

# Error analysis for discretizations of parabolic problems using continuous finite elements in time and mixed finite elements in space

Markus Bause<sup>1</sup>  · Florin A. Radu<sup>2</sup> ·  
Uwe Köcher<sup>1</sup> 

Received: 16 April 2015 / Revised: 10 March 2017 / Published online: 20 June 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** Variational time discretization schemes are getting of increasing importance for the accurate numerical approximation of transient phenomena. The applicability and value of mixed finite element methods in space for simulating transport processes have been demonstrated in a wide class of works. We consider a family of continuous Galerkin–Petrov time discretization schemes that is combined with a mixed finite element approximation of the spatial variables. The existence and uniqueness of the semidiscrete approximation and of the fully discrete solution are established. For this, the Banach–Nečas–Babuška theorem is applied in a non-standard way. Error estimates with explicit rates of convergence are proved for the scalar and vector-valued variable. An optimal order estimate in space and time is proved by duality techniques for the scalar variable. The convergence rates are analyzed and illustrated by numerical experiments, also on stochastically perturbed meshes.

**Mathematics Subject Classification** 65M12 · 65M60 · 76S05

---

✉ Markus Bause  
bause@hsu-hh.de

Florin A. Radu  
Florin.Radu@uib.no

Uwe Köcher  
koecher@hsu-hh.de

<sup>1</sup> Faculty of Mechanical Engineering, Helmut Schmidt University, Holstenhofweg 85, 220433 Hamburg, Germany

<sup>2</sup> Department of Mathematics, University of Bergen, Allégaten 41, 50520 Bergen, Norway

## 1 Introduction

Numerical simulations of time dependent single and multiphase phase flow and multicomponent transport processes in complex and porous media with strong heterogeneities and anisotropies are desirable in several fields of natural sciences and civil engineering as well as in a large number of branches of technology; cf. e.g., [22,29]. Typically, the discretization in space involves a significant set of complexities and challenges. MFEM (cf. [17,21]) have proved their potential and capability to approximate solutions with high accuracy and physical consistency; cf. e.g., [13,19]. So far, the temporal approximation of flows and transport phenomena in porous media have received relatively little interest (cf. e.g., [5,18,27,42–44,49] and the references therein) and have been limited to traditional non-adaptive first and second order methods, even if strong chemical reactions with high temporal variations in profiles are present. Rigorous studies of higher order time discretizations are still missing. The low-order implicit time discretization is of particular concern with respect to numerical diffusion for smooth solutions of transport problems (cf. [45] for a study on numerical diffusion for different temporal and spatial discretizations of a transport equation).

The Galerkin method is a well-recognised approach to solve time dependent problems; cf. e.g., [6,48]. However, until now it has rarely been used in practice for discretizing the time variable in approximations of initial-boundary value problems. Since recently, variational time discretization schemes based on continuous or discontinuous finite element techniques have been developed to the point that they can be put into use (cf. [30,31]) and demonstrate their significant advantages. Higher order methods are naturally embedded in these schemes and the uniform variational approach simplifies stability and error analyses. Further, goal-oriented error control [9] based on the dual weighted residual approach relies on variational space-time formulations and the concepts of adaptive finite element techniques for changing the polynomial degree as well as the length of the time intervals become applicable. Variational time discretization schemes that are combined with continuous or discontinuous finite element methods for the spatial variables are studied for flow and parabolic problems in, for instance [1–4,10,15,30–32,38,47] and for wave problems in, for instance [7,36,37]. In these works algebraic formulations of the variational time discretizations are developed [4,30,31,36,37,47], preconditioning techniques for the arising block matrix systems are addressed [4,10,32,37] and, finally, computational studies are performed.

