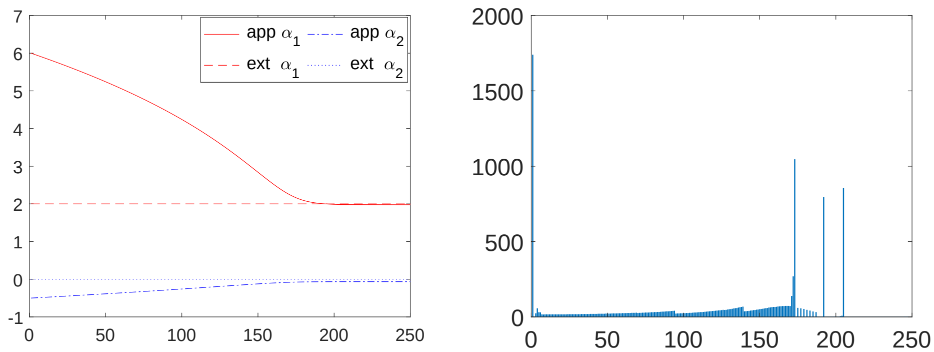

Figure 6.14: Reduced setting. Left: reconstructed parameter. Right: number of internal loops in each Landweber iteration.

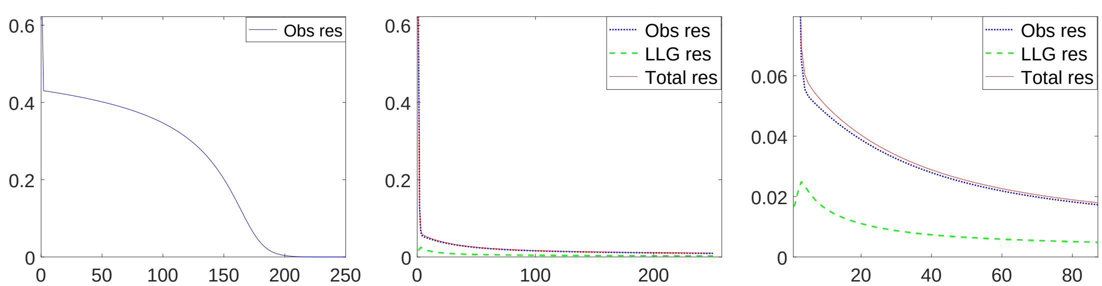

Figure 6.15: Residual. Left: reduced setting. Middle: all-at-once setting, first 250 iterations. Right: zoom of middle.

Table 6.3: Reconstruction with noisy data.

	AAO					RED				
δ	#it	r_{llg}	r_{obs}	e_{α_1}	e_{α_2}	#it	r_{llg}	r_{obs}	e_{α_1}	e_{α_2}
10%	259	0.0022	0.0619	0.292	0.090	49	3×10^{-6}	0.0703	0.030	0.034
5%	401	0.0011	0.0309	0.125	0.072	49	3×10^{-6}	0.0321	0.040	0.033
3%	564	0.0007	0.0186	0.040	0.062	49	3×10^{-6}	0.0200	0.044	0.033

Conclusion and outlook

In Chapter 5, it has been discussed that the map from the external magnetic field $\mathbf{h}$ to the magnetic moment vector $\mathbf{m}$ is essential to construct the particle distribution from the induced signals. This nonlinear map is described by the Landau-Lifshitz-Gilbert equation, which is realized to better yield the magnetization phenomenon since it takes in to account the relaxation effect. This chapter investigates the numerical approximation to the magnetic moment vector $\mathbf{m}$ as well as the reconstructed physical parameters α_1, α_2 through the LLG equation and measured signals. The numerical results show that our method is robust. Moreover, the computational schemes are practically preferable as all the steps involve solving only linear PDEs, also we provide multiple choices to the users: an all-at-once version and a reduced version. This work indicates a potential in helping to accelerate the concentration reconstruction algorithms in the next stage.

