ANALYSIS OF BACKWARD EULER PRIMAL DPG METHODS

THOMAS FÜHRER, NORBERT HEUER, AND MICHAEL KARKULIK

Abstract. We analyse backward Euler time stepping schemes for the primal DPG formulation of a class of parabolic problems. Optimal error estimates are shown in the natural norm and in the $L^2$ norm of the field variable. For the heat equation the solution of our primal DPG formulation equals the solution of a standard Galerkin scheme and, thus, optimal error bounds are found in the literature. In the presence of advection and reaction terms, however, the latter identity is not valid anymore and the analysis of optimal error bounds requires to resort to elliptic projection operators. It is essential that these operators be projections with respect to the spatial part of the PDE, as in standard Galerkin schemes, and not with respect to the full PDE at a time step, as done previously.

1. Introduction

In this work we analyse a backward Euler primal DPG time stepping scheme for the parabolic problem

\begin{align}
\dot{u} - \text{div} A \nabla u + \beta \cdot \nabla u + \gamma u &= f \quad \text{in } (0, T) \times \Omega, \\
u &= 0 \quad \text{in } (0, T) \times \partial \Omega, \\
u(0, \cdot) &= u_0 \quad \text{in } \Omega.
\end{align}

Here, $\Omega \subset \mathbb{R}^d$ ($d \in \{2, 3\}$) denotes a bounded Lipschitz domain. We make the standing assumptions that the coefficients $A, \beta,$ and $\gamma$ are time-independent, that they are essentially bounded in space, that $A$ is a symmetric positive matrix almost everywhere in $\Omega$, and that $(\beta \cdot \nabla v + \gamma v, v) \geq 0$ for all $v \in H^1_0(\Omega)$. The $L^2(\Omega)$ inner product is denoted with $\langle \cdot, \cdot \rangle$. We assume that the data satisfy $f(t, \cdot) \in L^2(\Omega)$ for all $t \in [0, T]$ and $u_0 \in L^2(\Omega)$.

The discontinuous Petrov–Galerkin method with optimal test functions (DPG) pertains to the class of minimum residual methods and was introduced in a series of papers [5, 7, 10]. It has been successfully applied to elliptic problems, see, e.g., [6, 8] for the Poisson problem and [18, 16] for fourth-order problems. Through the use of optimal test functions, the discrete problem inherits the stability of the continuous problem. This comes in advantageous for problems where robustness is one of the main challenges, e.g., singularly perturbed problems, see [23] for reaction-dominated diffusion problems and [11, 9] for the convection-dominated case. Space-time DPG methods have been studied previously, see, e.g., [9, 13, 12]. For other space-time minimum residual methods we refer to [21, 1, 29, 27]. Approaches employing the DPG methodology for the time discretization of parabolic and hyperbolic initial value problems have recently been investigated, cf. [21, 25]. On the other hand, time-stepping methods for ODEs are frequently employed in combination with standard Galerkin finite element methods in space, cf. the monograph [28] for parabolic equations, but less so with DPG methods. To the best of our knowledge there exist only two works in this direction, dealing with time-stepping and spatial DPG methods for the heat equation, namely [19] and [26].
In [19], a backward Euler method is used to discretise in time, and then the DPG methodology is applied to the ultraweak variational formulation of the resulting equations. The a priori analysis given there employs the Galerkin projection with respect to these very equations, and hence a spatial discretization error has to be accounted for in every time step. This gives rise to a theoretical error bound of order $O(h/k) + O(k)$ for lowest-order discretizations with $h$ and $k$ being the spatial mesh-width and time step, respectively. As numerical experiments from [26] indicate, this asymptotic error bound is not optimal. The authors of [26] study general spatial mesh-width and time step, respectively. As numerical experiments from [26] indicate, this fact that the field solution component of the primal DPG method is identical to the solution of the heat equation ($A$ is the identity, $\beta = 0$, and $\gamma = 0$) optimal error estimates in the $H^1(\Omega)$ norm. Section 4 is devoted to optimal error estimates in the $H^1(\Omega)$ norm. Section 4 is devoted to optimal error estimates in the $H^1(\Omega)$ norm. Section 4 is devoted to optimal error estimates in the $H^1(\Omega)$ norm. Section 4 is devoted to optimal error estimates in the $H^1(\Omega)$ norm.

Our motivation for the present work is to give a theoretically sound explanation of the optimal convergence rates seen in the numerical experiments from [26] for the backward Euler primal DPG method. We will consider general second order linear elliptic spatial differential operators. For the method as well as the necessary notation. We prove stability of the method and provide quasi-optimal error estimates in the $H^1(\Omega)$ norm. Section 4 is devoted to optimal error estimates in the $H^1(\Omega)$ norm. Section 4 is devoted to optimal error estimates in the $H^1(\Omega)$ norm. Section 4 is devoted to optimal error estimates in the $H^1(\Omega)$ norm.

The remainder of this paper is organised as follows: In Section 2 we introduce the fully discrete method as well as the necessary notation. We prove stability of the method and provide quasi-optimality results for the elliptic projection operator. In Section 3 we use these results to show optimal error estimates in the $H^1(\Omega)$ norm. Section 4 is devoted to optimal error estimates in weaker norms, particularly in $L^2(\Omega)$.

2. Time-stepping DPG formulation

2.1. Notation. The notation $a \lesssim b$ means that there exists a constant $C > 0$ that (possibly) depends on $\Omega$, $A$, $\beta$, $\gamma$, and $T$, but is independent of involved functions. We write $a \asymp b$ if $a \lesssim b$ and $b \lesssim a$. Furthermore, $a \gtrsim b$ stands for $b \lesssim a$.

We consider a time discretization $0 = t_0 < \cdots < t_N = T$, and for notational simplicity we use a uniform time step size $k = t_n - t_{n-1}$ (but we stress that this is not necessary). By $(\cdot, \cdot)$ and $\| \cdot \|$ we denote the inner product and norm in $L^2(\Omega)$. We consider spatial discretizations based on a shape-regular conforming simplicial mesh $T$ of $\Omega$. To any partition $T$ we associate its skeleton $S$ consisting of the boundaries of all elements, and the trace space

$$H^{-1/2}(S) := \left\{ \tilde{\sigma} \in \prod_{K \in T} H^{-1/2}(\partial K) \mid \exists \sigma \in H(\text{div}; \Omega) : \tilde{\sigma}|_{\partial K} = \sigma \cdot n_K|_{\partial K}, \ \forall K \in T \right\},$$

where $n_K$ denotes the unit outward normal vector on $\partial K$. This is a Hilbert space with norm

$$\| \tilde{\sigma} \|_{-1/2,k} := \inf \left\{ (\| \sigma \|^2 + k \| \text{div} \sigma \|^2)^{1/2} \mid \sigma \in H(\text{div}; \Omega), \tilde{\sigma}|_{\partial K} = \sigma \cdot n_K|_{\partial K}, \ \forall K \in T \right\}.$$
The trial space of our method will be
\[ U := H^1_0(\Omega) \times H^{-1/2}(S) \]
equipped with the norm
\[ \|(u, \tilde{\sigma})\|_{U,k}^2 := \|\nabla u\|^2 + \|\tilde{\sigma}\|_{-1/2,k}^2, \]
and the test space will be
\[ V := H^1(\mathcal{T}) := \prod_{K \in \mathcal{T}} H^1(K) \]
equipped with the norm
\[ \|v\|_{V,k}^2 := \frac{1}{k} \|v\|^2 + \|A^{1/2} \nabla_{\mathcal{T}} v\|^2. \]
Here, \( \nabla_{\mathcal{T}} v \) is the \( \mathcal{T} \)-piecewise gradient. For functions \( \tilde{\sigma} \in H^{-1/2}(S) \) we define
\[ \langle \tilde{\sigma}, v \rangle_S := (\text{div} \, \sigma, v) + (\sigma, \nabla_{\mathcal{T}} v) \quad \forall v \in H^1(\mathcal{T}), \]
where \( \sigma \in H(\text{div}; \Omega) \) with \( \sigma \cdot n_K|_{\partial K} = \tilde{\sigma}|_{\partial K} \) for all \( K \in \mathcal{T} \). We note that the definition is independent of the choice of \( \sigma \) and recall the following result.

