

A concise proof for existence and uniqueness of solutions of linear parabolic PDEs in the context of optimal control

Anton Schiela *

February 28, 2013

Abstract

We present a concise proof for existence and uniqueness of solutions of linear parabolic PDEs. It is based on an analysis of the corresponding differential operator and its adjoint in appropriate spaces and simple enough to be presented in the context of an introductory lecture on optimal control of PDEs. Our approach also clarifies some aspects in the structure of first order optimality conditions as illustrated at an example.

AMS MSC 2000: 35K15, 47N20, 49K20

Keywords: parabolic PDEs, optimal control, existence theory

1 Introduction

Optimal control of parabolic partial differential equations is a classical topic of optimal control theory (cf. e.g. [4, 5]) and taught regularly in lectures on this subject. While the theory of these equations is the basis for their optimal control, it is widely considered too difficult and time-consuming to present existence and uniqueness proofs for parabolic PDES to students in a lecture on optimal control. Classical approaches are the Galerkin method [7, 1, 6] and semi-group theory [2], two powerful techniques which apply to linear and nonlinear problems alike. However, their precise presentation is very time-consuming, and it involves many technical steps if applied in that generality that is needed for optimal control.

The purpose of this paper is to present a concise proof of existence and uniqueness of solutions for linear parabolic PDEs, which is straightforward and quite general in terms of regularity properties of the involved operators and data. Roughly speaking, we describe the differential operator on the left-hand side as a linear mapping A between appropriate Banach spaces and show that it is an isomorphism by analyzing A and its adjoint A^* . In the context of optimal control an analysis of A^* is necessary anyway, because it appears in the first order optimality conditions for optimal control problems. This is demonstrated in the following prototypical abstract result, which yields first order optimality conditions at a local minimizer:

Theorem 1.1. *Let Y , P , and U be Banach spaces, and P be reflexive. Assume that $A : Y \rightarrow P^*$ is an isomorphism and $B : U \rightarrow P^*$ is continuous. Let $f : Y \times U \rightarrow \mathbb{R}$ be a*

*Supported by the DFG Research Center MATHEON "Mathematics for key technologies", Berlin

functional. Assume that the problem

$$\min_{(y,u) \in Y \times U} f(y, u) \text{ s.t. } Ay - Bu = 0,$$

has a local minimizer $(\underline{y}, \underline{u})$. If f is Gâteaux differentiable at $(\underline{y}, \underline{u})$, then there exists a unique adjoint state $p \in P$ which satisfies:

$$\begin{aligned} \frac{\partial}{\partial y} f(\underline{y}, \underline{u}) + A^* p &= 0 \\ \frac{\partial}{\partial u} f(\underline{y}, \underline{u}) - B^* p &= 0. \end{aligned} \tag{1}$$

Proof. Let us define $X = Y \times U$ and $C := (A, -B) : X \rightarrow P^*$. Since $(\underline{y}, \underline{u})$ is a minimizer of f on $\ker C$, we can conclude that its Gâteaux derivative $f'(\underline{y}, \underline{u}) \in X^*$ satisfies $f'(\underline{y}, \underline{u})\delta x = 0$ for all $\delta x \in \ker C$. This means that

$$f'(\underline{y}, \underline{u}) \in (\ker C)^\perp := \{\ell \in X^* : \ell(\delta x) = 0 \quad \forall \delta x \in \ker C\}.$$

Since C is a surjective, continuous mapping between Banach spaces, we can apply the closed range theorem to conclude that $\text{ran } C^* = (\ker C)^\perp$, so that $f'(\underline{y}, \underline{u}) \in \text{ran } C^*$. This means, that there is $p \in P = P^{**}$, such that $f'(\underline{y}, \underline{u}) + C^* p = 0$, which is (1). \square

Remark 1.2. More generally, it would be sufficient to consider $A : Y \supset D \rightarrow P^*$ as a closed and surjective, possibly unbounded operator which is densely defined. Even more generally, for existence of (non-unique) p it would suffice that C as used in the proof has closed range.

Application of this abstract result to a concrete problem involves the following steps concerning A . First, A has to be defined as an operator between appropriately chosen spaces Y and P^* . Our choice of a dual space P^* for the codomain of A models the situation that A is an operator in weak form, i.e., the equation $Ay - Bu = 0$ in P^* is just a different notation for

$$(Ay - Bu)(p) = 0 \quad \forall p \in P.$$

Then one has to show that A is an isomorphism. Finally, it is desirable to give an interpretation of the adjoint operator A^* as a differential operator again, so that the first line of (1) can be interpreted as a partial differential equation. This is particularly interesting in the parabolic case.

In this work we will present a concise and clear way to perform this program for parabolic PDEs. In Section 2 we fix our functional analytic framework and provide auxiliary results. In Section 3 we define and analyze A and A^* in the case that they model linear parabolic equations and their adjoints. In Section 4 we apply our results to a class of optimal control problems with parabolic PDEs.