The next steps could be:

- Implement the proposed algorithm in higher dimensional field of view
- Apply the output system function (via the computed $\mathbf{m}$) to the final process: MPI reconstruction of the particle concentration.

Conclusion

The work in this thesis is primarily devoted to the analysis of inverse problems for parameter identification in time-dependent partial differential equations. Along with this, computational experiments supplement the analysis in order to show how the theory can be realized numerically. The key contributions of this research are:

- We parallelly formulated the inverse problem involving abstract evolution systems over a finite time line in two different settings: a classical reduced version and a new all-at-once version. The new approach avoids the construction of the parameter-to-state map, which is a conditional nonlinear map and in some cases unachievable. An avoidance of solving such nonlinear equations makes the all-at-once setting very efficient in practice by enabling treatment of a larger class of problems as well as facilitating and accelerating numerical computations. In addition, by parallelly working in two settings we achieved new results in the classical reduced version.

- The tangential cone condition and locally uniform boundedness of the derivative of the forward operator are two essential conditions, which shall grant access to the iterative regularization theory. We established these conditions in a rigorous way in the reduced setting as well as in its counterpart the all-at-once setting for some general classes of examples, also for targeted concrete practical examples.

- Eventually, we efficaciously applied the above theory to a very new promising application: Magnetic Particle Imaging.

In the connection of what has been obtained, we refer to the outlook in each of the chapters 2, 3, 5, 6 where we state several potential directions for future research.

Notation conventions

Physical quantities in MPI

$\mathbf{m}$	the magnetic moment vector
$\mathbf{H}_{\text{eff}}$	the effective magnetic field
$\mathbf{h}_{\text{ext}}$, $\mathbf{h}$	the external magnetic field
$\mathbf{B}$	the magnetic flux density
$\mathbf{E}$	the electric field
μ_0	the permeability in vacuum
M_S	the saturation magnetization
A	the exchange stiffness constant
α_D	the damping parameter
γ	the gyromagnetic constant
v	the voltage induced in the receive coil
$\mathbf{p}^R$	the coil sensitivity
$\widetilde{a}$	the transfer function, e.g, the band pass filter
c	the particle concentration
D	the particle core diameter
V_c	the particle core volume

Sets

Ω	a bounded, Lipschitz domain in $\mathbb{R}^d$
$\overline{\Omega}$	the closure of Ω
$\partial\Omega$	the boundary of Ω
$\mathcal{D}(F)$	the domain of a mapping F
$\mathcal{R}(F)$	the range of a mapping F

$\mathcal{N}(F)$	the null space of a mapping F
$X^\perp$	the orthogonal complement of a vector space X

Function spaces

$L^p(\Omega)$	the Lebesgue space of p-integrable functions f equipped by the norm $\|f\|_{L^p(\Omega)} = \left(\int_\Omega |f(x)|^p\,dx\right)^{\frac{1}{p}}$
$L^\infty(\Omega)$	the Lebesgue space of essentially bounded functions f equipped by the norm $\|f\|_{L^\infty(\Omega)} = \operatorname{ess\,sup}_{x\in\Omega} |f(x)|$
$W^{k,p}(\Omega)$	the Sobolev space of functions, whose weak derivatives up to k-th order belong to $L^p(\Omega)$, see p. 7
$W^{k,\infty}(\Omega)$	the Sobolev space of functions, whose weak derivatives up to k-th are essentially bounded, see p. 7
$H^k(\Omega)$	$W^{k,2}(\Omega)$
$H^k_0(\Omega)$	the Hilbert space $H^k(\Omega)$, whose trace on $\partial\Omega$ vanishes
$H^{-1}(\Omega)$	the dual space to $H^1_0(\Omega)$
V^*	the dual space to V
$L^p(0,T;V)$	space of measurable functions $u:[0,T]\to V$, see p. 8
$C(0,T;V)$	space of continuous functions $u:[0,T]\to V$, see p. 8
$W^{1,p,q}(0,T;V)$	the Sobolev-Bochner space, see p. 8
$L(V_1,V_2)$	the Banach space of linear continuous mappings $A: V_1\to V_2$ normed by $\|A\|_{L(V_1,V_2)} = \sup_{\|v\|_{V_1}\le 1}\|Av\|_{V_2}$