**Lemma 1.**
\[ \|\tilde{\sigma}\|_{-1/2,k} \approx \sup_{0 \neq v \in V} \frac{\langle \tilde{\sigma}, v \rangle_S}{\|v\|_{V,k}}. \]

**Proof.** The proof follows by now well-established arguments, see [4] or [17] for problems where the norms depend on parameters. Note that if we use \( \|A^{-1/2} \sigma\| \) instead of \( \|\sigma\|^2 \) in the definition of \( \|\cdot\|_{-1/2,k} \) then equality holds. \( \square \)

2.2. **Primal DPG formulation.** For a simpler notation, we use superindices to indicate time evaluations, i.e., for a time-dependent function \( v = v(t; \cdot) \) we use the notation \( v^n(\cdot) := v(t_n; \cdot) \). Moreover, we use similar notations for discrete quantities which however are not defined for all \( t \in [0, T] \). For instance, later on \( u^n_k \) will denote an approximation of \( u^n = u(t_n; \cdot) \) where \( u \) is the exact solution of the parabolic equation. We approximate \( \dot{u}^n \) by a backward difference, i.e., \( \dot{u}^n \approx (u^n - u^{n-1})/k \), this formally yields the elliptic PDE
\[ \frac{1}{k} u^n - \text{div} \, A \nabla u^n + \beta \cdot \nabla u^n + \gamma u^n = f^n + \frac{1}{k} u^{n-1}, \]
which admits a unique solution \( u^n \in H^1_0(\Omega) \). By testing with \( v \in V \) and introducing \( \tilde{\sigma}^n \) as
\[ \tilde{\sigma}^n|_{\partial K} := A \nabla u^n \cdot n_K|_{\partial K}, \quad \forall K \in \mathcal{T}, \]
and integrating by parts one obtains the primal DPG formulation
\[ \frac{1}{k} (u^n, v) + (A \nabla u^n, \nabla_{\mathcal{T}} v) + (\beta \cdot \nabla u^n, v) + (\gamma u^n, v) - \langle \tilde{\sigma}^n, v \rangle_S = (f^n, v) + \frac{1}{k} (u^{n-1}, v). \]

We introduce some bilinear forms and the right-hand side functional: For \( u = (u, \tilde{\sigma}) \in U, v \in V \) and \( w, g \in L^2(\Omega) \) set
\[ b(u, v) := (A \nabla u, \nabla_{\mathcal{T}} v) + (\beta \cdot \nabla u, v) + (\gamma u, v) - \langle \tilde{\sigma}, v \rangle_S, \]
\[ a(u, v) := \frac{1}{k} (u, v) + b(u, v), \]
\[ F(g, w; v) := (g, v) + \frac{1}{k} (w, v). \]
We recall that the inner product is given by

\[ \langle u, v \rangle = \langle f, u^n - 1; v \rangle, \quad \forall v \in V. \]

The (discrete) trial-to-test operator is given by

\[ a(u, v) = F(f^n, u^n - 1; v) \]

space. We consider the spaces

For the discretization of traces we use the space of facewise polynomials denoted by \( P^{p+1}(\mathcal{T}) \). Further, we note that \( a \) is not bounded independently of \( k \) using the above norms, but rather

\[ a(u, v) \lesssim \left( k^{-1/2} \| u \| + \| u \|_{U,k} \right) \| v \|_{V,k}, \]

where \( u = (u, \sigma) \).

2.2.1. Fully discrete scheme. By \( \mathcal{P}^p(K) \) we denote the space of polynomials of degree \( p \in \mathbb{N}_0 \), and

\[ \mathcal{P}^p(\mathcal{T}) := \{ v \in L^2(\Omega) : v|_K \in \mathcal{P}^p(K) \quad \forall K \in \mathcal{T} \} \]

For the discretization of traces we use the space of facewise polynomials denoted by \( \mathcal{P}^p(\mathcal{S}) \). In particular, we note that \( \mathcal{P}^p(\mathcal{S}) \) is the (normal-) trace space of the \( p \)-th order Raviart–Thomas space. We consider the spaces

\[ U_h := \mathcal{P}^{p+1}(\mathcal{T}) \cap H^1_0(\Omega) \times \mathcal{P}^p(\mathcal{S}), \]
\[ V_h := \mathcal{P}^{p+1}(\mathcal{T}). \]

The (discrete) trial-to-test operator is given by

\[ (\Theta_h u_h, v)_V = a(u_h, v) \quad \forall v \in V_h. \]

We recall that the inner product is given by

\[ (v, \delta v)_V = \frac{1}{k}(v, \delta v) + (A \nabla v, \nabla \delta v). \]

The fully discrete scheme then reads: Given \( u^n_h \in L^2(\Omega) \), solve

\[ a(u^n_h, \Theta_h w_h) = F(f^n, u^n_h - 1; \Theta_h w_h) \quad \forall w_h \in U_h, \quad n = 1, 2, \ldots. \]

The next lemma establishes coercivity of \( a(\cdot, \Theta_h \cdot) \) on \( U_h \).

Lemma 2. Let \( u_h = (u_h, \sigma_h) \in U_h \). Then,

\[ \frac{1}{k} \| u_h \|^2 + \| A^{1/2} \nabla u_h \|^2 \leq \| \Theta_h u_h \|^2_{V,k}. \]

If additionally, \(hk^{-1/2} \leq C_1 \) for some constant \( C_1 > 0 \), then,

\[ C_2 \| \sigma_h \|_{V,k} \leq C_3 \| \Theta_h(0, \sigma_h) \|_{V,k} \leq \| \Theta_h u_h \|_{V,k}. \]

Here, \( C_2, C_3 > 0 \) denote generic constants independent of \( h, k \) and \( u_h \).

Proof. For the first part we use that \( \mathcal{P}^{p+1}(\mathcal{T}) \cap H^1_0(\Omega) \subset V_h \) and \( (\beta \cdot \nabla w + \gamma w, w) \geq 0 \) for \( w \in H^1_0(\Omega) \): Let \( u_h = (u_h, \sigma_h) \in U_h \) be given. With the latter observation, the definition of the optimal test function and the fact that \( \langle \sigma_h, w \rangle_S = 0 \) for \( w \in H^1_0(\Omega) \) we get that

\[ \frac{1}{k} \| u_h \|^2 + \| A^{1/2} \nabla u_h \|^2 \leq \frac{1}{k} (u_h, u_h) + (A \nabla u_h, \nabla u_h) + (\beta \cdot \nabla u_h, u_h) + (\gamma u_h, u_h) \]
\[ = a(u_h, 0, u_h) = (\Theta_h u_h, u_h)_{V,k} \leq \| \Theta_h u_h \|_{V,k} \| u_h \|_{V,k} \]
\[ = \| \Theta_h u_h \|_{V,k} \left( \frac{1}{k} \| u_h \|^2 + \| A^{1/2} \nabla u_h \|^2 \right)^{1/2}. \]

This finishes the proof of the first estimate.
For the second one we use some results established in the literature on fully discrete DPG formulations (practical DPG), e.g. [22]. Let \( \Pi_h : V \to \tilde{V}_h \) be the Fortin operator defined in [22] Lemma 3.2 which in [22] is denoted by \( \Pi_{h}^{\text{grad}} \) with \( r = p + d \). By Lemma [1] and the Fortin property we have that

\[
\|\tilde{\sigma}_h\|_{1,2,k} \lesssim \sup_{0 \neq v \in H^1(T)} \frac{\langle \tilde{\sigma}_h, v \rangle_S}{\|v\|_{V,k}} = \sup_{0 \neq v \in H^1(T)} \frac{\langle \tilde{\sigma}_h, \Pi_h v \rangle_S}{\|\Pi_h v\|_{V,k}}.
\]

Moreover, from [22] proof of Lemma 3.2 we infer that

\[
k^{-1/2} \|\Pi_h v\| + \|\nabla \Pi_h v\| \lesssim k^{-1/2} \|v\| + (1 + k^{-1/2}) \|\nabla v\|.
\]

Thus, using the assumption \( hk^{-1/2} \lesssim 1 \) we see that

\[
\|\Pi_h v\|_{V,k} \lesssim \|v\|_{V,k} \quad \forall v \in V.
\]

Combining this with (4) we conclude that

\[
\|\tilde{\sigma}_h\|_{1,2,k} \lesssim \sup_{0 \neq v \in H^1(T)} \frac{\langle \tilde{\sigma}_h, \Pi_h v \rangle_S}{\|\Pi_h v\|_{V,k}} \lesssim \sup_{0 \neq v \in V} \frac{\langle \tilde{\sigma}_h, v \rangle_S}{\|v\|_{V,k}} = \|\Theta_h(0, \tilde{\sigma}_h)\|_{V,k}.
\]

Then, \( \|\Theta_h(0, \tilde{\sigma}_h)\|_{V,k} \leq \|\Theta_h(u_h, \tilde{\sigma}_h)\|_{V,k} + \|\Theta_h(u_h, 0)\|_{V,k} \). It remains to estimate \( \|\Theta_h(u_h, 0)\|_{V,k} \): We use the notation \( \Theta_h(u_h, 0) =: v_h \),

\[
\|\Theta_h(u_h, 0)\|_{V,k}^2 = a(u_h, 0, \Theta_h(u_h, 0)) = \frac{1}{k} \langle u_h, v_h \rangle + (A \nabla u_h, \nabla \tau v_h) + (\beta \cdot \nabla u_h + \gamma u_h, v_h)
\]

\[
\lesssim \left( \frac{1}{k} \|u_h\|^2 + \|A^{1/2} \nabla u_h\|^2 \right)^{1/2} \|v_h\|_{V,k} \leq \|\Theta_h u_h\|_{V,k} \|\Theta_h(u_h, 0)\|_{V,k}.
\]

Note that the involved constant only depends on the end time \( T \) and the coefficients, but is otherwise independent of \( k \) or \( h \).

