DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

A Parameter Identification Problem for a Nonlinear Parabolic Differential Equation

R. A. Ressel

Preprint 91

Edited by

AG Numerik/Optimierung
Fachbereich 12 - Mathematik und Informatik
Philipps-Universität Marburg
Hans-Meerwein-Str.
35032 Marburg

DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen”

A Parameter Identification Problem for a Nonlinear Parabolic Differential Equation

R. A. Ressel

Preprint 91

A parameter identification problem for a nonlinear parabolic differential equation

Rudolf A. Ressel*

January 20, 2011

The first part of this preprint proposes an analytical approach for a parameter identification problem involving a nonlinear parabolic PDE. The PDE is a reaction-diffusion equation modeling the gene expression of the embryogenesis of Drosophila melanogaster. The problem is formulated in an evolution equation setting, suitable solution and parameter spaces are introduced for that purpose. The continuous differentiability of the parameter-to-state operator is proved by a modified implicit function theorem. Continuity properties of the derivative of the parameter-to-state operator are elaborated. The adjoint of this derivative operator is formulated.
The second part concerns the applicability of Tikhonov regularization for the (non-Banach) setting.

1 Introduction

Over the course of the last three decades genetic research has produced a plenitude of data on gene expression in model systems such as Drosophila. It is therefore perfectly justifiable to say that genetic research has reached a stage where the application of mathematical methods permits extracting structural information from the data. A common object of study is the Drosophila fly since it represents a rather accessible way to experimentally investigate the effect and the mutual interaction of certain genes in the development of a model organism.
Despite all the compiled results and research techniques invented so far, conducting experiments in-vitro still poses an expensive challenge. It is hence desirable to probe

[1]Center for Industrial Mathematics / Fachbereich 3, University of Bremen, Postfach 33 04 40, 28334 Bremen, Germany, {ressel }@math.uni-bremen.de;
R. A. Ressel was supported by Deutsche Forschungsgemeinschaft, grant number Ma 1657/18-1

the possibilities of deducing critical parameters of interest in the animial's metabolism from the measured expression of genes at certain times of the lifecycle. The premise we take in our study is to consider gene product concentrations the state variables of an organism. The mutual interactions of genes we choose to reflect mathematically by the entirety of genes influencing the synthesis rate of each particular gene (as in the model proposed by [RS95]).
Mathematically these assumptions lead to an operator equation of the kind

$$\mathscr{D}(p) = y,$$

where p represents the sought-after control parameters and y are the concentrations of genes/gene products as state variables. The connection between p and y, the operator $\mathscr{D}$, will be given by the solution operator for some parabolic differential equation determining the evolution of the biological system.
In reality, our data is contaminated with noise, i.e. we have to work with y_{data} instead of y, with a noise level $\|y - y_{data}\| \leq \delta$. Another major impediment in identifying p from the data is the nonlinearity and ill-posedness of the parameter-to-state operator $\mathscr{D}$. Hence regularization techniques have to be employed. The reader unfamiliar with regularization may consult standard references such as [Lou89] or [EHN96].
To account for the noise in the data, Tikhonov regularization allows to reformulate our inverse problem as finding the minimizer of the functional

$$\|\mathscr{D}(p) - g\|^2 + J(p).$$

By the choice of the penalty term one has the liberty to enforce certain characterisics in the solution. One rather popular way is to assume the solution exhibits sparsity. The biology of the underlying problem in fact justifies this hypothesis: the action taking place is localized and the mutual influence of all genes on the synthesis of one particular genes is limited. Only few genes interfere with one particular gene. The pioneering work of [DDD04] provides a powerful numerical strategy to approximate the minimizer of the above Tikhonov functional for the definition $J(p) = \alpha \|p\|_{w,q}^q$, $1 \leq q \leq 2$, where the norm denotes a weighted Besov norm on a coefficient representiation of p with respect to some orthonormal basis. From among the generalizations of [DDD04] we will follow the course of the works [BBLM07, BLM09], since these treatments closely meet the pecularities of our problem of identifying the parameters in a nonlinear parabolic PDE. Similar iterative approaches for this minimizing task can be found in the publications [SMB$^+$08, Kaz11], a somewhat different iterative regularization procedure for nonlinear problems in Banach spaces is treated in [HK10].
The minimizing task as described above illustrates the problems that are being discussed in this paper: First of all one has to make an apt choice for the parameter and state spaces, taking into account the requirements of the forward problem (solvability of the PDE, analytical behavior of $\mathscr{D}$) and of the regularization procedure. Another major issue is to pick the type of penalty term J which best fits the nature of the problem. As a matter of course one wants to see the theory in action, ie an example needs to be programmed.

Some of the statements appearing in this work have already been published in the joint publication [DFM+11], albeit mostly without proofs and details. The purpose of this preprint is to supply the theoretical underpinning of the application presented in [DFM+11] and to serve as a progress report concerning the theory of the joint DFG SPP project.
We proceed as follows:

1. We will first be concerned with the construction of the parameter and solution spaces. This will enable us to check that our choice of such spaces makes the elliptic part of our differential operator continuously differentiable with respect to both the argument function g and the parameters D and λ. The reader less interested in this part may skip the next subsections, take the definitions of equation 5 and the preceding lines and start with subsection "PDE Theory".

2. Then follows a closer inspection of the nonlinear righthand side of our PDE: we first prove continuity for our setting and then the differentiability. Furthermore, we find some local kind of Hölder continuity for the nonlinear righthand side operator.

3. Then unique existence of a solution is presented, employing the results of [Grö89] and [HDR09, ACFP07, Grö92]. The theory of these references then gives more insight into the choice of the exponents for the parameter space.

4. By means of the implicit function theorem the preceding parts yield differentiability of $\mathcal{D}$. Further analysis of the derivative operator shows this derivative is even locally Hölder continuous for our choice of the parameter space.
   Next, the adjoint operator of the derivative is deduced as well as the action of the derivative of the discrepancy functional $\|\mathscr{D}(p) - g\|$.

5. The part on regularization is a brief outline of the algorithm we apply to our problem. We state (our adaption of) the gradient conditional gradient method and its numerical implementation as a soft-shrinkage procedure

2 Reaction-Diffusion equation, analysis of the forward problem

2.1 The classical initial-boundary value problem

The assumed causes for changing the concentration of gene products over time are direct regulation of the synthesis of one gene by the concentration of other genes; further causes are diffusive processes of gene products through the admissible domain and decay, i.e., consumption, of the respective gene products. The synthesis requires some regulating function in a manner that reflects saturation in the signal response. So the differential equation resulting from these assumptions must include a diffusion term accounting for the transport of gene products, with some spatially and temporally varying diffusion parameter D. Furthermore we have a decay term we choose to model linearly with some decay rate parameter λ. And finally and most interestingly, we have a synthesis term with a maximal synthesis rate R multiplied with some signal response (or regulation-expression) function. This signal response function takes as the input arguments the concentrations of all the gene products present and ranges from 0 to 1.

Our particular choice of this signal response function, the justification of which is given in the fundamental paper [MSR91], is of the following kind: $\Phi(Wg)$, i.e., the array of gene products g is multiplied with some parametric interaction matrix W, and each entry of this matrix-vector product is then the input into some sigmoidal function Φ. One notices that negative entries in W correspond to an inhibiting influence of one gene product on the other and positive ones represent a promoting effect. Precisely the identification of W, even before the other parameters D, λ and R, motivates the mathematical investigation of this real-world problem.

Taking some closed Lipschitz-bounded domain $U \subset \mathbb{R}^n$, $n = 2, 3$, such that $U = \overline{int(U)}$, as the physical domain of the metabolism, we denote by the scalar functions $g_i, i = 1, \ldots, d$ the gene product concentrations on U over the time interval $[0, T]$.

Then the partial differential equation governing the biochemical evolution considered reads:

$$\begin{aligned}
&\frac{\partial g_i}{\partial t} - \operatorname{div}(D_i(x,t)\operatorname{grad} g_i(x,t)) + \lambda_i(x,t) \cdot g_i(x,t) \\
&\qquad\qquad = R_i(x,t)\Phi_i((W(x,t)g(x,t))_i), \qquad (x,t) \in U_T = U \times (0,T] \\
&\frac{\partial g_i}{\partial \nu} = 0 \quad (x,t) \in \partial U \times [0,T] \\
&g(0) = g_0 \; on \; U \times \{0\}
\end{aligned} \tag{1}$$

where $i = 1, \ldots, d$, and the function $\Phi_i : \mathbb{R} \to \mathbb{R}$; $\quad \Phi_i(y) = \frac{1}{2}\left(\frac{y}{\sqrt{y^2+1}} + 1\right)$. This model PDE marks the starting point of the entire investigation. With the above formulation we have so far only made the translation from the physical model into somewhat loose mathematical parlance. In what mathematical way we understand this PDE has not yet been clearified. Therefore, in order to inspect our model PDE in a more rigorous fashion, we need to develop a functional analytic framework in which we can reformulate this PDE. The next section is devoted to this end.

2.2 Function spaces, Operators

In a first step we will motivate our choice of the parameter space. Then we will introduce the state space for the solutions of our PDE. We will not be utilizing the usual Hilbert space W_1^2 setting but rather exploit state of the art results on maximal parabolic regularity (see [HDR09]). This allows, at least partly, for subsets of reflexive spaces as admissible parameter sets (see 2.2.3). The introductory part 2.2.2 only presents basic tools such as multiplication spaces, spaces of Bochner integrable functions, and some types of derivatives (Frechet, distributiuonal, and metric space derivative). The reader acquainted with these ideas can skip this part.

2.2.1 The parameter space

The particular choice of the parameters must guarantee the existence and uniqueness of solutions of our PDE in some appropriate solution space. Hence, when choosing certain subspaces of Bochner integrable functions (ie generalized Sobolev spaces) as the setting for solutions (see subsection 2.2.3), the most straightforward and common choice for the parameters might be subsets of L_∞ spaces. However our theory does not stop at this point, since the application of generalized gradient methods involves handling the dual space of our parameter space. The L_∞-norm topology would then require dealing with the dual of some L_∞ product space. So whenever possible we aspire to to work with "better" topologies. Our workaround is to start with pointwise ae (almost everywhere) bounded sets of measurable functions, ie bounded subsets of L_∞ spaces (bounds are indicated in eqn. 2 right below). The underlying physical problem warrants these assumptions. The topology we then work with is induced by L_p norms. The global pointwise bounds are denoted by

$$0 < C_{\mathcal{P},1} \leq D, \lambda \leq C_{\mathcal{P},2}, \quad 0 \leq R \leq C_{\mathcal{P},2} \quad \|W\|_\infty \leq C_{\mathcal{P},2}. \tag{2}$$

The parameter space for D is defined as

$$\mathcal{P}_D = \{D \in L_{p_D}(U \times [0,T], \mathbb{R}^d) : 0 < C_{\mathcal{P},1} \leq D \leq C_{\mathcal{P},2}\}.$$

Accordingly, the spaces $\mathcal{P}_\lambda, \mathcal{P}_R, \mathcal{P}_W$ for λ, R, W are defined with the exponents of integrability $2 < p_\lambda, p_R, p_W$. For D we put $p_D = \infty$ right away. The other parameters, p_λ, p_R, p_W will be specified later to meet further analytical goals, eg the differentiability of the nonlinearity of the PDE.
The space of admissible parameter tuplets we denote by

$$\mathcal{P} = \mathcal{P}_D \times \mathcal{P}_\lambda \times \mathcal{P}_R \times \mathcal{P}_W$$

and endow this space with the product norm.
With this topology, $\mathcal{P}$ is not an open set. This poses some minor difficulties when one tries to differentiate maps on the parameter space. For this reason $\mathcal{P}$ can only be regarded as a metric space with a linear structure and translation invariant metric. At certain instances in our study this will require some elaboration on how the usual vector space concepts translate for metric spaces.

2.2.2 Differentiation and Integration

One important concept that will impact the choice of the topology of the parameter and solution spaces is the notion of multiplicator spaces from general integration theory. Our theory would certainly work without this short introduction. Nonetheless, since it motivates the ideas behind our construction the reader is given this short presentation.
Let $Z \subset \mathbb{R}$ be a closed and bounded domain. We make Z a finite measure space by using Lebesgue σ-algebra and the Lebesgue measure η. So all measurability will henceforth have to be understood as Lebesgue measurability (even if many of the statements below hold for more general measure spaces).
Some of the structural statements in this treatment of superposition operators are given for the scalar-valued case $d = 1$ only, simply to avoid indices and increase legibility. The reader will easily be able to see that the (1-D) claims also hold for the vector-valued case, which will be used when applying the general theorems to particular functions.
We start with a simple generalization of the well-known duality statement

$$(L_p(Z))' \cong L_{p'}(Z),$$

where $1/p + 1/p' = 1$.
Consider the space $(1 < q < p < \infty)$

$$\mathbb{M}_{p,q} = \{f \text{ measurable} : \forall g \in L_p : \ fg \in L_q, \ g \mapsto \int fg \text{ linear, continuous on } L_p\},$$

with the norm

$$\|f\|_{\mathbb{M}} = \sup_{\|g\|_{L_q}=1} (\|fg\|_{L_q}).$$

By theorem 47 in the appendix we have the isometric isomorphism

$$\mathbb{M}_{p,q} \cong L_r(Z), \quad r = \frac{pq}{p-q}. \tag{3}$$

This space is generally called a multiplicator space. For more details, in particular the treatment of $p < q$, consult [AZ90, pp. 91]. We merely remind the reader of this simple consequence of Hölder's theorem, because this concept pervades the rest of this publication, whether choices of integrability exponents are concerned or long chains of norm estimates are presented.

The other important concept involving integration is Bochner integrablity. We just give the basic definitions and assure the reader that "all the interesting statements" for scalar valued integrable functions theory hold also for Bochner integrable functions. For Bochner integrability further details may be found in [Sho97, Chap. III.1],[DU77, Chap. I-III] or [Yos80, Chap. V].

Definition 1. A function $f : Z \to Y$ is called simple, if there are finitely many Lebesgue measurable sets B_i of finite measure, such that $f(t) = y_i$ on each B_i and f vanishes outside $\bigcup_i B_i$. We define the Bochner integral for such functions as

$$\int_X f(t)dt \doteq \sum_{i=1}^n \eta(B_i)y_i.$$

Definition 2. A function $f : Z \to Y$ is called (Bochner-) measurable, if there is a sequence of simple funtions (f_j), such that

$$f_j(t) \to f(t), \text{ almost everywhere on Z}$$

If additionally

$$\lim_{j\to\infty} \int_X \|f_j(t) - f(t)\| \, dt = 0, \tag{4}$$

then we call f (Bochner-) integrable. The (Bochner-) integral we define as

$$\int_Z f(t)dt \doteq \lim_{j\to\infty} \int_X f_j(t)dt.$$

By standard theorems this is well-defined and the space of such functions is a Banach space, denoted by $L(Z, Y)$. Analogous to the scalar valued case, one defines the space $L_q(Z, Y)$ as the set of all Bochner integrable functions f such that $\int_Z \|f(t)\|_Y^q \, dt < \infty$. Convergence theorems (Egoroff's theorem, dominated convergence theorem) hold also for Bochner integrable functions. We remark that for $q \in [1, \infty)$, $Y \subset \mathbb{R}$ and $Z = L_q(U, \mathbb{R}^d)$ we have the isometric isometry

$$L_q(Y, L_q(U, \mathbb{R}^d)) \sim L_q(Y \times U, \mathbb{R}^d),$$

which will permit us to identify the two spaces occasionally without explicit mentioning.

Now we address different notions of differentiation: the Frechet derivative, derivatives on metric spaces and distributional derivatives.

Definition 3. Let X, Y be arbitrary Banach spaces and $U \subset X$ an open subset. A function $f : U \to Y$ is called (Frechet-)differentiable at some $x \in U$, if there is some $A(x) \in \mathcal{L}(X, Y)$, such that we have

$$\lim_{t\to 0} \sup_{\|e\|=1} \left\| \frac{f(x+te) - f(x)}{t} - A(x)e \right\| = 0.$$

In case the limit exists, $A(x)$ is called the derivative at x and we denote $Df(x) = A(x)$.

In our case we are only concerned with functions on an interval of the real line, so to ease notation, we fix X from now on to $[0,T]$, and differentiability is to be understood on the interior of the interval.
We introduce the notation $C^1(X,Y)$ for the space of differentiable functions from X to Y. Likewise for the repeated application of the definition, $C^m(X,Y)$ denote the space of m-times differentiable functions functions.
Since we will deal with rather peculiar admissible sets of parameters, we will introduce the notion of derivatives in metric spaces for the sake of mathematical rigor. However, the metric space $\mathcal{P}$ is a subset of a normed space, so the formulas and theorems down the road will look just like the respective ones for normed spaces. The theory presented relies on [Lan94]. The interested reader is referred to this source for more details. $(\mathcal{X},\rho_{\mathcal{X}})$, $(\mathcal{Y},\rho_{\mathcal{Y}})$ shall denote metric spaces, $B(x_0)$ a neighborhood in $\mathcal{X}$.

Definition 4 (Def. 2.1,[Lan94])**.** Two maps $f,g : B(x_0) \subset \mathcal{X} \to \mathcal{Y}$ are called tangent in x_0, if for all $\epsilon > 0$ we can find some $\delta > 0$ such that

$$\rho_{\mathcal{Y}}(f(x),g(x))) \leq \epsilon\rho_{\mathcal{X}}(x,x_0) \quad \text{for } x \in B(x_0),\ \rho_{\mathcal{X}}(x,x_0) < \delta.$$

The notation is $f \sim_{x_0} g$.

We remark that $\sim_{x_0}$ is an equivalence relation ([Lan94, Bem 2.1]).
The notion of differentiability we shall use for maps on our parameter space deviates from the original definition of [Lan94] (see [Lan94, Def. 2.2]). This serves to keep things close to the familiar vector space setting.
Assume $\mathcal{X}$ and $\mathcal{Y}$ to be metric spaces with a linear structure (eg convex subsets of vector spaces) and the respective metrics being translation invariant. Then we make the

Definition 5. A map $f : B(x_0) \to \mathcal{Y}$ is called differentiable, if there exists a Lipschitz continuous map D tangent to f, ie $D \sim_{x_0} f$, and D is affine and $D(x_0) = f(x_0)$. The set of all such maps is denoted by $f'(x_0)$.

