12 GALERKIN APPROACH FOR PARABOLIC PROBLEMS

To obtain a finite-dimensional approximation of (11.1), we need to discretize in both space and time: either separately (combining finite elements in space with a time stepping method for ordinary differential equations) or all-at-once (using a Galerkin approach with suitable discrete test spaces). Only a brief overview over the different approaches is given here.

12.1 TIME STEPPING METHODS

These approaches can be further discriminated based on the order of operations:

Method of lines This method starts with a discretization in space to obtain a system of ordinary differential equations, which are then solved with one of the vast number of available methods. In the context of finite element methods, we use a discrete space V_h of piecewise polynomials defined on the triangulation $\mathcal{T}_h$ of the domain Ω. Given a nodal basis $\{\varphi_j\}_{j=1}^{N_h}$ of V_h, we approximate the unknown solution as $u_h(t,x) = \sum_{j=1}^{N_h} U_j(t)\varphi_j(x)$. Letting $\mathcal{P}_h$ denote the L^2 projection on V_h and using the mass matrix $\mathrm{M}_{ij} = (\varphi_i, \varphi_j)$ and the (time-dependent) stiffness matrix $\mathrm{K}(t)_{ij} = a(t; \varphi_i, \varphi_j)$ yields the following linear system of ordinary differential equations for the coefficient vector $U(t) = (U_1(t), \dots U_{N_h}(t))^T$:

$$\begin{cases} \mathrm{M}\frac{d}{dt}U(t) + \mathrm{K}(t)U(t) = \mathrm{M}F(t), \\ \qquad\qquad U(0) = U_0, \end{cases}$$

where U_0 and $F(t)$ are the coefficients vectors of $\mathcal{P}_h u_0$ and $\mathcal{P}_h f(t)$, respectively. The choice of integration method for this system depends on the properties of K (such as its stiffness, which can lead to numerical instability). Some details can be found, e.g., in [Ern & Guermond 2004, Chapter 6.1].

Rothe's method This method consists in treating (11.1) as an ordinary differential equation in the Banach space V, which is discretized in time by replacing the time derivative $d_t u$ by a difference quotient:

- The *implicit Euler scheme* uses the backward difference quotient

$$d_t u(t+k) \approx \frac{u(t+k) - u(t)}{k}$$

for $k > 0$ at time $t+k$ to obtain for given $u(t)$ and unknown $u(t+k) \in V$ the *stationary* partial differential equation

$$\langle u(t+k), v\rangle_H + k\, a(t+k; u(t+k), v) = \langle u(t), v\rangle_H + k\,\langle f(t+k), v\rangle_{V^*,V}$$

for all $v \in V$.

- The *Crank–Nicolson scheme* uses the central difference quotient

$$d_t u(t + \tfrac{k}{2}) \approx \frac{u(t+k) - u(t)}{k}$$

for $k > 0$ at time $t + \frac{k}{2}$ to obtain

$$\langle u(t+k), v\rangle_H + \tfrac{k}{2}\, a(t+\tfrac{k}{2}; u(t+k), v) = \langle u(t), v\rangle_H - \tfrac{k}{2}\, a(t+\tfrac{k}{2}; u(t), v) + k\,\langle f(t+\tfrac{k}{2}), v\rangle_{V^*,V} \tag{12.1}$$

for all $v \in V$.

Starting with $t = 0$, these are then approximated and solved in turn for $u(t_m)$, $t_m := mk$, using a finite element discretization in space. This approach is discussed in detail in [Thomée 2006, Chapters 7–9]. The advantage of Rothe's method is that at each time step, a different spatial discretization can be used.

12.2 GALERKIN METHODS

Proceeding as in the stationary case, we can apply a Galerkin approximation to (11.2) by replacing X and Y with finite-dimensional spaces X_h and Y_h. Again, we can further discriminate between conforming and non-conforming approaches.

Conforming Galerkin methods In a conforming approach, we choose $X_h \subset X$ and $Y_h \subset Y$ and seek $u_h \in X_h$ such that

$$\int_0^T \langle d_t u_h(t), y_h(t)\rangle_{V^*,V} + a(t; u_h(t), y_h(t))\,dt = \int_0^T \langle f(t), y_h(t)\rangle_{V^*,V}\,dt \tag{12.2}$$

for all $y_h \in Y_h$. We now choose the discrete spaces as tensor products in space and time: Let

$$0 = t_0 < t_1 < \dots < t_N = T$$

and choose for each t_m, $1 \leq m \leq N$, a (possibly different) finite-dimensional subspace $V_m \subset V$. Let $P_r(t_{m-1}, t_m; V_m)$ denote the space of polynomials on the interval $[t_{m-1}, t_m]$ with degree up to r with values in V_m. Then we define

$$X_h = \left\{y_h \in C(0,T;V) : y_h|_{[t_{m-1},t_m]} \in P_r(t_{m-1}, t_m; V_m),\ 1 \leq m \leq N,\ y_h(0) = u_0\right\},$$
$$Y_h = \left\{y_h \in L^2(0,T;V) : y_h|_{[t_{m-1},t_m]} \in P_{r-1}(t_{m-1}, t_m; V_m),\ 1 \leq m \leq N\right\}.$$