Numerical analyses of semidiscretizations in time by variational methods and of variational space-time approaches can be found in, for instance [20,34,35,46,48]. In [48] discontinuous variational approximations of the time variable are studied for abstract parabolic problems whereas in [46] their continuous counterparts are analyzed. In [20,47] discontinuous variational approximations in time and space are studied and error estimates are proved. In [47] time-dependent domains are considered in an arbitrary Lagrangian Eulerian (ALE) framework and the advection–diffusion equation is written in mixed form as a system of first order equations in space. In [25] a discontinuous Galerkin method in time combined with a stabilized finite element approach in space for first order partial differential equations is investigated for static and dynamically changing meshes. Error estimates in the  $L^\infty(L^2)$  and  $L^2(L^2)$

norm are derived. In [34, 35] continuous space-time approximations for nonlinear wave equations with mesh modifications and for the Schrödinger equation are considered. Existence and uniqueness of the discrete solutions are discussed and error estimates are proved for the schemes.

As far as the MFE approximation of parabolic problems is concerned, in [48] an error estimate for the semidiscretization in space is given. However, for the flux variable an error estimate is proved only for the  $L^2$  norm. No estimate is provided for the error in divergence of the flux, that is part of the natural norm of the underlying function space  $\mathbf{H}(\text{div}; \Omega)$ . In [23, 33] similar error estimates, also in negative norms, are presented. In particular, estimates similar to the error estimates for conventional finite element approximations are established. The singular behavior of the error estimates as  $t \rightarrow 0$  for initial data in  $L^2(\Omega)$  is further included.

In this work a continuous Galerkin–Petrov (cGP) method is used for the discretization in time, whereas the MFEM [17, 21] is applied for the spatial discretization. Appreciable advantages of the MFEM are its local mass conservation property and the inherent approximation of the flux field as part of the formulation itself. In simulating coupled flow and transport processes in porous media the flux approximation of the flow problem is usually of higher practical interest than the approximation of the scalar variable itself. To the best of our knowledge, rigorous error estimates for fully discrete variational space-time discretization schemes that are based on MFE approximations are still missing. In our numerical analysis we split the temporal discretization error from the spatial one by introducing an auxiliary problem based on the semidiscretization in time. We firstly estimate the temporal discretization error and secondly the error between the semidiscrete and the fully discrete solution. The order of convergence estimates are derived in the natural norms of the variational space-time approach. They are summarized in Theorem 4.6. For the scalar variable of the MFE approach one of the given error estimates, measured in the norm of  $L^2(0, T; L^2(\Omega))$ , is optimal in space and time if a certain regularity assumption is supposed to be satisfied. For constant scalar-valued diffusion coefficients an error estimate for the flux variable in the norm of  $L^2(0, T; L^2(\Omega))$  is further provided. It is optimal in space and suboptimal in time. In the Gaussian quadrature points of the temporal discretization optimal order error estimates for the flux variable in  $L^2(\Omega)$  are even obtained for heterogeneous diffusion matrices. The existence and uniqueness of the semidiscrete and fully discrete solution is further established. Even though a prototype model problem is studied here only, we believe that the techniques for analyzing mixed variational space-time approximation schemes can be applied similarly to more complex flow and transport problems in porous media.

This work is organized as follows. In Sect. 2 our fully discrete variational space-time method is developed. In Sect. 3 we address the semidiscrete problem by proving existence and uniqueness of its solution and error estimates for the semidiscretization in time. In Sect. 4 we study the fully discrete problem and show the existence and uniqueness of its solution. The error between the semidiscrete and fully discrete problem is estimated. In Theorem 4.6 an error estimate for the simultaneous space-time discretization is provided by combining the before-given estimates of the temporal and spatial discretization. In Sect. 5 we illustrate and validate our derived error estimates by numerical experiments. We end our work with some conclusions in Sect. 6.