We remark that by requiring the derivative maps to be affine and $D(x_0) = f(x_0)$, we ensure $f'(x_0)$ is a singleton. Otherwise uniqueness cannot be ensured (compare [Lan94, Def. 2.2]).
Another important concept are distributional derivatives. A definition and standard findings for distributions can be found in [GGZ74, Kap. III] or in [Fol99, Ch. 9.1]. Let $\mathscr{D}(U)$ denote the set of test functions $C_c^\infty(0,T)$ with the usual topology of convergence on compact subsets of K (see [Fol99, p. 282]). Let Y be another Banach space. Then the space of all Y valued continuous functionals (with respect to the topology of $\mathscr{D}(U)$) is called the space of Y valued distributions $\mathscr{D}'(U)$.
The Y valued distribution $\mathscr{T}$ defined by

$$\mathscr{T}(\rho) = -\int_0^T g(t)\rho(t)dt$$

for some function $g \in L_s(0, T, Y)$, is called the distributional derivative of g and shall be denoted by g'.
From now on we let $Z = [0, T]$, some $T > 0$. Now we are able to define the solution space for our PDE

Theorem 6. Let V be a Banach space that can be densely and continuously embedded into the dual V'. We define the space $\mathscr{W} = \{u \in L_s(0, T; V) : u' \in L_{s'}(0, T; V')\}$ with norm $\|u\|^2 = \|u\|^2_{L_{s'}(0,T;V)} + \|u'\|^2_{L_s(0,T;V')}$, $s \geq 2$. Then $\mathscr{W}_s$ is a Banach space and we have a continuous embedding $\mathscr{W}_s \hookrightarrow C([0, T], (V, V')_{(1/s^*,s)})$.

Notice that our definition is a particular case of the space $W^1_s(0, T, V') \cap L_s(0, T, V)$, compare eg [ACFP07, ch.1]. References to proofs for the claims can be found right there. We introduce the notation $G = (V, V')_{(1/s^*,s)}$.

To get closer to the final definition of our solution space, we define for $dim(U) = n$ the following function space:

$$V_q = W^1_q(U, \mathbb{R}^d) \; q \in (n, n + \epsilon).$$

The exponent q and the number $\epsilon > 0$ will be specified later on.
With this partiular choice we will collect some well-known facts. Recall first that the interpolation space G is a Besov space: $G = B^m_{q,q}(U, \mathbb{R}^d)$, where $m = 1 - 2/q$, see eg [Tri95, 2.4.2 Rem 2 b), 4.3.1 Thm 2]. The extension of this statement to our type of domain (Lipschitz domain) follows analogously to the result presented in [HDR09, Thm 3.4], as outlined in the remark [HDR09, Rem 3.6].
By [Ada78, 7.37 Thm.], we have the continuous embedding $G \hookrightarrow C(\overline{U}, \mathbb{R}^d)$. So we find

$$\mathscr{W}_s \hookrightarrow C([0, T], C(\overline{U}, \mathbb{R}^d)).$$

With the embedding of $\mathscr{W}_s$ in theorem 6 and the embedding $G \hookrightarrow C(\overline{U}, \mathbb{R}^d)$ we are able to formulate a more general regularity result for our space $\mathcal{V}$ concerning the existence of higher moments.

Theorem 7. We can continuously embed $I_M : \mathscr{W} \hookrightarrow L_r(0, T; L_r(U, \mathbb{R}^d))$, for any $r \in (q, \infty]$.

Proof. Clear by the embedding $\mathscr{W}_s \hookrightarrow C([0, T], C(\overline{U}, \mathbb{R}^d))$. □

In order to differentiate nonlinearities of the kind we are dealing with one needs to have it map from a space of higher integrability to one of lower integrability. The last theorem will therefore enable us to differentiate the nonlinear right-hand side between the spaces we shall define (see subsection 2.2.4).

A theorem analogue to the Sobolev embedding theorems is the following.

Theorem 8 (Lions-Aubin)**.** Let B_0, B, B_1 be Banach spaces such that we have the dense embeddings $B_0 \hookrightarrow B$ and $B \hookrightarrow B_1$, where the first embedding is compact and the second is continuous. Let B_0 and B_1 be reflexive. If we define

$$E = \{u \in L_s(0, T; B_0) : u' \in L_{s'}(0, T; B_1)\},$$

then we obtain the compact embedding $E \hookrightarrow L_s(0, T; B)$.

Proof. See [Sho97, III., Prop 1.3] □

The next remark gives a compact embedding into a space of less regularity than above.

Corollary 9. We have the compact embedding

$$\mathscr{W}_s \to L_s(0, T, G).$$

Proof. Notice that $V \hookrightarrow G$ compactly, since $H^1_q \hookrightarrow H^m_q$ compactly and $H^m_q \to G$ continuously by [Tri95, 4.6.1 Thm b)]. Also, $G \hookrightarrow V'$ continuously. Apply the last theorem 8 of Lions-Aubin. □

2.2.3 The linear differential operator

We fix some $\mathfrak{q}$ such that $\mathfrak{q} \in (n, n+\epsilon)$, specifying the space $V_q = W^1_{\mathfrak{q}}$. We also pick some exponent $r > \mathfrak{q}$ as it appears in theorem 7 and denote this choice by $\mathfrak{r}$. For notational convenience, we introduce the space $V_{\mathfrak{q}'} = W^1_{\mathfrak{q}'}$ $\mathcal{V}_{\mathfrak{q}'} = L_{\mathfrak{q}'}(0, T, V_{\mathfrak{q}'})$ and (by Philipps theorem) the dual space $(\mathcal{V}_{\mathfrak{q}'})' = L_{\mathfrak{q}}(0, T, V'_{\mathfrak{q}'})$ of $\mathcal{V}_{\mathfrak{q}}$. As an aside we remark that $(\mathcal{V}_{\mathfrak{q}'})' \subsetneq (\mathcal{V}_{\mathfrak{q}})'$.
Letting $s = \mathfrak{q}$, we introduce the space $W(0, T) = \mathscr{W}_{\mathfrak{q}}$ as our solution space. This choice for s is somewhat arbitrary, yet it allows to avoid a second set of exponents of integrability for the time domain. The reader can easily find the generalization for other choices of s, introducing whenever necessary exponents for the time domain and adapt the proofs.
Now that the solution space of the forward problem is at hand, we specify the parameter space, namely its exponents of integrability. To see the dependence of the exponents of integrability, one proceeds as follows:
Having chosen $\mathfrak{q}$ as mentioned above, this choice induces the exponent $\mathfrak{r}$ as defined above, ie $\mathfrak{r} = \mathfrak{r}(\mathfrak{q})$.
Then we set the topologies of the parameter spaces by picking the exponents p_D, p_λ, p_R, p_W, such that

$$\frac{1}{p_R} + \frac{1}{p_W} + \frac{1}{\mathfrak{r}} < \frac{1}{\mathfrak{q}}, \qquad \frac{1}{p_\lambda} + \frac{1}{\mathfrak{r}} \le \frac{1}{\mathfrak{q}} \tag{5}$$

holds.
The rationale behind the above imposition on p_W and p_R will be elucidated in subsection 2.2.4, where we investigate the nonlinear right-hand side. The guiding principle will be the repeated use of Hölder estimates and the ideas concerning multiplicator spaces (see eqn. (3) and the pertaining remarks)
The motivation for the requirement on p_λ can be seen in the lines right below. We define the differential operator $\frac{d}{dt} + \mathscr{A} : \mathcal{P} \times W(0, T) \to (\mathcal{V}_{\mathfrak{q}'})'$ by

$$\begin{aligned}(\frac{d}{dt} + \mathscr{A})(\pi, u) = f,\ f(v) = \int_0^T \int_U uv'dxdt + \int_0^T \int_U D\nabla u \nabla v dxdt \\ + \int_0^T \int_U \lambda uv dxdt, \qquad v \in C^1_0([0, T] \times U, \mathbb{R}^d),\ \pi = (D, \lambda, R, W) \in \mathcal{P}.\end{aligned}$$

Theorem 10. By our choice of the space $W(0,T)$ and $\mathcal{P}$ and the global boundedness of the elements in $\mathcal{P}$ this operator is linear and continuous (and hence C^1).

Proof. The linearity of the operator is obvious.
The continuity of $\frac{d}{dt}$ follows directly from the fact that $u' \in (\mathcal{V}_{\mathfrak{q}'})'$ by $u \in W(0,T)$.
By the choice of the parameter p_λ and the inclusion $W(0,T) \subset L^{\mathfrak{r}}$, the map

$$\mathcal{P} \times W(0,T) \to (\mathcal{V}_{\mathfrak{q}'})', \ (\lambda, u) \mapsto \lambda u$$

is boundedly continuous and by its bilinearity also C^1-differentiable.
The map

$$\mathcal{P} \times W(0,T) \to (\mathcal{V}_{\mathfrak{q}'})', \ (D, u) \mapsto -div(D\nabla u)$$

is bilinear, (locally Lipschitz) continuous and therefore also C^1-differentiable:

$$\begin{aligned}
\|-\nabla \cdot (D_1 \nabla u_1) + -div(D_2 \nabla u_2)\|_{(\mathcal{V}_{\mathfrak{q}'})'} &\leq \|-div(D_1 \nabla u_1) + -div(D_1 \nabla u_2)\|_{(\mathcal{V}_{\mathfrak{q}'})'} \\
&\quad + \|-div(D_1 \nabla u_2) + -div(D_2 \nabla u_2)\|_{(\mathcal{V}_{\mathfrak{q}'})'} \\
&\leq C_{\mathcal{P},2} \|div(\nabla(u_1 - u_2))\|_{(\mathcal{V}_{\mathfrak{q}'})'} \\
&\quad + \|D_1 - D_2\|_{\mathcal{P}_W} \|div \nabla u_2\|_{(\mathcal{V}_{\mathfrak{q}'})'} \\
&\leq C_{\mathcal{P},2} \|u_1 - u_2\|_{W(0,T)} \\
&\quad + C \|D_1 - D_2\|_{\mathcal{P}_W} \|u_2\|_{\mathcal{V}_{\mathfrak{q}}}.
\end{aligned}$$

□

A technical lemma we shall use in later chapters about regularization is

Lemma 11. Let z_n be a sequence from $W(0,T)$ that converges weakly: $z_n \rightharpoonup 0$. Assume furthermore that $z_n(\tilde{t}) = y \ \forall n$, some $\tilde{t} \in [0,T]$. Then we find $y \cong 0$.

Proof. By Theorem 8 we instantly know $z_n \to 0$ in $L_{\mathfrak{q}}(0,T,G)$. By a standard result of integration theory (which applies also to Bochner integrals), we can find a subsequence, also denoted by (z_n), which converges ae to 0 (ie $\|z_n\|_G \to 0$).
By the inlcusion of $W(0,T) \hookrightarrow C([0,T],G)$, this convergence must then occur everywhere on $[0,T]$. Therefore (z_k) converges pointwise everywhere to 0 and hence $y \cong 0$. □

2.2.4 Superposition Operators

Since our PDE contains a nonlinear right-hand side we need to introduce some concepts to handle such nonlinearities. For a comprehensive treatment of these nonlinearities, the so-called superposition operators, the interested reader may consult the standard reference [AZ90]. We will present under which prerequisites we can expect continuity, Hölder continuity and differentiability and verify that our nonlinearity satisfies all parts mentioned. At that point, the choice of the topology (compare eqns. (5)) for p_W and p_R

will become clear. Namely the differentiability imposes the heaviest restrictions on the choice of the parameters, enforcing a setup which yields the "excess integrability" we had hinted to. As a side remark, the ideas concerning multiplication spaces will cross our path in this subsection, where we will define a number of auxiliary exponents.
We do not investigate maps on open sets of normed spaces but rather on a metric space $\mathcal{P}$. For all statements concerning continuity and differentiability on this metric space this poses no problems since the metric is induced by a norm and $\mathcal{P}$ has a linear structure. Computations in the proofs run just the same as for open sets of vector spaces.
Before we start with the actual topic, we make a technical remark

Remark 12. The restriction of φ to the real line $\mathbb{R}$ is globally Lipschitz continuous and so are all the derivatives. In particular, they suffice a decay of the kind $\varphi^{(n)}(x) = o(|x|^n)$ for $|x| \to \infty$.

Proof. φ can be defined as a function on a suitable strip in $\mathbb{C}$ and then shown to be analytical. The statement concerning properties of the derivatives follows by Cauchy's integral formula. We additionally want the solution of the PDE to depend differentiably on the parameters.

□

Therefore the function $\Phi : \mathbb{R} \to \mathbb{R}$ is smooth and its derivatives are globally Lipschitz continuous.
As an aside we mention that if one chooses a different sigmoidal response function, as was hinted in the introduction, that response function needs to suffice these smoothness, boundedness and Lipschitz conditions as well. For rational sigmoidal functions or tan and tanh this is clearly the case.

Now we return to our quest of analyzing superposition operators. In all the generic statements and claims below, the space Z is a finite measure space.

Definition 13. A function $f : Z \times \mathbb{R} \to \mathbb{R}$, $(z, u) \mapsto f(z, u)$ which is measurable in (z) and continuous in u shall be called a Caratheodory function.

Remark 14. Let f be a Caratheodory function as above. Let $v : Z \to \mathbb{R}$ be measurable function. Then $f(\cdot, v(\cdot))$ is a measurable function.

Proof. Let (v_n) be a sequence of simple functions converging pointwise ae to v, where $v_n = \sum_i a_i^{(n)} \chi_{B_{n,i}}^{(n)}$, the $B_{n,i}$ being disjoint measurable sets and $a_1^{(n)} = 0$, $B_{n,1} = [(\cup_{i>1} B_{n,i})]^C$. Then for some measurable set $A \subset \mathbb{R}$ notice first that $C_{n,i} = (f(\cdot, a_i^{(n)}))^{-1}(A)$ is measurable. Then $D_{n,i} = C_{n,i} \cap B_{n,i}$ is measurable and therefore $\bigcup_i D_{n,i} = (f(\cdot, v_n(\cdot)))^{-1}(A)$ is measurable. This implies the measurability of $f(\cdot, v_n(\cdot))$ and by the continuity of f in the second argument and the pointwise ae convergence of v_n we get $f(\cdot, v(\cdot)) = \lim f(\cdot, v_n(\cdot))$ almost everywhere. This implies measurability of the limit function as desired. □

Definition 15. Let $1 \le q, p \le \infty$ and f a Caratheodory function as defined above. Suppose that for f we have the growth estimate

$$|f(z, u(z))| \le C(h(z) + |u(z)|^{\frac{p}{q}}) \quad \text{ae on } Z,$$

where $h \ge 0$ is from $L_q(Z)$. Then we define formally the operator $\mathscr{B}(u) = f(z, u(z))$.

We suppose stronger conditions than usual in the theory of superposition operators since the Caratheodory functions we will encounter are quite well-behaved.

Theorem 16. The operator $\mathscr{B}$ defined above is bounded, maps $L_p(Z) \to L_q(Z)$, and is continuous with the estimate

$$\|\mathscr{B}(u)\|_{L_q(Z)} \le C(\|h\|_{L_q(Z)} + \|u\|^{\frac{p}{q}}_{L_p(Z)}).$$

Proof. Measurability was treated above.
Concerning the boundedness and the estimate, conclude from the growth estimate for $p, q \le \infty$

$$\int_Z |\mathscr{B}(u)|^q dz \le C \int_Z (h^q(z) + |u(z)|^q) dz \le C(\|h\|^q_{L^q(Z)} + \|u\|^p_{L^p(Z)}).$$

For $p < q = \infty$ the growth estimate becomes a global boundedness condition for f. So we get by the finiteness of the measure space

$$\int_Z |\mathscr{B}(u)|^q dz \le C \int_Z (h^q(z) + \|u(z)\|_{L^\infty(Z)}) dz \le C(\|h\|^q_{L^q(Z)} + \|u\|^p_{L^p(Z)}).$$

Now let (u_k) be a sequence from L^p that converges to u in L^p. By general results from integration theory we can refine any subsequence of (u_k) to a further subsubsequence (u_k) such that for (u_k) we have ae convergence and ae uniform boundedness: $|u_k(z)| \le w(z)$ for some nonnegative $w \in L^p$.
We find

$$\begin{aligned} |\mathscr{B}(z, u_k(z)) - \mathscr{B}(z, u(z))|^q &\le 2^q(|\mathscr{B}(z, u_k(z))|^q + |\mathscr{B}(z, u(z))|^q) \\ &\le C(h^q(z) + w^p(z) + |u(z)|^p). \end{aligned}$$

So the sequence $|\mathscr{B}(z, u_k(z)) - \mathscr{B}(z, u(z))|^q$ is dominated by an integrable function. Pointwise ae convergence of this sequence is clear by the ae convergence of (u_k) and continuity of $\mathscr{B}$ in the last entry.
So by the dominated convergence theorem we obtain $\mathscr{B}(u_k) \to \mathscr{B}(u)$ in L^q. By the usual subsubsequence argument this convergence also holds for the entire sequence.
This establishes continuity. □

As an application we make the following

Remark 17. Let p_W, p_R be as before. Take the parameter function R and the signal response function Φ as defined before, as well as the domain $U \times [0,T]$. Define the exponents $\mathfrak{t} = (1/p_W + 1/\mathfrak{r})^{-1}$ and $\mathfrak{d} = \mathfrak{q}p_R/(p_R - \mathfrak{q})$ (notice the concept of multiplication spaces coming into play).
Define formally the superposition operator

$$\tilde{F} : L_{\mathfrak{t}}([0,T], L_{\mathfrak{t}}(U,\mathbb{R})) \to L_{\mathfrak{d}}([0,T], L_{\mathfrak{d}}(U,\mathbb{R}))$$

by $\tilde{F}_i(v) = \Phi(v(x,t))$, $i = 1,\dots,d$, where $v(x,t) \in L_{\mathfrak{t}}([0,T], L_{\mathfrak{t}}(U,\mathbb{R}))$ arbitrary.
Then $\tilde{F}_i$ is a continuous superposition operator between the stated spaces.