Since this is a conforming approximation, we can deduce well-posedness of the corresponding discrete problem in the usual fashion (noting that $d_t u_h \in Y_h$ for $u_h \in X_h$). (Since functions in X – and hence in X_h are continuous in time by Theorem 11.1, this approach is often called *continuous Galerkin* or $cG(r)$ method.)

This approach is closely related to Rothe's method. Consider the case $r = 1$ (i.e., piecewise linear in time) and, for simplicity, a time-independent bilinear form. Since functions in X_h are continuous at $t = t_m$ for all $0 \leq m \leq N$ and linear on each intervall $[t_{m-1}, t_m]$, we can write

$$u_h(t) = \frac{t_m - t}{t_m - t_{m-1}} u_h(t_{m-1}) + \frac{t - t_{m-1}}{t_m - t_{m-1}} u_h(t_m), \qquad t \in [t_{m-1}, t_m]$$

with coefficients $u_h(t_{m-1}), u_h(t_m) \in V_m$. (For $t_0 = 0$, we fix $u_h(t_0) = u_0$.) Similarly, functions in Y_h are constant and thus

$$y_h(t) = y_h(t_{m-1}) =: v_h \in V_m.$$

Inserting this into (12.2) and setting $k_m := t_m - t_{m-1}$ yields for all $v_h \in V_m$ that

$$\langle u_h(t_m) - u_h(t_{m-1}), v_h\rangle_{V^*,V} + \frac{k_m}{2} a(u_h(t_{m-1}) + u_h(t_m), v_h) = \int_{t_{m-1}}^{t_m} \langle f(t), v_h\rangle_{V^*,V}\, dt,$$

which is a modified Crank–Nicolson scheme (which, in fact, can be obtained by approximating the integral on the right-hand side using the midpoint rule, which is exact for $y_h \in Y_h$). For this method, one can show error estimates of the form[1]

$$\|u_h(t_m) - u(t_m)\|_{L^2(\Omega)} \leq C(h^s \|u_0\|_{H^s(\Omega)} + k^2 \|u_0\|_{H^4(\Omega)}),$$

for $f = 0$ and $u_0 \neq 0$, where s depends on the accuracy of the spatial discretization, and $k = \max k_m$.

Discontinuous Galerkin methods Instead of enforcing continuity of the discrete solution u_h through the definition of X_h, we can also use $X_h = Y_h$ and modify the bilinear form. Let $J_m := (t_{m-1}, t_m]$ denote the half-open interval between two time steps of length $k_m = t_m - t_{m-1}$. Then we set for $r \geq 0$

$$X_h = Y_h = \left\{y_h \in L^2(0,T;V) : y_h|_{J_m} \in P_r(t_{m-1}, t_m; V_m),\ 1 \leq m \leq N\right\} \subset Y,$$

[1] [Thomée 2006, Theorem 7.8]

where V_m is again a finite-dimensional subspace of V. Note that functions in X_h can be discontinuous at the points t_m but are continuous from the left with limits from the right, and so we will write for $u_h \in X_h$

$$u_m := u_h(t_m) = \lim_{\varepsilon \to 0} u_h(t_m - \varepsilon), \qquad u_m^+ := \lim_{\varepsilon \to 0} u_h(t_m + \varepsilon)$$

and

$$[\![u_h]\!]_m = u_m^+ - u_m.$$

Similarly to the stationary case, we now define the discrete bilinear form

$$b_h(u_h, y_h) = \sum_{m=1}^{N} \int_{J_m} \langle d_t u_h(t), y_h(t)\rangle_H + a(t; u_h(t), y_h(t))\, dt + \sum_{m=1}^{N} \langle [\![u_h]\!]_{m-1}, y_{m-1}^+\rangle_H$$

(which can be derived by integration by parts on each interval J_m and rearranging the jump terms). Note that as $0 \notin J_1$, we will need to specify $u_h(0) = u_0$ separately, which we do by setting $[\![u_h]\!]_0 := u_0^+ - u_0$. We then search for $u_h \in X_h$ satisfying

$$b_h(u_h, y_h) = \langle f, y_h\rangle_{Y^*,Y} \qquad \text{for all } y_h \in X_h. \tag{12.3}$$

Since the exact solution $u \in X$ is continuous and satisfies $u(0) = u_0$, we have

$$b_h(u, y_h) = b(u, y_h) = \langle f, y_h\rangle_{Y^*,Y} \qquad \text{for all } y_h \in X_h,$$

and hence this is a consistent approximation.