## 2 The fully discrete variational scheme

### 2.1 Notation and preliminaries

Throughout this paper, standard notations are used. A summary of the notations used in this work is presented in Appendix B. Let  $\Omega \subset \mathbb{R}^d$ , with  $d = 2$  or  $d = 3$ , be a polygonal or polyhedral bounded domain. We denote by  $H^p(\Omega)$  the Sobolev space of  $L^2$  functions with derivatives up to order  $p$  in  $L^2(\Omega)$  and by  $\langle \cdot, \cdot \rangle$  the inner product in  $L^2(\Omega)$ . Sobolev spaces of vector-valued functions are written in bold letters. Further, let  $H_0^1(\Omega) = \{u \in H^1(\Omega) \mid u = 0 \text{ on } \partial\Omega\}$  and  $H^{-1}(\Omega)$  denote its dual space. For the norms of the Sobolev spaces the notation is

$$\| \cdot \| := \| \cdot \|_{L^2(\Omega)}, \quad \| \cdot \|_p := \| \cdot \|_{H^p(\Omega)}, \quad \text{for } p \in \mathbb{N}, p \geq 1.$$

For the mixed problem formulation we use the abbreviations

$$V = \mathbf{H}(\text{div}; \Omega) = \left\{ \mathbf{q} \in L^2(\Omega) \mid \nabla \cdot \mathbf{q} \in L^2(\Omega) \right\}, \quad W = L^2(\Omega),$$

and

$$\| \mathbf{v} \|_V := \left( \| \mathbf{v} \|^2 + \| \nabla \cdot \mathbf{v} \|^2 \right)^{1/2}.$$

Let  $X_0 \subseteq X \subseteq X_1$  be three reflexive Banach spaces with continuous embeddings. Then we consider the following set of Banach space valued function spaces,

$$\begin{aligned} C(\bar{I}; X) &= \{w: [0, T] \rightarrow X \mid w \text{ is continuous}\}, \\ L^2(I; X) &= \left\{ w: (0, T) \rightarrow X \mid \int_I \|w(t)\|_X^2 dt < \infty \right\}, \\ H^1(I; X_0, X_1) &= \left\{ w \in L^2(I; X_0) \mid \partial_t w \in L^2(I; X_1) \right\}, \end{aligned}$$

that are equipped with their natural norms (cf. [24]) and where the time derivative  $\partial_t$  is understood in the sense of distributions on  $(0, T)$ . In particular, every function in  $H^1(I; X_0, X_1)$  is continuous on  $[0, T]$  with values in  $X$ ; cf. [24]. For  $X_0 = X = X_1$  we simply write  $H^1(I; X)$ . Moreover, we put  $H_0^1(I; X) = \{u \in H^1(I; X) \mid u(0) = 0\}$ .

For  $u \in H_0^1(\Omega)$  let  $A: H_0^1(\Omega) \mapsto H^{-1}(\Omega)$  be defined uniquely by

$$\langle Au, v \rangle = a(u, v) \quad \text{for all } v \in H_0^1(\Omega) \tag{2.1}$$

with

$$a(u, v) := \langle \mathbf{D}\nabla u, \nabla v \rangle,$$

where the matrix  $\mathbf{D} = \mathbf{D}(\mathbf{x}) = (d_{ij}(\mathbf{x}))_{i,j=1}^d$  satisfies  $d_{ij} \in L^\infty(\Omega)$  and is elliptic with

$$D_M |\xi|^2 \geq \xi^\top \mathbf{D}(\mathbf{x}) \xi \geq D_m |\xi|^2, \quad \theta_M |\xi|^2 \geq \xi^\top \mathbf{D}(\mathbf{x})^{-1} \xi \geq \theta_m |\xi|^2, \quad (2.2)$$

for almost every  $\mathbf{x} \in \Omega$ , all  $\xi \in \mathbb{R}^d$  and some constants  $0 < D_m \leq D_M < \infty$ . In (2.2) we put  $\theta_m := D_M^{-1}$  and  $\theta_M := D_m^{-1}$ . Under the previous assumptions it holds that

$$a(v, v) \geq \alpha \|v\|_1^2 \quad \text{for all } v \in H_0^1(\Omega), \quad (2.3)$$

$$|a(u, v)| \leq \beta \|u\|_1 \|v\|_1 \quad \text{for all } u, v \in H_0^1(\Omega). \quad (2.4)$$

Thus,  $A: H_0^1(\Omega) \mapsto H^{-1}(\Omega)$  is a linear and continuous operator. For a subspace  $D(A) \subset H_0^1(\Omega)$  let  $A: D(A) \mapsto H^{-1}(\Omega)$  be a bijective linear continuous operator. For instance, if  $\Omega$  is a convex polygonal or polyhedral bounded domain and  $d_{ij} \in W^{1,\infty}(\Omega)$ , for  $i, j = 1, \dots, d$ , is satisfied, then the operator  $A$  is a bijective linear continuous operator from  $D(A) = H^2(\Omega) \cap H_0^1(\Omega)$  to  $L^2(\Omega)$ ; cf. [28].