Proof. Notice the parameter R is ae in L_{p_R}. The function Φ is continuous, so clearly $\tilde{F}_i : U \times [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $(x,t,u) \mapsto R_i(x,t)\Phi(v(x,t))$ is a Caratheodory function.
By the global boundedness of Φ we get the estimate

$$|\Phi(v)| \le C_\Phi \le C + |v|^{\mathfrak{t}/\mathfrak{d}}$$

for any real $r \ge 1$. □

Clearly, the multiplications

$$\mathcal{P}_W \times L_{\mathfrak{r}}([0,T], L_{\mathfrak{r}}(U,\mathbb{R}^d)) \to L_{\mathfrak{t}}([0,T], L_{\mathfrak{t}}(U,\mathbb{R})),\ (W,u) \mapsto Wu$$

and

$$\mathcal{P}_R \times L_{\mathfrak{d}}([0,T], L_{\mathfrak{d}}(U,\mathbb{R}^d)) \to L_{\mathfrak{q}}([0,T], L_{\mathfrak{q}}(U,\mathbb{R})),\ (R,\Phi(Wu)) \mapsto R\Phi(Wu)$$

are continous by the usual Hölder inequality (even continuously differentiable).
The nonlinearity F is indeed a function of both u and parameters R, W, ie $F = F(R,W,u)$. As an abuse of notation we will frequently write $F(u)$, $F(R)$, $F(W)$ instead of $F(R,W,u)$, whenever it is clear how the other arguments are fixed. This said, we conclude that the compositions

$$u \mapsto F(u), \text{ where } F_i(u) = R_i\tilde{F}_i((Wu)_i) = R_i\Phi((Wu)_i)$$

and

$$W \mapsto F(W), \text{ where } F_i(u) = R_i\tilde{F}_i((Wu)_i) = R_i\Phi((Wu)_i)$$

are continuous.

In the course of developping our regularization technique we will not only want our superposition operator to be continuously differentiable, but also prove a stronger kind of continuity for the derivative. To this end we now present two structural theorems and then apply these.
The first stronger type of continuity is given under the prerequisites of the following

Theorem 18. Let $\mathscr{B} : L_p(Z) \to L_q(Z)$ be a continuous superposition operator as in the preceding theorem 16 generated by a Caratheodory function f, $1 \le q \le p < \infty$. f is assumed to be globally Lipschitz continuous in the entry u, such that we have an ae pointwise Lipschitz estimate of the kind

$$|f(z,u) - f(z,v)| \le T(z) \, \|u - v\|_{\mathbb{R}^d} \,,$$

where $T \ge 0$, $T(z) \in L_r(Z)$, $r = \frac{pq}{p-q}$.
Then the operator $\mathscr{B}$ is globally Lipschitz continuous with Lipschitz constant $\|T\|_r$.

Proof. By the given estimate we have for $u, v \in \mathbb{R}^d$

$$|f(z,u) - f(z,v)|^q \le T^q(z) \, \|u - v\|^q_{\mathbb{R}^d} \,.$$

This implies

$$\int_Z |f(z,u) - f(z,v)|^q dz \le \|T(z) \, \|u - v\|_{\mathbb{R}^d}\|^q_q \le \|T\|_r \int_Z \|u - v\|^q_{\mathbb{R}^d}$$

and thus for some $u, v \in L^p(Z)$

$$\|\mathscr{B}(u) - \mathscr{B}(v)\|_{L_q(Z)} \le \|u - v\|_{L_p(Z)}$$

□

For the case $p < q$, we cannot expect to get such a general result (compare [AZ90, Thm. 3.10 c)]). The deeper reason lies in the fact the multiplicator space must be taken from (compare eqn. (3) and the reference given there) does not contain any measurable functions. Nevertheless, another stronger type of continuity can be established under certain assumptions on the Caratheodory function f even in this case.

Theorem 19. Let $\mathscr{B} : L_p(Z) \to L_q(Z)$ be a continuous superposition operator as in the preceding theorem 16 generated by a Caratheodory function f, $1 < p < q < \infty$. f is assumed to be globally Lipschitz continuous and globally bounded in the entry u. Assume furthermore an ae pointwise Lipschitz estimate of the kind

$$|f(z,u) - f(z,v)| \le T(z) \, \|u - v\|_{\mathbb{R}^d} \,,$$

where $T(z) \in L_r(Z)$, $r = \frac{p}{p-1}$.
Then the operator $\mathscr{B}$ is Hölder continuous with Hölder exponent $\gamma = 1/q$ and constant $\|T\|^{1/q}_{\frac{p}{p-1}}$.

Proof. By the given estimate we have for $u, v \in \mathbb{R}^d$

$$|f(z,u) - f(z,v)| \le T(z) \, \|u - v\|_{\mathbb{R}^d} \,.$$

By the global boundedness of f we also find

$$|f(z,u) - f(z,v)|^q \le C_{max} |f(z,u) - f(z,v)|,$$

where $C_{max} = \max\{1, \sup |f(z,u) - f(z,v)|^{q-1}|\}$.
This implies

$$\int_Z |f(z,u) - f(z,v)|^q dz \leq C_{max} \left\|T \left\|u - v\right\|_{\mathbb{R}^d}\right\|$$

and thus for some $u, v \in L^p(Z)$

$$\left\|\mathscr{B}(u) - \mathscr{B}(v)\right\|_{L_q(Z)} \leq \|T\|^{1/q}_{\frac{p}{p-1}} \left\|u - v\right\|^{1/q}_{L_p(Z)}$$

□

A similar statement (yet without proof) can be found in [AZ90, Rem. p. 105] .
The promised application concerns what will later turn out to be the derivative of F. A first step towards this finding is to define this operator formally and prove its continuity properties. Define the exponents

$$\mathfrak{p} = \left(1/p_R + 1/p_W + 1/\mathfrak{r}\right)^{-1}, \; \mathfrak{a} = \frac{\mathfrak{p}\mathfrak{q}}{\mathfrak{p} - \mathfrak{q}}.$$

One can easily recongnize these conventions are motivated by the definition of the multiplication space $L_{\mathfrak{a}}(0, T, L_{\mathfrak{a}}(U, \mathbb{R}^d)) = \mathbb{M}_{\mathfrak{p},\mathfrak{q}}$. This space we shall use in the next technical

Lemma 20. The superposition operators

$$L_{\mathfrak{r}}(0, T, L_{\mathfrak{r}}(U, \mathbb{R}^d)) \to L_{\mathfrak{a}}(0, T, L_{\mathfrak{a}}(U, \mathbb{R}^d)), \; u \mapsto \Phi'(Wu)$$

and

$$\mathcal{P}_W \to L_{\mathfrak{a}}(0, T, L_{\mathfrak{a}}(U, \mathbb{R}^d)), \; W \mapsto \Phi'(Wu)$$

are Hölder continuous with Hölder exponent $\gamma = 1/\mathfrak{a}$.

Proof. By the global boundedness of Φ' and 14 the superposition operator maps into the stated space.
The continuity we can infer by remarking that the image is a globally bounded and by using the estimate

$$|\Phi'(v)| \leq C_{\Phi'} \leq C + |v|^{\mathfrak{r}/\mathfrak{a}}.$$

Concerning the application of Thm. 20 one remembers Φ' is globally Lipschitz continuous and globally bounded. Then since $W \in L_\infty$ elementwise, in particular $W \in L_{\mathfrak{r}/\mathfrak{r}-1}$. The Lipschitz estimate is clear.
Noteworthy in the proof for the second operator is to remark that in case $p_W > \mathfrak{a}$ we use theorem 18 and notice that on bounded sets Lipschitz continuity includes Hölder continuity.
In case $p_W < \mathfrak{a}$ we are in the same setting as for the first operator and only confirm $u \in L_{p_W/p_W-1}$ since $p_W/p_W - 1 \leq 2 \leq \mathfrak{r}$. The rest of the proof follows analogously. □

A similar statement (yet without proof) can be found in [AZ90, Rem. p. 105] . With the same motivation as above we define the multiplicator spaces

$$\mathbb{M}_{\mathfrak{r},\mathfrak{q}} = \mathscr{L}(L_{\mathfrak{r}}(0,T,L_{\mathfrak{r}}(U,\mathbb{R}^d)),L_{\mathfrak{q}}(0,T,L_{\mathfrak{q}}(U,\mathbb{R}^d)))$$

and

$$\mathbb{M}_{p_W,\mathfrak{q}} = \mathscr{L}(L_{p_W}(0,T,L_{p_W}(U,\mathbb{R}^{d\times d})),L_{\mathfrak{q}}(0,T,L_{\mathfrak{q}}(U,\mathbb{R}^d))).$$

so we can map into the appropriate multiplication spaces:

Corollary 21. The superposition operators

$$L_{\mathfrak{r}}(0,T,L_{\mathfrak{r}}(U,\mathbb{R}^d)) \to \mathbb{M}_{\mathfrak{r},\mathfrak{q}},\ u \mapsto (h \mapsto R\Phi'(Wu)Wh)$$

and

$$\mathcal{P}_W \to \mathbb{M}_{p_W,\mathfrak{q}},\ W \mapsto (\tilde{W} \mapsto R\Phi'(Wu)\tilde{W}u)$$

are Hölder continuous with Hölder exponent $\gamma = 1/\mathfrak{a}$.

We remark that $\mathbb{M}_{\mathfrak{r},\mathfrak{q}}$ is an $L_{\mathfrak{b}}$ space and $\mathbb{M}_{p_W,\mathfrak{q}}$ is an $L_{\mathfrak{c}}$ space, where

$$\mathfrak{b} = \frac{\mathfrak{q}\mathfrak{r}}{\mathfrak{r}-\mathfrak{q}},\ \mathfrak{c} = \frac{\mathfrak{q}p_W}{p_W-\mathfrak{a}}.$$

Concerning the differentiability of a superposition operator we will not give a statement of maximal generality (which can be found at [AZ90, Thm 3.13]). We will rather make stronger assumptions to obtain a sufficiency statement only and prove this by hand. That way we can dispense with all the heavy machinery of [AZ90] and keep our treatment self-contained.

Theorem 22. Let $1 \leq q < r < \infty$ and put $b = \frac{rq}{r-q}$. Let f be a Caratheodory function. Suppose additionally that f is continously differentiable in the entry u, and f' a Caratheodory function.
For f we assume a growth estimate as in the preceding theorem with auxiliary function $h \geq 0$ in L_p and denote the pertaining superposition operator by $\mathscr{B}$.
Assume that the superposition operator formally defined by $\mathscr{B}'(u) = f'(z,u(z))$ is a continuous map $\mathscr{B}' : L_r \to L_b (\hookrightarrow L_q)$ (eg when f' satisfies a growth condition of the type $|f'(z,u(z))| \leq C(h(z) + |u(z)|^{\frac{p}{r}})$ ae on Z). Finally suppose the existence of a set of functions $(a_\beta) \subset L_q(Z)$ such that $\|a_\beta\|_q = o(\beta^{\frac{-\epsilon}{1-\epsilon}})$ for $\beta \to \infty$. Assume for all a_β the joint growth condition

$$|f(z,u(z)+v) - f(z,u(z)) - f'(z,u(z))v| \leq \beta^{-1}(a_\beta(z) + C\beta^{r/q}|v(z)|^{r/q}).$$

Therefore the superposition operator $\mathscr{B} : L_r(Z) \to L_q(Z)$, $\mathscr{B}(u) = f(z,u(z))$ is continuously differentiable. The derivative has the form

$$\mathscr{B}'(u)h(z) = f'(z,u(z))h(z).$$

Proof. Set $\beta = \frac{1}{\|v\|_p^{1-\epsilon}}$. From the growth estimate we gather

$$|f(z,u(z)+v) - f(z,u(z)) - f'(z,u(z))v|^q / \|v\|_p^q \leq \|v\|_p^{-q\epsilon} (a_\beta(z) + C\beta^{p/q}|v(z)|^{p/q})^q$$

which implies (using the Hölder inequality and monotonicity of $(\cdot)^q$)

$$\|G(z,v(z))\|_q^q / \|v\|_p^q \leq \|v\|_p^{-q\epsilon} (\|a_\beta(z)\|_q + C\beta^{p/q} \|v(z)\|_p^{p/q})^q.$$

Remarking $\|v\|_p^{-\epsilon} = \beta^{\epsilon/1-\epsilon}$, we find

$$\|G(z,v(z))\| / \|v\|_p \leq \|v\|_p^{-\epsilon} (\|a_\beta(z)\|_q + C\beta^{p/q} \|v(z)\|_p^{p/q}) = \|a_\beta(z)\|_q / \beta^{\epsilon/1-\epsilon} + C \|v\|_p^{\epsilon p/q}.$$

By $p > q$ the second summand in the last estimate vanishes as $\|v\|_p \to 0$. For the first summand we invoke the presumed convergence property of a_β for $\beta \to \infty$ to establish convergence to 0 for $\|v\|_p \to 0$. So $\|G(z,v(z))\|_q = o(\|v\|_p)$.
Therefore the superposition operator $\mathscr{B} : L^p(Z) \to L^q(Z)$, $\mathscr{B}(u) = f(z,u(z))$ is continuously differentiable. The derivative has the form

$$\mathscr{B}'(u)h(z) = f'(z,u(z))h(z).$$

□

Take the parameter function W, the signal response function Φ, and $\mathfrak{q}, p_W, p_R$ as before. Then we obtain

Theorem 23. The operator $u \mapsto F(u)$ as defined above is continuously differentiable and the (locally Hölder continuous) derivative operator is

$$F' : L_\mathfrak{r}([0,T], L_\mathfrak{r}(U,\mathbb{R}^d)) \to \mathbb{M}_{\mathfrak{r},\mathfrak{q}},$$
$$F'(u) = R\Phi'(Wu)W$$

Proof. We only need to check the growth condition of the preceding theorem.
By the smoothness of φ we have (defining g)

$$\begin{aligned}|g(x,t)| &= |\varphi((W(x,t)(u(x,t)+v(x,t)))_i) - \varphi((W(x,t)(u(x,t))_i) \\ &\quad - \varphi'((W(x,t)(u(x,t))_i)(W(x,t)v(x,t))_i| \\ &\leq |\varphi''(\xi(x,t)||(W(x,t)v(x,t))_i|^2 \\ &= C|(W(x,t)v(x,t))_i|^2\end{aligned}$$

By the global boundedness of φ and φ' we get

$$\begin{aligned}|\varphi((W(x,t)(u(x,t)+v(x,t)))_i) - \varphi((W(x,t)(u(x,t))_i)| \\ \leq C_{\varphi'}|(W(x,t)(v(x,t))_i)|\end{aligned}$$

and

$$|\varphi'((W(x,t)(u(x,t))_i)(W(x,t)v(x,t))_i| \leq C_{\varphi'}|(W(x,t)v(x,t))_i|.$$

Thus

$$|g(x,t)| \leq C_{\varphi'}|(W(x,t)(v(x,t))_i)| + C_{\varphi'}|(W(x,t)v(x,t))_i| = 2C_{\varphi'}|(W(x,t)(v(x,t))_i)|.$$

This implies for any $\theta \in (0,1)$ by interpolation of exponents

$$\begin{aligned}|g(x,t)| &= C|(W(x,t)v(x,t))_i|^{2\theta} \cdot |(W(x,t)(v(x,t))_i)|^{1-\theta} \\ &\leq C|W(x,t)v(x,t))_i|^{1+\theta} \\ &\leq C\sum_j (|W_{i,j}(x,t)||v_j(x,t)|)^{1+\theta} \quad \text{by Jensen‘s inequality.}\end{aligned}$$

Since we have $\mathfrak{r} > \mathfrak{q}$, we can pick $\theta > 0$ such that $\mathfrak{r} - \mathfrak{q}(1+\theta) > 0$. A simple computation shows

$$p_W \geq (1+\theta)\tilde{q} = \frac{\mathfrak{r}}{\mathfrak{r} - (1+\theta)\mathfrak{q}} \Leftrightarrow \frac{1}{\mathfrak{q}\tilde{q}} + \frac{1+\theta}{p_W} \leq \frac{1}{(1+\theta)\mathfrak{q}}$$

and the last inequality is true by eqn. 5, whenever θ is chosen sufficiently small. Therefore $|W_{i,j}(x,t)| = C(x,t)$ is a $L_{\tilde{q}}$ function.
Now use the Young inequality (see remark 5.3, $\tilde{p} = \mathfrak{r}/\mathfrak{q}\frac{1}{1+\theta}$, $\tilde{q} = s$ and $\epsilon = \beta^{\mathfrak{r}/\mathfrak{q}-1}$) on the last estimate

$$\begin{aligned}C(x,t)|v_j(x,t)|^2 &\leq (\beta^{\mathfrak{r}/\mathfrak{q}-1})|v_j(x,t)|^{\mathfrak{r}/\mathfrak{q}} + (\beta^{\mathfrak{r}/\mathfrak{q}-1})^{-\tilde{q}/\tilde{p}} C_{\tilde{p}} C(x,t)^{\tilde{q}} \\ &= (\beta^{\mathfrak{r}/\mathfrak{q}-1})|v_j(x,t)|^{\mathfrak{r}/\mathfrak{q}} + \beta^{-(\tilde{p}(1+\theta)-1)(\tilde{q}-1)} C_{\tilde{p}} C(x,t)^{\tilde{q}} \\ &= (\beta^{\mathfrak{r}/\mathfrak{q}-1})|v_j(x,t)|^{\mathfrak{r}/\mathfrak{q}} + \beta^{-1-\theta\tilde{p}(\tilde{q}-1)} C_{\tilde{p}} C(x,t)^{\tilde{q}} \\ &= (\beta^{\mathfrak{r}/\mathfrak{q}-1})|v_j(x,t)|^{\mathfrak{r}/\mathfrak{q}} + \beta^{-(1+\delta)} C_{\tilde{p}} C(x,t)^{\tilde{q}}\end{aligned}$$

since clearly $\theta\tilde{p}(\tilde{q}-1) > 0$.
Therefore $a_\beta = \beta^{-\delta} C_{\tilde{p}} C(x,t)^{\tilde{q}}$ yields a function as in the preceding theorem and the proof is complete. □

Another close look at the last proof confirms that the roles of u and W are interchangeable. So analogously:

Theorem 24. The superposition operator $W \mapsto F(Wu)$ mapping between spaces $\mathcal{P}_W \to L_\mathfrak{q}(0,T,L_\mathfrak{q}(U,\mathbb{R}^d))$ as defined above is continuously differentiable and the (locally Hölder continuous) derivative operator is the respective one stated in Cor. 21.

The continuity of $F(R)$ trivial, since $R \mapsto F(R)$ is just a linear (and by eqn. (5) continuous) map, which entails even continuous differentiability.
So we have ultimately shown that the right-hand side operator is totally differentiable with respect to both u and the parameters W and R and the derivative is at least Hölder continous.