To prove well-posedness of the discrete problem, we proceed as in Theorem 8.2. Define the discrete norm

$$|\!|\!|u_h|\!|\!|^2 = \sum_{m=1}^{N} \int_{J_m} \|d_t u_h(t)\|_H^2 + \|u_h(t)\|_V^2\, dt + \|[\![u_h]\!]_m\|_H^2.$$

Theorem 12.1. *Under the assumptions of Theorem 11.3, there exists a unique solution $u_h \in X_h$ to (12.3) satisfying*

$$|\!|\!|u_h|\!|\!| \leq C\left(\int_0^T \|f(t)\|_{Y^*}^2\, dt + \|u_0\|_H^2\right).$$

Proof. Continuity of b_h with respect to $|\!|\!|\cdot|\!|\!|$ follows from the definition. It remains to show injectivity of $B_h : u \mapsto b_h(u, \cdot)$ (which suffices for bijectivity since $X_h = Y_h$ are finite-dimensional). Instead of verifying the inf-sup-condition, we do this directly. Let $u_h \in X_h$ satisfy $b_h(u_h, y_h) = 0$ for all $y_h \in X_h$ with $u_0 = 0$. Since functions in X_h can be discontinuous at the time points t_m, we can insert $y_h = \chi_{J_m} u_h \in X_h$ for each $1 \leq m \leq N$, where $\chi_{J_m}(t) = 1$

if $t \in J_m$ and zero else. We start with $J_1 = (t_0, t_1]$. Since χ_{J_1} is constant on J_1 and zero outside J_1, we have using $u_0 = 0$ that

$$\begin{aligned} 0 &= b_h(u_h, \chi_{J_1} u_h) \\ &= \int_{J_1} \langle d_t u_h(t), u_h(t)\rangle_{V^*,V} + a(t; u_h(t), u_h(t))\, dt + \langle u_0^+ - u_0, u_0^+\rangle_H \\ &\geq \frac{1}{2}\|u_1\|_H^2 - \frac{1}{2}\|u_0^+\|_H^2 + \alpha \int_{J_1} \|u_h(t)\|_V^2\, dt + \|u_0^+\|_H^2 \\ &\geq \frac{1}{2}\|u_1\|_H^2 + \alpha \int_{J_1} \|u_h(t)\|_V^2\, dt. \end{aligned}$$

Hence, $u_h|_{J_1} = 0$ and $u_1 = 0$, and we can proceed in a similar way for $J_2, J_3, \dots, J_N$ to deduce that $u_h = 0$. The estimate then follows from bijectivity using the closed range theorem. □

We next show a stability result for the discontinuous Galerkin approximation. For simplicity, we assume from now on that the bilinear form a is time-independent and symmetric, and that $V_1 = \cdots = V_N = V_h$. Let $A : V \to V^*$ again denote the operator corresponding to the bilinear form a, i.e., $\langle Au, v\rangle_H = a(u, v)$ for all $u, v \in V$. We also assume for the sake of presentation that the discrete solution u_h is sufficiently regular that $Au_h(t) \in H$.

Theorem 12.2. *For given $f \in L^2(0, T; H)$ and $u_0 \in H$, the solution u_h of (12.3) satisfies*

$$\sum_{m=1}^{N} \int_{J_m} \|d_t u_h(t)\|_H^2 + \|Au_h(t)\|_H^2\, dt + k_m^{-1} \|[\![u_h]\!]_{m-1}\|_H^2 \leq C\left(\int_0^T \|f(t)\|_H^2\, dt + \|u_0\|_H^2\right).$$

Proof. We estimate in turn each term on the left-hand side by inserting suitable test functions y_h in (12.3).

Step 1. To estimate $\|Au_h(t)\|_H$, we set $y_h = \chi_{J_m} Au_h$ for $1 \leq m \leq N$ to obtain

$$\int_{J_m} \langle d_t u_h(t), Au_h(t)\rangle_H + \|Au_h(t)\|_H^2\, dt + \langle [\![u_h]\!]_{m-1}, (Au_h)_{m-1}^+\rangle_H = \int_{J_m} \langle f(t), Au_h(t)\rangle_H\, dt.$$

Due to the bilinearity and symmetry of a, we have

$$\begin{aligned} \int_{J_m} \langle d_t u_h(t), Au_h(t)\rangle_H\, dt &= \int_{J_m} a(u_h(t), d_t u_h(t))\, dt = \int_{J_m} \frac{d}{dt}\left(\frac{1}{2} a(u_h(t), u_h(t))\right) dt \\ &= \frac{1}{2} a(u_m, u_m) - \frac{1}{2} a(u_{m-1}^+, u_{m-1}^+). \end{aligned}$$