Operators

$\dot{u}$	the first order time derivative of u
$\ddot{u}$	the second order time derivative of u
Δ	the spatial Laplace operator
Δ_N	the spatial Laplace operator incorporating Neumann boundary condition (see p. 117): $\langle -\Delta_N u, v\rangle_{H^1(\Omega)^*,H^1(\Omega)} = (\nabla u,\nabla v)_{L^2(\Omega)} \quad \forall u,v\in H^1(\Omega)$
∇	the spatial gradient operator
$F^\star$	the Hilbert space adjoint of F
F^*	the Banach space adjoint of F

F'	the Gâteaux or Fréchet derivative of F
$\\|\cdot\\|_X$	a norm in Banach space X
$\|\cdot\|_X$	a seminorm in Banach space X
$\|\cdot\|$	the modulus in $\mathbb{R}^d$
$(\cdot,\cdot)$	the inner product on Hilbert spaces
$\langle\cdot,\cdot\rangle_{V,V^*}$	the paring between dual spaces V, V^*
$\times$	the cross product
$\hookrightarrow$	a continuous embedding
id	the identity operator
$a \succeq b$	$a \geq b$ with strict inequality if $b = 0$
$\vec{a} \cdot \vec{b}$	the Euclidean inner product between vectors $\vec{a}, \vec{b}$
$A : B$	the Frobenius inner product between matrices A, B
$\nabla \cdot \vec{v}$	the divergence of vector field $\vec{v}$
$\lfloor r \rfloor$	the largest integer lower or equal to r

Constants

C	a generic positive constant
$C_{X,Y}, C^{\Omega}_{X,Y}$	norm of the continuous embedding $X(\Omega) \hookrightarrow Y(\Omega)$, see p. 7
C_{PF}	the constant in the Poincaré-Friedrichs inequality: $\forall u \in W^{1,p}_0(\Omega), 1 \leq p \leq \infty : \\|u\\|_{L^p(\Omega)} \leq C_{PF}\\|\nabla u\\|_{L^p(\Omega)}$
p^*	$p^* = \frac{p}{p-1}$, the dual index of $p \in [1, \infty]$

Notation for regularization

$\mathcal{B}^X_\rho(x)$	the closed ball of radius ρ around x in X
$F^\dagger$	the Moore-Penrose generalized inverse of operator F
$x^\dagger$	the minimal-norm solution
(R_α, α)	a regularization method
y^δ	a noisy version of y
x_0	an initial guess
k_*	a stopping index
c_{tc}, c_{tcc}	the constant in the tangential cone condition.

Index

Bibliography

[1] S. M. Allen and J. W. Cahn, *Ground state structures in ordered binary alloys with second neighbor interactions*, Acta Metallurgica, 20 (1972), pp. 423–433.

[2] Alouges and François, *A new finite element scheme for Landau–Lifchitz equations*, Discrete Contin. Dyn. Syst. Ser. S, 1 (2008), pp. 187–196.

[3] F. Alouges, E. Kritsikis, J. Steiner, and J.-C. Toussaint, *A convergent and precise finite element scheme for Landau—Lifschitz—Gilbert equation*, Numerische Mathematik, 128 (2014), pp. 407–430.

[4] A. B. Bakushinsky and M. Y. Kokurin, *Iterative Methods for Approximate Solution of Inverse Problems*, vol. 577 of Mathematics and Its Applications, Springer, Dordrecht, 2004.

[5] L. Baňas, M. Page, and D. Praetorius, *A convergent linear finite element scheme for the Maxwell-Landau-Lifshitz-Gilbert equations*, Electronic Transactions on Numerical Analysis, 44 (2015), pp. 250—-270.