2.3 Solvability

A weak formulation of our model PDE is stated in the following.

Definition 25. $u \in W(0,T)$ is a weak solution of the above PDE, if

$$u' + \mathscr{A}u = F(u) \text{ in } (\mathcal{V}_{\mathfrak{q}'})', \quad u(0) = u_0 \in G \tag{6}$$

In order to develop existence and uniqueness of a solution we will first fix the argument u to some arbitrary $w \in C([0,T]; G)$ in the nonlinear righthandside function Φ to obtain a righthandside function $F(t, w(t)) = f(t) \in (\mathcal{V}_{\mathfrak{q}'})'$. So we may restate the definition of a solution for this simplified Cauchy problem:

$$u' + \mathscr{A}u = f \text{ in } (\mathcal{V}_{\mathfrak{q}'})', \quad u(0) = u_0 \in G. \tag{7}$$

Recent results of maximal parabolic regularity ([HDR09],[ACFP07]) permit the following statement, for which a sketched proof and further references can be found in the appendix (Theorem 56).

Theorem 26. There exists some $\tilde{\epsilon} = \tilde{\epsilon}(U, C_{\mathcal{P},1}, C_{\mathcal{P},2})$, such that for $\mathfrak{q} \in (3, 3+\tilde{\epsilon})$ and the pertaining definitions of $\mathcal{P}$ and $W(0,T)$ we can find a unique solution u for the problem in equation 7 with any right-hand side $f \in (\mathcal{V}_{\mathfrak{q}'})'$. This solution depends continuously on the right-hand side

$$\|u\|_{W(0,T)} \leq C(\|f\|_{(\mathcal{V}_{\mathfrak{q}'})'} + \|u_0\|_G)$$

and the constant $C = C(U, C_{\mathcal{P},1}, C_{\mathcal{P},2})$.

After treating the Cauchy problem with simplified right-hand side, we will now address the original problem. The admissible range of the parameters shall be $\mathcal{P}$ as before. We rephrase our original (possibly multidimensional) model equation (1) as a Cauchy problem

$$u \in L_2(0,T;V): \quad u' + \mathscr{A}u = F(u) \text{ in } L_2(0,T;V'), \quad u(0) = u_0. \tag{8}$$

To establish existence and uniqueness for the nonlinear setting, one makes use of Banach's fixed point theorem in the Banach space

$$C([0,T]; L_2(U, \mathbb{R}^d)),$$

noting that we had the continuous embedding $W(0,T) \hookrightarrow C([0,T]; G)$ and $G \hookrightarrow L_2(U, \mathbb{R}^d)$.
Let $B : C([0,T]; L_2(U, \mathbb{R}^d)) \to C([0,T]; L_2(U, \mathbb{R}^d))$ be the map which assigns the unique solution u of the afore-mentioned linearized problem with fixed righthandside $F(w)$ to the argument w, ie $w \mapsto u$. One establishes that at least on a shortend interval B is a contraction to obtain a unique fixed point and via the compactness of the interval $[0,T]$ arrives at the claim

Theorem 27. The Cauchy problem as stated in equation 8 has a unique solution.

Proof. See [Eva08, pp. 500]. □

3 The control-to-state map

3.1 Continuity and differentiability of the control-to-state map

In order to prove existence, uniqueness and differentiability of the original problem, we will employ the well-known implicit function theorem by exploiting the given smoothness of the right-hand side function Φ. All the major ingredients we have gathered in the preceding sections. Continuous differentiability was covered in 2.2.3 and 2.2.4. The solvability theory for PDEs will supply the invertibility part for the implicit function theorem. The reason to choose the implicit function theorem is that it allows to play differentiability proof back to mentioned differentiability results rather than doing the direct proof by lengthy norm estimates. The theorem also permits to elegantly obtain an explicit formula for the derivative. This formula we will then dissect further to arrive at a stronger continuity statement we need for regularization. Another pleasant result of this dissection is the straightforward formula we get for the adjoint operator of the parameter-to-state operator. This is also of pratical importance for any numerical implementation.

The notation we use for parameters and functionals in this subsection is to be understood with the specifications of eqns. (5) and subsection 2.3. Then we introduce the operator

$$\mathcal{C} : \mathcal{P} \times W(0,T) \to G \times (\mathcal{V}_{\mathsf{q}'})' \tag{9}$$

$$\mathcal{C} : ((D,\lambda,R,W),u) \mapsto (u(0) - u_0, u' + \mathscr{A}u - F(u)). \tag{10}$$

Clearly this operator is well-defined: the first component of the image exists by theorem 6, the second component of $\mathcal{C}$ is well-defined by the assumptions on $u, \mathscr{A}$, and F.
For the next lemma we fix the first argument and show continous differentiability.

Lemma 28. The map $\mathcal{S} = \mathcal{C}(p_0,\cdot) : W(0,T) \to G \times (\mathcal{V}_{\mathsf{q}'})'$ is continously differentiable and the differential at any $y \in W(0,T)$ is an isomorphism from $W(0,T)$ to $G \times (\mathcal{V}_{\mathsf{q}'})'$.

Proof. For the first component of the image of $\mathcal{S}$ we see that the evaluation map $u \mapsto u(0)$ is a linear map. It is furthermore continuous:

$$|u(0)| \le \|u\|_{C(0,T;G)} \le C_{embed} \|u\|_{V_{\mathsf{q}}}.$$

For the second component we consider summands separately.
The elliptic operator $\mathscr{A}$ and the map $u \mapsto u'$ we know to be C^1 by thm. 10.
For the third summand $u \mapsto F(u) = R \cdot \Phi(Wu)$ we refer to the concluding statement at the end of subsection 2.2.4.
So we may deduce that $\mathcal{S}$ is a continuously differentiable map.
For the derivative $\mathcal{S}'$ in the direction of some $h \in W$, we obtain $\mathcal{S}'(y)h = (h(0), h' + \mathscr{A}h - F'(y)h)$. Define $\tilde{\lambda} = \lambda - F'(y)$ and denote the resulting linear differential operator by $\tilde{\mathscr{A}}$. It is then clear that by the (componentwise) boundedness of F', $\tilde{\lambda}$ satisfies the requirements of subsection 2.3. Therefore we may apply the findings of that subsection to

$\mathcal{S}'(u)$. For arbitrary $(v_0, f) \in G \times \mathcal{V}'$ there exists a solution h for $(h(0), h' + \mathscr{A} h - F'(y)h) = (v_0, f)$. This ensures surjectivity. The uniqueness of this solution ensures injectivity of $\mathcal{S}'(y)$. The stability statement in section 2.3 guarantees the continuity of the inverse map.
So we may conclude that for any $y \in W(0,T)$, $\mathcal{S}'(y)$ is an isomorphism from $W(0,T)$ to $G \times \mathcal{V}'$. □

Lemma 29. The map $\mathscr{P} = \mathcal{C}(\cdot, u) : \mathcal{P} \to G \times (\mathcal{V}_{\mathsf{q}'})'$ is continuously differentiable.

Proof. Just as above $(D, \lambda) \mapsto \mathscr{A}$ is clearly linear and bounded. It is therefore in particular continuously differentiable.
Again by the concluding statement in subsection 2.2.4 we know the continuous differentiability of the operator $(R, W) \mapsto F$.
Since partial continuous differentiability in all the parameters is thus confirmed, we may conclude that $\mathscr{P}$ is totally continuously differentiable. □

The reader may take a look at the reasoning from above theorems to verify the equations

$$\frac{\partial \mathcal{C}}{\partial u}(p_0, u)(h) = (h(0), h' + \mathscr{A}_0 h - R_0 \Phi'(W_0 u) W_0 h)$$

and

$$\frac{\partial \mathcal{C}}{\partial p}(p_0, u)(p_1) = (0,\ (\lambda_1 u - \nabla \cdot (D_1 \nabla u) - R_1 \Phi(W_0 u) - R_0 \Phi'(W_0 u) W_1 u)).$$

Here $\mathscr{A}_0$ denotes the usual operator $\mathscr{A}$ formed by using the parameters λ_0 and D_0. Analogous notation $\mathscr{A}_1$ is used for the respective operator with different parameters λ_1 and D_1.
The last two lemmata enable us to state the following application of the implicit function theorem

Corollary 30. The control to state map $\mathscr{D} : \mathcal{P} \to W(0,T), \quad (D, \lambda, R, W) \mapsto u$, assigning the unique solution of our PDE 6 to each tuplet of parameters (D, λ, R, W), is continuously differentiable and the derivative is given by the following formula:

$$\mathscr{D}'(p_0)(p_1) = -\left(\frac{\partial \mathcal{C}}{\partial u}(p_0, u)\right)^{-1} \circ \frac{\partial \mathcal{C}}{\partial p}(p_0, u)(p_1) = \ v \ \text{ solving the Cauchy problem}$$
$$v(0) = 0 : v' + \mathscr{A}_0 v - R_0 \Phi'(W_0 u) W_0 v = -\mathscr{A}_1 u + R_1 \Phi(W_0 u) + R_0 \Phi'(W_0 u) W_1 u$$

where $u = u(p_0) = \mathscr{D}(p_0)$.

Proof. The existence proof from theorem 27 ensures the existence of a zero of the map $\mathcal{C}$. The last two lemmata supply the other part of the conditions in the implicit function theorem in the formulation of [Lan94, Satz 2.2,Satz 3.1]. □

3.2 Properties of the derivative of the control to state map

The last result, namely the explicit formula for the derivative of the control to state map at some p_0, enable us to investigate further interesting and useful properties of $\mathcal{D}'$.
Our inspection will be divided into several lemmata, which then allow us to show the Hölder continuity of the operators $\mathcal{D}$ and $\mathcal{D}'$ on bounded sets.
In longer chains of estimates, constants will often be merged into one constant C, which may therefore change from one step to the next. This is merely intended to increase legibility. In the norm estimates to follow, the guiding concept will be the use of Hölder estimates and the ideas we mentioned for multiplicator spaces.

Lemma 31. The map $\mathcal{D}'(\cdot) : \mathcal{P} \to \mathscr{L}(\mathcal{P}, W(0,T))$ is bounded on bounded sets.

Proof. Let $\tilde{\mathscr{A}}$ denote the linear differential operator induced by the parameters D_0 and $\tilde{\lambda} = \lambda - R_0\Phi'(W_0u)W_0$, and $r(p_0)(p_1) = -\mathscr{A}_1 u + R_1\Phi(W_0u) + R_0\Phi'(W_0u)W_1u$. Then $\mathcal{D}'(p_0)(p_1)$ is the solution of the Cauchy problem

$$v(0) = 0,\ v' + \tilde{\mathscr{A}}v = r(p_0)(p_1).$$

By the continuous dependence of the PDE solution on the right-hand side (see subsection 2.3, thm. 26), we have the linear stabitlity estimate

$$\|v\|_{\mathcal{V}_q} \leq C\,\|r\|_{(\mathcal{V}_{q'})'}\,.$$

Denote $u = \mathcal{D}(p_0)$. Likewise, such a stability estimate holds for the solution of our model PDE (eqn. (6)):

$$\|u\|_{\mathcal{V}_q} \leq C\,\|R_0\Phi(W_0u)\|_{\mathcal{V}'} \leq C\,\|R_0\|_{(\mathcal{V}_{q'})'} \leq C\,\|R_0\|_{\mathcal{P}_R} \leq C\,\|p_0\|_{\mathcal{P}}\,.$$

Now use the last estimate in the following:

$$\begin{aligned}\|[r(p_0)](p_1)\|_{(\mathcal{V}_{q'})'} &\leq \|\mathscr{A}_1u\|_{(\mathcal{V}_{q'})'} + \|R_1\Phi(W_0u)\|_{(\mathcal{V}_{q'})'} + \|R_0\Phi'(W_0u)W_1u\|_{(\mathcal{V}_{q'})'} \\ &\leq C\,\|u\|_{\mathcal{V}_q}\,\|p_1\|_{\mathcal{P}} + C_\Phi\,\|R_1\|_{\mathcal{P}} + \|R_0\|_{\mathcal{P}_R}\,\|W_1\|_{\mathcal{P}_W}\,\|u\|_{\mathcal{V}_q} \\ &\leq C\,\|p_0\|_{\mathcal{P}}\,\|p_1\|_{\mathcal{P}} + C\,\|p_1\|_{\mathcal{P}} + C\,\|p_0\|_{\mathcal{P}}\,\|p_1\|_{\mathcal{P}}\,\|p_0\|_{\mathcal{P}}\,.\end{aligned}$$

This implies the boundedness of the operator $\mathcal{D}'$:

$$\|\mathcal{D}'(p_0)\|_{\mathscr{L}(\mathcal{P},W(0,T))} \leq C(1 + \|p_0\|_{\mathcal{P}} + \|p_0\|_{\mathcal{P}}^2).$$

□

Using the mean value theorem (compare [Wer00, Satz III.5.4 b)]) on $\mathcal{D}$ we obtain for bounded sets

Lemma 32. The map $\mathcal{D} : \mathcal{P} \to W(0,T)$ is Lipschitz continuous on convex, bounded sets with uniform Lipschitz constant.

Proof. By the mean value theorem in finite dimensions we have for some $p_0, p_2 \in \mathcal{P}$, and any $p_\theta = \theta p_0 + (1-\theta)p_2$, $\theta \in (0,1)$

$$\|\mathcal{D}(p_0) - \mathcal{D}(p_2)\|_{W(0,T)} \leq \sup_{\theta\in(0,1)} \left(\|\mathcal{D}'(p_\theta)\|_{\mathscr{L}(\mathcal{P},W(0,T))}\right) \|p_0 - p_2\|_{\mathcal{P}} .$$

The sup in the estimate exists by the preceding lemma and is bounded uniformly by the preceding lemma. □

Notice that this type of Lipschitz continuity is stronger than the raywise Lipschitz continuity which follows from differentiability of $\mathcal{D}$.

Next we consider the superposition operators $\frac{\partial \mathcal{C}}{\partial p}$ and $\frac{\partial \mathcal{C}}{\partial u}$.

Lemma 33. The operator $\frac{\partial \mathcal{C}}{\partial u} : \mathcal{P} \times W(0,T) \to \mathscr{L}(W(0,T), G \times (\mathcal{V}_{\mathfrak{q}'})')$, $(p_0, u) \mapsto (h \mapsto (h(0), h' + \mathscr{A}_0 h - R_0\Phi'(W_0 u)W_0 h)$ is Hölder continuous on bounded sets with Hölder constant depending on Φ' and the bound of the set. The Hölder exponent is γ as in lemma 20.

Proof. The first component is the map $h \mapsto h(0)$, mapping $W(0,T) \to G$, so it is independent of the argument (p_0, u) and therefore Hölder continuous as desired.
Considering the second component of the image we start with the summand $(p_0, u) \mapsto (h \mapsto h' + \mathscr{A}_0 h)$. This is a bounded linear operator and hence globally Lipschitz.
For the last summand $(p_0, u) \mapsto (h \mapsto R_0\Phi'(W_0 u)W_0 h)$ we refer to theorem 20, take the points $(p_0, u), (p_2, v)$ and get the estimate

$$\begin{aligned}
&\|[R_2\Phi'(W_2 v)W_2 - R_0\Phi'(W_0 u)W_0]h\|_{(\mathcal{V}_{\mathfrak{q}'})'} \\
&\leq C(\|R_0 - R_2\|_{p_R} + (\|W_2 - W_0\|_{p_W} + \|u_2 - u_0\|_{W(0,T)})^{\gamma})\, \|h\|_{L^{\mathfrak{r}}(0,T,L^{\mathfrak{r}}(UT],\mathbb{R}^d))}
\end{aligned}$$

Since we are on a bounded set, this implies Hölder continuity as desired. The constant C in the last estimate depends on the bound of the considered domain and on $\sup(\Phi')$. □

A helpful technical lemma is the following:

Lemma 34. On the domain $\mathcal{P} \times W(0,T)$ the map $\frac{\partial \mathcal{C}}{\partial u}^{-1}$ is bounded by some constant C which depends only on $C_{\mathcal{P},1}, C_{\mathcal{P},2}$ and the domains U and $[0,T]$.

Proof. This follows by the findings of thm. 56 in the appendix. □

Using corollary 60 in the appendix, we can deduce the following

Corollary 35. The map $\left(\frac{\partial \mathcal{C}}{\partial u}(\cdot,\cdot)\right)^{-1} : \mathcal{P} \times W(0,T) \to \mathscr{L}(G \times (\mathcal{V}_{\mathfrak{q}'})', W(0,T))$ is locally Hölder continuous. The Hölder constant is uniform for all neighborhoods of Hölder continuity and these neighborhoods are balls of uniform radius.

Proof. By the preceding theorem we can take for (p_0, u) any bounded neighborhood (denoted by A), such that

$$\left\| \frac{\partial \mathcal{C}}{\partial u}(p_0, u) - \frac{\partial \mathcal{C}}{\partial u}(p_2, v) \right\| \leq C \left\| (p_0, u) - (p_2, v) \right\| < C\epsilon,$$

where $\epsilon > 0$ is arbitrary. Let B be a neighborhood which satisfies the requirements of corollary 60 in the appendix. Then we find

$$\begin{aligned} \left\| \frac{\partial \mathcal{C}}{\partial u}^{-1}(p_0, u) - \frac{\partial \mathcal{C}}{\partial u}^{-1}(p_2, v) \right\| &\leq C \left\| \frac{\partial \mathcal{C}}{\partial u}^{-1}(p_0, u) \right\|^2 \left\| \frac{\partial \mathcal{C}}{\partial u}(p_0, u) - \frac{\partial \mathcal{C}}{\partial u}(p_2, v) \right\| \\ &\leq C \left\| (p_0, u) - (p_2, v) \right\|^{\gamma}. \end{aligned}$$

In the last line we used the preceding lemma. □

Lemma 36. The operator

$$\begin{aligned} &\frac{\partial \mathcal{C}}{\partial p} : \mathcal{P} \times W(0,T) \to \mathscr{L}(\mathcal{P}, G \times (\mathcal{V}_{\mathfrak{q}'})') \\ &\frac{\partial \mathcal{C}}{\partial p}(p, u)(p_1) = (0, \mathscr{A}_1 u + R_1 \Phi(W_0 u) + R_0 \Phi'(W_0 u) W_1 u \end{aligned}$$

is Hölder continuous on bounded sets with Hölder constant depending on the bound of the considered set and Φ'.