Similarly, since A is time-independent,

$$\begin{aligned} \langle [\![u_h]\!]_{m-1}, (Au_h)_{m-1}^+\rangle_H &= a([\![u_h]\!]_{m-1}, u_{m-1}^+) \\ &= \frac{1}{2} a([\![u_h]\!]_{m-1}, u_{m-1}^+ + u_{m-1} + [\![u_h]\!]_{m-1}) \\ &= \frac{1}{2} a(u_{m-1}^+, u_{m-1}^+) - \frac{1}{2} a(u_{m-1}, u_{m-1}) + \frac{1}{2} a([\![u_h]\!]_{m-1}, [\![u_h]\!]_{m-1}). \end{aligned}$$

Inserting these into the bilinear form $b_h(u_h, y_h)$ yields

$$a([\![u_h]\!]_{m-1}, [\![u_h]\!]_{m-1}) + a(u_m, u_m) - a(u_{m-1}, u_{m-1}) + 2\int_{J_m} \|Au_h(t)\|_H^2\, dt = 2\int_{J_m} \langle f, Au_h(t)\rangle_H\, dt. \tag{12.4}$$

Summing over all $1 \leq m \leq N$ yields

$$\sum_{m=1}^{N} a([\![u_h]\!]_{m-1}, [\![u_h]\!]_{m-1}) + \sum_{m=1}^{N} \int_{J_m} 2\|Au_h(t)\|_H^2\, dt \leq \sum_{m=1}^{N} \int_{J_m} 2\langle f(t), Au_h(t)\rangle_H\, dt + a(u_0, u_0). \tag{12.5}$$

For $2 \leq m \leq N$, we can simply use coercivity of a to eliminate the jump terms and apply Young's inequality to $\langle f(t), Au_h(t)\rangle_H$ to absorb the norm of Au_h on J_m in the left-hand side. For $m = 1$, we use that

$$a([\![u_h]\!]_0, [\![u_h]\!]_0) - a(u_0, u_0) = a(u_0^+, u_0^+) - 2a(u_0, u_0^+)$$

and for $\varepsilon > 0$ the generalized Young's inequality

$$a(u_0, u_0^+) = \langle u_0, Au_0^+\rangle_H \leq \frac{\varepsilon}{2}\|Au_0^+\|_H^2 + \frac{1}{2\varepsilon}\|u_0\|_H^2.$$

Since $t \mapsto \|Au_h(t)\|_H^2$ is a polynomial in t of degree up to $2r$ on J_1, we have the estimate

$$k_1 \|Au_0^+\|_H^2 \leq C \int_{t_0}^{t_1} \|Au_h(t)\|_H^2\, dt.$$

Choosing $\varepsilon > 0$ small enough such that $\varepsilon C k_1^{-1} < 1$ yields

$$\sum_{m=1}^{N} \int_{J_m} \|Au_h(t)\|_H^2\, dt \leq C\left(\int_0^T \|f(t)\|_H^2\, dt + \|u_0\|_H^2\right). \tag{12.6}$$

Step 2. For the bound on $d_t u_h$, we use the inverse estimate

$$\int_{J_m} \|y_h(t)\|_H^2\, dt \leq C k_m^{-1} \int_{J_m} (t - t_{m-1})\, \|y_h(t)\|_H^2\, dt$$

for all $y_h \in P_r(t_{m-1}, t_m; V_h)$, which follows from a scaling argument in time and equivalence of norms on the finite-dimensional space $P_r(0,1;V_h)$. Now choose $y_h = \chi_{J_m}(t - t_{m-1})d_t u_h$ for $1 \le m \le N$. Since $y_{m-1}^+ = 0$, we have using the Cauchy–Schwarz inequality that

$$\begin{aligned}\int_{J_m} (t - t_{m-1})\,\|d_t u_h(t)\|_H^2\,dt &= \int_{J_m} (t - t_{m-1})\,\langle f(t) - Au_h(t), d_t u_h(t)\rangle_H\,dt\\ &\le \left(\int_{J_m} (t - t_{m-1})\,\|f(t) - Au_h(t)\|_H^2\,dt\right)^{\frac12}\\ &\quad\cdot \left(\int_{J_m} (t - t_{m-1})\,\|d_t u_h(t)\|_H^2\,dt\right)^{\frac12}.\end{aligned}$$

Applying the inverse estimate for $y_h = d_t u_h$, the Cauchy–Schwarz inequality for the first integral and estimating the norm there using (12.6) yields

$$\sum_{m=1}^{N} \int_{J_m} \|d_t u_h(t)\|_H^2\,dt \le C\left(\int_0^T \|f(t)\|_H^2\,dt + \|u_0\|_H^2\right). \tag{12.7}$$