It should be noted that the condition \( hk^{-1/2} \lesssim 1 \) is only needed to get an \( k \) independent estimate for the trace norm. If this condition is not satisfied then the constant depends on \( k \).

**Theorem 3.** Problem (3) is well posed. In particular, the solutions are stable in the sense that

\[
\|u_h^n\| \leq \left( \|u_h^n\|^2 + k \|A^{1/2} \nabla u_h^n\|^2 \right)^{1/2} \leq \sum_{j=1}^n k \|f^j\| + \|u_h^0\|.
\]

**Proof.** According to Lemma 2 Problem (3) admits unique solutions \( u_h^n \in U_h, n = 1, 2, \ldots \). Then,

\[
\|\Theta_h u_h^n\|_{V,k} = a(u_h^n, \Theta_h u_h^n) = F^n(u_h^n, u_h^{n-1}, \Theta h u_h^n) = (f^n + k^{-1} u_h^{n-1}, \Theta_h u_h^n)
\]

\[
\leq k^{1/2} \|f^n + k^{-1} u_h^{n-1}\| \|u_h^n\| \lesssim k^{1/2}\|f^n + k^{-1} u_h^{n-1}\| \|\Theta_h u_h^n\|_{V,k}.
\]

This together with the estimate from Lemma 2 shows that

\[
\left( \|u_h^n\|^2 + k \|A^{1/2} \nabla u_h^n\|^2 \right)^{1/2} \leq k^{1/2}\|\Theta_h u_h^n\|_{V,k} \leq k \|f^n\| + \|u_h^{n-1}\|.
\]

Iterating the arguments concludes the proof. \( \square \)
2.2.2. Remark: Heat equation — the trivial case. Let us consider the simplest model which is the heat equation where $A$ is the identity, $\beta = 0$, $\gamma = 0$.

It is straightforward to see that for $(u_h, 0) \in U_h$, $\Theta_h(u_h, 0) = u_h$ since

$$
(u_h, v_h)_{V,k} = \frac{1}{k}(u_h, v_h) + (\nabla u_h, \nabla_T v_h) = a((u_h, 0), v_h) = (\Theta_h(u_h, 0), v_h)_{V,k}.
$$

Using the test functions $\Theta_h(v_h, 0)$ in (3) and the fact that $\langle \hat{\sigma}_h, v_h \rangle_S = 0$ we see that

$$
a(u_h^n, \Theta_h(v_h, 0)) = \frac{1}{k}(u_h^n, v_h) + (\nabla u_h^n, \nabla v_h) = (f^n + k^{-1}u_h^{n-1}, v_h) = F(f^n, u_h^{n-1}; \Theta_h(v_h, 0))
$$

for all $v_h \in P^{p+1}(T) \cap H^1_0(\Omega)$. Let $u_h^n = (u_h^n, \hat{\sigma}_h^n) \in U_h$ denote the solution to (3), then the solution component $u_h$ satisfies

$$
\frac{1}{k}(u_h^n, v_h) + (\nabla u_h^n, \nabla v_h) = (f^n, v_h) + \frac{1}{k}(u_h^{n-1}, v_h) \quad \forall v_h \in P^{p+1}(T) \cap H^1_0(\Omega),
$$

which is the standard Galerkin FEM, see [28 Section 1]. Thus the primal DPG solution component $u_h$ is identical to the standard Galerkin FEM solution. In particular, optimal error estimates in $L^\infty(L^2)$ and $L^\infty(H^1_0)$, see [28], are valid for the primal DPG solution. In [26] optimal $L^\infty(H^1_0)$ have been observed in numerical experiments, which can be explained with the observation above.

Note that this remark is true only if we consider $\beta = 0$, $\gamma = 0$.

2.3. Elliptic projection-type operator. To obtain optimal error estimates we introduce an elliptic projection. The idea goes back to [30] to obtain optimal $L^2(\Omega)$ a priori error estimates and is extensively used for Galerkin methods, but has not been studied for least-squares methods until only very recently in [20]. For DPG methods an additional difficulty arises since test norms are mesh dependent. A main difference to Galerkin methods is that, although the elliptic part of the parabolic PDE might be symmetric, our elliptic projection operator always corresponds to a non-symmetric problem. The reason is that we need to use optimal test functions used in (3) in combination with the bilinear form $b(\cdot, \cdot)$, which models the elliptic part.

We define the elliptic projection operator $\mathcal{E}_h : U \to U_h$ by

$$
b(\mathcal{E}_h u, \Theta_h w_h) = b(u, \Theta_h w) \quad \forall w_h \in U_h.
$$

In the proofs below and in some results we will use the (semi-)norm

$$
\|u\|_V^2 := \|A^{1/2}\nabla u\|^2 + \|\Theta_h(0, \hat{\sigma})\|_{V,k}^2.
$$

Note that from the results of Lemma 2 we infer that $\|u_h\|_k \approx \|u_h\|_{U,k}$ for $u_h \in U_h$ and the equivalence constants are independent of $h, k$ if $hk^{-1/2} = O(1)$. The next lemma establishes boundedness and an inf-sup condition of $b(\cdot, \Theta_h \cdot)$. Note that this is not as trivial as it seems since $\Theta_h$ is calculated using the bilinear form $a(\cdot, \cdot)$ and the inner product in $V$. Recall that $a(\cdot, \cdot)$ is not bounded independently of $k$ and that $(\cdot, \cdot)_{V,k}$ includes terms weighted with negative powers of the time step size $k$.

Lemma 4. The bilinear form $b(\cdot, \Theta_h \cdot)$ is bounded,

$$
|b(u, \Theta_h w_h)| \lesssim \|u\|_{U,k}\|w_h\|_{U,k},
$$

for all $u \in U$, $w_h \in U_h$, and fulfills the inf-sup condition

$$
\|u_h\|_k \lesssim \sup_{0 \neq v_h \in U_h} \frac{b(u_h, \Theta_h v_h)}{\|v_h\|_k}
$$

for all $u_h \in U_h$. In particular, we conclude that Problem 3 admits a unique solution.
Proof. We show only the second boundedness estimate, as the first one follows from $\|\cdot\|_k \lesssim \|\cdot\|_{U,k}$. To that end let $v := \Theta_h w_h = \Theta_h(w_h, \tilde{\chi}_h)$. We use the splitting $v = w_h + \tilde{w}_h$. From the definition of the optimal test function we see that

$$
\frac{1}{k}(w_h, \delta v) + (A\nabla w_h, \nabla\tau \delta v) + \frac{1}{k} (\tilde{w}_h, \delta v) + (A\nabla\tau \tilde{w}_h, \nabla\tau \delta v) \\
= \frac{1}{k}(w_h, \delta v) + (A\nabla w_h, \nabla\tau \delta v) + (\beta \cdot \nabla w_h, \delta v) + (\gamma w_h, \delta v) - \langle \tilde{\chi}_h, \delta v \rangle_S.
$$

Therefore the component $\tilde{w}_h \in V_h$ satisfies

$$
\frac{1}{k}(\tilde{w}_h, \delta v) + (A\nabla\tau \tilde{w}_h, \nabla\tau \delta v) = (\beta \cdot \nabla w_h, \delta v) + (\gamma w_h, \delta v) - \langle \tilde{\chi}_h, \delta v \rangle_S \quad \forall \delta v \in V_h.
$$

We note that $a((0, \tilde{\chi}_h), \delta v) = -\langle \tilde{\chi}_h, \delta v \rangle_S$ and therefore the definition of the optimal test function yields

$$
|\langle \tilde{\chi}_h, \delta v \rangle_S| \leq \|\Theta_h(0, \tilde{\chi}_h)\|_{V,k} \|\delta v\|_{V,k}.
$$

Standard estimates then show that

$$
\|\tilde{w}_h\|_{V,k}^2 \leq k^{1/2}\|\beta \cdot \nabla w_h + \gamma w_h\|k^{-1/2}\|\tilde{w}_h\| + \|\Theta_h(0, \tilde{\chi}_h)\|_{V,k} \|\tilde{w}_h\|_{V,k}.
$$