Due to the properties (2.3), (2.4) of the bilinear form  $a(\cdot, \cdot)$  the lemma of Lax–Milgram ensures that the operator  $A: H_0^1(\Omega) \mapsto H^{-1}(\Omega)$  defined in (2.1) is invertible and satisfies in the corresponding operator norm the stability estimates

$$\|A\| \leq \beta \quad \text{and} \quad \|A^{-1}\| \leq \alpha.$$

Moreover, for all  $g \in H^{-1}(\Omega)$  it holds that

$$\langle g, A^{-1}g \rangle = \langle AA^{-1}g, A^{-1}g \rangle \geq \alpha \|A^{-1}g\|_1^2 \geq \frac{\alpha}{\beta^2} \|g\|_{H^{-1}(\Omega)}^2. \quad (2.5)$$

As usual, by  $c > 0$  we denote a generic constant throughout the paper.

### 2.2 Problem formulation

As a prototype model for more sophisticated multiphase flow and multicomponent reactive transport systems in porous media (cf. e.g. [22,29]) we study in this work

$$\partial_t u - \nabla \cdot (\mathbf{D} \nabla u) = f \quad \text{in } \Omega \times I, \quad (2.6)$$

$$u = 0 \quad \text{on } \partial\Omega \times I, \quad (2.7)$$

$$u(\cdot, 0) = u_0 \quad \text{in } \Omega, \quad (2.8)$$

equipped with homogeneous Dirichlet boundary conditions for simplicity only, where  $I = (0, T]$  with final time  $T > 0$  and the diffusion matrix  $\mathbf{D}$  satisfies the assumptions made in the previous subsection.

Let  $f \in L^2(I; W)$  and  $u_0 \in H_0^1(\Omega)$  be given. Then the existence of a unique weak solution

$$u \in L^2(I; H_0^1(\Omega)) \cap H^1(I; W) \cap C(\bar{I}; W) \quad (2.9)$$

to (2.6)–(2.8) is ensured; cf. [26, p. 382, Thm. 5]. We note that (2.9) already provides an improved regularity for the weak solution of (2.6)–(2.8); cf. [26, p. 378, Thm. 3].

In order to derive our family of discretization schemes, we first define the auxiliary flux variable  $\mathbf{q} := -\mathbf{D}\nabla u$  for the weak solution  $u$  of (2.6)–(2.8) that is given by (2.9). Since  $\partial_t u \in L^2(I; W)$  is satisfied by (2.9) and  $f \in L^2(I; W)$  holds by assumption, it directly follows that  $\mathbf{q} \in L^2(I; \mathbf{V})$ . The pair  $\{u, \mathbf{q}\} \in H^1(I; W) \cap C(\bar{I}; W) \times L^2(I; \mathbf{V})$  is then also the unique solution to the set of variational equations

$$\int_0^T \langle \partial_t u, w \rangle dt + \int_0^T \langle \nabla \cdot \mathbf{q}, w \rangle dt = \int_0^T \langle f, w \rangle dt, \tag{2.10}$$

$$\int_0^T \langle \mathbf{D}^{-1} \mathbf{q}, \mathbf{v} \rangle dt - \int_0^T \langle u, \nabla \cdot \mathbf{v} \rangle dt = 0 \tag{2.11}$$

for all  $w \in L^2(I; W)$  and  $\mathbf{v} \in L^2(I; \mathbf{V})$  and satisfies the initial condition  $u(0) = u_0$ . To find (2.11), integration by parts was used. The global problem formulation (2.10), (2.11) motivates our semidiscretization in time.