Proof. The first component (in the image) of $\frac{\partial \mathcal{C}}{\partial p}$ is clear. For the second component take $u, v \in W(0,T)$ and $p_0, p_2 \in \mathcal{P}$, p_1 such that $p_0 \pm p_1, p_2 \pm p_1 \in \mathcal{P}$ then consider

$$\begin{aligned} &\left\| [\mathscr{A}_1 u - R_1 \Phi(W_0 u)] - [\mathscr{A}_1 v - R_1 \Phi(W_0 v)] \right\|_{\mathcal{V}'} \\ &\leq \|p_1\|_{\mathcal{P}} \|u - v\|_{W(0,T)} + \|p_1\| C_{\Phi} (\|W_0 - W_2\|_{\mathcal{G}} + \|u - v\|_{W(0,T)})^{\gamma} \end{aligned}$$

Similar to the proof of the last lemma we invoke theorem 20 in

$$\begin{aligned} &\|R_0 \Phi'(W_0 u) W_1 u - R_2 \Phi'(W_2 v) W_1 v\|_{(\mathcal{V}_{\mathfrak{q}'})'} \\ &\leq \|R_0 \Phi'(W_0 u) W_1 u - R_0 \Phi'(W_2 u) W_1 u\|_{(\mathcal{V}_{\mathfrak{q}'})'} \\ &\ + \|R_0 \Phi'(W_2 u) W_1 u - R_0 \Phi'(W_2 u) W_1 v\|_{(\mathcal{V}_{\mathfrak{q}'})'} \\ &\ + \|R_0 \Phi'(W_2 u) W_1 v - R_0 \Phi'(W_2 v) W_1 v\|_{(\mathcal{V}_{\mathfrak{q}'})'} \\ &\ + \|R_0 \Phi'(W_2 v) W_1 v - R_2 \Phi'(W_2 v) W_1 v\|_{(\mathcal{V}_{\mathfrak{q}'})'} \\ &\leq \|p_0\|_{\mathcal{P}} L_{\Phi'} \left[\|W_0 - W_2\|_{\mathcal{P}} \|u\|_{L_{\mathfrak{r}}(U \times [0,T], \mathbb{R}^d)} \right]^{\gamma} \|p_1\|_{\mathcal{P}} \|u\|_{W(0,T)} \\ &\ + \|p_0\|_{\mathcal{P}} \sup(\Phi) \|p_1\|_{\mathcal{P}} \|u - v\|_{W(0,T)} \\ &\ + \|p_0\|_{\mathcal{P}} L_{\Phi'} \left[\|W_2\|_{\mathcal{P}} \|u - v\|_{L_{\mathfrak{r}}(U \times [0,T], \mathbb{R}^d)} \right]^{\gamma} \|p_1\|_{\mathcal{P}} \|v\|_{W(0,T)} \\ &\ + \|p_0 - p_2\|_{\mathcal{P}} \sup(\Phi) \|p_1\|_{\mathcal{P}} \|v\|_{W(0,T)} . \\ &\leq C(\|u\|_{W(0,T)}, \|v\|_{W(0,T)}, \|p_0\|_{\mathcal{P}}) \|p_1\|_{\mathcal{P}} \left[\|u - v\|_{W(0,T)} + \|p_0 - p_2\|_{\mathcal{P}} \right]^{\gamma} . \end{aligned}$$

From all these estimates we obtain Hölder continuity of $\frac{\partial \mathcal{C}}{\partial p}$ on bounded sets with the mentioned dependence of the Hölder constant. □

An elementary statement is the following

Lemma 37. Given some bounded subsets X_1, X_2 of different Banach spaces , Y a Banach space, and some Hölder continous function $f : X_1 \times X_2 \to Y$, some Lipschitz continuous function $g : X_1 \to X_2$, then the composition $h : X_1 \to X_2, h(x) = f(x, g(x))$ is Hölder continuous (with exponent γ).

Proof.

$$\begin{aligned}\|h(x_1) - h(x_2)\| &= \|f(x_1, g(x_1)) - f(x_2, g(x_2))\| \leq C \|(x_1, g(x_1)) - (x_2, g(x_2))\|^\gamma \\ &= C(\|x_1 - x_2\| + \|g(x_1) - g(x_2)\|)^\alpha \leq C(\|x_1 - x_2\| + C\|x_1 - x_2\|)^\gamma \\ &\leq C\|x_1 - x_2\|^\gamma .\end{aligned}$$

□

Equipped with the results of lemmata 36,37 and corollary 35 we are able to establish the following local property of $\mathcal{D}$

Theorem 38. The map $\mathcal{D}'$ is locally Hölder continous.

Proof. Take p_0, p_1, p_2 as in the above proofs. Then for $\|p_0 - p_2\|$ sufficiently small, by Hölder continuity the distance $\left\|(\frac{\partial \mathcal{C}}{\partial u})^{-1}(p_0, \mathcal{D}(p_0)) - (\frac{\partial \mathcal{C}}{\partial u})^{-1}(p_2, \mathcal{D}(p_2))\right\|$ becomes small enough such that such that we may apply corollary 35. This yields the estimate

$$\begin{aligned}&\|[\mathcal{D}'(p_0) - \mathcal{D}'(p_2)]\| \\ &\leq \left\|\left[-(\frac{\partial \mathcal{C}}{\partial u})^{-1}(p_0, \mathcal{D}(p_0)) \circ \frac{\partial \mathcal{C}}{\partial p}(p_0, \mathcal{D}(p_0)) + (\frac{\partial \mathcal{C}}{\partial u})^{-1}(p_2, \mathcal{D}(p_2)) \circ \frac{\partial \mathcal{C}}{\partial p}(p_2, \mathcal{D}(p_2))\right]\right\| \\ &\leq \left\|-(\frac{\partial \mathcal{C}}{\partial u})^{-1}(p_0, \mathcal{D}(p_0)) \circ \frac{\partial \mathcal{C}}{\partial p}(p_0, \mathcal{D}(p_0)) + (\frac{\partial \mathcal{C}}{\partial u})^{-1}(p_2, \mathcal{D}(p_2)) \circ \frac{\partial \mathcal{C}}{\partial p}(p_0, \mathcal{D}(p_0))\right\| \\ &+ \left\|-(\frac{\partial \mathcal{C}}{\partial u})^{-1}(p_2, \mathcal{D}(p_2)) \circ \frac{\partial \mathcal{C}}{\partial p}(p_0, \mathcal{D}(p_0)) + (\frac{\partial \mathcal{C}}{\partial u})^{-1}(p_2, \mathcal{D}(p_2)) \circ \frac{\partial \mathcal{C}}{\partial p}(p_2, \mathcal{D}(p_2))\right\| \\ &\leq \left\|-(\frac{\partial \mathcal{C}}{\partial u})^{-1}(p_0, \mathcal{D}(p_0)) + (\frac{\partial \mathcal{C}}{\partial u})^{-1}(p_2, \mathcal{D}(p_2))\right\| \left\|\frac{\partial \mathcal{C}}{\partial p}(p_0, \mathcal{D}(p_0))\right\| \\ &+ \left\|(\frac{\partial \mathcal{C}}{\partial u})^{-1}(p_2, \mathcal{D}(p_2))\right\| \left\|\frac{\partial \mathcal{C}}{\partial p}(p_0, \mathcal{D}(p_0)) - \frac{\partial \mathcal{C}}{\partial p}(p_2, \mathcal{D}(p_2))\right\| \\ &\leq C(\|p_0 - p_2\|_{\mathcal{P}} + \|\mathcal{D}(p_0) - \mathcal{D}(p_2)\|_{W(0,T)})^\gamma \left\|\frac{\partial \mathcal{C}}{\partial p}(p_0, \mathcal{D}(p_0))\right\| \\ &+ C(\|p_0 - p_2\|_{\mathcal{P}} + \|\mathcal{D}(p_0) - \mathcal{D}(p_2)\|_{W(0,T)})^\gamma\end{aligned}$$

The last estimate follows by the statements on the continuity and boundedness of $\frac{\partial \mathcal{C}}{\partial p}$ and $\left(\frac{\partial \mathcal{C}}{\partial u}(\cdot, \cdot)\right)^{-1}$.
The Lipschitz continuity of $\mathcal{D}$ yields the claim. □

3.3 The derivative of the discrepancy term

The underlying physical experiment consists of sampling concentrations of fluids on a bounded domain. In reality data can only be gathered at finitely many points in time. Each such snapshot is itself an averaging process over a time interval, however short it may be. Similarly, in space we will also gather concentration values only on a finite grid, which has the same consequence of measuring averaged concentrations instead of point evaluations. So the concentrations we intend to measure or simulate will be not point evaluations but approximations of these. Apart from resembling the physical reality of data sampling more accurately, this also allows to obtain measurable functions as right-hand sides in the resulting adjoint PDE (see below). Point evaluations as observation operators would result in atomic measures ("delta functions") as right-hand sides so the regularity of the solution would be lost.
These practical issues necessitate the projection of our simulated solution into some appropriate observation space $\mathcal{O}$ when we want to compute the discrepancy between the simulated evolution and measured data. More precisely we are interested in minimizing the functional

$$1/2 \left\| \mathcal{Obs} \circ \mathcal{D}(p) - y_{data} \right\|_{\mathcal{O}}^2 ,$$

where $\mathcal{Obs} : W(0,T) \to \mathcal{O}$ is the observation map on the solution. For the observation space we put $\mathcal{O} = (L_{\mathfrak{q}}(\tilde{U}, \mathbb{R}^d))^N$, where $\tilde{U} = 1, \ldots, M$ is the measurement grid of patches U_j such that $U = \cup_j U_j$, $j = 1, \ldots, M$. Each of the N snapshots is obtained by averaging over mutually disjoint short intervals $I_i \subset [0,T]$, $i = 1, \ldots, N$ at each grid patch U_j, with integration weights c_{ij}. For the simulated solution this amounts to the map

$$v \mapsto \left(c_{ij} \int_0^T \int_U \chi_{ij}(t) v(x,t) dx dt \right), \; i = 1, \ldots, N, \; j = 1, \ldots, M$$

where $\chi_{ij} : U \times [0,T] \to \mathbb{R}$ is the characteristic function of $U_j \times I_i$ and c_{ij} the normalizing weight. The attentive reader realizes that we merely approximate a point evaluation
An intermediate step in the minizing task from above is to find the derivative of the above functional. Formally the derivative reads

$$\left(\mathcal{D}'(p_0)\right)^* (\mathcal{Obs})^* \, \mathcal{Obs} \left(\mathcal{D}(p_0) - y_{data}\right) .$$

Clearly,

$$\left(\mathcal{D}'(p_0)\right)^* = - \left(\frac{\partial \mathcal{C}}{\partial p}(p_0, u) \right)^* \left(\frac{\partial \mathcal{C}}{\partial u}^{-1} (p_0, u) \right)^* ,$$

which amounts to the adjoint of a multiplication operator and of a linear solution operator of a parabolic PDE. So we need to identify the action of $\left(\frac{\partial \mathcal{C}}{\partial p}(p_0,u)\right)^*$, $\left(\frac{\partial \mathcal{C}}{\partial u}^{-1}(p_0,u)\right)^*$, and $\mathcal{Obs}^*$.

Concerning $\mathcal{Obs}$

An application of the definition of the adjoint operator reveals

$$\begin{aligned}&\mathcal{Obs}^*:\mathcal{O}'\to(W(0,T))',\\&\qquad(f_1,\dots,f_{N_T})\mapsto\sum_{ij}c_{ij}\chi_{ij}(x,t)\cdot f_{ij}\in L_{\mathsf{q}'}(0,T,L_{\mathsf{q}'}(\tilde{U},\mathbb{R}^d)),\ (f_{ij})_j\in L_{\mathsf{q}}(\tilde{U},\mathbb{R}^d)\end{aligned}$$

because $(\mathcal{Obs}^*(f),v)_{W(0,T))',W(0,T)}=(f,\mathcal{Obs}(v))_{\mathcal{O}',\mathcal{O}}=\sum_{ij}c_{ij}\int_{U_j}\int_{I_i}f_{ij}v(x,t)dxdt$, for any $v\in W(0,T)$.

Concerning $(\frac{\partial\mathcal{C}}{\partial p})$

Denote $A=(\frac{\partial\mathcal{C}}{\partial p})(p_0,u)):\mathcal{P}\to G\times\mathcal{V}'$.
Let $\tilde{g}=(g_0,g)\in G\times\mathcal{V}$, $p_1\in\mathcal{P}$. For the adjoint operator we obtain

$$\begin{aligned}(A^*(\tilde{g}),p_1)_{\mathcal{P}',\mathcal{P}}&=(\tilde{g},A(p_1))_{(G,G)\times(\mathcal{V}',\mathcal{V})}\\&=0+\int_0^T\int_U g[\mathscr{A}_1u-R_1\Phi(W_0u)-R_0\Phi'(W_0u)W_1u]dxdt.\end{aligned}$$

Thus the adjoint operator A^* maps

$$\tilde{g}=(g_0,g)\mapsto(\langle\nabla g_i,\nabla u_i\rangle,g_iu_i,-g_i\cdot\Phi((W_0u)_i),r_{ij},\ i=1,\dots,d),$$

where $(r_{ij})_{ij}=(-g_i\cdot R_{0,i}\Phi'((W_0u)_i)u_j)_{ij},\ i,j=1,\dots,d$.The action of the last entry as an element of $L_{p'_W}(0,T,L_{p'_W}(U,\mathbb{R}^{d\times d}))$ on an admissible matrix $\tilde{W}$ then reads as:

$$\tilde{W}\mapsto\int_0^T\int_U\sum_i\sum_j r_{ij}\tilde{W}_{ij}dxdt.$$

Concerning $(\frac{\partial\mathcal{C}}{\partial u})$

The operator $(\frac{\partial\mathcal{C}}{\partial u})^{-1}(p_0,u))$ is a linear solution operator for a problem of the type as eqn. (7). Denote by $B=(\frac{\partial\mathcal{C}}{\partial u})(p_0,u))$. Then for the dual we find $B^*:G\times\mathcal{V}\to(W(0,T))'$. So we pick some $\tilde{g}=(g_0,g)\in G\times\mathcal{V}_{\mathsf{q}}$. The elements in $(W(0,T))'$ we need to consider are of the kind $w\in L_{\mathsf{q}'}(0,T,L_{\mathsf{q}'}(U,\mathbb{R}^d))$ (see above).
To ease notation we denote the differential operator $v\mapsto\mathscr{A}v-R_0\Phi(W_0u)W_0v$ by $\mathscr{K}v$. With these ideas and conventions in mind we investigate more closely how the adjoint operator acts.
So if $B^*(\tilde{g})=w$ in $(W(0,T))'$, we test with some $h\in W(0,T)$, $h(0)=0$ to obtain

$$(h'+\mathscr{A}_0h-R_0\Phi'(W_0u)W_0h,v)_{\mathcal{V}',\mathcal{V}}=(w,h)_{L_{\mathsf{q}'},L_{\mathsf{q}}}.$$

Thus by the denseness of $\{h:h(0)=0,\ h\in W\}$ in $\mathcal{V}$ this is in the distributional sense

$$-v'+\mathscr{K}v=w\quad\mathcal{V}',\quad v(T)=0.$$

By rem. 62 in the appendix $(B^*)^{-1} = (B^{-1})^*$, so we have identified the dual of $(\frac{\partial \mathcal{C}}{\partial u})^{-1}$ as the solution map of an adjoint PDE.

Theorem 39. Putting the results of this subsection together, the action of the derivative of

$$\|\mathcal{Obs} \circ \mathcal{D}(p_0) - y_{data}\|_{\mathcal{O}}^2$$

for given $p_0 \in \mathcal{P}$ is the following sequence of operations:

1. Solve $u' + \mathscr{A}_0 u = F(u),\ u(0) = u_0$.
2. Get the projection $\mathcal{Obs}(u)$ of the solution u into $\mathcal{O}$.
3. Compute $w = \mathcal{Obs}(u) - y_{data}$.
4. Generate $\mathcal{Obs}^*(w)$.
5. Solve $-v' + \mathscr{A}_0 v - F'(u)v = \mathcal{Obs}^*(w),\ v(T) = 0$
6. Compute $(\langle \nabla v_i, \nabla u_i \rangle, v_i u_i, -v_i \cdot \Phi((W_0 u)_i), -v_i \cdot R_{0,i}\Phi'((W_0 u)_i)u_j),\ i, j = 1, \ldots, d)$.

4 Regularization

4.1 Tikhonov regularization in the parameter space

In order to solve the present inverse problem by soft shrinkage techniques, one has to ensure Tikhonov regularization is applicable in the given setting. To do so, we will prove the basic theorem of Tikhonov regularization for our setup.
The standard regularization theory will be introduced in a rather abstract setting. After that we will show that our particular problem fits this setting.

Definition 40 (Tikhonov Funktional)**.** Let X, Y closed subsets of Banach spaces, such that X can be equipped with a topology τ. X is assumed to be τ closed. The unit ball in Y is assumed to be weakly compact. Let $J : X \to \mathbb{R}^+$ be a τ lower semicontinuous functional with τ precompact level sets. Assume the set $J(X) \cap \mathbb{R}$ is nonempty. Let furthermore $\mathscr{D} : X \to Y$ be a (norm-norm) continuous map, whose graph $\mathcal{G}$ is (τ, weak) closed. For some data sample g we have $g \in Y$. Then we define the Tikhonov functional by

$$K_{\alpha,g}(u) = \|\mathscr{D}(u) - g\|^2 + \alpha J(u) \tag{11}$$

Theorem 41 (Existence of a minimizer)**.** Let $K_{\alpha,g}(u)$ be as defined above. Then there exists a minimizer for the Tikhonov functional, which lies in $\overline{X}$.

Proof. See [Bur08, ch. 4.1,4.2]. □

Theorem 42 (Continuity of the regularized solution)**.** Let the Tikhonov functional be as above. Let $y_n \to y^\delta$ in Y and x_n, x^δ the minimizers for K_{α,y_n}, K_{α,y^δ} respectively. Then we can find a subnet (x_{n_k}), that converges in τ towards $x \in X$, where x is a minimizer of K_{α,y^δ}.