[6] L. Baňas, M. Page, D. Praetorius, and J. Rochat, *A decoupled and unconditionally convergent linear FEM integrator for the Landau-Lifshitz-Gilbert equation with magnetostriction*, IMA Journal of Numerical Analysis, 34 (2014), pp. 1361–1385.

[7] H. T. Banks and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*, Birkhauser, Boston, Basel, Berlin, 1989.

[8] S. Bartels and A. Prohl, *Convergence of an implicit finite element method for the Landau-Lifshitz-Gilbert equation*, SIAM J. Numer. Anal., 44 (2006), pp. 1405–1419.

[9] ——, *Convergence of an implicit, constraint preserving finite element discretization of p-harmonic heat flow into spheres*, Numerische Mathematik, 109 (2008), pp. 489—-507.

[10] A. Battermann and M. Heinkenschloss, *Preconditioners for Karush-Kuhn-Tucker matrices arising in the optimal control of distributed systems*, in Control and Estimation of Distributed Parameter Systems: International Conference in Vorau, Austria, July 14-20, 1996, W. Desch, F. Kappel, and K. Kunisch, eds., Birkhäuser Basel, Basel, 1998, pp. 15–32.

[11] A. Battermann and E. Sachs, *Block preconditioner for KKT systems in PDE-governed optimal control problems*, in Workshop on Fast Solutions of Discretized Optimization Problems, R. Hoppe, K.-H. Hoffmann, and V. Schulz, eds., Birkhäuser, 2001, pp. 1–18.

[12] J. Baumeister, B. Kaltenbacher, and A. Leitão, *On Levenberg-Marquardt-Kaczmarz iterative methods for solving systems of nonlinear ill-posed equations*, Inverse Problems and Imaging, 4 (2010), pp. 335—-350.

[13] R. Boiger, J. Hasenauer, S. Hross, and B. Kaltenbacher, *Integration based profile likelihood calculation for PDE constrained parameter estimation problems*, Inverse Problems, 32 (2016). Art. ID 125009.

[14] L. Borcea, *Electrical impedance tomography*, Inverse Problems, 18 (2002), pp. R99–R136.

[15] L. Bronsard and B. Stoth, *The Ginzburg-Landau equations of superconductivity and the one-phase Stefan problem*, Ann. Inst. Henri Poincaré, 15, No. 3 (1998), pp. 371–397.

[16] F. Bruckner, D. Suess, M. Feischl, T. Führer, P. Goldenits, M. Page, D. Praetorius, and M. Ruggeri, *Multiscale modeling in micromagnetics: Existence of solutions and numerical integration*, Mathematical Models and Methods in Applied Sciences, 24 (2014), pp. 2627–2662.

[17] ——, *Multiscale modeling in micromagnetics: Existence of solutions and numerical integration*, Mathematical Models and Methods in Applied Sciences, 24 (2014), pp. 2627–2662.

[18] M. Burger and W. Mühlhuber, *Iterative regularization of parameter identification problems by sequential quadratic programming methods*, Inverse Problems, 18 (2002), pp. 943–969.

[19] ——, *Numerical approximation of an SQP-type method for parameter identification*, SIAM J. Numer. Anal., 40 (2002), pp. 1775–1797.

[20] E. Casas, C. Ryll, and F. Tröltzsch, *Sparse optimal control of the Schlögl and Fitzhugh-Nagumo systems*, Computational Methods in Applied Mathematics, 13 (2013), pp. 415–442.

[21] ——, *Second order and stability analysis for optimal sparse control of the Fitzhugh-Nagumo equation*, SIAM J. Control Optim., 53 (2015), pp. 2168–2202.

[22] ——, *Optimal control of a class of reaction diffusion equations*, Computational Optimization and Applications, 70 (2018), pp. 677–707.

[23] G. Chavent and K. Kunisch, *On weakly nonlinear inverse problems*, SIAM J. Appl. Math., 56 (1996), pp. 542–572.

[24] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, Springer New York, 2013.

[25] L. R. Croft, P. W. Goodwill, and S. M. Conolly, *Relaxation in x-space magnetic particle imaging*, IEEE transactions on medical imaging, 31 (2012), pp. 2335–2342.