2 Function Spaces for Time-dependent problems

In this section we recapitulate basic results on the function spaces involved. They can either be found easily in the literature (see e.g. [7, 1, 6]) or have a short proof, which is presented here.

2.1 Basic results on vector valued functions

Let $(X, \|\cdot\|)$ be a separable Banach space and $[0, T]$ a bounded proper interval in \mathbb{R} . Let us recall some standard definitions and results on vector valued functions.

The space $C([0, T]; X)$ is the space of all continuous functions $f : [0, T] \rightarrow X$ with the usual norm $\|f\|_{C([0, T], X)} = \max_{t \in [0, T]} \|f(t)\|$. By $C_0^\infty([0, T]; X)$ we denote the space of infinitely often differentiable functions with compact support in $]0, T[$. For $1 \leq q < \infty$, $L_q(0, T; X)$ denotes the space of all (equivalence classes w.r.t. zero measure) of Bochner integrable functions $f :]0, T[\rightarrow X$ such that

$$\|f\|_{L_q(X)} = \|f\|_{L_q(0, T; X)} := \left(\int_0^T \|f(t)\|^q dt \right)^{1/q} < \infty.$$

If X is reflexive and separable, $1 < q < \infty$, and $q^{-1} + (q')^{-1} = 1$, then $L_q(0, T; X)$ is reflexive and separable, and $L_q(0, T; X)^* \cong L_{q'}(0, T; X^*)$.

Denote by χ_I the characteristic function of an interval $I \subset [0, T]$. By a step function $s : [0, T] \rightarrow X$, we denote the finite sum

$$s := \sum_{k=1}^n \chi_{I_k} x_k, \quad (2)$$

where $I_k \subset [0, T]$ are intervals and $x_k \in X$ for $k = 1 \dots n$. It is well known (cf. e.g. [1, Kap. IV Lemma 1.3]) that the step functions are dense in $L_q(0, T; X)$.

Lemma 2.1. $C_0^\infty([0, T]; X)$ is dense in $L_q(0, T; X)$ for $1 \leq q < \infty$.

Proof. This is an easy consequence of the density of step functions: each characteristic function χ_I is contained in $L_q(0, T; \mathbb{R})$ and can thus be approximated by a smooth function in $\chi_I^\varepsilon \in C_0^\infty([0, T], \mathbb{R})$ to arbitrary accuracy. Thus, each step function, defined as in (2) can be approximated to arbitrary accuracy by

$$s^\varepsilon := \sum_{k=1}^n \chi_{I_k}^\varepsilon x_k$$

Since step functions can in turn be used to approximate elements of $L_q(0, T; X)$ our assertion follows. \square

2.2 Evolution triples and generalized time derivatives

Let H be a separable Hilbert space, V a separable and reflexive Banach space and $V \hookrightarrow H$ be continuously densely embedded. Then we can write $H^* \cong H$ via the Riesz isomorphism $h^* \sim \langle h, \cdot \rangle_H$. This also yields the adjoint embedding $H \hookrightarrow V^*$. The triple $V \hookrightarrow H \hookrightarrow V^*$ is called *evolution triple*. In the following we will always identify H and H^* via its Riesz isomorphism. In applications, H is usually an L_2 -space.

If $v \in L_1(0, T; V)$, $w \in L_1(0, T; W)$ then w is called generalized (or distributional) time derivative of v , if

$$\int_0^T \varphi'(t)v(t) dt = - \int_0^T \varphi(t)w(t) dt \quad \forall \varphi \in C_0^\infty([0, T], \mathbb{R}) \quad (3)$$

and we write $w = v'$. The generalized time derivative is unique.

Proposition 2.2. $v \in L_q(0, T; V)$ has the generalized time derivative $v' = l \in L_{q'}(0, T; V^*)$, if and only if

$$\int_0^T \varphi'(t) \langle v(t), w \rangle_H dt = - \int_0^T \varphi(t) l(t)(w) dt \quad \forall w \in V, \varphi \in C_0^\infty([0, T], \mathbb{R}). \quad (4)$$

Proof. Cf. [7, Prop. 23.20(b)]. □

Let in the following $V \hookrightarrow H \hookrightarrow V^*$ be a fixed evolution triple. Recall the following standard definitions and results (c.f. e.g. [1, IV§1.5]). The normed space

$$\begin{aligned} W([0, T]) &:= \{v \in L_2(0, T; V) : v' \in L_2(0, T; V^*)\} \\ \|v\|_{W([0, T])} &:= \|v\|_{L_2(0, T; V)} + \|v'\|_{L_2(0, T; V^*)} \end{aligned}$$

is complete and there is the continuous embedding:

$$W([0, T]) \hookrightarrow C([0, T], H).$$

For $v, w \in W([0, T])$ the formulas of integration by parts hold:

$$\langle v(t), w(t) \rangle_H - \langle v(s), w(s) \rangle_H = \int_s^t v'(\tau)(w(\tau)) + w'(\tau)(v(\tau)) d\tau \quad (5)$$

For later purpose, we prove a special density result.