*Remark 2.1* • Below, in order apply Lagrange interpolation in time to the function  $f$ , we need the stronger assumption that  $f \in C([0, T]; W)$  is satisfied.

- Below, we introduce a semidiscrete approximation in time of the flux  $\mathbf{q}$  in a subspace of  $C([0, T]; \mathbf{V})$ . For this we need to assume that  $\mathbf{D}\nabla u_0 \in \mathbf{V}$  holds.
- Higher order regularity of weak solutions to (2.6)–(2.8), that is needed below for the proof of higher order convergence rates, can be obtained under further technical assumptions about the data, coefficients and the boundary of the domain  $\Omega$ . For the prototype model problem (2.6)–(2.8) such higher order regularity results are well-known; cf. [26, p. 386, Thm. 6]. For (elliptic) regularity results in domains with non-smooth boundaries we refer to, e.g., [28, 39]. Below, we tacitly assume that the required assumptions about the data and  $\partial\Omega$  are satisfied such that the existence of a sufficiently regular solution can be assumed. Without such an assumption the application of higher order methods is not meaningful.

### 2.3 Variational discretization in time by a continuous Galerkin method

For the discretization in time we decompose the time interval  $(0, T]$  into  $N$  subintervals  $I_n = (t_{n-1}, t_n]$ , where  $n \in \{1, \dots, N\}$  and  $0 = t_0 < t_1 < \dots < t_{n-1} < t_n = T$ . Further  $\tau$  denotes the discretization parameter in time and is defined as the maximum time step size  $\tau = \max_{1 \leq n \leq N} \tau_n$ , where  $\tau_n = t_n - t_{n-1}$ . We introduce the function spaces of piecewise polynomials of order  $r$  in time,

$$\begin{aligned} \mathcal{X}^r(X) &:= \left\{ u_\tau \in C(\bar{I}; X) \mid u_\tau|_{\bar{I}_n} \in \mathbb{P}_r(\bar{I}_n; X), \forall n \in \{1, \dots, N\} \right\}, \\ \mathcal{Y}^r(X) &:= \left\{ w_\tau \in L^2(I; X) \mid w_\tau|_{I_n} \in \mathbb{P}_r(I_n; X), \forall n \in \{1, \dots, N\} \right\}, \end{aligned}$$

where

$$\mathbb{P}_r(J; X) = \left\{ p: J \rightarrow X \mid p(t) = \sum_{j=0}^r \xi_n^j t^j, \xi_n^j \in X, j = 0, \dots, r \right\}$$

and  $\mathcal{X}^r(X) \subset H^1(0, T; W)$ . We let

$$\mathcal{X}_0^r(X) = \left\{ u_\tau \in \mathcal{X}^r(X) \mid u_\tau(0) = 0 \right\}.$$

Further, we put

$$\mathcal{W} = X_0^r(W) \times X^r(V) \quad \text{and} \quad \mathcal{V} = Y^{r-1}(W) \times Y^{r-1}(V).$$

We equip the function spaces  $\mathcal{W}$  and  $\mathcal{V}$  with their natural norms being defined by

$$\begin{aligned} \|\{u_\tau, \mathbf{q}_\tau\}\|_{\mathcal{W}}^2 &= \|u_\tau\|_{L^2(I; W)}^2 + \|\partial_t u_\tau\|_{L^2(I; W)}^2 + \|\mathbf{q}_\tau\|_{L^2(I; V)}^2, \\ \|\{w_\tau, \mathbf{v}_\tau\}\|_{\mathcal{V}}^2 &= \|w_\tau\|_{L^2(I; W)}^2 + \|\mathbf{v}_\tau\|_{L^2(I; V)}^2. \end{aligned} \tag{2.12}$$