Step 3. It remains to estimate the jump terms. For this, we set $y_h = \chi_{J_m} [\![u_h]\!]_{m-1}$ for $1 \le m \le N$. This yields

$$\begin{aligned}\left\|[\![u_h]\!]_{m-1}\right\|_H^2 &= \int_{J_m} \langle f(t) - Au_h(t), [\![u_h]\!]_{m-1}\rangle_H - \langle d_t u_h(t), [\![u_h]\!]_{m-1}\rangle_H\,dt\\ &\le \frac{k_m}{2}\int_{J_m} \|f(t) - Au_h(t) - d_t u_h(t)\|_H^2\,dt + \frac{1}{2k_m}\int_{J_m} \left\|[\![u_h]\!]_{m-1}\right\|_H^2\,dt,\end{aligned}$$

where we have used the generalized Young's inequality. Since $[\![u_h]\!]_{m-1}$ is constant in time, we have

$$\int_{J_m} \left\|[\![u_h]\!]_{m-1}\right\|_H^2\,dt = k_m \left\|[\![u_h]\!]_{m-1}\right\|_H^2.$$

From (12.6) and (12.7), we thus obtain

$$\sum_{m=1}^{N} k_m^{-1} \left\|[\![u_h]\!]_{m-1}\right\|_H^2 \le C\left(\int_0^T \|f(t)\|_H^2\,dt + \|u_0\|_H^2\right),$$

which completes the proof. □

Before we address a priori error estimates, we discuss how to formulate discontinuous Galerkin methods as time stepping methods. For simplicity, we assume from now on that the bilinear form a is time-independent and that $V_1 = \cdots = V_N = V_h$. First consider the case $r = 0$, i.e., piecewise constant functions in time. Then, $d_t(u_h|_{J_m}) = 0$ and $u_h|_{J_m} = u_m =$

$u_{m-1}^+ \in V_h$. Using as test functions $y_h = \chi_{J_m} v_h$ for arbitrary $v_h \in V_h$ and $m = 1, \dots, N$, we obtain

$$\langle u_m, v_h\rangle_H + k_m\, a(u_m, v_h) = \langle u_{m-1}, v_h\rangle_H + \int_{J_m} \langle f(t), v_h\rangle_{V^*,V}\,dt$$

for all $v_h \in V_h$, which is a variant of the implicit Euler scheme.[2] For $r = 1$ (piecewise linear functions), we make the ansatz

$$u_h|_{J_m}(t) = u_m^0 + \frac{t - t_{m-1}}{k_m} u_m^1$$

for $u_m^0, u_m^1 \in V_h$. Again, we choose for each J_m test functions which are zero outside J_m; specifically, we take $\chi_{J_m}(t)v_h$ and $\chi_{J_m}(t)\frac{t-t_{m-1}}{k_m} w_h$ for arbitrary $v_h, w_h \in V_h$. Inserting these in turn into the bilinear form and computing the integrals yields the coupled system

$$\begin{aligned}\langle u_m^0, v_h\rangle_H + k_m\, a(u_m^0, v_h) + \langle u_m^1, v_h\rangle_H + \frac{k_m}{2} a(u_m^1, v_h)\\ = \langle u_{m-1}, v_h\rangle_H + \int_{J_m} \langle f(t), v_h\rangle_{V^*,V}\,dt,\\ \frac{k_m}{2} a(u_m^0, w_h) + \frac{1}{2}\langle u_m^1, w_h\rangle_H + \frac{k_m}{3} a(u_m^1, w_h)\\ = \frac{1}{k_m}\int_{J_m} (t - t_{m-1})\,\langle f(t), w_h\rangle_{V^*,V}\,dt\end{aligned}$$

for all $v_h, w_h \in V_h$. By solving this system successively at each time step and setting $u_m = u_m^0 + u_m^1$, we obtain the approximate solution u_h.[3]

12.3 A PRIORI ERROR ESTIMATES

As before, we will estimate the error $u - u_h$ using the approximation properties of the space X_h. Due to the discontinuity of the functions in X_h, we can use a local projection on each time intervall J_m to bound the approximation error. It will be convenient to split this error into two parts: one due to the temporal and one due to the spatial discretization.

We first consider the temporal discretization error. Let

$$X_r = \left\{y_r \in L^2(0,T;V) : y_r|_{J_m} \in P_r(t_{m-1}, t_m; V),\ 1 \le m \le N\right\}$$

and consider the local projection $\pi_r u \in X_r$ of $u \in X$ defined by $\pi_r u(t_0) = u(t_0)$ and

$$\begin{cases} \pi_r u(t_m) = u(t_m), & \\ \displaystyle\int_{J_m} (u(t) - \pi_r u(t))\varphi(t)\,dt = 0 & \text{for all } \varphi \in P_{r-1}(J_m; V),\end{cases}$$

[2] If the discrete spaces are different for each time interval, we need to use the H-projection of u_{m-1} on V_m.