Lemma 2.3. *The mapping*

$$\begin{aligned} E_{0, T} : W([0, T]) &\rightarrow L_2(0, T; V) \times H \times H \\ w &\mapsto (w, w(0), w(T)) \end{aligned}$$

is continuous and has dense range. Its adjoint

$$\begin{aligned} E_{0, T}^* : L_2(0, T; V^*) \times H \times H &\rightarrow W([0, T])^* \\ E_{0, T}^*(l, l_0, l_T)(w) &= \int_0^1 l(t)(w(t)) dt + \langle l_0, w(0) \rangle_H + \langle l_T, w(T) \rangle_H \end{aligned} \quad (6)$$

is thus injective.

Proof. Continuity of $E_{0, T}$ follows from the continuity of the embeddings $W([0, T]) \hookrightarrow L_2(0, T; V)$ and $W([0, T]) \hookrightarrow C([0, T]; H)$.

Let $(v, v_0, v_T) \in L_2(0, T; V) \times H \times H$ and $\varepsilon > 0$ a given desired accuracy of approximation. Then by density of V in H we can approximate (v_0, v_T) by $(\tilde{v}_0, \tilde{v}_T) \in V \times V$ to an accuracy of $\varepsilon/3$, each. Further, by Lemma 2.1 we can approximate the function $v - (T-t)\tilde{v}_0 - t\tilde{v}_T$ by $\tilde{v} \in C_0^\infty([0, T], V)$ to an accuracy of $\varepsilon/3$. Then

$$w := \tilde{v} + (T-t)\tilde{v}_0 + t\tilde{v}_T \in W([0, T])$$

is constructed in such a way that $(w, w(0), w(T)) = (w, \tilde{v}_0, \tilde{v}_T)$ approximates (v, v_0, v_T) to an accuracy of ε .

Finally, injectivity of $E_{0, T}^*$ is implied by the following well known the general formula which holds for all continuous operators $Q : X \rightarrow R$ in normed spaces.

$$\overline{\text{ran } Q} = \ker Q_\perp^* := \{r \in R : r^*(r) = 0 \forall r^* \in \ker Q^*\}.$$

In this case $\overline{\text{ran } Q} = R$, and thus $\ker Q_\perp^* = \{0_{R^*}\}$. □

Injectivity of $E_{0,T}^*$ means that the right hand side of (6), which is an element of $W([0, T])^*$ has a *unique* representation by $(l, l_0, l_T) \in L_2(0, T; V^*) \times H \times H$.

Remark 2.4. Obviously, a similar density result holds, if we choose finitely many $t_i \in [0, T]$ for $i = 0 \dots n$ and define

$$\begin{aligned} E_{t_0, \dots, t_n} : W([0, T]) &\rightarrow L_2(0, T; V) \times (H)^{n+1} \\ w &\mapsto (w, w(t_0), \dots, w(t_n)). \end{aligned}$$

2.3 A special time derivative operator

Consider the linear and operator

$$\begin{aligned} D : W([0, T]) &\rightarrow L_2(0, T; V)^* \times H^* \\ (Dw)(v, v_0) &= \int_0^T w'(t)(v(t)) dt + \langle w(0), v_0 \rangle_H. \end{aligned} \quad (7)$$

It is continuous by definition of $W([0, T])$ and the embedding $W([0, T]) \hookrightarrow C([0, T], H)$. The corresponding adjoint operator reads:

$$\begin{aligned} D^* : L_2(0, T; V) \times H &\rightarrow W([0, T])^* \\ D^*(v, v_0)(w) &= (Dw)(v, v_0) = \int_0^T w'(t)(v(t)) dt + \langle w(0), v_0 \rangle_H. \end{aligned} \quad (8)$$

The integration by parts formula (5) yields for $w \in W([0, T])$:

$$(Dw)(w, w(0)) = \int_0^T w'(t)(w(t)) dt + \|w(0)\|_H^2 = \frac{1}{2} (\|w(T)\|_H^2 + \|w(0)\|_H^2). \quad (9)$$

The following result plays a pivoting role in our analysis. It characterizes the range of the restriction of D^* to $W([0, T]) \times H$.

Proposition 2.5. *The following two assertions are equivalent:*

(i) *There exist $(l, l_0, l_T) \in L_2(0, T; V^*) \times H \times H$ such that for all $w \in W([0, T])$*

$$D^*(v, v_0)(w) = \int_0^T -l(t)w(t) dt + \langle l_0, w(0) \rangle_H + \langle l_T, w(T) \rangle_H \quad (10)$$

(ii) $v \in W([0, T])$.

If one of these assertions holds, and hence both of them, then $l = v'$, $l_0 = v_0 - v(0)$, $l_T = v(T)$.