Proof. See [Bur08, ch. 4.1,4.2]. □

Theorem 43 (Tikhonov regularization)**.** Considering the Tikhonov functional and the entire notation from above, let $y \in rg(A)$ and J possess a minimizer x among all solutions of $\mathscr{D}(u) = y$. Choose the parameter $\alpha = \alpha(\delta, y^\delta)$, such that

$$\alpha \to 0, \quad \frac{\delta^2}{\alpha} \to 0, \quad \text{for} \quad \delta \to 0.$$

Then for every net of parameters (α_n) there is a τ convergent net $x_{\alpha_n}^{\delta_n}$, whose τ-limit is a minimizer of J and a solution of $\mathscr{D}(u) = y$.

Proof. See [Bur08, ch. 4.1,4.2]. □

Before we proceed, we remind the reader that weak* compactness and weak* sequential compactness are in general not equivalent (unlike for weak compactness by the Eberlein-Smulian theorem ([Con00, thm. 13.1])), which is why we cannot avoid nets here. The reader unfamiliar with this concept may either confer [Fol99, ch 4.3] or consider nets as generalized sequences that operate analogously to sequences.

Now the theory will be applied to our setting: We equip the admissible parameter set $\mathcal{P}_\lambda$, $\mathcal{P}_R$, and $\mathcal{P}_W$ with the L_∞ norm. Hence we obtain closed, bounded subsets of Banach spaces and denote these spaces as $\tilde{\mathcal{P}}_\lambda$, $\tilde{\mathcal{P}}_R$, and $\tilde{\mathcal{P}}_W$. The weak* topology shall then take the role of the τ topology in def. 40 for these three parameters, ie we have τ_λ, τ_R, and τ_W topologies on the respective spaces. The τ closedness for the spaces is clear by the weak* compactness of closed balls (Alaoglu's theorem, cf. [Fol99, thm. 5.18]).
Concerning the parameter D, we will now restrict the admissible set. Define $\tilde{\mathcal{P}}_D = \mathcal{P}_D \cap \mathcal{B}$, where $\mathcal{B} = \{f \in L_{\mathfrak{q}'}(0,T,W^1_{\mathfrak{q}'}(U,\mathbb{R}^d)) : \|f\|_{L_{\mathfrak{q}'}(0,T,W^1_{\mathfrak{q}'}(U,\mathbb{R}^d))} \leq C_{\mathcal{P},3}\}$ for some (possibly large) constant $C_{\mathcal{P},3} > 0$. The norm of $\tilde{\mathcal{P}}_D$ shall be the L_∞ norm just as for $\mathcal{P}_D$.
For the τ_D closedness of $\tilde{\mathcal{P}}_D$ consider the following

Lemma 44. The set $\tilde{\mathcal{P}}_D$ is τ compact and therefore in particular τ closed.

Proof. Given an arbitrary net D_k in $\tilde{\mathcal{P}}_D$. By the reflexivity of $L_{\mathfrak{q}'}(0,T,W^1_{\mathfrak{q}'}(U,\mathbb{R}^d))$ (cf. [Sho97, thm. III.1.5]) and Alaoglu's theorem $\mathcal{B}$ is weakly compact. So we can find a subnet p_n that converges weakly to some $D \in L_{\mathfrak{q}'}(0,T,W^1_{\mathfrak{q}'}(U,\mathbb{R}^d))$. By the weak lower semicontinuity of the norm we find $\|p\|_{L_{\mathfrak{q}'}(0,T,W^1_{\mathfrak{q}'}(U,\mathbb{R}^d))} \leq C_{\mathcal{P},3}$.
Now by the weak* compactness of $\tilde{\mathcal{P}}_D$ we can find a subnet D_m of D_n, that converges in the weak* topology to some $\tilde{D}$. One easily verifies that the $L_{\mathfrak{q}'}(0,T,W^1_{\mathfrak{q}'}(U,\mathbb{R}^d))$ weak convergence implies $L_{\mathfrak{q}'}(0,T,L_{\mathfrak{q}'}(U,\mathbb{R}^d))$ weak convergence to the same limit D. Similarly, weak* convergence of D_m implies $L_{\mathfrak{q}'}(0,T,L_{\mathfrak{q}'}(U,\mathbb{R}^d))$ weak convergence to the same limit D by the finiteness of the underlying measure space. Therefore $\tilde{D} = D$. So we have constructed a subnet that converges in $\tilde{\mathcal{P}}_D$. □

The product of these respective admissible parameter sets with mentioned norms we shall denote with $\mathcal{P}^\infty$ and equip with the product norm. The τ topology is then induced as the product of the respective τ topologies on $\tilde{\mathcal{P}}_{...}$. The considered subset $\mathcal{P}^\infty$ is closed and bounded and τ closed as as desired in def. 40.
The space $\mathcal{O}$, a finite dimensional space, satisfies the requirements of the space Y in def. 40.
The map $\mathscr{D}$ from above, as the notation suggests, is defined by $\mathit{Obs} \circ \mathcal{D} \circ I : \mathcal{P}^\infty \to \mathcal{O}$, where $I : \mathcal{P}^\infty \to \mathcal{P}$ is the natural embedding. One easily verifies differentiability.
The penalty term J we do not specify further at this point and only assume all requirements of the theory to be satisfied. Prominent examples, which we will use later, are Besov penalty terms.
The only major task that remains to prove is the (weak*, weak) closedness of the graph of $\mathscr{D}$.

Lemma 45. The graph $\mathcal{G} = \{(p,u) \in \mathcal{P}^\infty \times W(0,T) : u = \mathscr{D}(p)\}$ of $\mathscr{D}$ is (weak*,weak) closed.

Proof. Let $(p_k) = (\lambda_k, D_k, R_k, W_k)$ be a net of parameters from $\mathcal{P}^\infty$, such that $p_k \rightharpoonup^* p_0$ and $\mathscr{D}(p_k) = u_k \rightharpoonup u$. Define $\mathcal{D}(p_k) = z_k$, $\mathcal{D}(p_0) = z$, i.e. $u_k = \mathit{Obs}(z_k)$.
We need to show $\mathscr{D}(p_0) = u$. To arrive at this result we will proceed in two steps.

1. We will first show we can find some weakly convergent subnet z_n with weak limit z such that $\mathcal{O}bs(z) = u$.

2. Then we prove $\mathcal{D}(p_0) = z$, since this implies $u = \mathcal{O}bs(z) = \mathcal{O}bs(\mathcal{D}(p_0)) = \mathscr{D}(p_0)$. To this end we recur to the implicit definition of $\mathcal{D}$ via the operator $\mathcal{C}$.

The reader is to mind the difference of $\mathscr{D}$ and $\mathcal{D}$ in the following.
By Banach-Steinhaus' theorem ([Fol99, thm. 5.13]) the net (p_k) is norm bounded and hence by lemma 32 the net (z_k) is $W(0,T)$ norm bounded. Therefore we can find a subnet that converges in the $W(0,T)$ weak topology to some $z \in W(0,T)$ by Alaoglu's theorem. Furthermore, by theorem 8 this subnet, denoted by (z_n), converges $L_{\mathsf{q}}(0,T,L_{\mathsf{q}}(U,\mathbb{R}^d))$ strongly, since compactness of a linear operator implies complete continuity ([Con00, VI,§3]). That implies for any subdomains $I \subset [0,T]$, $U_j \subset U$

$$\int_0^T \int_U z_n(x,t)\chi_{U_j \times I} dxdt \to \int_0^T \int_U z(x,t)\chi_{U_j \times I} dxdt.$$

So we gather $\mathcal{O}bs(z_n) \to \mathcal{O}bs(z)$. Then this limit $\mathcal{O}bs(z)$ must coincide with the original (weak) limit u.
Secondly, let the elliptic operator $\mathscr{A}_n$ be induced by p_n and $F_n(\cdot) = R_n\Phi(W_n\cdot)$, $n = 0,1,\ldots$ denote the functionals as they appear in our PDE (eqn. 6). By section 3.1 we know $\mathcal{C}(p_n, z_n) = (z_n(0) - u_0, z_n' + \mathscr{A}_n z_n - F_n(z_n)) = (0,0)\ \forall n$.
If we can show for any arbitrary $y \in \mathcal{V}_{\mathsf{q}}$

$$\mathcal{C}(p_0, z) = (0,0),$$

then z is the unique solution to the Cauchy problem $v \in W(0,T): \ v' + \mathscr{A}_0 v - F_0(v) = 0 \text{ in } \mathcal{V}, \quad v(0) = u_0$, i.e. $z = \mathcal{D}(p_0)$.
We start out by inspecting the first entry of $\mathcal{C}(p_0, z)$. Since $z_n - z \rightharpoonup 0$ $W(0,T)$ weakly and $z_n(0) - z(0) = u_0 - z(0)$, we use lemma 11 to see that $\lim z_n(0) = z(0)$ in G. Since $z_n(0) = u_0\ \forall n$ that implies $z(0) = u_0$ (i.e. $\mathcal{C}_1(p_0, z) = 0$).
Next we prove $z' + \mathscr{A}_0 z - F_0(z) = 0$:
Pick some arbitrary $y \in C(U \times [0,T]) \cap \mathcal{V}_{\mathsf{q}'}$. Consider the estimate

$$\begin{aligned} |\mathcal{C}_2(p_m, z_m)(y) - \mathcal{C}_2(p_0, z)(y)| &= |(z_m - z)'(y) + (\mathscr{A}_n z_n - \mathscr{A} z)(y) + (F_n(z_n) - F_n(z))(y)| \\ &\le |(z_n - z)'(y)| + |([\mathscr{A}_n - \mathscr{A}]z)(y)| \\ &\quad + |(\mathscr{A}_n[z_n - z])(y)| + |(F_n(z_n) - F(z))(y)| \end{aligned}$$

It is straightforward to see that the first two summands in the last estimate converge to 0 by the assumed weak and weak* convergence (the sup on $\|p_n\|$ exists by the Banach-Steinhaus theorem).
For the third summand we find

$$|(\mathscr{A}_n[z_n - z])(y)| = \left| \int_0^T \int_U D_n \nabla(z_n - z)\nabla y + \lambda_n(z_n - z) y dxdt \right|.$$

By thm. 8, we have strong $L_\mathfrak{q}(0,T,L_\mathfrak{q}(U,\mathbb{R}^d))$ convergence of $z_n - z$ to zero. Therefore

$$\int_0^T \int_U \lambda_n(z_n - z) y dx dt \le \|\lambda_n\|_{\mathcal{P}_\lambda} \|z_n - z\|_{L_\mathfrak{q}(0,T,L_\mathfrak{q}(U,\mathbb{R}^d))} \|y\|_{L_{\mathfrak{q}'}(0,T,L_{\mathfrak{q}'}(U,\mathbb{R}^d))}$$

converges to 0, since by Banach-Steinhaus $\|\lambda_n\|_{\mathcal{P}_\lambda}$ is bounded for all indices n.
Concerning the diffusion part of the integral, use Gauss' formula to arrive at

$$\begin{aligned}\left|\int_0^T \int_U D_n \nabla(z_n - z)\nabla y dx dt\right| = &\left|\int_0^T \int_U -\nabla(D_n \nabla y)(z_n - z) dx dt\right| \\ &+ \left|\int_0^T \int_{\partial U} \langle (D\nabla y, \nu)\rangle (z_n - z) d\sigma\right|.\end{aligned}$$

Precisely at this point we need the extra regularity of D, which necessitated the restriction of $\mathcal{P}_D$ to $\tilde{\mathcal{P}}_D$. Since $D \in \mathcal{B}$ and due to the smoothness of y, the operator $-\nabla(D_n \nabla y)$ is uniformly bounded in $L_{\mathfrak{q}'}(0,T,L_{\mathfrak{q}'}(U,\mathbb{R}^d))$. So as before, the strong $L_\mathfrak{q}(0,T,L_\mathfrak{q}(U,\mathbb{R}^d))$ convergence of $(z_n - z)$ lets the first integral converge to 0. For the boundary integral over $[0,T] \times \partial U$ we invoke thm. 8 together with [Ada78, 6.3 Thm.,Part III] to see the trace operator is compact and hence completely continuous. Therefore we also have strong convergence for the trace of $(z_n - z)$ and the boundary integral converges to 0.

For the term $|(F_n(z_n) - F(z))(y)|$ we find:

$$\begin{aligned}|(F_n(z_n) - F(z))(y)| &\le \|R_n\| \, |[\Phi(W_n z_n) - \Phi(Wz)](y)| + |(R_n - R)[\Phi(Wz)(y)]| \\ &\le \sup_n(\|p_n\|) \sup(\Phi') |[(W_n z_n) - (Wz)](y)| + |(R_n - R)[\Phi(Wz)(y)]|.\end{aligned}$$

The last summand in the last estimate converges by weak convergence of the parameters. For the first summand, considerations analogous to those for the differential operator part above show convergence.
We have established

$$\lim_n \mathcal{C}_2(p_n, z_n) = \mathcal{C}_2(p_0, z).$$

Clearly, $\mathcal{C}_2(p_n, z_n) = 0 \ \forall n$, so $\mathcal{C}_2(p_0, z) = 0$ as desired. □

Now we will give an example for the functional J which will return in the numerical applications. Let $(\psi_i)_i$ denote some orthonormal basis of $\mathcal{H} = \left(L_2(0,T,L_2(U,\mathbb{R}^d))\right)^3 \times L_2(0,T,L_2(U,\mathbb{R}^{d\times d}))$, $(w_i)_i$ nonnegative weights. Put

$$J(u) = \sum_i w_i |\langle p, \psi_i\rangle|^p.$$

The absolute of the scalar product in $\mathcal{H}$ is certainly $\mathcal{H}$ weakly lower semicontinuous and any $\mathcal{P}^\infty$ weak* convergent net $(p_n)_n$ is also $\mathcal{H}$ weakly convergent (since $\mathcal{P}^\infty \subset \mathcal{H}$). Therefore we directly infer the $\mathcal{P}$ weak* lower semicontinuity of J.
Concerning precompactness of level sets, we just remark that the space $\mathcal{P}^\infty$ is itself weak* compact by lemma 44. Thus sublevel sets are trivially weak* precompact.

So we have established weak* precompactness of sub levelsets for our choice of J.
Notice that the implications in the three classical theorems make no topological statement but in one case where the existence of a $\mathcal{P}^\infty$ weak* convergent is proved. As the readers can easily verify on their own, this implies as well the $\mathcal{P}$ weak convergence of that sequence. This bridges the gap to the weak $\mathcal{P}$ topology more natural for the norm topology defined on $\mathcal{P}$.

4.2 Sparsity and Implementation

The preceding subsection on Tikhonov regularization enabled us to abstractly formulate how we can recover an approximate solution to our inverse problem when dealing with noisy data. A rather tricky issue we face in the numerical computation of a minimizer for the functional of eqn. (11). The work of [DDD04] has become a popular approach for this task. Since we are dealing with a nonlinear problem, we shall therefore use the generalized version of this iterated soft shrinkage approach as outlined in [BLM09]. Our setting differs from [BLM09] in the fact that our solution and parameter spaces are subsets of Banach spaces rather than Hilbert spaces. By replacing inner products in the claims and proofs of [BLM09] a lot of the statements carry over without major modification. In particular, the requirements the authors of [BLM09] impose on the derivative of the forward operator we can ensure by our findings on the continuity properties (subsection 3.2). Details of this generalization to our setting will appear in a future publication.
For an outline of the implementation of this generalized conditional gradient method the reader is recommended to confer [DFM$^+$11] and [BLM09]. A few building blocks necessary to modify the structure for our setting still need to be worked out in detail. Therefore results on this part will be presented to the public in the near future.

5 Appendix

5.1 Integration Theory

In this subsection any underlying measure space, even if not stated explicitly, is assumed to be a finite measure space.

Theorem 46. If $0 < p < q < r \leq \infty$, then $L^p\hat{L}^r \subset L^q$ and $\|f\|_{L^q} \leq \|f\|_{L^p}^{\lambda} \|f\|_{L^r}^{1-\lambda}$, where $\lambda \in (0,1)$ is defined by

$$\lambda = \frac{1/q - 1/r}{1/p - 1/r}.$$

Proof. See [Fol99][Prop. 6.10]. □

As a useful application the Hölder inequality, we have

Theorem 47. Given $1 \leq q < p < \infty$, $r = \frac{pq}{p-q}$, some finite measure space $(\Omega, \mathcal{S}, \mu)$, $f \in L^r$, and $g \in L^p$ we obtain

$$\|fg\|_{L^q} \leq \|f\|_{L^r} \|g\|_{L^p},$$

and there is some g for which equality holds (see also [AZ90, pp. 91]).

Proof. Clearly $fg \in L_q \Leftrightarrow |fg|^q \in L_1$. Since $|g|^q \in L_{p/q}$, any such $f \in \mathbb{M}_{p,q}$ must be such that $|f|^q \in L_s$ by duality theory, where s is the conjugate Hölder exponent for p/q, ie $1/s + q/p = 1$. This implies $f \in L_r$, where $r = sq$.
By Hölder we find

$$\|fg\|_{L_q} \leq \|f\|_{L_r} \|g\|_{L_p},$$

so

$$\|f\|_{\mathbb{M}} \leq \|f\|_{L_r}.$$

Since the Hölder estimate is sharp, we can find some $g \in L_p$, such that equality holds. So we have

$$\|f\|_{\mathbb{M}} = \|f\|_{L_r}.$$

Therefore $\mathbb{M}_{p,q} \subset L_r$ isometrically and the converse isometric inclusion is obvious by the Hölder estimate.
This establishes the claim. □

Theorem 48. On a finite measure space (X, μ) we have for $1 \leq p < q \leq \infty$ $L^q(\mu) subset L^p(\mu)$ and

$$\|f\|_{L^p} \leq \|f\|_{L^q} \mu(X)^{1/p-1/q}.$$

Proof. See [Fol99][Prop. 6.12]. □

Theorem 49. Assume we are in a finite measure space. Given a sequence $f_n \to f$ in L_p, $|f_n| \leq M$ globally ae, $p \in [1, \infty)$, and a sequence $g_n \to g$ in L_q, $q \in [1, \infty)$, we find the convergence

$$f_n g_n \to fg \text{ in } L_q.$$

Proof. By the global ae. boundedness of the functions f_n we have

$$\|f_n g_n - f_n g\|_q \leq M \|g_n - g\|_q \to 0.$$

Again by global boundedness of the f_n , each element of the sequence $(f_n g) \subset L_q$ is ae pointwise bounded by the L_q function $M|g|$. By standard arguments from measure theory we can find an ae pointwise convergent subsequence (f_k), ie $(f_k g)$ is ae pointwise convergent. We may therefore apply the dominated convergence theorem to find

$$f_k g \to fg \text{ in } L_q.$$

By standard subsequence arguments we find convergence for the whole sequence $(f_n g)$. This yields the desired result

$$\|f_n g_n - fg\|_q \leq M \|g_n - g\|_q + \|f_n g - fg\|_q \to 0.$$

□

As a corollary we show

Corollary 50. Let $X_i = \{f \in L_{r_i}(V) : \|f\| \leq M_i\}, i = 1, \dots, n$, where V is a finite measure space and $M_i > 0$ are constant numbers. For the exponents assume $\sum_i 1/r_i + 1/s = 1/t \leq 1$. Define $K : X_1 \times \ldots \times X_n \times L_s(V) \to L_s(V)$ by $(f_1, \dots, f_n, f) \mapsto \left\|\int_V f_1 \cdot \ldots \cdot f_n \cdot f\right\|_{L_t(V)}$. Then K is a continuous map.