[3] Similarly, discontinuous Galerkin methods for $r \ge 2$ lead to $(r+1)$-stage implicit Runge–Kutta time-stepping schemes.

for all $1 \le m \le N$. (For $r = 0$, the second condition is void.) This projection is well-defined since $u \in X$ is continuous in time, and hence the interpolation conditions make sense. Using the Bramble–Hilbert lemma and a scaling argument, we obtain for sufficiently smooth u the following error estimate for every $t \in J_m$, $1 \le m \le N$:

$$\|u(t) - \pi_r u(t)\|_H \le C k_m^{r+1} \int_{J_m} \left\|d_t^{r+1} u(\tau)\right\|_H \, d\tau.$$

Similarly, we assume that for each $t \in [0,T]$ the spatial interpolation error in V_h satisfies the estimate

$$\|u(t) - \mathcal{I}_h u(t)\|_H + h\,\|u(t) - \mathcal{I}_h u(t)\|_V \le C h^{s+1} \|u(t)\|_{H^{s+1}(\Omega)}.$$

(This is the case, e.g., if $H = L^2(\Omega)$, $V = H_0^1(\Omega)$, and V_h consists of continuous piecewise polynomials of degree $s \ge 1$; see Theorem 5.9.)

Finally, we will make use of a duality argument, which requires considering for given $\varphi \in H$ the solution of the adjoint equation

$$b_h(y_h, z_h) = 0 \qquad \text{with} \quad z_N = \varphi.$$

Integrating by parts on each interval J_m and rearranging the jump terms, we can express the adjoint equation as

$$\sum_{m=1}^{N} \int_{J_m} -\langle y_h(t), d_t z_h(t)\rangle_H + a(y_h(t), z_h(t))\, dt + \sum_{m=1}^{N-1} \left\langle y_m, [\![z_h]\!]_m \right\rangle_H + \langle y_N, z_N\rangle_H = \langle y_N, \varphi\rangle_H . \tag{12.8}$$

This can be interpreted as a backwards in time equation with "initial value" $z_h(t_N) = \varphi$. Making the substitution $\tau = t_N - t$, we can apply Theorem 12.2 to obtain

$$\sum_{m=1}^{N} \int_{J_m} \|d_t z_h(t)\|_H^2 + \|A^* z_h(t)\|_H^2 \, dt + \sum_{m=1}^{N} \left\|[\![z_h]\!]_{m-1}\right\|_H^2 \le C \|\varphi\|_H^2, \tag{12.9}$$

where A is again the operator corresponding to the bilinear form a.

Now everything is in place to show the following a priori estimate for the discrete solution at each time step.[4]

[4]It is possible – though more involved – to show error estimates for arbitrary $t \in [0,T]$; see, e.g., [Thomée 2006, Theorem 12.2].

Theorem 12.3. *For $r = 0$, the solutions $u \in X$ to (11.2) and $u_h \in X_h$ to (12.3) satisfy*

$$\|u(t_m) - u_m\|_H \le C \max_{1\le n\le m} \left(h^{s+1} \sup_{t\in J_n} \|u(t)\|_{H^{s+1}(\Omega)} + k_n \int_{J_n} \|d_t u\|_H \, dt \right)$$

for all $1 \le m \le N$.

Proof. We write the error $e(t)$ at each time t as

$$\begin{aligned} e(t) = u(t) - u_h(t) &= (u(t) - \mathcal{I}_h \pi_r u(t)) + (\mathcal{I}_h \pi_r u(t) - u_h(t)) \\ &=: e_1(t) + e_2(t). \end{aligned}$$

For $t = t_m$, we have $\pi_r u(t_m) = u(t_m)$ by construction, and hence

$$\|e_1(t_m)\|_H = \|\mathcal{I}_h u(t_m) - u(t_m)\|_H \le C h^{s+1} \|u(t_m)\|_{H^{s+1}(\Omega)}. \tag{12.10}$$

To bound $e_2(t_m)$, we use the duality trick. For arbitrary $\varphi \in H$, let z_h denote the solution of (12.8) with $N = m$. Since we have a consistent approximation, we can use the Galerkin orthogonality to deduce

$$0 = b_h(e, y_h) = b_h(e_1, y_h) + b_h(e_2, y_h) \quad \text{for all } y_h \in X_h.$$