Proof. (i) \Rightarrow (ii): Take $w = w(x) \in V$ and $\varphi = \varphi(t) \in C_0^\infty([0, T]; \mathbb{R})$ so that $\varphi w \in W([0, T])$ and compute $D^*(v, v_0)(\varphi w)$:

$$\begin{aligned} \int_0^T -l(t)(w)\varphi(t) dt &= \int_0^T -l(t)(\varphi(t)w) dt + \underbrace{\langle l_0, \varphi(0)w \rangle_H}_{=0} + \underbrace{\langle l_T, \varphi(T)w \rangle_H}_{=0} \\ &= D^*(v, v_0)(\varphi w) = D(\varphi w)(v, v_0) \\ &= \int_0^T (\varphi(t)w)'(v(t)) dt + \underbrace{\langle \varphi(0)w, v_0 \rangle_H}_{=0} = \int_0^T \varphi'(t)\langle w, v \rangle_H dt. \end{aligned}$$

Hence, by Proposition 2.2 $l \in L_2(0, T; V^*)$ is the weak derivative of v , and thus $v \in W([0, T])$.

(ii) \Rightarrow (i): If $v \in W([0, T])$, integration by parts (5) yields for $w \in W([0, T])$:

$$\begin{aligned} D^*(v, v_0)(w) &= \int_0^T w'(t)(v(t)) dt + \langle w(0), v_0 \rangle_H \\ &= \int_0^T -v'(t)(w(t)) dt + \langle v_0 - v(0), w(0) \rangle_H + \langle v(T), w(T) \rangle_H. \end{aligned}$$

Hence, we conclude that (10) holds, together with the equations $l = v' \in L_2(0, T; V^*)$, $l_0 = v_0 - v(0) \in H$, $l_T = v(T) \in H$. By Lemma 2.3 l , l_0 , and l_T are uniquely defined by these equations. \square

3 Linear parabolic PDEs and their adjoint equations

Having studied the time derivative, we add an elliptic operator, from which we assume only very mild properties. Consider a continuous operator

$$\Lambda : L_2(0, T; V) \rightarrow L_2(0, T; V^*)$$

that satisfies the ellipticity condition:

$$\exists \lambda > 0 : (\Lambda y)(y) \geq \lambda \|y\|_{L_2(V)}^2 \quad (11)$$

The most prominent example is the following operator for $V = H_0^1(\Omega)$, which yields the classical heat equation:

$$(\Lambda y)(v) := \int_0^T \int_{\Omega} \nabla y \cdot \kappa(t, \omega) \nabla v d\omega. \quad (12)$$

To obtain continuity and ellipticity of Λ in this example we have to impose that $\kappa(t, \omega)$ is a uniformly bounded measurable function from $[0, T] \times \Omega$ into the set of real matrices, which satisfies a uniform ellipticity condition $v \cdot \kappa(t, \omega) v \geq c|v|^2$.

Remark 3.1. However, our assumptions include much more general settings. In particular, Λ may be non-local in time. With that one can use a similar trick as in [1, Kap. VII] to cover certain linear PDEs of second order in time.

Now we define the linear and continuous operator A that corresponds to our parabolic PDE:

$$\begin{aligned} A : W([0, T]) &\rightarrow L_2(0, T; V^*) \times H^* \\ (Ay)(p, p_0) &:= (Dy)(p, p_0) + (\Lambda y)(p) \end{aligned} \quad (13)$$

$$(Ay)(p, p_0) = \int_0^T y'(t)(p(t)) dt + \langle y(0), p_0 \rangle_H + (\Lambda y)(p).$$

A parabolic PDE with data $(l, y_0) \in L_2(0, T; V^*) \times H$ can then be written in the following way:

$$(Ay)(p, p_0) = \int_0^T l(t)(p(t)) dt + \langle y_0, p_0 \rangle_H \quad \forall (p, p_0) \in L_2(0, T; V) \times H.$$

Testing with $(p, 0)$ and $(0, p_0)$ we obtain the two equations

$$\begin{aligned} \int_0^T y'(t)(p(t)) dt + (\Lambda y)(p) &= \int_0^T l(t)(p(t)) dt \quad \forall p \in L_2(0, T; V) \\ \langle y(0), p_0 \rangle_H &= \langle y_0, p_0 \rangle_H \quad \forall p_0 \in H. \end{aligned} \quad (14)$$

The first line is the parabolic PDE in weak form, the second line is equivalent to the initial condition $y(0) = y_0$.

Its adjoint operator reads:

$$\begin{aligned} A^* : L_2(0, T; V) \times H &\rightarrow W([0, T])^* \\ A^*(p, p_0)(w) &= (Aw)(p, p_0). \end{aligned} \quad (15)$$

Interestingly, the domain of A^* consists of pairs (p, p_0) of variables. We will give an interpretation of A^* as a differential operator in Section 3.2 below. In Section 4 we will apply Theorem 1.1 to A and A^* in order to analyze an optimal control problem. That will also clarify the role of p_0 .

3.1 Existence and uniqueness of solutions

In this section we present the main result of our work. Our proof is divided into two parts. The first part is purely functional analytic and describes a well-known technique. In the second part we verify the abstract assumptions of the first.