[26] B. D. Cullity and C. D. Graham, *Introduction to Magnetic Materials*, John Wiley & Sons, 2011.

[27] W. Demtroeder, *Experimentalphysik 2*, Springer Berlin Heidelberg, 2013.

[28] S. S. Dragomir, *Some Gronwall Type Inequalities and Applications*, Victoria University of Technology, Melbourne City MC, 2002.

[29] F. Dunker and T. Hohage, *On parameter identification in stochastic differential equations by penalized maximum likelihood*, Inverse Problems, 30 (2014), p. 095001 (20 pages). arXiv:1404.0651 [stat.CO].

[30] T. Dunst, M. Klein, A. Prohl, and A. Schäfer, *Optimal control in evolutionary micromagnetism*, IMA Journal of Numerical Analysis, 35 (2015), pp. 1342–1380.

[31] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht/Boston/London, 1996.

[32] L. C. Evans, *Partial Differential Equations*, Graduate Studies in Mathematics 19. AMS, Providence, RI, 1998.

[33] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Springer, USA, 2006.

[34] J. Flemming, *Theory and examples of variational regularization with non-metric fitting functionals*, J. Inv. Ill-Posed Problems, 18 (2010), pp. 677—-699.

[35] A. S. Fokas and G. A. Kastis, *Mathematical Methods in PET and SPECT Imaging*, Springer New York, New York, NY, 2015, pp. 903–936.

[36] J. Francu, *Monotone operators: A survey directed to applications to differential equations*, Aplikace Matematiky, 35 (1990), pp. 257–301.

[37] T. L. Gilbert, *A phenomenological theory of damping in ferromagnetic materials*, IEEE Transactions on Magnetics, 40 (2004), pp. 3443–3449.

[38] B. H. Gilding and R. Kersner, *Travelling Waves in Nonlinear Diffusion-Convection Reaction*, Springer Basel AG, Switzerland, 2004.

[39] B. Gleich and J. Weizenecker, *Tomographic imaging using the nonlinear response of magnetic particles*, Nature, 435 (2005), pp. 1214–1217.

[40] P. Grisvard, *Elliptic problems in nonsmooth domains*, Pitman Advanced Pub. Program Boston, 1985.

[41] C. W. Groetsch, *Inverse Problems in the Mathematical Sciences*, Vieweg + Teubner Verlag, 1993.

[42] B. Guo and M.-C. Hong, *The Landau-Lifshitz equation of the ferromagnetic spin chain and harmonic maps*, Calc. Var., 1 (1993), pp. 311–334.

[43] B. Guo and M.-C. Hong, *The Landau-Lifshitz equation of the ferromagnetic spin chain and harmonic maps*, Calculus of Variations and Partial Differential Equations, 1 (1993), pp. 311–334.

[44] E. Haber and U. M. Ascher, *Preconditioned all-at-once methods for large, sparse parameter estimation problems*, Inverse Problems, 17 (2001), p. 1847.

[45] M. Haltmeier, R. Kowar, A. Leitao, and O. Scherzer, *Kaczmarz methods for regularizing nonlinear ill-posed equations II: Applications*, Inverse Problems and Imaging, 1 (2007), pp. 507–523.

[46] M. Haltmeier, A. Leitao, and O. Scherzer, *Kaczmarz methods for regularizing nonlinear ill-posed equations I: Convergence analysis*, Inverse Problems and Imaging, 1 (2007), pp. 289–298.

[47] M. Hanke, *A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems*, Inverse Problems, 13 (1997), pp. 79–95.

[48] M. Hanke, *Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems*, Num. Funct. Anal. Optim, 18 (1997), pp. 971–933.

[49] M. Hanke, A. Neubauer, and O. Scherzer, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems*, Numer. Math., 72 (1995), pp. 21–37.