Using that $k \leq T$ we infer that

(6)

$$
\|\tilde{w}_h\|_{V,k} \lesssim k^{1/2}\|\nabla w_h\| + \|\Theta_h(0, \tilde{\chi}_h)\|_{V,k} \lesssim \|w_h\|_{k}.
$$

Observe that $\langle \tilde{\sigma}_h, w_h \rangle_S = 0$ since $w_h \in H_0^1(\Omega)$. Finally,

$$
|b(u, v)| \leq |b(u, w_h)| + |b(u, \tilde{w}_h)| \\
\lesssim \|(A\nabla u, \nabla w_h) + (\beta \cdot \nabla u + \gamma u, w_h)\| + \|u\|_{U,k} \|\tilde{w}_h\|_{V,k} \lesssim \|u\|_{U,k} \|w_h\|_{k}.
$$

This shows boundedness. In order to show the inf-sup condition, we first establish the coercivity estimate

(7)

$$
\|A^{1/2} \nabla u_h\| \leq 2b(u_h, \Theta_h u_h).
$$

Recall that the optimal test function is characterised by

$$
\frac{1}{k}(\Theta_h u_h, \delta v) + (A\nabla \Theta_h u_h, \nabla\tau \delta v) = a(u_h, \delta v) = \frac{1}{k}(u_h, \delta v) + b(u_h, \delta v) \quad \forall \delta v \in V_h.
$$

Setting $\delta v = v = \Theta_h u_h$ the last identity together with Young’s inequality proves that

(8)

$$
b(u_h, v) = \frac{1}{k}\|v\|^2 + \|A^{1/2} \nabla v\|^2 - \frac{1}{k}(u_h, v) \geq \frac{1}{k}\|v\|^2 + \|A^{1/2} \nabla v\|^2 - \frac{1}{2k}\|u_h\|^2 - \frac{1}{2k}\|v\|^2.
$$

We are going to estimate $k^{-1/2}\|u_h\|$. From Lemma [2] we get that $k^{-1}\|u_h\|^2 + \|A^{1/2} \nabla u_h\| \leq k^{-1}\|v\|^2 + \|A^{1/2} \nabla v\|^2$, hence,

$$
\|A^{1/2} \nabla u_h\|^2 \leq \|A^{1/2} \nabla v\|^2 + \frac{1}{k}\|v\|^2 - \frac{1}{k}\|u_h\|^2.
$$

Combining this with (8) yields

$$
b(u_h, v) \geq \frac{1}{k}\|v\|^2 + \|A^{1/2} \nabla v\|^2 - \frac{1}{2k}\|u_h\|^2 - \frac{1}{2k}\|v\|^2 \\
= \left(\frac{1}{2}\|A^{1/2} \nabla v\|^2 + \frac{1}{2k}\|v\|^2 - \frac{1}{2k}\|u_h\|^2\right) + \frac{1}{2}\|A^{1/2} \nabla v\|^2 \\
\geq \frac{1}{2}\|A^{1/2} \nabla v\|^2 + \frac{1}{2}\|A^{1/2} \nabla u_h\|^2 \geq \frac{1}{2}\|A^{1/2} \nabla u_h\|^2.
$$
Now we are in position to establish the inf-sup condition
\[ \left\| u_h \right\|_k \lesssim \sup_{0 \neq v_h \in U_h} \frac{b(u_h, \Theta_h v_h)}{\left\| v_h \right\|_k} : \]

With the boundedness estimates for \( b(\cdot, \cdot) \) we have that
\[
\left\| \Theta_h(0, \tilde{\sigma}_h) \right\|_{V,K}^2 = a((0, \tilde{\sigma}_h), \Theta_h(0, \tilde{\sigma}_h)) = b((0, \tilde{\sigma}_h), \Theta_h(0, \tilde{\sigma}_h)) \\
= b(u_h, \Theta_h(0, \tilde{\sigma}_h) - b(u_h, 0, \Theta_h(0, \tilde{\sigma}_h)) \\
\lesssim \sup_{0 \neq v_h \in U_h} \frac{b(u_h, \Theta_h v_h)}{\left\| v_h \right\|_k} \left\| \Theta_h(0, \tilde{\sigma}_h) \right\|_{V,K} + \left\| A^{1/2} \nabla u_h \right\| \left\| \Theta_h(0, \tilde{\sigma}_h) \right\|_{V,K}.
\]

Applying the coercivity estimate (7) for the second term yields
\[
\left\| A^{1/2} \nabla u_h \right\|^2 \lesssim b(u_h, \Theta_h u_h) = \frac{b(u_h, \Theta_h u_h)}{\left\| u_h \right\|_k} \left\| u_h \right\|_k.
\]

Combining the latter two estimates shows that
\[
\left\| u_h \right\|_k^2 \lesssim \left( \sup_{0 \neq v_h \in U_h} \frac{b(u_h, \Theta_h v_h)}{\left\| v_h \right\|_k} \right)^2 + \sup_{0 \neq v_h \in U_h} \frac{b(u_h, \Theta_h v_h)}{\left\| v_h \right\|_k} \left\| u_h \right\|_k.
\]

Young’s inequality finishes the proof of the inf–sup condition. \( \square \)

**Lemma 5.** Let \( E_h u := (u_h, \tilde{\sigma}_h) \in U_h \) be the solution of Problem (5). Then it holds
\[
\left\| \nabla(u - u_h) \right\| \lesssim \left\| u - u_h \right\|_k \lesssim \inf_{v_h \in U_h} \left\| u - v_h \right\|_{U,K}.
\]

If additionally \( hk^{-1/2} = O(1) \) (see Lemma 2) then
\[
\left\| \tilde{\sigma} - \tilde{\sigma}_h \right\|_{-1/2,K} \lesssim \inf_{v_h \in U_h} \left\| u - v_h \right\|_{U,K}.
\]

**Proof.** The best approximation properties follow from standard arguments. We give the details only for sake of completeness: Let \( w_h \in U_h \) be arbitrary, then
\[
\left\| u - u_h \right\|_k \lesssim \left\| u - w_h \right\|_k + \left\| u_h - w_h \right\|_k,
\]
and using (5) and the preceding lemma,
\[
\left\| u_h - w_h \right\|_k \lesssim \sup_{0 \neq v_h \in U_h} \frac{b(u_h - w_h, \Theta_h v_h)}{\left\| v_h \right\|_k} \lesssim \sup_{0 \neq v_h \in U_h} \frac{b(u - w_h, \Theta_h v_h)}{\left\| v_h \right\|_k} \lesssim \left\| u - w_h \right\|_{U,K}.
\]

Since \( \left\| u \right\|_k \gtrsim \left\| \nabla u \right\| \) this shows the first of the claimed best-approximation estimates. The second follows by noting that \( \left\| \tilde{\sigma}_h \right\|_{-1/2,K} \lesssim \left\| \Theta_h(0, \tilde{\sigma}_h) \right\|_{V,K} \) if \( hk^{-1/2} = O(1) \) (see Lemma 2) and considering
\[
\left\| \tilde{\sigma} - \tilde{\sigma}_h \right\|_{-1/2,K} \lesssim \left\| u - u_h \right\|_{U,K} \lesssim \left\| u - w_h \right\|_{U,K} + \left\| u_h - w_h \right\|_{U,K} \lesssim \left\| u - w_h \right\|_{U,K} + \left\| u_h - w_h \right\|_{U,K}.
\]

The last term is handled as before which concludes the proof. \( \square \)

Combining the latter quasi-best approximation result with standard approximation properties yields

**Corollary 6.** Suppose that the components of \( u = (u, \tilde{\sigma}) \) are sufficiently smooth, then
\[
\left\| u - E_h u \right\|_k = O(h^{p+1}).
\]

If additionally \( hk^{-1/2} = O(1) \) then
\[
\left\| u - E_h u \right\|_{U,K} = O(h^{p+1}).
\]
3. Optimal error estimate in energy norm

This section is devoted to prove optimal error estimates in the $H^1(\Omega)$ norm of the $n$-th solution of the backward Euler method (3). We make the same assumptions on the regularity of solutions as in [28].