Proof. First of all the reader notices that by the pointwise ae bound in the $f_1^{(k)}, \dots, f_n^{(k)}$ the product of all the functions is in $L_t(V)$, so K is well-defined.
Let $= (f_1^{(k)}, \dots, f_n^{(k)}, f^{(k)})$ be a sequence converging to $(f_1, \dots, f_n f)$ in the respective norms. Clearly the product $g_k \Pi_i f_i^{(k)}$ is in $L_{\tilde{r}}(V)$, where $1/\tilde{r} = \sum_i 1/r_i$. The sequence (g_k) is also ae bounded pointwise. So we may apply the above theorem.

□

5.2 Sobolev spaces

Theorem 51. Let u be in $W(0,T)$. Then the truncation $\tilde{u} = \max\{k, u\}$ (meaning componentwise max) is also from $W(0,T)$ for any $k \in \mathbb{R}^d$.

Proof. See the proof of Lemma 4.4 in chapter II of [LSU68]. □

Theorem 52. Let $U \subset \mathbb{R}^n$ be an open, Lipschitz bounded domain. Then we have the following (continuous) embeddings. For $n = mp$ we have

$$W^{m,p}(U, \mathbb{R}^d) \hookrightarrow L^q(U, \mathbb{R}^d)$$

where $p \leq q \leq \infty$.
For $mp < n$ and $n - mp < k \leq n$ we have

$$W^{m,p}(U, \mathbb{R}^d) \hookrightarrow L^q(U, \mathbb{R}^d)$$

where $p \leq q \leq np/(n-mp)$.
For $mp > n > (m-1)p$, $j \geq 0$ we have the continuous embedding

$$W_p^{m+j}(U, \mathbb{R}^d) \hookrightarrow C_\alpha^j(\overline{U}, \mathbb{R}^d),$$

where $\alpha \in (0, m-(n/p))$ and $C_\alpha^j(\overline{U}, \mathbb{R}^d)$ denotes the j times differentiable maps with α-Hölder continuous j-th derivative.

Proof. See [Ada78, V., 5.4] and adapt to the multidimensional image space $\mathbb{R}^d$. □

5.3 Young inequality

Remark 53. For $a, b, \epsilon, \tilde{p} > 0$ we have the estimate

$$ab \leq \epsilon a^{\tilde{p}} + C(\epsilon) b^{\tilde{q}},$$

where $C(\epsilon) = (\epsilon p)^{-\tilde{q}/p} q^{-1}$ and $\frac{1}{\tilde{p}} + \frac{1}{\tilde{q}} = 1$.

.

Proof. See [Eva08][p. 622]. □

5.4 Maximal parabolic regularity

Another theorem which is interesting in its own right is the following

Remark 54. A bounded, connected Lipschitz domain G with nonempty interior is in $\mathcal{R}_q$ for some $q \in (2, \infty)$. In case the domain is even C^1 bounded, then $G \in \mathcal{R}_q$ for any $q \in [2, \infty)$.
$\mathcal{R}_q$ is the collection of all regular sets in the sense of Gröger (see [Grö89]), such that the duality map $Id + div\nabla : W_q^1(G) \to (W_{q'}^1(G)$ is surjective.

Proof. The proof for the statement concerning Lipschitz domains can be found in [Grö89, Thm. 3], the statement concerning C^1 domains is in [Grö89, Rem. 7]. □

If our domain $U \subset \mathbb{R}^3$, we have in particular:

Remark 55. Take U closed, Lipschitz bounded, then there is some $q > 3$ that $U \in \mathcal{R}_q,\ q > 3$

Proof. See [HDR09, Prop. 7.1,vi)] □

These remarks ensure the applicability of the maximal regularity results of [HDR09] for our domain, including the case where U is an L-shaped domain of dimension $n = 2$ or 3.

Theorem 56. There exists some $\tilde{\epsilon} = \tilde{\epsilon}(U, C_{\mathcal{P},1}, C_{\mathcal{P},2})$, such that for $\mathfrak{q} \in (n, n+\tilde{\epsilon})$ and the pertaining definitions of $\mathcal{P}$ and $W(0,T)$ we can find a unique solution u for the problem in equation 7 with any right-hand side $f \in (\mathcal{V}_{\mathfrak{q}'})'$. This solution depends continuously on the right-hand side

$$\|u\|_{W(0,T)} \leq C(\|f\|_{(\mathcal{V}_{\mathfrak{q}'})'} + \|u_0\|_G)$$

and the constant $C = C(U, C_{\mathcal{P},1}, C_{\mathcal{P},2}, T)$.

Proof. Define $\tilde{W}(0,T) = \{u \in W(0,T) : u(0) = 0\}$. Let $L = \frac{d}{dt} + \Delta : \tilde{W}(0,T) \to (\mathcal{V}_{\mathfrak{q}'})'$, then by [HDR09, Thm. 5.4; 5.5,i); Prop. 7.1,vi)] this becomes an isomorphism, ie the problem

$$u' + C_{\mathcal{P},1}\Delta u = f, \quad f \in (\mathcal{V}_{\mathfrak{q}'})' u(0) = 0$$

has a unique solution u in $\tilde{W}(0,T)$.
The rest of the proof mimicks the proof of the two statements [ACFP07, Lem. 1.2, Prop. 1.3]. Similar to [ACFP07, Lem. 1.2] we find that since $-L$ is a sectorial operator (see [Lun95, Def. 2.0.1]), there is some $\omega > 0$ such that for all $\theta > \omega$ we have

$$\|\theta + L\| \leq \frac{M}{\theta},$$

where $M > 0$ depends on L. Then define $B(t) : \tilde{W}(0,T) \to (\mathcal{V}_{\mathfrak{q}'})'$ as the perturbation operator through

$$B(t)u(v) = \sum_i \int_0^T \int_U (D_i(x,t) - C_{\mathcal{P},1})(\nabla u_i(x,t), \nabla v_i(x,t)) dx dt + \int_0^T \int_U (\lambda_i(x,t) - C_{\mathcal{P},1}) u_i(x,t) v_i(x,t)) dx dt.$$

Clearly,

$$\|B(t)u\|_{(\mathcal{V}_{\mathfrak{q}'})'} \leq (C_{\mathcal{P},2} - C_{\mathcal{P},1}) \|u\|_{\tilde{W}(0,T)}.$$

By choosing θ sufficiently large, we obtain $(C_{\mathcal{P},2} - C_{\mathcal{P},1}) \leq \theta/2M$. So the crucial estimate in part b) of the proof of [ACFP07, Prop. 1.3] is

$$\left\|B(\theta + L)^{-1} f\right\|_{(\mathcal{V}_{\mathfrak{q}'})'} \leq \theta/2M \left\|(\theta + L)^{-1} f\right\|_{\tilde{W}(0,T)} \leq 1/2 \|f\|_{(\mathcal{V}_{\mathfrak{q}'})'}.$$

All other parts of the proof are literally as in [ACFP07, Prop. 1.3].
Concerning the bound on the norm of $(L+B)^{-1}$ we take a closer look at the proof again: $\theta + L + B = (I + B(\theta + L)^{-1})(\theta + L)$. From this we infer that the norm of $(\theta + L + B)^{-1}$ depends solely on $M, C_{\mathcal{P},1}, C_{\mathcal{P},1}$ (θ depends on $C_{\mathcal{P},1}, C_{\mathcal{P},2}$). Since the underlying PDE is only the exponentially shifted Cauchy problem, the norm of $(L + B)^{-1}$ only differs by an exponential rescaling factor $exp(\theta T)$. So as before, the norm depends only on $M, C_{\mathcal{P},1}, C_{\mathcal{P},1}$.
For the nonhomogeneous initial condition $u(0) = x$, part c) of the proof of [ACFP07,

Prop. 1.3] together with the norm equivalence $\|x\|_G = inf\{\|w\|_{W(0,T)} : w \in W(0,T), w(0) = x\}$ (see [Lun95, Chap. 1]) yields for the solution u of

$$u' + \mathscr{A}u = f \in (\mathcal{V}_{q'})', \quad u(0) = x \in G$$

the estimate

$$\|u\|_{W(0,T)} \leq C(\|f\|_{(\mathcal{V}_{q'})'} + \|x\|_G),$$

with $C = C(U, C_{\mathcal{P},1}, C_{\mathcal{P},2})$.

□

5.5 Inversion of operators

We present some standard theory on the inversion of bounded, linear operators between to Banach spaces X and Y. Proofs are taken from or mimick those in [Ber74, ch. 50]. The indices for the norms are omitted where it is clear which norm is taken. The i

Theorem 57. Let x be from $\mathscr{L}(X,X)$ (or $\mathscr{L}(Y,Y)$), such that $\|\mathscr{I}_X - x\| < 1$ then x is invertible and $\|x^{-1}\| \leq \frac{1}{1-\|\mathscr{I}_X - x\|}$.

Proof. See [Ber74, Corollary 50.3]. □

If $x \in \mathscr{L}(X,Y)$ an invertible element, ie its inverse $x^{-1} \in \mathscr{L}(Y,X)$ is bounded, we denote the set of such x by $G\mathscr{L}(X,Y)$.

Lemma 58. Let $x \in G\mathscr{L}(X,Y)$ an invertible element. Let $y \in \mathscr{L}(X,Y)$ be from the set $U = \{z : \|z - x\| < \|x^{-1}\|^{-1}\}$. Then $y \in G\mathscr{L}(X,Y)$.

Proof. Consider

$$\left\|\mathscr{I}_Y - yx^{-1}\right\|_{\mathscr{L}(Y,Y)} = \left\|(x-y)x^{-1}\right\|_{\mathscr{L}(Y,Y)} \leq \|x-y\|_{\mathscr{L}(X,Y)} \left\|x^{-1}\right\|_{\mathscr{L}(Y,X)} < 1.$$

This implies yx^{-1} is invertible by the preceding theorem. So, since x is bijective by assumption, $y = (yx^{-1})x$ must be as well. By the open mapping theorem y must then have a continous inverse. □

Theorem 59. The inversion map $Inv : G\mathscr{L}(X,Y) \to G\mathscr{L}(Y,X),\ x \mapsto x^{-1}$ is locally Lipschitz continuous.

Proof. Assume $x, y \in \mathscr{L}(X,Y)$. Then $y^{-1} - x^{-1} = x_n^{-1}(x-y)x^{-1}$, so we get

$$\left\|y^{-1} - x^{-1}\right\| \leq \left\|y^{-1}\right\| \|x-y\| \left\|x^{-1}\right\|.$$

Therefore it remains to show that for y sufficiently close to x we can bound the inverse y^{-1}.
Consider

$$\left\|\mathscr{I}_Y - x^{-1}y\right\| = \left\|x^{-1}(x-y)\right\| \leq \left\|x^{-1}\right\| \|x-y\|,$$

whence for y with $\|x-y\| < 1/2 \left\|x^{-1}\right\|^{-1}$ we see that $x^{-1}y \in \mathscr{L}(X,X)$ is invertible by the first theorem of this subsection and

$$\left\|(x^{-1}y^{-1})^{-1}\right\| \leq 2.$$

Rewritten, this implies

$$\left\|y^{-1}x\right\| \leq 2$$

and thus

$$\left\|y^{-1}\right\| = \left\|(y^{-1}x)x^{-1}\right\| \leq 2\left\|x^{-1}\right\|.$$

So for all y with $\|y-x\| \leq 1/2 \left\|x^{-1}\right\|^{-1}$

$$\left\|y^{-1} - x^{-1}\right\| \leq 2\left\|x^{-1}\right\|^2 \|x-y\|.$$

and continuity is established. □

Corollary 60. Assume $\|x^{-1}\| \leq \tilde{C}$ for all $x \in D$, where D is some bounded set, then for every $x \in D$

$$\left\|x^{-1} - y^{-1}\right\| \leq \tilde{C}^2 \|x-y\|, \quad B_{1/2\tilde{C}}(x) = \{y : \ \|y-x\| \leq 1/2\tilde{C}\}$$

where the constant $\tilde{C}$ is uniform for all $x \in D$.

Proof. By the estimate $\|x^{-1}\| \leq C$ we find $B_{1/2\tilde{C}}(x) \subset \{y : \ \|y-x\| \leq 1/2\|x^{-1}\|\}$ for any $x \in D$. So the Lipschitz estimate follows by the statement in the preceding theorem. □

Theorem 61. The inversion map Inv from above is even differentiable with derivative

$$Inv'(x)(h) = -x^{-1}hx^{-1}.$$

Proof. We only need to show

$$\left\|(x+h)^{-1} - x^{1-} + x^{-1}hx^{-1}\right\| = o(\|h\|)$$

for $h \to 0$.
To this end observe

$$\begin{aligned}
&(x+h)^{-1} - x^{1-} + x^{-1}hx^{-1} \\
&= [\mathscr{I}_X - x^{1-}(x+h) + x^{-1}hx^{-1}(x+h)](x+h)^{-1} \\
&= [\mathscr{I}_X - \mathscr{I}_X - x^{-1}h + x^{-1}h(\mathscr{I}_X + x^{-1}h)](x+h)^{-1} \\
&= [x^{-1}hx^{-1}h)](x+h)^{-1}
\end{aligned}$$

By the preceding theorem the last factor $(x+h)^{-1}$ remains bounded for bounded h. Therefore $\|(x+h)^{-1} - x^{1-} + x^{-1}hx^{-1}\| = o(\|h\|)$ as desired. □

Concerning the dual of the inverse of some invertible bounded linear operator $A : X \to Y$, X, Y being Banach spaces, the following interesting remark holds

Remark 62. Let $A : X \to Y$be a bounded linear operator, X, Y being reflexive Banach spaces. Assume A has a bounded linear inverse. Then for the dual opartor $A^* : Y' \to X'$ the identity

$$(A^{-1})^* = (A^*)^{-1} : X' \to Y'$$

holds.

Proof. Let $X^\perp$ denote the subspace of X' which vanishes on all of X, and $(X')^\perp$ denote the subspace of X that vanishes on all of X'.
By [Wer00, Thm IV.5.1] $Range(A^*)^\perp = Kernel(A) = \{0\}$, $Kernel(A^*) = Range(A))^\perp = \{0\}$ and $Range(A^*))$ is closed. So A^* is bijective as well. By the open mapping theorem therefore A^* has a bounded inverse.
Let $x' = A^*y' \in X'$, $Ax = y \in Y$. Then

$$\begin{aligned} ((A^{-1})^*x', y)_{Y',Y} &= ((A^{-1})^*A^*y', y)_{Y',Y} \\ &= (y', y)_{Y',Y} \\ &= ((A^*)^{-1}x', y)_{Y',Y}. \end{aligned}$$

This implies $(A^{-1})^* = (A^*)^{-1}$ on $Range(A^*)$. $Range(A^*))^\perp = Kernel(A) = \{0\}$ implies denseness of $Range(A^*)$ in Y' and by the closedness of $Range(A^*)$ this emcompasses all of Y'. □

References

[ACFP07] W. Arendt, R. Chill, S. Fornaro, and C. Poupaud, *L_p-maximal regularity for non-autonomous evolution equations*, J. Differential Equations **237** (2007), 1–26.

[Ada78] R. Adams, *Sobolev spaces*, Academic Press, San Diego, Cal., 1978.

[AZ90] J. Appell and P. Zabrejko, *Nonlinear Superposition Operators*, Cambridge University Press, Cambridge, UK, 1990.

[BBLM07] T. Bonesky, K. Bredies, D. A. Lorenz, and P. Maass, *A generalized conditional gradient method for nonlinear operator equations with sparsity constraints* , Inverse Problems **23** (2007), 2041–2058.

[Ber74] S. Berberian, *Lectures in Functional Analysis and Operator Theory*, Springer, New York, Heidelberg, Berlin, 1974.

[BLM09] K. Bredies, D. A. Lorenz, and P. Maass, *A generalized conditional gradient method and its connection to an iterative shrinkage method*, Comput. Optim. Appl. **42** (2009), 173–193.

[Bur08] M. Burger, *Inverse Problems (Lecture Notes)*, WWU Münster, Münster, 2008.

[Con00] J. Conway, *A Course in Functional Analysis*, Springer, New York, 2000.

[DDD04] I. Daubechies, M. Defrise, and C. DeMol, *An iterative thresholding algorithm for linear inverse problems with sparsity constraint*, Commun. Pure Appl. Math. **57** (2004), 1413–1457.

[DFM+11] S. Dahlke, U. Friedrich, P. Maass, T. Raasch, and R. Ressel, *An adaptive wavelet solver for a nonlinear parameter identification problem for a parabolic differential equation*, Preprint 86 within DFG Priority Program 1324 (2011).

[DU77] J. Diestel and J. Uhl, *Vector measures*, American Mathematical Society, Providence, Rhode Island, 1977.

[EHN96] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, Kluwer, Amsterdam, 1996.

[Eva08] L. C. Evans, *Partial Differential Equations*, American Mathematical Society, Providence, Rhode Island, 2008.

[Fol99] G. B. Folland, *Real analysis. Modern techniques and their applications. 2nd ed.*, Pure and Applied Mathematics. A Wiley-Interscience Series of Texts, Monographs, and Tracts. New York, NY: Wiley. xiv, 1999.

[GGZ74] H. Gajewski, K. Gröger, and K. Zacharias, *Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen*, Akademie-Verlag, Berlin, 1974.

[Grö89] K. Gröger, *A $W^{1,p}$ estimate for solutions to mixed boundary value problems for second order elliptic differential equations*, Math. Ann. **283** (1989), 679–687.

[Grö92] ______, *$W^{1,p}$ estimates of solutions to evolution equations corresponding to nonsmooth second order elliptic differential operators*, Nonlinear Anal., Theory Methods Appl. **18** (1992), 569–577.