[50] D. N. Hào, N. V. Duc, and N. V. Thang, *Backward semi-linear parabolic equations with time-dependent coeffcients and locally Lipschitz source*, Inverse Problems, 34, 5 (2018). Art. ID 055010.

[51] W. Hinshaw and A. Lent, *An introduction to NMR imaging: From the Bloch equation to the imaging equation*, Proc. IEEE, 71 (1983), pp. 338–350.

[52] B. Hofmann, *On the degree of ill-posedness for nonlinear problems*, J. Inverse Ill-Posed Prob., 2 (1994), pp. 61–76.

[53] B. Hofmann and M. Yamamoto, *On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems*, Applicable Analysis, 89 (2010), pp. 1705—1727.

[54] T. Hohage and F. Weidling, *Verification of a variational source condition for acoustic inverse medium scattering problems*, Inverse Problems, 31 (2015). Art. ID 075006.

[55] S. Hubmer, E. Sherina, A. Neubauer, and O. Scherzer, *Lamé parameter estimation from static displacement field measurements in the framework of nonlinear inverse problems*, SIAM Journal on Imaging Sciences, 11 (2018), pp. 1268–1293.

[56] V. Isakov, *Inverse Problems for Partial Differential Equations*, Applied Mathematical Sciences, Springer, USA, 2006.

[57] B. Kaltenbacher, *Regularization based on all-at-once formulations for inverse problems*, SIAM Journal of Numerical Analysis, 54 (2016), pp. 2594–2618.

[58] B. Kaltenbacher, *All-at-once versus reduced iterative methods for time dependent inverse problems*, Inverse Problems, 33 (2017). Art. ID 064002.

[59] B. Kaltenbacher, A. Kirchner, and B. Vexler, *Goal oriented adaptivity in the IRGNM for parameter identification in PDEs II: all-at once formulations*, Inverse Problems, 30 (2014). 045002.

[60] B. Kaltenbacher and A. Klassen, *On convergence and convergence rates for Ivanov and Morozov regularization and application to some parameter identification problems in elliptic PDEs*, Inverse Problems, 34 (2018).

[61] B. Kaltenbacher, A. Neubauer, and O. Scherzer, *Iterative Regularization Methods for Nonlinear Problems*, de Gruyter, Berlin, New York, 2008. Radon Series on Computational and Applied Mathematics.

[62] B. Kaltenbacher, A. Neubauer, and O. Scherzer, *Iterative Regularization Methods for Nonlinear Ill-posed Problems*, Radon Series on Computational and Applied Mathematics, de Gruyter, Berlin, 2008.

[63] B. Kaltenbacher, A. Neubauer, and O. Scherzer, *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*, De Gruyter, 2008.

[64] B. Kaltenbacher, T. T. N. Nguyen, and O. Scherzer, *The tangential cone condition for some coefficient identification model problems in parabolic PDEs*, Springer volume on "Time-dependent Problems in Imaging and Parameter Identification", (to appear).

[65] B. Kaltenbacher, T. T. N. Nguyen, A. Wald, and T. Schuster, *Parameter identification for the Landau-Lifshitz-Gilbert equation in magnetic particle imaging*, Springer volume on "Time-dependent Problems in Imaging and Parameter Identification", (to appear).

[66] B. Kaltenbacher and M. L. Previatti de Souza, *Convergence and adaptive discretization of the IRGNM Tikhonov and the IRGNM Ivanov method under a tangential cone condition in Banach space*, Numerische Mathematik, (2018).

[67] B. Kaltenbacher, F. Schöpfer, and T. Schuster, *Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems*, Inverse Problems, 25 (2009), p. 065003.

[68] S. Kindermann, *Convergence of the gradient method for ill-posed problems*, Inverse Problems and Imaging, 11 (2017), pp. 703–720. arXiv:1606.00274 [math.NA].

[69] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, Springer New York Dordrecht Heidelberg London, 2011.

[70] A. Kirsch and A. Rieder, *Inverse problems for abstract evolution equations with applications in electrodynamics and elasticity*, Inverse Problems, 32 (2016), p. 085001.