Lemma 3.2. *Let X, R be Banach spaces. Let $Q : X \rightarrow R$ be a continuous linear operator that satisfies*

- (i) $\exists C$ such that $\|x\| \leq C\|Qx\|$ for all $x \in X$
- (ii) $\ker Q^* = \{0_{R^*}\}$

Then Q and Q^ are isomorphisms.*

Proof. Using (ii), a simple corollary of the Hahn-Banach theorem yields that Q has dense range:

$$\overline{\text{ran } Q} = (\ker Q^*)_{\perp} := \{r \in R : r^*(r) = 0 \ \forall r^* \in \ker Q^*\} = \{0_{R^*}\}_{\perp} = R.$$

By continuity of Q and (i), $Q : X \rightarrow \text{ran } Q$ is an isomorphism. This implies by completeness of X that $\text{ran } Q$ is a complete subspace of R and thus closed in R . Hence, $\text{ran } Q = \overline{\text{ran } Q} = R$, and thus $Q : X \rightarrow R$ is an isomorphism. So Q^* is an isomorphism, as well. \square

Remark 3.3. This result can be generalized in the following way. If $Q : X \supset D \rightarrow R$ is a closed, densely defined operator, then under (i) and (ii) one can show that Q and Q^* have a continuous inverse.

Let us now come back to our concrete setting of parabolic PDEs:

Theorem 3.4. *A as defined in (13) and A^* are both continuously invertible. In particular, the problem of solving the parabolic PDE (14) is well posed.*

Proof. We will verify Assumption (i) and (ii) of Lemma 3.2 for A as defined in (13), which then yields our desired result. Our key is the following *a-priori estimate* for $v \in W([0, T])$ which holds due to (9) and (11):

$$\begin{aligned} (Av)(v, v(0)) &= (Dv)(v, v(0)) + (\Lambda v)(v) \\ &= \frac{1}{2} (\|v(T)\|_H^2 + \|v(0)\|_H^2) + (\Lambda v)(v) \geq \frac{1}{2} \|v(0)\|_H^2 + \lambda \|v\|_{L_2(V)}^2 \\ &\geq \tilde{\lambda} \|(v, v(0))\|_{L_2(V) \times H}^2, \quad \text{where } \tilde{\lambda} = \min\{\lambda, 1/2\}. \end{aligned} \quad (16)$$

Ad (i) : From (16) we conclude for all $y \in W([0, T])$:

$$\tilde{\lambda} \|(y, y(0))\|_{L_2(V) \times H}^2 \leq (Ay)(y, y(0)) \leq \|Ay\|_{L_2(V)^* \times H^*} \|(y, y(0))\|_{L_2(V) \times H},$$

which yields

$$\|y\|_{L_2(V)} \leq \|(y, y(0))\|_{L_2(V) \times H} \leq \tilde{\lambda}^{-1} \|Ay\|_{L_2(V)^* \times H^*}. \quad (17)$$

Next we write $y' = -\Lambda y + A_1 y$, where $A_1 y \in L_2(0, T; V^*)$ shall be the first component of $Ay \in L_2(0, T; V^*) \times H^*$, and compute,

$$\begin{aligned} \|y'\|_{L_2(V^*)} &= \|- \Lambda y + A_1 y\|_{L_2(V^*)} \\ &\leq \|\Lambda\| \|y\|_{L_2(V)} + \|A_1 y\|_{L_2(V^*)} \\ &\leq (\|\Lambda\| \tilde{\lambda}^{-1} + 1) \|Ay\|_{L_2(V)^* \times H^*}. \end{aligned}$$

Adding (17) we obtain the desired estimate (i):

$$\begin{aligned} \|y\|_{W([0, T])} &= \|y'\|_{L_2(V^*)} + \|y\|_{L_2(V)} \\ &\leq (\|\Lambda\| \tilde{\lambda}^{-1} + 1 + \tilde{\lambda}^{-1}) \|Ay\|_{L_2(V)^* \times H^*}. \end{aligned}$$

Ad (ii) : Consider $(p, p_0) \in \ker A^* \subset L_2(0, T; V) \times H$, i.e.,

$$D^*(p, p_0)(w) = (-\Lambda^* p)(w) \quad \forall w \in W([0, T]).$$

Then in view of Proposition 2.5 we can write

$$D^*(p, p_0) = (\Lambda^* p, 0, 0) = (l, l_0, l_T) \in L_2(0, T; V^*) \times H \times H,$$

and obtain $p \in W([0, T])$, $0 = l_0 = p_0 - p(0)$ and $0 = l_T = p(T)$. Thus we are allowed to test $A^*(p, p_0)$ with p and compute:

$$0 = A^*(p, p_0)(p) = (Ap)(p, p_0) = (Ap)(p, p(0)).$$

This implies $(p, p_0) = (p, p(0)) = 0$ due to (16) and hence $\ker A^* = \{0\}$. \square