**Theorem 7.** Let $u^n_h = (u^n_h, \tilde{\sigma}^n_h) \in U_h$ denote the solution of (3). Suppose that the components of $u(t; \cdot)$ are sufficiently regular. Under these assumptions there exists $k_0$ (independent of $h$) such that for $k < k_0$ the solution satisfies

$$\|\nabla u^n - \nabla u^n_h\| = \mathcal{O}(h^{p+1} + k) + \mathcal{O}(\|\nabla (u^0 - u^0_h)\|).$$

If additionally, $hk^{-1/2} = \mathcal{O}(1)$ then

$$\|\tilde{\sigma}^n - \tilde{\sigma}^n_h\|_{-1/2,k} = \mathcal{O}(h^{p+1} + k) + \mathcal{O}(\|\nabla (u^0 - u^0_h)\|).$$

**Proof.** With the elliptic projection operator $E_h$ (see (3)) we consider the splitting

$$u^n - u^n_h = (u^n - E_h u^n) + (E_h u^n - u^n_h).$$

We may also use the norm $\| \cdot \|_k$ defined in the proof of Lemma 5

**Step 1.** By Corollary 6 we have

$$\|u^n - E_h u^n\|_k = \mathcal{O}(h^{p+1}).$$

**Step 2.** We derive error equations: First, write $\tilde{E}_h u^n = (\tilde{E}^1_h u^n, \tilde{E}^2_h u^n)$. Then, by (3)

$$a(\tilde{E}_h u^n, v_{opt}) = \frac{1}{k} (\tilde{E}^1_h u^n, v_{opt}) + b(\tilde{E}_h u^n, v_{opt}) = \frac{1}{k} (\tilde{E}^1_h u^n, v_{opt}) + b(u^n, v_{opt})$$

$$= (k^{-1} \tilde{E}^1_h u^n + f^n - \dot{u}^n, v_{opt}) \quad \forall v_{opt} \in \Theta_h(U_h).$$

Second, by (3)

$$a(u^n_h, v_{opt}) = (f^n + k^{-1} u^n_h - 1, v_{opt}) \quad \forall v_{opt} \in \Theta_h(U_h).$$

Third, combining both identities and writing $w^n = (w^n, \tilde{\chi}^n) := \tilde{E}_h u^n - u^n_h$ yields

$$\frac{1}{k} (w^n, v_{opt}) + b(w^n, v_{opt}) = \frac{1}{k} (\tilde{e}^1_h + w^n - 1, v_{opt}) \quad \forall v_{opt} \in \Theta_h(U_h),$$

where

$$\tilde{e}^1_h := \tilde{E}^1_h (u^n - u^n_h) - k \dot{u}^n.$$

Putting the term with $w^n - 1$ on the left-hand side yields

$$\frac{1}{k} (w^n - w^n - 1, v_{opt}) + b(w^n, v_{opt}) = \frac{1}{k} (\tilde{e}^1_h, v_{opt}) \quad \forall v_{opt} \in \Theta_h(U_h).$$

**Step 3.** We use the test function $v_{opt} = \Theta_h(v_h, 0) = v_h + \bar{v}_h$ with $v_h = w^n - w^n - 1$. Reordering the terms in the error equations (11),

$$\frac{1}{k} \|w^n - w^n - 1\|^2 + \frac{1}{2} \|A^{1/2} \nabla w^n\|^2 - \frac{1}{2} \|A^{1/2} \nabla w^n - 1\|^2 + \frac{1}{2} \|A^{1/2} \nabla (w^n - w^n - 1)\|^2$$

$$= \frac{1}{k} (\tilde{e}^1_h, v_h + \bar{v}_h) - \frac{1}{k} (w^n - w^n - 1, \bar{v}_h)$$

$$- b(w^n, \bar{v}_h) - (\beta \cdot \nabla w^n, v_h) - (\gamma w^n, v_h)$$

$$= r_1 + r_2 + r_3 + r_4 + r_5.$$
Step 4. We estimate the contributions $r_j$ of the right-hand side of the error equation (12). Recall that $v_h = w^n - w^{n-1}$. Throughout we use Young’s inequality with a parameter $\delta > 0$ and the estimate from (10), i.e., $k^{-1/2}\|\tilde{v}_h\| \leq \|\tilde{v}_h\|_{V,k} \lesssim k^{1/2}\|\nabla (w^n - w^{n-1})\|$. First,
\[
|r_1| = \frac{1}{k} (e^n_h, v_h + \tilde{v}_h) \leq k^{-1/2}\|e^n_h\|k^{-1/2}\|v_h\| + k^{-1/2}\|e^n_h\|k^{-1/2}\|\tilde{v}_h\|
\lesssim \delta^{-1}k^{-1}\|e^n_h\|^2 + \delta k^{-1}\|w^n - w^{n-1}\|^2 + \delta k\|A^{1/2}\nabla (w^n - w^{n-1})\|^2.
\]
Second,
\[
|r_2| \lesssim k^{-1/2}\|w^n - w^{n-1}\|k^{-1/2}\|\tilde{v}_h\| \leq \delta k^{-1}\|w^n - w^{n-1}\|^2 + \delta^{-1}k\|A^{1/2}\nabla (w^n - w^{n-1})\|^2.
\]
Third, recall the definition of the (discrete) optimal test function, i.e.,
\[
|\langle \widehat{\chi}^n, v_h \rangle| = |a((0, \widehat{\chi}^n), \tilde{v}_h)| = |(\Theta_h(0, \widehat{\chi}^n), \tilde{v}_h)_{V,k}|
\]
which together with the Cauchy–Schwarz inequality shows that
\[
|r_3| = |b(w^n, \tilde{v}_h)| \lesssim (\|\nabla w^n\| + \|\Theta_h(0, \widehat{\chi}^n)\|_{V,k})k^{1/2}\|\tilde{v}_h\|
\leq \|\nabla (w^n - w^{n-1})\|k^{1/2}\|\tilde{v}_h\| + k^{1/2}\|w^{n-1}\|\|\tilde{v}_h\| + k^{1/2}\|\Theta_h(0, \widehat{\chi}^n)\|_{V,k}\|\tilde{v}_h\|.
\]
We need to estimate $\|\Theta_h(0, \widehat{\chi}^n)\|_{V,k}$: Using the error equation (9)
\[
\frac{1}{k}(w^n, \Theta_h(0, \widehat{\chi}^n)) + b(w^n, \Theta_h(0, \widehat{\chi}^n)) = \frac{1}{k}(e^n_h + w^{n-1}, \Theta_h(0, \widehat{\chi}^n))
\]
and the definition of $\Theta_h$ we get that
\[
\frac{1}{k}(\Theta_h(0, \widehat{\chi}^n), w^n) + (A\nabla_{\tau} \Theta_h(0, \widehat{\chi}^n), \nabla_{\tau} w^n) + (\beta \cdot \nabla w^n + \gamma w^n, \Theta_h(0, \widehat{\chi}^n)) + \|\Theta_h(0, \widehat{\chi}^n)\|_{V,k}^2
\]
\[
= -\langle \widehat{\chi}^n, w^n \rangle_{S=0} = \frac{1}{k}(e^n_h + w^{n-1}, \Theta_h(0, \widehat{\chi}^n)).
\]
Using the definition of $\Theta_h$ another time shows that
\[
\frac{1}{k}(w^{n-1}, \Theta_h(0, \widehat{\chi}^n)) = -(A\nabla_{\tau} \Theta_h(0, \widehat{\chi}^n), \nabla w^{n-1}) - \langle \widehat{\chi}^n, w^{n-1} \rangle_{S=0} \lesssim \|\Theta_h(0, \widehat{\chi}^n)\|_{V,k}\|\nabla w^{n-1}\|.
\]
Combining the latter two estimates together with standard estimates gives
\[
\|\Theta_h(0, \widehat{\chi}^n)\|_{V,k} \lesssim k^{-1/2}\|e^n_h\| + \|\nabla w^{n-1}\| + k^{1/2}\|\nabla w^n\|.
\]
Using the error equations again together with Lemma 2 it is easy to see that $k^{1/2}\|\nabla w^n\| \lesssim \|e^n_h\| + \|\nabla w^{n-1}\|$. Therefore,
\[
(13) \quad \|\Theta_h(0, \widehat{\chi}^n)\|_{V,k} \lesssim k^{-1/2}\|e^n_h\| + \|\nabla w^{n-1}\|.
\]
For the term $r_3$ we thus get the estimate
\[
|r_3| \lesssim (k^{1/2} + \delta)\|A^{1/2}\nabla (w^n - w^{n-1})\|^2 + k\delta^{-1}\|\nabla w^{n-1}\|^2 + \delta^{-1}\|e^n_h\|^2.
\]
Fourth,
\[
|r_4| \lesssim k^{1/2}\|\nabla w^n\|k^{-1/2}\|v_h\| \lesssim k\delta^{-1}\|A^{1/2}\nabla (w^n - w^{n-1})\| + k\delta^{-1}\|A^{1/2}\nabla w^{n-1}\|^2 + \delta k^{-1}\|w^n - w^{n-1}\|^2.
\]
Fifth, using Poincaré’s inequality,
\[
|r_5| \lesssim \|w^n\|\|v_h\| \lesssim k\delta^{-1}\|\nabla w^n\|^2 + \delta k^{-1}\|v_h\|^2
\lesssim k\delta^{-1}\|A^{1/2}\nabla (w^n - w^{n-1})\|^2 + k\delta^{-1}\|A^{1/2}\nabla w^{n-1}\|^2 + \delta k^{-1}\|w^n - w^{n-1}\|^2.
\]
Step 5. Using the bounds for the \( r_j \) in the error equation (12), choosing first \( \delta \) sufficiently small independently of \( k \), and then \( k \leq k_0 \) sufficiently small in dependence of \( \delta \), and taking into account the norm equivalence \( \| A^{1/2} \nabla (\cdot) \| \approx \| \nabla (\cdot) \| \), we obtain
\[
\| A^{1/2} \nabla w^n \|^2 - \| A^{1/2} \nabla w^{n-1} \|^2 \lesssim k^{-1} \| e_h^k \|^2 + k \| \nabla w^{n-1} \|^2.
\]
Iterating this estimate, applying the discrete Gronwall inequality [28, Lemma 10.5], \( \sum_{j=1}^n k \leq T \) and norm equivalence \( \| A^{1/2} \nabla (\cdot) \| \approx \| \nabla (\cdot) \| \) once again show that
\[
\| \nabla w^n \|^2 \lesssim \sum_{j=1}^n \frac{1}{k} \| e_h^j \|^2 + \| \nabla w^0 \|^2.
\]