With respect to these norms the space  $\mathcal{W}$  is a Banach space and the space  $\mathcal{V}$  is a reflexive Banach space. Further, we define the space-time bilinear form  $a_\tau \in \mathcal{L}(\mathcal{W} \times \mathcal{V}; \mathbb{R})$  by means of

$$\begin{aligned} a_\tau(\{u_\tau, \mathbf{q}_\tau\}, \{w_\tau, \mathbf{v}_\tau\}) &= \int_0^T \left( \langle \partial_t u_\tau, w_\tau \rangle + \langle \nabla \cdot \mathbf{q}_\tau, w_\tau \rangle \right) dt \\ &\quad + \int_0^T \langle \mathbf{D}^{-1} \mathbf{q}_\tau, \mathbf{v}_\tau \rangle dt - \int_0^T \langle u_\tau, \nabla \cdot \mathbf{v}_\tau \rangle dt \end{aligned}$$

for  $\{u_\tau, \mathbf{q}_\tau\} \in \mathcal{W}$  and  $\{w_\tau, \mathbf{v}_\tau\} \in \mathcal{V}$ . Obviously, the mapping  $a_\tau: \mathcal{W} \times \mathcal{V} \mapsto \mathbb{R}$  is linear and continuous, i.e.

$$|a_\tau(\{u_\tau, \mathbf{q}_\tau\}, \{w_\tau, \mathbf{v}_\tau\})| \leq c \|\{u_\tau^0, \mathbf{q}_\tau\}\|_{\mathcal{W}} \|\{w_\tau, \mathbf{v}_\tau\}\|_{\mathcal{V}} \tag{2.13}$$

with some constant  $c > 0$  independent of  $\tau$  and  $T$ .

For the family of continuous variational time discretization schemes the spaces  $\mathcal{X}^r(X)$  of continuous functions act as spaces for the solution whereas the spaces  $\mathcal{Y}^{r-1}(X)$  consisting of piecewise polynomials that are discontinuous at the end points of the time intervals are used as test spaces. Since the spaces of the trial and test functions differ here, a discretization of Galerkin–Petrov type is thus obtained.

A semidiscrete variational approximation of the mixed form of problem (2.6)–(2.8), referred as the exact form of cGP( $r$ ), is then defined by solving the variational equations (2.10), (2.11) in discrete subspaces: Find  $\{u_\tau, \mathbf{q}_\tau\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(V)$  such that

$$\int_0^T \langle \partial_t u_\tau, w_\tau \rangle dt + \int_0^T \langle \nabla \cdot \mathbf{q}_\tau, w_\tau \rangle dt = \int_0^T \langle f, w_\tau \rangle dt, \tag{2.14}$$

$$\int_0^T \langle \mathbf{D}^{-1} \mathbf{q}_\tau, \mathbf{v}_\tau \rangle dt - \int_0^T \langle u_\tau, \nabla \cdot \mathbf{v}_\tau \rangle dt = 0, \tag{2.15}$$

for all  $w_\tau \in \mathcal{Y}^{r-1}(W)$  and  $\mathbf{v}_\tau \in \mathcal{Y}^{r-1}(V)$  with the initial conditions that  $u_\tau(0) := u_0$  and  $\mathbf{q}_\tau(0) := -\mathbf{D}\nabla u_0$  (cf. Remark 2.1).

We refer to the solution of Eqs. (2.14), (2.15) as the continuous Galerkin–Petrov method with piecewise polynomials of order  $r$  and use the notation cGP( $r$ ). To ensure the existence and uniqueness of solutions to (2.14), (2.15), it is sufficient to use the test spaces  $\mathfrak{Y}^{r-1}(W)$  and  $\mathfrak{Y}^{r-1}(V)$  with piecewise polynomials of order  $r - 1$ , since the continuity constraint at the discrete time points  $t_n, n = 0, \dots, N - 1$ , that is implied by the definition of the solution spaces  $\mathcal{X}^r(W)$  and  $\mathcal{X}^r(V)$ , yields a further condition. By using discontinuous test basis functions  $w_\tau(t) = w\psi_{n,i}(t)$  and  $\mathbf{v}_\tau = \mathbf{v}\psi_{n,i}(t)$ , for  $i = 1, \dots, r$ , with arbitrary time independent functions  $w \in W$  and  $\mathbf{v} \in V$ , respectively, and piecewise polynomial functions  $\psi_{n,i}: I \mapsto \mathbb{R}$  that are of order  $r - 1$  on  $I_n$  and vanish on  $I \setminus \bar{I}_n$ , we can recast the variational equations (2.14), (2.15) as a time marching scheme: For  $n = 1, \dots, N$  find  $u_\tau|_{\bar{I}_n} \in \mathbb{P}_r(\bar{I}_n; W)$  and  $\mathbf{q}_\tau|_{\bar{I}_n} \in P_r(\bar{I}_n; V)$  such that