Corollary 3.5. *Theorem 3.4 still holds if instead of (11) the following weaker assumption is imposed:*

$$\exists \omega < \infty, \lambda > 0 : \left(\omega \|y\|_{L_2(H)}^2 + (\Lambda y)(y) \right) \geq \lambda \|y\|_{L_2(V)}^2.$$

Proof. This follows by the usual trick with the transformation $\hat{y} = e^{-\omega t} y$, see e.g. [1, Beweis von Satz VI.1.3]. \square

3.2 The adjoint equation as a backwards-in-time equation

It is a well known fact in optimal control that in many cases the adjoint equation, defined by A^* can be interpreted as a backwards-in-time parabolic equation (cf. e.g. [5, Chap. 3]). While we already know continuous invertibility of A^* by Theorem 3.4 a precise characterization, when this interpretation is possible requires a little additional work.

Consider the backward time-derivative:

$$\begin{aligned} D_- : W([0, T]) &\rightarrow L_2(0, T; V^*) \times H^* \\ (D_- w)(v, v_T) &= \int_0^T -w'(t)(v(t)) dt + \langle w(T), v_T \rangle_H. \end{aligned} \quad (18)$$

It owns its name to the fact that $D_- w(t) = Dw(T - t)$ via the chain rule. Then one can define

$$\begin{aligned} A_- := D_- + \Lambda^* : W([0, T]) &\rightarrow L_2(0, T; V^*) \times H^* \\ (A_- p)(y, y_T) &= \int_0^T -p'(t)(y(t)) dt + \langle p(T), y_T \rangle_H + (\Lambda y)(p). \end{aligned} \quad (19)$$

In analogy to the forward equation, the operator equation $A_- p = (l, p_T)$ corresponds to the following backward-in-time problem with final value

$$\begin{aligned} \int_0^T -p'(t)(v(t)) dt + (\Lambda v)(p) &= \int_0^T l(t)(v(t)) dt \quad \forall v \in L_2(0, T; V) \\ \langle p(T), v_T \rangle_H &= \langle p_T, v_T \rangle_H \quad \forall v_T \in H. \end{aligned} \quad (20)$$

Arguing with the above transformation of time, we immediately see that A_- is an isomorphism due to Theorem 3.4, because $(\Lambda^* y)(y) = (\Lambda y)(y)$ and hence Λ^* satisfies (11) as well.

Proposition 3.6. *The following identity holds for all $p, y \in W([0, T])$:*

$$A^*(p, p(0))(y) = (A_- p)(y, y(T)). \quad (21)$$

Proof. We compute as follows, using integration by parts (5) in the final step:

$$\begin{aligned} &A^*(p, p(0))(y) - (A_- p)(y, y(T)) \\ &= (Ay)(p, p(0)) - (A_- p)(y, y(T)) \\ &= (Dy)(p, p(0)) + (\Lambda y)(p) - (D_- p)(y, y(T)) - (\Lambda y)(p) \\ &= (Dy)(p, p(0)) - (D_- p)(y, y(T)) \\ &= \int_0^T y'(t)(p(t)) + p'(t)(y(t)) dt + \langle y(0), p(0) \rangle_H - \langle p(T), y(T) \rangle_H = 0. \end{aligned}$$

□

If we define – in the spirit of Lemma 2.3 – the embeddings $E_0 : W([0, T]) \rightarrow L_2(0, T; V) \times H$ via $E_0 p := (p, p(0))$ and similarly $E_T : W([0, T]) \rightarrow L_2(0, T; V) \times H$ via $E_T y := (y, y(T))$, then the above relation can be written as:

$$A^* E_0 = E_T^* A_- : W([0, T]) \rightarrow W([0, T])^*. \quad (22)$$

This makes the statement “the adjoint operator can be interpreted as a backward-in-time operator” precise within our framework.

Via the range of the embedding E_0 we may view $W([0, T])$ as a dense subspace of $L_2(0, T; V) \times H$. Here density follows as in Lemma 2.3. Since A^* is continuous and coincides with $E_T^* A_-$ on $W([0, T])$ by (22), we conclude that A^* is the unique continuous extension of $E_T^* A_-$ from $W([0, T])$ to $L_2(0, T; V) \times H$. Note here that E_T^* is injective, since E_T has dense range (cf. Lemma 2.3). However, by definition of A^* , solutions of the equation $A^*(p, p_0) = \ell$ consist of a pair (p, p_0) and do neither have any initial values, nor are they continuous in time for general right hand sides $\ell \in W([0, T])^*$.

Remark 3.7. Of course, a similar extension can be performed for the forward operator A via the relation $A_- E_T = E_0^* A$ which follows from adjoining (22). We see that now A_- is the unique continuous extension of $E_0^* A$ from $W([0, T])$ to $L_2(0, T; V) \times H$. The resulting notion of solution has similarities with the approach of [3], which states that $y \in L_2(0, T; V)$ is a weak solution of a parabolic equation, if

$$\int_0^T -y(t)v'(t) dt + (\Lambda y)(v) = \int_0^T l(t)v(t) dt + \langle y_0, v(0) \rangle_H$$

$$\forall v \in L_2(0, T; V) \text{ with } v' \in L_2(0, T; V), v(T) = 0.$$

However, in [3] the space of possible right-hand sides is smaller than $W([0, T])^*$.