From this and $d_t(z_h|_{J_n}) = 0$ we obtain with $y_h = e_2 \in X_h$ that

$$\begin{aligned} \langle e_2(t_m), \varphi\rangle_H &= b_h(e_2, z_h) = -b_h(e_1, z_h) \\ &= -\sum_{n=1}^{m} \int_{J_n} a(e_1(t), z_h(t))\, dt - \sum_{n=1}^{m-1} \left\langle e_1(t_n), [\![z_h]\!]_n \right\rangle_H - \langle e_1(t_m), \varphi\rangle_H . \end{aligned}$$

Introducing $\langle A e_1, z_h(t)\rangle_H = a(e_1, z_h)$ as above and estimating e_1 by its pointwise in time maximum yields

$$|\langle e_2(t_m), \varphi\rangle_H| \le \left(\sup_{t\le t_m} \|e_1(t)\|_H\right) \left(\sum_{n=1}^{m} \int_{J_n} \|A^* z_h(t)\|_H \, dt + \sum_{n=1}^{m-1} \left\|[\![z_h]\!]_n\right\|_H + \|\varphi\|_H \right).$$

From the dual definition of the norm in H and estimate (12.9), we obtain

$$\|e_2(t_m)\|_H \le C \max_{1\le n\le m} \sup_{t\in J_n} \|e_1(t)\|_H . \tag{12.11}$$

It remains to bound $e_1(t)$ for arbitrary $t \in J_n$, which we do by estimating

$$\begin{aligned} \|e_1(t)\|_H &= \|u(t) - \mathcal{I}_h \pi_r u(t)\|_H \\ &\le \|u(t) - \pi_r u(t)\|_H + \|\pi_r u(t) - \mathcal{I}_h \pi_r u(t)\|_H \\ &\le C k_n \int_{J_n} \|d_t u(\tau)\|_H \, d\tau + C h^{s+1} \|u(t)\|_{H^{s+1}(\Omega)} . \end{aligned} \tag{12.12}$$

Combining (12.10), (12.11) and (12.12) yields the claim. □

For $r = 1$, one can proceed similarly (using that $d_t z_h|_{J_m} \in P_{r-1}(J_m, V_h)$, and hence that $\int_{J_n} \langle d_t z_h(t), u(t) - \pi_r u(t)\rangle_H \, dt$ vanishes by definition of π_r) to obtain[5]

Theorem 12.4. *For $r = 1$, the solutions $u \in X$ to (11.2) and $u_h \in X_h$ to (12.3) satisfy*

$$\|u(t_m) - u_m\|_H \leq C \max_{1\leq n\leq m} \left(h^{s+1} \sup_{t\in J_n} \|u(t)\|_{H^{s+1}(\Omega)} + k_n^3 \int_{J_n} \|d_t^2 u(t)\|_{H^2(\Omega)} \, dt \right)$$

for all $1 \leq m \leq N$.

The general case (including time-dependent bilinear form a and different discrete spaces V_m) can be found in [Chrysafinos & Walkington 2006].

[5] e.g., [Thomée 2006, Theorem 12.7]

BIBLIOGRAPHY

R. A. Adams & J. J. F. Fournier (2003), *Sobolev Spaces*, 2nd ed., Amsterdam: Academic Press.

D. Arnold et al. (2002), Unified analysis of discontinuous Galerkin methods for elliptic problems, *SIAM Journal on Numerical Analysis* 39(5), 1749–1779, doi: 10.1137/S0036142901384162.

W. Bangerth, R. Hartmann & G. Kanschat (2013), deal.II Differential Equations Analysis Library, Technical Reference, url: http://www.dealii.org/.

D. Boffi, F. Brezzi & M. Fortin (2013), *Mixed and Finite Element Methods and Applications*, vol. 44, Springer Series in Computational Mathematics, New York: Springer, doi: 10.1007/978-3-642-36519-5.

D. Braess (2007), *Finite Elements*, 3rd ed., Cambridge: Cambridge University Press, doi: 10.1017/CBO9780511618635.

J. H. Bramble & S. R. Hilbert (1970), Estimation of linear functionals on Sobolev spaces with application to Fourier transforms and spline interpolation. *SIAM J. Numer. Anal.* 7, 112–124, doi: 10.1137/0707006.

S. C. Brenner & L. R. Scott (2008), *The Mathematical Theory of Finite Element Methods*, 3rd ed., vol. 15, Texts in Applied Mathematics, New York: Springer, doi: 10.1007/978-0-387-75934-0.

K. Chrysafinos & N. J. Walkington (2006), Error estimates for the discontinuous Galerkin methods for parabolic equations, *SIAM J. Numer. Anal.* 44(1), 349–366 (electronic), doi: 10.1137/030602289.