[HDR09] R. Haller-Dintelmann and J. Rehberg, *Maximal parabolic regularity for divergence operators including mixed boundary conditions*, J. Differential Equations **247** (2009), 1354–1396.

[HK10] T. Hein and K. S. Kazimierski, *Modified Landweber iteration in Banach spaces - convergence and convergence rates*, Numerical Functional Analysis and Optimization **31** (2010), 1158 – 1189.

[Kaz11] K. S. Kazimierski, *Minimization of the Tikhonov functional in Banach spaces smooth and convex of power type by steepest descent in the dual*, Computational Optimization and Applications **48** (2011), 309–324.

[Lan94] A. Langenbach, *Über L-Diffbarkeit auf metrischen Räumen und implizite Funktionen*, Math. Nachr. **166** (1994), 55–65.

[Lou89] A. K. Louis, *Inverse und schlechtgestellte Probleme*, Teubner, Stuttgart, 1989.

[LSU68] O. A. Ladyzhenskaya, V. A. Solonnikov, and N. N. Uraltseva, *Linear and quasi-linear equations of parabolic type*, American Mathematical Society, Providence, Rhode Island, 1968.

[Lun95] A. Lunardi, *Analytic Semigroups and Optimal Regularity in Parabolic Problems*, Birkhäuser, Basel, 1995.

[MSR91] E. Mjolsness, D. Sharp, and J. Reinitz, *A Connectionist Model of Development*, Journal of Theoretical Biology **152** (1991), 429–453.

[RS95] J. Reinitz and D. Sharp, *Mechanism of Eve Stripe formation*, Mech. Dev. **49** (1995), 133–158.

[Sho97] R. E. Showalter, *Monotone operators in Banach space and nonlinear partial differential equations*, Mathematical Surveys and Monographs. 49. Providence, RI: American Mathematical Society, 1997.

[SMB+08] T. Schuster, P. Maass, T. Bonesky, K. S. Kazimierski, and F. Schöpfer, *Minimization of Tikhonov Functionals in Banach Spaces*, Abstract and Applied Analysis **2008** (2008).

[Tri95] H. Triebel, *Interpolation Theory, Function Spaces, Differential Operators*, Johann Ambrosius Barth Verlag, Heidelberg, 1995.

[Wer00] D. Werner, *Funktionalanalysis*, Springer, Berlin, Heidelberg, 2000.

[Yos80] K. Yosida, *Functional Analysis*, Springer, Berlin, Heidelberg, 1980.

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

[1] R. Ramlau, G. Teschke, and M. Zhariy. A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems. Preprint 1, DFG-SPP 1324, September 2008.

[2] G. Plonka. The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-dimensional Data. Preprint 2, DFG-SPP 1324, September 2008.

[3] E. Novak and H. Woźniakowski. Optimal Order of Convergence and (In-) Tractability of Multivariate Approximation of Smooth Functions. Preprint 3, DFG-SPP 1324, October 2008.

[4] M. Espig, L. Grasedyck, and W. Hackbusch. Black Box Low Tensor Rank Approximation Using Fibre-Crosses. Preprint 4, DFG-SPP 1324, October 2008.

[5] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch. Adaptive Wavelet Methods and Sparsity Reconstruction for Inverse Heat Conduction Problems. Preprint 5, DFG-SPP 1324, January 2009.

[6] E. Novak and H. Woźniakowski. Approximation of Infinitely Differentiable Multivariate Functions Is Intractable. Preprint 6, DFG-SPP 1324, January 2009.

[7] J. Ma and G. Plonka. A Review of Curvelets and Recent Applications. Preprint 7, DFG-SPP 1324, February 2009.

[8] L. Denis, D. A. Lorenz, and D. Trede. Greedy Solution of Ill-Posed Problems: Error Bounds and Exact Inversion. Preprint 8, DFG-SPP 1324, April 2009.

[9] U. Friedrich. A Two Parameter Generalization of Lions' Nonoverlapping Domain Decomposition Method for Linear Elliptic PDEs. Preprint 9, DFG-SPP 1324, April 2009.

[10] K. Bredies and D. A. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. Preprint 10, DFG-SPP 1324, April 2009.

[11] K. Bredies and D. A. Lorenz. Regularization with Non-convex Separable Constraints. Preprint 11, DFG-SPP 1324, April 2009.

[12] M. Döhler, S. Kunis, and D. Potts. Nonequispaced Hyperbolic Cross Fast Fourier Transform. Preprint 12, DFG-SPP 1324, April 2009.

[13] C. Bender. Dual Pricing of Multi-Exercise Options under Volume Constraints. Preprint 13, DFG-SPP 1324, April 2009.

[14] T. Müller-Gronbach and K. Ritter. Variable Subspace Sampling and Multi-level Algorithms. Preprint 14, DFG-SPP 1324, May 2009.

[15] G. Plonka, S. Tenorth, and A. Iske. Optimally Sparse Image Representation by the Easy Path Wavelet Transform. Preprint 15, DFG-SPP 1324, May 2009.

[16] S. Dahlke, E. Novak, and W. Sickel. Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings IV: Errors in L_2 and Other Norms. Preprint 16, DFG-SPP 1324, June 2009.

[17] B. Jin, T. Khan, P. Maass, and M. Pidcock. Function Spaces and Optimal Currents in Impedance Tomography. Preprint 17, DFG-SPP 1324, June 2009.

[18] G. Plonka and J. Ma. Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing. Preprint 18, DFG-SPP 1324, June 2009.

[19] G. Teschke and C. Borries. Accelerated Projected Steepest Descent Method for Nonlinear Inverse Problems with Sparsity Constraints. Preprint 19, DFG-SPP 1324, July 2009.

[20] L. Grasedyck. Hierarchical Singular Value Decomposition of Tensors. Preprint 20, DFG-SPP 1324, July 2009.

[21] D. Rudolf. Error Bounds for Computing the Expectation by Markov Chain Monte Carlo. Preprint 21, DFG-SPP 1324, July 2009.

[22] M. Hansen and W. Sickel. Best m-term Approximation and Lizorkin-Triebel Spaces. Preprint 22, DFG-SPP 1324, August 2009.

[23] F.J. Hickernell, T. Müller-Gronbach, B. Niu, and K. Ritter. Multi-level Monte Carlo Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 23, DFG-SPP 1324, August 2009.

[24] S. Dereich and F. Heidenreich. A Multilevel Monte Carlo Algorithm for Lévy Driven Stochastic Differential Equations. Preprint 24, DFG-SPP 1324, August 2009.

[25] S. Dahlke, M. Fornasier, and T. Raasch. Multilevel Preconditioning for Adaptive Sparse Optimization. Preprint 25, DFG-SPP 1324, August 2009.

[26] S. Dereich. Multilevel Monte Carlo Algorithms for Lévy-driven SDEs with Gaussian Correction. Preprint 26, DFG-SPP 1324, August 2009.

[27] G. Plonka, S. Tenorth, and D. Roşca. A New Hybrid Method for Image Approximation using the Easy Path Wavelet Transform. Preprint 27, DFG-SPP 1324, October 2009.

[28] O. Koch and C. Lubich. Dynamical Low-rank Approximation of Tensors. Preprint 28, DFG-SPP 1324, November 2009.

[29] E. Faou, V. Gradinaru, and C. Lubich. Computing Semi-classical Quantum Dynamics with Hagedorn Wavepackets. Preprint 29, DFG-SPP 1324, November 2009.

[30] D. Conte and C. Lubich. An Error Analysis of the Multi-configuration Time-dependent Hartree Method of Quantum Dynamics. Preprint 30, DFG-SPP 1324, November 2009.

[31] C. E. Powell and E. Ullmann. Preconditioning Stochastic Galerkin Saddle Point Problems. Preprint 31, DFG-SPP 1324, November 2009.

[32] O. G. Ernst and E. Ullmann. Stochastic Galerkin Matrices. Preprint 32, DFG-SPP 1324, November 2009.

[33] F. Lindner and R. L. Schilling. Weak Order for the Discretization of the Stochastic Heat Equation Driven by Impulsive Noise. Preprint 33, DFG-SPP 1324, November 2009.

[34] L. Kämmerer and S. Kunis. On the Stability of the Hyperbolic Cross Discrete Fourier Transform. Preprint 34, DFG-SPP 1324, December 2009.

[35] P. Cerejeiras, M. Ferreira, U. Kähler, and G. Teschke. Inversion of the noisy Radon transform on $SO(3)$ by Gabor frames and sparse recovery principles. Preprint 35, DFG-SPP 1324, January 2010.

[36] T. Jahnke and T. Udrescu. Solving Chemical Master Equations by Adaptive Wavelet Compression. Preprint 36, DFG-SPP 1324, January 2010.

[37] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Irregular Shearlet Frames: Geometry and Approximation Properties. Preprint 37, DFG-SPP 1324, February 2010.

[38] G. Kutyniok and W.-Q Lim. Compactly Supported Shearlets are Optimally Sparse. Preprint 38, DFG-SPP 1324, February 2010.

[39] M. Hansen and W. Sickel. Best m-Term Approximation and Tensor Products of Sobolev and Besov Spaces – the Case of Non-compact Embeddings. Preprint 39, DFG-SPP 1324, March 2010.

[40] B. Niu, F.J. Hickernell, T. Müller-Gronbach, and K. Ritter. Deterministic Multi-level Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 40, DFG-SPP 1324, March 2010.

[41] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Construction of Compactly Supported Shearlet Frames. Preprint 41, DFG-SPP 1324, March 2010.

[42] C. Bender and J. Steiner. Error Criteria for Numerical Solutions of Backward SDEs. Preprint 42, DFG-SPP 1324, April 2010.

[43] L. Grasedyck. Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorization. Preprint 43, DFG-SPP 1324, April 2010.

[44] M. Hansen und W. Sickel. Best m-Term Approximation and Sobolev-Besov Spaces of Dominating Mixed Smoothness - the Case of Compact Embeddings. Preprint 44, DFG-SPP 1324, April 2010.

[45] P. Binev, W. Dahmen, and P. Lamby. Fast High-Dimensional Approximation with Sparse Occupancy Trees. Preprint 45, DFG-SPP 1324, May 2010.

[46] J. Ballani and L. Grasedyck. A Projection Method to Solve Linear Systems in Tensor Format. Preprint 46, DFG-SPP 1324, May 2010.

[47] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence Rates for Greedy Algorithms in Reduced Basis Methods. Preprint 47, DFG-SPP 1324, May 2010.

[48] S. Kestler and K. Urban. Adaptive Wavelet Methods on Unbounded Domains. Preprint 48, DFG-SPP 1324, June 2010.

[49] H. Yserentant. The Mixed Regularity of Electronic Wave Functions Multiplied by Explicit Correlation Factors. Preprint 49, DFG-SPP 1324, June 2010.

[50] H. Yserentant. On the Complexity of the Electronic Schrödinger Equation. Preprint 50, DFG-SPP 1324, June 2010.

[51] M. Guillemard and A. Iske. Curvature Analysis of Frequency Modulated Manifolds in Dimensionality Reduction. Preprint 51, DFG-SPP 1324, June 2010.

[52] E. Herrholz and G. Teschke. Compressive Sensing Principles and Iterative Sparse Recovery for Inverse and Ill-Posed Problems. Preprint 52, DFG-SPP 1324, July 2010.

[53] L. Kämmerer, S. Kunis, and D. Potts. Interpolation Lattices for Hyperbolic Cross Trigonometric Polynomials. Preprint 53, DFG-SPP 1324, July 2010.

[54] G. Kutyniok and W.-Q Lim. Shearlets on Bounded Domains. Preprint 54, DFG-SPP 1324, July 2010.

[55] A. Zeiser. Wavelet Approximation in Weighted Sobolev Spaces of Mixed Order with Applications to the Electronic Schrödinger Equation. Preprint 55, DFG-SPP 1324, July 2010.

[56] G. Kutyniok, J. Lemvig, and W.-Q Lim. Compactly Supported Shearlets. Preprint 56, DFG-SPP 1324, July 2010.

[57] A. Zeiser. On the Optimality of the Inexact Inverse Iteration Coupled with Adaptive Finite Element Methods. Preprint 57, DFG-SPP 1324, July 2010.

[58] S. Jokar. Sparse Recovery and Kronecker Products. Preprint 58, DFG-SPP 1324, August 2010.

[59] T. Aboiyar, E. H. Georgoulis, and A. Iske. Adaptive ADER Methods Using Kernel-Based Polyharmonic Spline WENO Reconstruction. Preprint 59, DFG-SPP 1324, August 2010.

[60] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the Convergence of Generalized Polynomial Chaos Expansions. Preprint 60, DFG-SPP 1324, August 2010.

[61] S. Holtz, T. Rohwedder, and R. Schneider. On Manifolds of Tensors of Fixed TT-Rank. Preprint 61, DFG-SPP 1324, September 2010.

[62] J. Ballani, L. Grasedyck, and M. Kluge. Black Box Approximation of Tensors in Hierarchical Tucker Format. Preprint 62, DFG-SPP 1324, October 2010.

[63] M. Hansen. On Tensor Products of Quasi-Banach Spaces. Preprint 63, DFG-SPP 1324, October 2010.

[64] S. Dahlke, G. Steidl, and G. Teschke. Shearlet Coorbit Spaces: Compactly Supported Analyzing Shearlets, Traces and Embeddings. Preprint 64, DFG-SPP 1324, October 2010.

[65] W. Hackbusch. Tensorisation of Vectors and their Efficient Convolution. Preprint 65, DFG-SPP 1324, November 2010.

[66] P. A. Cioica, S. Dahlke, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Spatial Besov Regularity for Stochastic Partial Differential Equations on Lipschitz Domains. Preprint 66, DFG-SPP 1324, November 2010.

[67] E. Novak and H. Woźniakowski. On the Power of Function Values for the Approximation Problem in Various Settings. Preprint 67, DFG-SPP 1324, November 2010.

[68] A. Hinrichs, E. Novak, and H. Woźniakowski. The Curse of Dimensionality for Monotone and Convex Functions of Many Variables. Preprint 68, DFG-SPP 1324, November 2010.

[69] G. Kutyniok and W.-Q Lim. Image Separation Using Shearlets. Preprint 69, DFG-SPP 1324, November 2010.

[70] B. Jin and P. Maass. An Analysis of Electrical Impedance Tomography with Applications to Tikhonov Regularization. Preprint 70, DFG-SPP 1324, December 2010.

[71] S. Holtz, T. Rohwedder, and R. Schneider. The Alternating Linear Scheme for Tensor Optimisation in the TT Format. Preprint 71, DFG-SPP 1324, December 2010.

[72] T. Müller-Gronbach and K. Ritter. A Local Refinement Strategy for Constructive Quantization of Scalar SDEs. Preprint 72, DFG-SPP 1324, December 2010.

[73] T. Rohwedder and R. Schneider. An Analysis for the DIIS Acceleration Method used in Quantum Chemistry Calculations. Preprint 73, DFG-SPP 1324, December 2010.

[74] C. Bender and J. Steiner. Least-Squares Monte Carlo for Backward SDEs. Preprint 74, DFG-SPP 1324, December 2010.

[75] C. Bender. Primal and Dual Pricing of Multiple Exercise Options in Continuous Time. Preprint 75, DFG-SPP 1324, December 2010.

[76] H. Harbrecht, M. Peters, and R. Schneider. On the Low-rank Approximation by the Pivoted Cholesky Decomposition. Preprint 76, DFG-SPP 1324, December 2010.

[77] P. A. Cioica, S. Dahlke, N. Döhring, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Adaptive Wavelet Methods for Elliptic Stochastic Partial Differential Equations. Preprint 77, DFG-SPP 1324, January 2011.

[78] G. Plonka, S. Tenorth, and A. Iske. Optimal Representation of Piecewise Hölder Smooth Bivariate Functions by the Easy Path Wavelet Transform. Preprint 78, DFG-SPP 1324, January 2011.

[79] A. Mugler and H.-J. Starkloff. On Elliptic Partial Differential Equations with Random Coefficients. Preprint 79, DFG-SPP 1324, January 2011.

[80] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. A Derandomization of the Euler Scheme for Scalar Stochastic Differential Equations. Preprint 80, DFG-SPP 1324, January 2011.

[81] W. Dahmen, C. Huang, C. Schwab, and G. Welper. Adaptive Petrov-Galerkin methods for first order transport equations. Preprint 81, DFG-SPP 1324, January 2011.

[82] K. Grella and C. Schwab. Sparse Tensor Spherical Harmonics Approximation in Radiative Transfer. Preprint 82, DFG-SPP 1324, January 2011.

[83] D.A. Lorenz, S. Schiffler, and D. Trede. Beyond Convergence Rates: Exact Inversion With Tikhonov Regularization With Sparsity Constraints. Preprint 83, DFG-SPP 1324, January 2011.

[84] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: Approximation by empirical measures. Preprint 84, DFG-SPP 1324, January 2011.

[85] S. Dahlke and W. Sickel. On Besov Regularity of Solutions to Nonlinear Elliptic Partial Differential Equations. Preprint 85, DFG-SPP 1324, January 2011.

[86] S. Dahlke, U. Friedrich, P. Maass, T. Raasch, and R.A. Ressel. An adaptive wavelet method for parameter identification problems in parabolic partial differential equations. Preprint 86, DFG-SPP 1324, January 2011.

[87] A. Cohen, W. Dahmen, and G. Welper. Adaptivity and Variational Stabilization for Convection-Diffusion Equations. Preprint 87, DFG-SPP 1324, January 2011.

[88] T. Jahnke. On Reduced Models for the Chemical Master Equation. Preprint 88, DFG-SPP 1324, January 2011.

[89] P. Binev, W. Dahmen, R. DeVore, P. Lamby, D. Savu, and R. Sharpley. Compressed Sensing and Electron Microscopy. Preprint 89, DFG-SPP 1324, March 2011.

[90] P. Binev, F. Blanco-Silva, D. Blom, W. Dahmen, P. Lamby, R. Sharpley, and T. Vogt. High Quality Image Formation by Nonlocal Means Applied to High-Angle Annular Dark Field Scanning Transmission Electron Microscopy (HAADF-STEM). Preprint 90, DFG-SPP 1324, March 2011.

[91] R. A. Ressel. A Parameter Identification Problem for a Nonlinear Parabolic Differential Equation. Preprint 91, DFG-SPP 1324, May 2011.