Finally, we give a necessary and sufficient condition for $p \in W([0, T])$ in terms of $A^*(p, p_0)$:

Proposition 3.8. *The following assertions are equivalent:*

(i) *There exist $(l, l_0, l_T) \in L_2(0, T; V^*) \times H \times H$ such that for all $v \in W([0, T])$ it holds:*

$$A^*(p, p_0)v = \int_0^T l(t)(v(t)) dt + \langle l_0, v(0) \rangle_H + \langle l_T, v(T) \rangle_H. \quad (23)$$

(ii) $p \in W([0, T])$.

If one, and hence both of these assertions hold, then $p(T) = l_T$ and $p_0 - p(0) = l_0$.

Proof. By definition of A^* we obtain

$$D^*(p, p_0)(v) = \langle l_T, v(T) \rangle_H + \langle l_0, v(0) \rangle_H + (l - \Lambda^* p)(v),$$

where $l - \Lambda^* p \in L_2(0, T; V^*)$. Thus, Application of Proposition 2.5 is possible and yields the desired equivalence and $l_T = p(T)$ and $p_0 - p(0) = l_0$. \square

Thus, in case that condition (i) of Proposition 3.8 holds we can compute via (21) for $v \in W([0, T])$:

$$\begin{aligned} A^*(p, p_0)(v) &= A^*(p, p(0))(v) + A^*(0, p_0 - p(0))(v) \\ &= (A_- p)(v, v(T)) + \langle p_0 - p(0), v(0) \rangle_H \\ &= (A_- p)(v, v(T)) + \langle l_0, v(0) \rangle_H. \end{aligned}$$

Hence, by our density result Lemma 2.3, instead of (23) we may write equivalently:

$$(A_- p)(v, v_T) = \int_0^T l(t)(v(t)) dt + \langle l_0, v_T \rangle_H \quad \forall (v, v_T) \in L_2(0, T; V) \times H$$

$$p_0 - p(0) = l_0.$$

Again, p is given via a backward-in-time parabolic PDE. Moreover, p_0 can be computed from $p(0)$ and l_0 .

4 Application to an optimal control problem

Let us now apply our abstract result Theorem 1.1 to a parabolic optimal control problem. Here we choose $Y = W([0, T])$, $P = L_2(0, T; V) \times H$. For the purpose of illustration it is best to consider initial value control $u \in U := H$. All spaces are complete, and even reflexive under our standing assumption that V is separable and reflexive and H is a separable Hilbert space.

Defining the parabolic operator A as in (13) and $B : U \rightarrow P^*$ via

$$(Bu)(v, v_0) := \langle u, v_0 \rangle_H,$$

our abstract operator equation

$$Ay - Bu = 0 \quad \text{in } P^*$$

can be written down as follows in a concrete way: for all $(v, v_0) \in P$ we have

$$\int_0^T y'(t)(v(t)) dt + \langle y(0), v_0 \rangle_H + (\Lambda y)(v) = \langle u, v_0 \rangle_H. \quad (24)$$

Thus, u indeed enters as an initial value $y(0) = u$. Clearly, B is continuous, and we have shown in Theorem 3.4 that A is an isomorphism. We denote again by $f : Y \times U \rightarrow \mathbb{R}$ a differentiable objective functional and consider the problem

$$\min f(y, u) \text{ s.t. } Ay - Bu = 0. \quad (25)$$

Theorem 1.1 applied to this setting yields existence of a unique pair $(p, p_0) \in P$, such that

$$\frac{\partial}{\partial y} f(\underline{y}, \underline{u}) + A^*(p, p_0) = 0 \quad (26a)$$

$$\frac{\partial}{\partial u} f(\underline{y}, \underline{u}) - B^*(p, p_0) = 0, \quad (26b)$$

if $(\underline{y}, \underline{u})$ is a minimizer of (25). Let us now specify a simple objective functional by

$$f(y, u) := \frac{1}{2} \|y(0) - y_{d,0}\|_H^2 + \frac{1}{2} \int_0^T \|y(t) - y_d(t)\|_H^2 dt + \frac{1}{2} \|y(T) - y_{d,T}\|_H^2 + \frac{\alpha}{2} \|u\|_H^2, \quad (27)$$

where $y_{d,0} \in H$, $y_d \in L_2(0, T; H)$, and $y_{d,T} \in H$ are given data, and $\alpha > 0$ is a positive real parameter. Our choice is slightly non-standard, because it includes $y(0)$ and thus observation at the initial time. As we will see, this illustrates a case, where $p_0 \neq p(0)$, and thus the role of this additional variable.