[71] T. Kluth, *Mathematical models for magnetic particle imaging*, Inverse Problems, 34 (2018), p. 083001.

[72] T. Kluth and P. Maass, *Model uncertainty in magnetic particle imaging: Nonlinear problem formulation and model-based sparse reconstruction*, International Journal on Magnetic Particle Imaging, 3 (2017).

[73] T. Knopp and T. M. Buzug, *Magnetic Particle Imaging: an Introduction to Imaging Principles and Scanner Instrumentation*, Springer Berlin Heidelberg, 2012.

[74] T. Knopp, N. Gdaniec, and M. Möddel, *Magnetic particle imaging: from proof of principle to preclinical applications*, Physics in Medicine & Biology, 62 (2017), p. R124.

[75] R. Kowar and O. Scherzer, *Convergence analysis of a Landweber—Kaczmarz method for solving nonlinear ill-posed problems*, Ill-Posed and Inverse Problems ed V G Romanov et al, 23 (2002), pp. 69—-90.

[76] M. Kruzík and A. Prohl, *Recent Developments in the Modeling, Analysis, and Numerics of Ferromagnetism*, SIAM Review, 48 (2006), pp. 439–483.

[77] P. Kügler, *Online parameter identification in time-dependent differential equations as a non-linear inverse problem*, European Journal of Applied Mathematics, 19.5 (2008), pp. 479—-506.

[78] P. Kügler, *Online parameter identification without Ricatti-type equations in a class of time-dependent partial differential equations: an extended state approach with potential to partial observations*, Inverse Problems, 26 (2010). Art. ID 035004.

[79] K. Kunisch and E. W. Sachs, *Reduced SQP methods for parameter identification problems*, SIAM Journal on Numerical Analysis, 29 (1992), pp. 1793–1820.

[80] F. Kupfer and E. Sachs, *Numerical solution of a nonlinear parabolic control problem by a reduced SQP method*, Computational Optimization and Applications, 1 (1992), pp. 113–135.

[81] O. A. Ladyzhenskaya, V. Solonnikov, and N. N. Ural'tseva, *Linear and Quasilinear Equations of Parabolic Type*, Izd. Nauka, Moscow, 1967. (Engl. Transl.: AMS, Providence, 1968.).

[82] O. A. Ladyzhenskaya and N. N. Ural'tseva, *Linear and Quasilinear Equations of Elliptic Type*, Izd. Nauka, Moscow, 1964. (Engl. Transl.: Acad. Press, New York, 1968.).

[83] L. Landau and E. Lifshitz, *3 - On the theory of the dispersion of magnetic permeability in ferromagnetic bodies Reprinted from Physikalische Zeitschrift der Sowjetunion 8, Part 2, 153, 1935.*, in Perspectives in Theoretical Physics, L. PITAEVSKI, ed., Pergamon, Amsterdam, 1992, pp. 51–65.

[84] L. Landweber, *An iteration formula for Fredholm integral equations of the first kind*, American Journal of Mathematics, 73 (1951), pp. 615–624.

[85] A. Lechleiter and A. Rieder, *Newton regularizations for impedance tomography: convergence by local injectivity*, Inverse Problems, 24 (2008), pp. 065009, 18.

[86] F. Leibfritz and E. W. Sachs, *Inexact SQP interior point methods and large scale optimal control problems*, SIAM Journal on Control and Optimization, 38 (1999), pp. 272–293.

[87] T. Malthus, *An Essay on the Principles of Population*, J. Johnson, London, 1798.

[88] T. März and A. Weinmann, *Model-based reconstruction for magnetic particle imaging in 2D and 3D*, Inverse Problems & Imaging, 10 (2016), pp. 1087 – 1110.

[89] J. Nagumo, S. Yoshizawa, and Arinomoto, *Bistable transmission lines*, IEEE Transactions On Circuit Theory, 12 (1965), pp. 400–412.

[90] F. Natterer, *The Mathematics of Computerized Tomography*, Vieweg+Teubner Verlag, 1986.