Step 6. The basic ideas to estimate the sum \( \sum_{j=1}^n \frac{1}{k} \| e_h^j \|^2 \) are the same as given in [28, Section 1] with some modifications, i.e., the use of elliptic projection operators and its properties defined in the present work. For the sake of completeness we recall the main steps where we closely follow the presentation of our own work [20]: We write the error \( e_h^j \) defined in (10) as
\[
e_h^j := \varepsilon_h^{1,j} + \varepsilon_h^{2,j} := [\varepsilon_h^{1,j}(u^j - u^{j-1}) - (u^j - u^{j-1})] + [u^j - u^{j-1} - k\dot{u}^j].
\]
We write \( u^j - u^{j-1} = \int_{t_{j-1}}^{t_j} \dot{u}(s) \, ds \). Then, the first term is estimated with Corollary 6 and the Cauchy–Schwarz inequality in the time variable, yielding
\[
\| e_h^{1,j} \| \lesssim \int_{t_{j-1}}^{t_j} h^{p+1} C(\dot{\bar{u}}(s)) \, ds \leq k^{1/2} \left( \int_{t_{j-1}}^{t_j} h^{2(p+1)} C(\dot{\bar{u}}(s))^2 \, ds \right)^{1/2}.
\]
Here, \( C(\dot{\bar{u}}(s)) \) depends on higher-order Sobolev norms of \( \dot{\bar{u}}(s) \). Similar arguments together with a Taylor expansion show for the second contribution \( e_h^{2,j} \) that
\[
\| e_h^{2,j} \| = \| u^j - u^{j-1} - k \dot{u}^j \| \leq k \int_{t_{j-1}}^{t_j} \| \dot{\bar{u}}(s) \| \, ds \leq k^{3/2} \left( \int_{t_{j-1}}^{t_j} \| \dot{\bar{u}}(s) \|^2 \, ds \right)^{1/2}.
\]
Combining the latter two estimates we conclude that
\[
\frac{1}{k} \sum_{j=1}^n \| e_h^j \|^2 = O(h^{2(p+1)} + k^2).
\]

Step 7. The trace estimate follows from
\[
\| \Theta(0, \tilde{x}^n) \|_{V,k} \lesssim k^{-1/2} \| e_h^n \| + \| \nabla w^{n-1} \|,
\]

4. Optimal Error Estimate in Weaker Norms

Throughout this section we assume that the coefficients \( A, \beta, \gamma \) and the domain \( \Omega \) are such that for given data \( f, g \in L^2(\Omega) \) the unique solutions \( w, v \in H_0^1(\Omega) \) of the primal problem
\begin{align}
\tag{14a} &-\text{div} \, A \nabla w + \beta \cdot \nabla w + \gamma w = f \quad \text{in } \Omega, \\
\tag{14b} &w|_\Gamma = 0,
\end{align}

and the dual problem
\begin{align}
\tag{15a} &-\text{div} (A \nabla v + \beta v) + \gamma v = g \quad \text{in } \Omega, \\
\tag{15b} &v|_\Gamma = 0,
\end{align}

11
satisfy the regularity estimates

\[ \|w\|_{H^2(\Omega)} + \|A\nabla w\|_{H^1(T)} \lesssim \|f\|, \quad \|v\|_{H^2(\Omega)} \lesssim \|g\|. \]

4.1. Further analysis of elliptic projection. We need to take a closer look at the elliptic projection operator in order to show higher rates in the \( L^2(\Omega) \) norm. To that end we will consider duality arguments (Aubin–Nitsche trick), see [2] for the primal DPG method for the Poisson problem. Duality arguments for the DPG method with ultra-weak formulation have been explored in [14, 15]. Contrary to the mentioned works, here we need to develop duality arguments for a non-symmetric bilinear form that defines the elliptic projection. Additionally, the norms in this work depend on the time step which can be arbitrarily small, making the analysis more delicate.

Recall the definition of the elliptic projection from (5): Given \( u \in U \), define \( \mathcal{E}_h u := u_h \in U_h \) by

\[ b(u_h, \Theta_h w_h) = b(u, \Theta_h w_h) \quad \forall w_h \in U_h, \]

where the discrete trial-to-test operator \( \Theta_h : U_h \to V_h \) is defined through

\[ (\Theta_h w_h, \delta v)_{V,k} = a(w_h, \delta v) \quad \forall \delta v \in V_h. \]

For the analysis we will use an equivalent representation as a mixed system:

**Lemma 8.** Problem (17) is equivalent to the mixed system: Find \((v_h, u_h) \in V_h \times U_h\) such that

\[ (v_h, \delta v)_{V,k} + b(u_h, \delta v) = b(u, \delta v), \]

\[ a(\delta w, v_h) = 0 \]

for all \((\delta v, \delta w) \in V_h \times U_h\).

In particular,

\[ \|v_h\|_{V,k} \lesssim \|u - u_h\|_k. \]

**Proof.** Let \( u_h \in U_h \) be the solution of (17). Define \( v_h \in V_h \) by \((v_h, \delta v)_{V,k} = b(u - u_h, \delta v) \) for all \( \delta v \in V_h \). Then,

\[ a(\delta w, v_h) = (\Theta_h \delta w, v_h)_{V,k} = b(u - u_h, \Theta_h \delta w) = 0 \]

for all \( \delta w \in U_h \).

To see the other direction, suppose that \((v_h, u_h) \in V_h \times U_h\) solves (18). With the same consideration as above we get that

\[ 0 = a(\delta w, v_h) = (\Theta_h \delta w, v_h)_{V,k} = b(u - u_h, \Theta_h \delta w), \]

for all \( \delta w \in U_h \) which means that \( u_h \in \hat{U}_h \) satisfies (17).

The final estimate follows from (18a) and the definition of the norm \( \| \cdot \|_k \),

\[ \|v_h\|^2_{V,k} = b(u - u_h, v_h) = b((u - u_h, 0), v_h) - (\hat{\sigma} - \hat{\sigma}_h, v_h) = b((u - u_h, 0), v_h) + a((0, \hat{\sigma} - \hat{\sigma}_h), v_h) \]

\[ = b((u - u_h, 0), v_h) + (\Theta_h(0, \hat{\sigma} - \hat{\sigma}_h), v_h)_{V,k} \lesssim \|u - u_h\|\|v_h\|_{V,k}. \]

This concludes the proof. \( \square \)

**Theorem 9.** Suppose that problems (14), (15) satisfy the regularity estimates in (16). Then, for given \( u = (u, \hat{\sigma}) \in U \), the elliptic projection \( u_h := (u_h, \hat{\sigma}_h) := \mathcal{E}_h u \in \hat{U}_h \) converges at a higher rate in the \( L^2(\Omega) \) norm, i.e.,

\[ \|u - u_h\| \lesssim h\|u - u_h\|_k. \]
Proof. Step 1. Let \( v \in H^1_0(\Omega) \) denote the solution of the dual problem (15) with \( g = u - u_h \). Integration by parts and the definition of the bilinear form \( b(\cdot, \cdot) \) yield
\[
\|u - u_h\|^2 = \langle u - u_h, -\text{div}(A\nabla v + \beta v) + \gamma v \rangle \\
= \langle A\nabla (u - u_h), \nabla v \rangle + (\beta \cdot \nabla (u - u_h) + \gamma (u - u_h), v) = b(u - u_h, v).
\]

Step 2. We characterise \( v = \Theta w \) where \( \Theta: U \rightarrow V \) is the continuous trial-to-test operator defined through
\[
(\Theta w, \delta v)_V = a(w, \delta v) \quad \forall \delta v \in V.
\]

From the latter definition we obtain
\[
(v, \delta v)_V = \frac{1}{k}(v, \delta v) + (A\nabla v, \nabla \delta v) = a((v, 0), \delta v) - (\beta \cdot \nabla v + \gamma v, \delta v) \quad \forall \delta v \in V.
\]
It follows that \( w = (v, 0) + w^* \) where \( w^* = (\hat{\chi}^*, \hat{\chi}^*) \in U. \) The component \( w^* \in H^1_0(\Omega) \) solves
\[
-\text{div} A\nabla w^* + \beta \cdot \nabla w^* + \gamma w^* + \frac{1}{k} w^* = -\beta \cdot \nabla v - \gamma v
\]
and \( \hat{\chi}^* \) is given by \( \hat{\chi}^*|_{\partial K} = A\nabla w^* \cdot n_K|_{\partial K} \) for all \( K \in \mathcal{T} \).