Theorem 4.1. *The following system of equations has a unique solution $(\underline{y}, \underline{u}, (p, p_0)) \in Y \times U \times P$:*

$$\int_0^T \langle y(t) - y_d(t), w(t) \rangle_H dt + \int_0^T -p'(t)(w(t)) dt + (\Lambda w)(p) = 0 \quad \forall w \in L_2(0, T; V) \quad (28a)$$

$$(y(T) - y_{d,T}) + p(T) = 0 \text{ in } H$$

$$(y(0) - y_{d,0}) + (p_0 - p(0)) = 0 \text{ in } H \quad (28b)$$

$$\alpha u - p_0 = 0 \text{ in } H \quad (28c)$$

$$\int_0^T -y'(t)(v(t)) dt + (\Lambda y)(v) = 0 \quad \forall v \in L_2(0, T; V) \quad (28d)$$

$$y(0) - u = 0 \text{ in } H.$$

Moreover, $(\underline{y}, \underline{u})$ is the unique minimizer of the optimal control problem (25) with f given by (27).

Proof. Existence of a minimizer follows from standard weak convergence arguments, exploiting continuity and convexity of f , coercivity of f w.r.t u , continuous invertibility of A , and reflexivity of U . By strict convexity of the problem, this minimizer is also unique. Also by convexity and Theorem 1.1, $(y, u, (p, p_0))$ being a solution of the combined system (24) and (26a)-(26b), is equivalent to (y, u) being a minimizer of the above problem. Thus, it remains to show that (28a)-(28d) is equivalent to (26a)-(26b), and (24).

First of all, the state equation (28d) is a direct consequence of (24). Next, (28c) follows from (26b). Indeed, for our choice of f and B (26b) reads:

$$\alpha \langle u, \delta u \rangle_H - \langle \delta u, p_0 \rangle_H = 0 \quad \forall \delta u \in H.$$

By the Riesz-isomorphisms in H (28c) follows.

It remains to show that (28a) and (28b) follow from (26a). For this, we compute for all $w \in W([0, T])$:

$$\frac{\partial}{\partial y} f(y, u)w = \underbrace{\langle y(0) - y_{d,0}, w(0) \rangle_H}_{l_0} + \int_0^T \underbrace{\langle y(t) - y_d(t), w(t) \rangle_H}_{l(t)} dt + \underbrace{\langle y(T) - y_{d,T}, w(T) \rangle_H}_{l_T}.$$

Comparing this expression with (23), we observe that condition (i) in Proposition 3.8 is valid. This yields $p \in W([0, T])$, and (28b), as well as the second line of (28a). Finally, the first line of (28a) follows from Proposition 3.6 and the concrete form of A_- , given in (19). \square

Thus, only the distinction between p_0 and $p(0)$ allows us to write down this optimality system in correct form. In particular, (28b) is only correct with p_0 . In other cases, however, the variable p_0 does not enter explicitly. In particular, if f does not explicitly depend on $y(0)$, then $l_0 = 0$ in Proposition 3.8 and thus $p_0 = p(0)$. Further, for a different type of control than initial value control, p_0 usually does not enter into the computation of the control. In both cases it is possible to derive optimality conditions, which do not involve p_0 .

5 Conclusion

Let us conclude this work with recapitulating the two main insights gained from our analysis. First, we have seen that an existence and uniqueness proof for linear parabolic PDEs can be presented in a concise and straightforward way. This may be of interest in lectures on optimal control of PDEs, applied functional analysis or introductory lectures on PDEs. Of course, the well established techniques of proof cannot be replaced by our approach, since they are much more general in terms of nonlinearity and often constructive. Moreover, we have seen a way of extending the notion of solution of parabolic PDEs by considering the adjoint of a time reversed parabolic problem.

The second insight concerns optimal control theory. The adjoint state, which appears in the first order optimality conditions is actually a pair (p, p_0) of variables in the setting of parabolic problems. This corresponds to the fact that the data that defines an inhomogeneous parabolic initial value problem is also a pair of variables (l, y_0) consisting of the source term l and the initial value y_0 . The first component p corresponds to the source term, second corresponds p_0 to the initial value of the parabolic PDE.

References

- [1] H. Gajewski, K. Gröger, and K. Zacharias. *Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen*. Akademie-Verlag, Berlin, 1974.
- [2] E. Hille and R. S. Phillips. *Functional analysis and semigroups*. Amer. Math. Soc. Colloq. Publ., Providence, 1957.
- [3] O. A. Ladyzhenskaya, V. A. Solonnikov, and N. N. Ural'ceva. *Linear and Quasilinear Equations of Parabolic Type*. American Math. Society, Providence, R.I., 1968.
- [4] J. L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations*. Springer-Verlag, Berlin, 1971.
- [5] F. Tröltzsch. *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*. 2010.
- [6] J. Wloka. *Partial differential equations*. Cambridge University Press, 1987.
- [7] E. Zeidler. *Nonlinear Functional Analysis and its Applications*, volume II/A. Springer, New York, 1990.