P. G. Ciarlet (2002), *The Finite Element Method for Elliptic Problems*, vol. 40, Classics in Applied Mathematics, Reprint of the 1978 original [North-Holland, Amsterdam], Philadelphia, PA: Society for Industrial & Applied Mathematics (SIAM), doi: 10.1137/1.9780898719208.

R. Courant (1943), Variational methods for the solution of problems of equilibrium and vibrations, *Bull. Amer. Math. Soc.* 49, 1–23, doi: 10.1090/S0002-9904-1943-07818-4.

D. A. Di Pietro & A. Ern (2012), *Mathematical Aspects of Discontinuous Galerkin Methods*, vol. 69, Mathématiques et Applications, New York: Springer, doi: 10.1007/978-3-642-22980-0.

R. E. Edwards (1965), *Functional Analysis. Theory and Applications*, Rinehart & Winston, New York: Holt.

A. Ern & J.-L. Guermond (2004), *Theory and Practice of Finite Elements*, vol. 159, Applied Mathematical Sciences, New York: Springer, doi: 10.1007/978-1-4757-4355-5.

L. C. Evans (2010), *Partial Differential Equations*, 2nd ed., vol. 19, Graduate Studies in Mathematics, Providence, RI: American Mathematical Society, doi: 10.1090/gsm/019.

P. Grisvard (2011), *Elliptic Problems in Nonsmooth Domains*, Classics in Applied Mathematics 69, Society for Industrial & Applied Mathematics, doi: 10.1137/1.9781611972030.

O. A. Ladyzhenskaya & N. N. Ural'tseva (1968), *Linear and Quasilinear Elliptic Equations*, Translated from the Russian by Scripta Technica, Inc. Translation editor: Leon Ehrenpreis, New York: Academic Press.

A. Logg, K.-A. Mardal, G. N. Wells, et al. (2012), *Automated Solution of Differential Equations by the Finite Element Method*, Springer, doi: 10.1007/978-3-642-23099-8,

J.-C. Nédélec (1980), Mixed finite elements in $\mathbb{R}^3$, *Numerische Mathematik* 35(3), 315–341, doi: 10.1007/BF01396415.

R. Rannacher (2008), Numerische Mathematik 2, Lecture notes, url: http://numerik.iwr.uni-heidelberg.de/~lehre/notes/num2/numerik2.pdf.

P. Raviart & J. Thomas (1977), A mixed finite element method for 2-nd order elliptic problems, in: *Mathematical Aspects of Finite Element Methods*, ed. by I. Galligani & E. Magenes, vol. 606, Lecture Notes in Mathematics, Berlin: Springer, 292–315, doi: 10.1007/BFb0064470.

M. Renardy & R. C. Rogers (2004), *An Introduction to Partial Differential Equations*, 2nd ed., vol. 13, Texts in Applied Mathematics, New York: Springer, doi: 10.1007/b97427.

R. E. Showalter (1997), *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, vol. 49, Mathematical Surveys and Monographs, Providence, RI: American Mathematical Society, doi: 10.1090/surv/049.

G. Strang (1972), Variational crimes in the finite element method, in: *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations (Proc. Sympos., Univ. Maryland, Baltimore, Md., 1972)*, New York: Academic Press, 689–710, doi: 10.1016/B978-0-12-068650-6.50030-7.

E. Süli (2011), Finite Element Methods for Partial Differential Equations, Lecture notes, url: http://people.maths.ox.ac.uk/suli/fem.pdf.

V. Thomée (2006), *Galerkin Finite Element Methods for Parabolic Problems*, 2nd ed., vol. 25, Springer Series in Computational Mathematics, Berlin: Springer, doi: 10.1007/3-540-33122-0.

G. M. Troianiello (1987), *Elliptic Differential Equations and Obstacle Problems*, The University Series in Mathematics, New York: Plenum Press, doi: 10.1007/978-1-4899-3614-1.

R. Verfürth (2013), *A Posteriori Error Estimation Techniques for Finite Element Methods*, Numerical Mathematics and Scientific Computation, Oxford: Oxford University Press, doi: 10.1093/acprof:oso/9780199679423.001.0001.

J. Wloka (1987), *Partial Differential Equations*, Translated from the German by C. B. Thomas and M. J. Thomas, Cambridge: Cambridge University Press, doi: 10.1017/CBO9781139171755.

E. Zeidler (1995a), *Applied Functional Analysis, Applications to mathematical physics*, vol. 108, Applied Mathematical Sciences, New York: Springer, doi: 10.1007/978-1-4612-0815-0.

E. Zeidler (1995b), *Applied Functional Analysis, Main principles and their applications*, vol. 109, Applied Mathematical Sciences, New York: Springer, doi: 10.1007/978-1-4612-0821-1.

M. Zlámal (1968), On the finite element method, *Numer. Math.* 12, 394–409, doi: 10.1007/BF02161362.