Step 3. Let \( v_h \in V_h \) be such that \( (v_h, u_h) \in V_h \times U_h \) is the solution of the equivalent mixed system (18). Then, we see with Step 1 and using (18) that
\[
|b(u - u_h, v) - b(u - u_h, v) - (v, v_h)h_V| + a(w, v_h)
\]
(21)
\[
|b(u - u_h, v) - b(u - u_h, v) - (v, v_h)h_V| + a(w - \delta w, v_h)
\]

for all \( (\delta v, \delta w) \in V_h \times U_h \).

We choose \( \delta v \in \mathcal{P}^{p+1}(\Omega) \cap H^1_0(\Omega) \) to be the best approximation of \( v \) in the \( H^1_0(\Omega) \) norm and \( \delta w = (\delta v, 0) + \delta w^* \) where \( \delta w^* \in U_h \) will be chosen below. The first term is estimated (note that \( v - \delta v \in H^1_0(\Omega) \)) using the regularity of \( v \) by
\[
|b(u - u_h, v - \delta v)| \lesssim \|\nabla(u - u_h)\| \|\nabla(v - \delta v)\| \lesssim h\|v\|_{H^2(\Omega)} \|\nabla(u - u_h)\| \lesssim h\|u - u_h\|_k.
\]

For the remaining terms, note that \( w - \delta w = (v - \delta v, 0) + (w^* - \delta w^*) \). Then, using the estimate \( \|v_h\|_{V,k} \lesssim \|u - u_h\|_k \) from Lemma 8 and boundedness (2) of \( a \),
\[
|b(u - u_h, v - \delta v) - a(w - \delta w, v_h)|
\]
\[
\lesssim h^{1/2} \|\nabla(v - \delta v)\| \|v_h\|_V + h^{-1/2} \|w^* - \delta w^*\|_{U,k} \|u_h\|_V
\]
\[
\lesssim h\|u - u_h\|_k + h^{-1/2} \|w^* - \delta w^*\|_{U,k} \|u - u_h\|_k.
\]
To finish the proof it remains to show the estimate \( (k^{-1/2} \|w^* - \delta w^*\| + \|w^* - \delta w^*\|_{U,k}) \lesssim h\|u - u_h\|_k \).

Basically, this follows from the weak form of the PDE defined in Step 2: Let \( \delta w^* = (\delta \chi^*, \delta \chi_K) \in U_h \) with \( \delta w^* \) being the best approximation of \( w^* \) with respect to \( \|\nabla(\cdot)\| \) and \( \delta \chi^*|_{\partial K} = \delta \chi^* \cdot n_K|_{\partial K} \) for all \( K \in \mathcal{T} \). Here, \( \delta \chi^* \) denotes the Raviart–Thomas projection of \( \chi^* = A\nabla w^* \) into the Raviart–Thomas space of order \( p \).

The weak form of PDE (19) shows that
\[
\frac{1}{k} \|w^*\|^2 + \|A^{1/2}w^*\|^2 \le \frac{1}{k} \langle w^*, w^* \rangle + \langle A\nabla w^*, \nabla w^* \rangle + (\beta \cdot \nabla v + \gamma v, w^*)
\]
\[
= (-\beta \cdot \nabla v - \gamma v, w^*) \lesssim \|\nabla v\| \|w^*\|
\]
which implies first \( k^{-1} \|w^*\| \lesssim \|\nabla v\| \). Bootstrapping this estimate then shows
\[
k^{-1} \|w^*\| + k^{-1/2} \|\nabla w^*\| \lesssim \|\nabla v\| \lesssim \|u - u_h\|_k.
\]
Moreover, PDE (19) together with our regularity assumptions (16), and using the latter estimate show that
\[ \| \mathbf{w}^* \|_{H^2(\Omega)} \lesssim \| - \beta \cdot \nabla v - \gamma v - k^{-1} \mathbf{w}^* \| \lesssim \| \nabla v \| \lesssim \| u - u_h \| . \]

Therefore, standard approximation results and the aforegoing stability analysis prove that
\[ k^{-1/2} \| \mathbf{w}^* - \delta \mathbf{w}^* \| + \| \nabla (\mathbf{w}^* - \delta \mathbf{w}^*) \| \lesssim h k^{-1/2} \| \nabla \mathbf{w}^* \| + h \| \mathbf{w}^* \|_{H^2(\Omega)} \lesssim h \| u - u_h \|. \]

By our regularity assumptions we have that \( \chi^* = A \nabla \mathbf{w}^* \in H^1(\mathbb{T})^d \cap H(\text{div}; \Omega) \) so that the trace terms can be estimated in a similar fashion as in [14, Theorem 5]: From Lemma 1 we know that
\[ \Pi \left(\frac{22}{22}\right) \]
Iterating the arguments yields
\[ \text{Then, integration by parts yields} \]
\[ \langle \mathbf{\hat{w}}^* - \delta \mathbf{\hat{w}}^*, \mathbf{v} \rangle_{\mathcal{S}} = \langle \mathbf{\hat{x}}^* - \delta \mathbf{\hat{x}}^*, \nabla \mathbf{\hat{v}} \rangle + \langle \text{div} \mathbf{\hat{x}}^*, (1 - \Pi^p) \mathbf{\hat{v}} \rangle \lesssim (h \| \mathbf{\hat{x}}^* \|_{H^1(\mathbb{T})} + h \| \text{div} \mathbf{\hat{x}}^* \| \| \nabla \mathbf{\hat{v}} \| \]

where \( \Pi^p \) is the \( L^2 \) projection on \( \mathcal{P}^p(\mathbb{T}) \). For details on the arguments used we refer to [14, Proof of Theorem 5]. We stress that the estimates are independent of \( k \). Using the bounds from Step 3, this shows that
\[ \min_{\delta \mathbf{\hat{x}} \in H^{-1/2}(\mathcal{S})} \| \mathbf{\hat{x}}^* - \delta \mathbf{\hat{x}} \|_{-1/2, k} \lesssim h \| u - u_h \| , \]
which concludes the proof. \( \square \)

4.2. Error analysis in the \( L^2(\Omega) \) norm.

**Theorem 10.** Let \( \mathbf{u}_h^n \in U_h \) denote the solution of (3). Suppose that the components of \( \mathbf{u}(t; \cdot) \) are sufficiently regular. Under the additional regularity assumptions (16) the solution satisfies
\[ \| \mathbf{u}_h^n - \mathbf{u}_h^n \| = \mathcal{O}(h^{p+2} + k) + \mathcal{O}(\| \mathbf{u}_h^0 - \mathbf{u}_h^0 \|). \]

**Proof.** With the elliptic projection operator \( \mathcal{E}_h \) (see (3)) we consider the splitting
\[ \mathbf{u}_h^n - \mathbf{w}_h^n = (\mathbf{u}_h^n - \mathbf{\hat{u}}_h^n) + (\mathcal{E}_h \mathbf{u}_h^n - \mathbf{w}_h^n). \]

By Theorem 9 and Corollary 6 we get that
\[ \| \mathbf{u}_h^n - \mathcal{E}_h \mathbf{u}_h^n \| = \mathcal{O}(h^{p+2}). \]

Writing \( \mathbf{w}_h^n = (\mathbf{w}_h^n, \mathbf{\hat{w}}_h^n) := \mathcal{E}_h \mathbf{u}_h^n - \mathbf{w}_h^n \) we recall the error equation from (11),
\[ \frac{1}{k}(\mathbf{w}_h^n, \mathbf{v}_{\text{opt}}) + b(\mathbf{w}_h^n, \mathbf{v}_{\text{opt}}) = \frac{1}{k}(\mathbf{e}_h^n + \mathbf{w}_h^{n-1}, \mathbf{v}_{\text{opt}}) \quad \forall \mathbf{v}_{\text{opt}} \in \Theta_h(U_h). \]