[91] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, SIAM, Philadelphia, 2001.

[92] A. A. Nepomnyashchy, *Coarsening versus pattern formation*, Elsevier, Comptes Rendus Physique, 16 (2015), pp. 1–14.

[93] T. T. N. Nguyen, *Landweber-Kaczmarz for parameter identification in time-dependent inverse problems: All-at-once versus reduced version*, Inverse Problems, 35 (2019). Art. ID. 035009.

[94] C. E. Orozco and O. N. Ghattas, *A reduced SAND method for optimal design of non-linear structures*, International Journal for Numerical Methods in Engineering, 40 (1997), pp. 2759–2774.

[95] C. V. Pao, *Nonlinear Parabolic and Elliptic Equations*, Plenum Press, New York and London, 1992.

[96] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer, Verlag New York, 1983.

[97] B. Pedretscher, B. Kaltenbacher, and O. Bluder, *Parameter Identification in Stochastic Differential Equations to Model the Degradation of Metal Films*, PAMM · Proc. Appl. Math. Mech, 17 (2017), pp. 775–776.

[98] D. B. Reeves and J. B. Weaver, *Approaches for modeling magnetic nanoparticle dynamics*, Critical Reviews in Biomedical Engineering, 42 (2014).

[99] A. Rieder, *On the regularization of nonlinear ill-posed problems via inexact Newton iterations*, Inverse Problems, 15 (1999), pp. 309–327.

[100] T. Roubíček, *Nonlinear Partial Differential Equations with Applications*, International Series of Numerical Mathematics, Springer Basel, 2013.

[101] O. Scherzer, *Convergence criteria of iterative methods based on Landweber iteration for nonlinear problems*, J. Math. Anal. Appl., 194 (1995), pp. 911–933.

[102] J. Schöberl and W. Zulehner, *Symmetric indefinite preconditioners for saddle point problems with applications to PDE-constrained optimization problems*, SIAM Journal on Matrix Analysis and Applications, 29 (2007), pp. 752–773.

[103] F. Schöpfer, A. K. Louis, and T. Schuster, *Nonlinear iterative methods for linear ill-posed problems in Banach spaces*, Inverse Problems, 22 (2006), pp. 311–329.

[104] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. Kazimierski, *Regularization Methods in Banach Spaces*, de Gruyter, Berlin, New York, 2012. Radon Series on Computational and Applied Mathematics.

[105] A. R. Shenoy, M. Heinkenschloss, and E. M. Cliff, *Airfoil design by an all-at-once method*, International Journal for Computational Fluid Mechanics, 11 (1998), pp. 3–25.

[106] L. Shepp and Y. Vardi, *Maximum likelihood reconstruction for emission tomography*, IEEE Trans. Med. Imag., 1 (1982), pp. 113–122.

[107] S. Ta'asan, *"one shot" methods for optimal control of distributed parameter systems: I finite dimensional control*, tech. rep., Institute for Computer Applications in Science and Engineering : NASA Langley Research Center, 1991.

[108] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill Posed Problems*, V. H. Winston, Washington, 1977.

[109] F. Tröltzsch, *Optimal Control of Partial Differential Equations Theory, Methods and Applications*, Graduate Studies in Mathematics, American Mathematical Society, 2010.

[110] N. H. Tuan and D. D. Trong, *On a backward parabolic problem with local Lipschitz source*, Journal of Mathematical Analysis and Applications, 414 (2014), pp. 678–692.

[111] T. van Leeuwen and F. J. Herrmann, *A penalty method for PDE-constrained optimization in inverse problems*, Inverse Problems, 32 (2016), p. 015007.

[112] A. Wald and T. Schuster, *Sequential subspace optimization for nonlinear inverse problems*, Journal of Inverse and Ill-posed Problems, 25 (2016), pp. 99–117.

[113] ——, *Tomographic terahertz imaging using sequential subspace optimization*, New Trends in Parameter Identification, Springer Series Trends in Mathematics, (2018).