We test with \( \mathbf{v}_{\text{opt}} = \Theta_h \mathbf{w}_h^n \) and using Lemma 2 we infer that
\[ k^{-1} \| \mathbf{w}_h^n \|^2 \leq a(\mathbf{w}_h^n, \Theta_h \mathbf{w}_h^n) = \frac{1}{k}(\mathbf{e}_h^n + \mathbf{w}_h^{n-1}, \Theta_h \mathbf{w}_h^n) \leq k^{-1/2} \| \mathbf{e}_h^n + \mathbf{w}_h^{n-1} \| a(\mathbf{w}_h^n, \Theta_h \mathbf{w}_h^n)^{1/2}, \]
and further that
\[ \| \mathbf{w}_h^n \| \leq \| \mathbf{e}_h^n \| + \| \mathbf{w}_h^{n-1} \|. \]

Iterating the arguments yields
\[ \| \mathbf{w}_h^n \| \leq \sum_{j=1}^n \| \mathbf{e}_h^j \| + \| \mathbf{w}_h^0 \|. \]

The last term is estimated with the triangle inequality together with Theorem 9 and Corollary 6 to obtain that
\[ \| \mathbf{w}_h^0 \| \leq \| \mathbf{u}_h^0 - \mathbf{u}_h^0 \| + \| \mathbf{u}_h^0 - \mathcal{E}_h \mathbf{u}_h^0 \| = \| \mathbf{u}_h^0 - \mathbf{u}_h^0 \| + \mathcal{O}(h^{p+2}). \]
The estimate
\[ \sum_{j=1}^{n} \|e^j_h\| = O(h^{p+2} + k) \]
is shown by following the very same argumentation as in [28, Theorem 1.5]. For the sake of completeness we repeat the main steps which are also similar to the ones already presented in Step 7 of the proof of Theorem 7.

From Step 7 of the proof of Theorem 7 we recall the splitting
\[ e^j_h := e^j_h^1 + e^j_h^2 := [e^j_h^1(u^j - u^{j-1}) - (u^j - u^{j-1})] + [u^j - u^{j-1} - k\dot{u}^j]. \]
With the same arguments (without using a Cauchy–Schwarz inequality in the time variable) we get that
\[ \|e^j_h^2\| \leq k \int_{t_{j-1}}^{t_j} \|\dddot{u}(s)\| \, ds. \]
Similarly, for the first contribution we apply Corollary [6] and Theorem [9] to get that
\[ \|e^j_h^1\| \lesssim h^{p+2} \int_{t_{j-1}}^{t_j} C(\dddot{\tilde{u}}(s)) \, ds, \]
where \( C(\dddot{\tilde{u}}(s)) \) depends on higher-order Sobolev norms of \( \dddot{\tilde{u}}(s) \). Summing over \( j \) finishes the proof. \( \square \)

References

[1] R. Andreev. Stability of sparse space-time finite element discretizations of linear parabolic evolution equations. *IMA J. Numer. Anal.*, 33(1):242–260, 2013.
[2] T. Bouma, J. Gopalakrishnan, and A. Harb. Convergence rates of the DPG method with reduced test space degree. *Comput. Math. Appl.*, 68(11):1550–1561, 2014.
[3] D. Broersen and R. Stevenson. A robust Petrov-Galerkin discretisation of convection-diffusion equations. *Comput. Math. Appl.*, 68(11):1605–1618, 2014.
[4] C. Carstensen, L. Demkowicz, and J. Gopalakrishnan. Breaking spaces and forms for the DPG method and applications including Maxwell equations. *Comput. Math. Appl.*, 72(3):494–522, 2016.
[5] L. Demkowicz and J. Gopalakrishnan. A class of discontinuous Petrov-Galerkin methods. Part I: the transport equation. *Comput. Methods Appl. Mech. Engrg.*, 199(23-24):1558–1572, 2010.
[6] L. Demkowicz and J. Gopalakrishnan. Analysis of the DPG method for the Poisson problem. *SIAM J. Numer. Anal.*, 49(5):1788–1809, 2011.
[7] L. Demkowicz and J. Gopalakrishnan. A class of discontinuous Petrov-Galerkin methods. II. Optimal test functions. *Numer. Methods Partial Differential Equations*, 27(1):70–105, 2011.
[8] L. Demkowicz and J. Gopalakrishnan. A primal DPG method without a first-order reformulation. *Comput. Math. Appl.*, 66(6):1058–1064, 2013.
[9] L. Demkowicz, J. Gopalakrishnan, S. Nagaraj, and P. Sepúlveda. A spacetime DPG method for the Schrödinger equation. *SIAM J. Numer. Anal.*, 55(4):1740–1759, 2017.
[10] L. Demkowicz, J. Gopalakrishnan, and A. H. Niemi. A class of discontinuous Petrov-Galerkin methods. Part III: Adaptivity. *Appl. Numer. Math.*, 62(4):396–427, 2012.
[11] L. Demkowicz and N. Heuer. Robust DPG method for convection-dominated diffusion problems. *SIAM J. Numer. Anal.*, 51(5):2514–2537, 2013.
[12] L. Diening and J. Storn. A Space-Time DPG Method for the Heat Equation. arXiv:2012.13229, arXiv.org, 2020.
[13] J. Ernesti and C. Wieners. Space-time discontinuous Petrov-Galerkin methods for linear wave equations in heterogeneous media. *Comput. Methods Appl. Math.*, 19(3):465–481, 2019.
[14] T. Führer. Superconvergence in a DPG method for an ultra-weak formulation. *Comput. Math. Appl.*, 75(5):1705–1718, 2018.
[15] T. Führer. Superconvergent DPG methods for second-order elliptic problems. *Comput. Methods Appl. Math.*, 19(3):483–502, 2019.
[16] T. Führer, A. Haberl, and N. Heuer. Trace operators of the bi-Laplacian and applications. *IMA J. Numer. Anal.*, in print (published online: https://doi.org/10.1093/imanum/draa012), 2020.
[17] T. Führer and N. Heuer. A robust DPG method for large domains. arXiv.org. arXiv:2011.12098, 2020.

[18] T. Führer, N. Heuer, and A. H. Niemi. An ultraweak formulation of the Kirchhoff–Love plate bending model and DPG approximation. Math. Comp., 88(318):1587–1619, 2019.

[19] T. Führer, N. Heuer, and J. Sen Gupta. A time-stepping DPG scheme for the heat equation. Comput. Methods Appl. Math., 17(2):237–252, 2017.

[20] T. Führer and M. Karkulik. New a priori analysis of first-order system least-squares finite element methods for parabolic problems. Numer. Methods Partial Differential Equations, 35(5):1777–1800, 2019.

[21] T. Führer and M. Karkulik. Space-time least-squares finite elements for parabolic equations. accepted for publication in CAMWA, preprint: arXiv:1911.01942, 2021.

[22] J. Gopalakrishnan and W. Qiu. An analysis of the practical DPG method. Math. Comp., 83(286):537–552, 2014.

[23] N. Heuer and M. Karkulik. A robust DPG method for singularly perturbed reaction-diffusion problems. SIAM J. Numer. Anal., 55(3):1218–1242, 2017.

[24] J. Muñoz Matute, D. Pardo, and L. Demkowicz. A DPG-based time-marching scheme for linear hyperbolic problems. Comput. Methods Appl. Mech. Engrg., 373:113539, 18, 2021.

[25] J. Muñoz Matute, D. Pardo, and L. Demkowicz. Equivalence between the DPG method and the exponential integrators for linear parabolic problems. J. Comput. Phys., 429:110016, 2021.

[26] N. V. Roberts and S. Henneking. Time-stepping DPG formulations for the heat equation. Comput. Math. Appl., 2021. in press.

[27] R. Stevenson and J. Westerdiep. Stability of galerkin discretizations of a mixed space-time variational formulation of parabolic evolution equations. arXiv:1902.06279, arXiv.org, 2019.

[28] V. Thomée. Galerkin finite element methods for parabolic problems, volume 25 of Springer Series in Computational Mathematics. Springer-Verlag, Berlin, second edition, 2006.

[29] K. Voronin, C. S. Lee, M. Neumüller, P. Sepulveda, and P. S. Vassilevski. Space-time discretizations using constrained first-order system least squares (CFOSLS). J. Comput. Phys., 373:863–876, 2018.

[30] M. F. Wheeler. A priori $L_2$ error estimates for Galerkin approximations to parabolic partial differential equations. SIAM J. Numer. Anal., 10:723–759, 1973.

Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Santiago, Chile
Email address: tofuhrer@mat.uc.cl

Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Santiago, Chile
Email address: nheuer@mat.uc.cl

Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile
Email address: michael.karkulik@usm.cl