$$\int_{I_n} \langle \partial_t u_\tau, w \rangle \psi_{n,i}(t) dt + \int_{I_n} \langle \nabla \cdot \mathbf{q}_\tau, w \rangle \psi_{n,i}(t) dt = \int_{I_n} \langle f, w \rangle \psi_{n,i}(t) dt, \tag{2.16}$$

$$\int_{I_n} \langle \mathbf{D}^{-1} \mathbf{q}_\tau, \mathbf{v} \rangle \psi_{n,i}(t) dt - \int_{I_n} \langle u_\tau, \nabla \cdot \mathbf{v} \rangle \psi_{n,i}(t) dt = 0 \tag{2.17}$$

for all  $w \in W$  and  $\mathbf{v} \in V$  and  $i = 1, \dots, r$  with the continuity constraints  $u_\tau|_{I_n}(t_{n-1}) = u_\tau|_{I_{n-1}}(t_{n-1})$  and  $\mathbf{q}_\tau|_{I_n}(t_{n-1}) = \mathbf{q}_\tau|_{I_{n-1}}(t_{n-1})$  for  $n \geq 2$  and the initial conditions  $u_\tau|_{I_n}(t_{n-1}) := u_0, \mathbf{q}_\tau|_{I_n}(t_{n-1}) := -\mathbf{D}\nabla u_0$  for  $n = 1$ .

To determine  $u_\tau|_{\bar{I}_n}$  and  $\mathbf{q}_\tau|_{\bar{I}_n}$ , we represent them in terms of basis functions, with respect to the time variable, of the spaces  $\mathcal{X}^r(W)$  and  $\mathcal{X}^r(V)$  such that

$$u_\tau|_{\bar{I}_n}(t) = \sum_{j=0}^r U_n^j \varphi_{n,j}(t) \quad \text{and} \quad \mathbf{q}_\tau|_{\bar{I}_n}(t) = \sum_{j=0}^r \mathbf{Q}_n^j \varphi_{n,j}(t), \quad \text{for } t \in I_n, \tag{2.18}$$

with coefficient functions  $U_n^j \in W$  and  $\mathbf{Q}_n^j \in V$  for  $j = 0, \dots, r$  and polynomial basis functions  $\varphi_{n,j} \in \mathbb{P}_r(\bar{I}_n; \mathbb{R})$  that are Lagrange functions with respect to  $r + 1$  nodal points  $t_{n,j} \in I_n$  satisfying the conditions  $\varphi_{n,j}(t_{n,i}) = \delta_{i,j}$  for  $i, j = 0, \dots, r$ . For the treatment of the continuity constraint in time we put  $t_{n,0} = t_{n-1}$ . The other points  $t_{n,1}, \dots, t_{n,r}$  are chosen as the quadrature points of the  $r$ -point Gaussian quadrature formula on  $I_n$  which is exact if the function to be integrated is a polynomial of degree less or equal to  $2r - 1$ . The basis functions  $\varphi_{n,j} \in \mathbb{P}_r(\bar{I}_n; \mathbb{R})$  of (2.18), for  $j = 0, \dots, r$ , are defined, as usual in the finite element framework, via the affine reference transformation onto  $\hat{I} = [0, 1]$ . The test basis functions  $\psi_{n,i} \in P_{r-1}(\bar{I}_n; \mathbb{R})$  with  $\psi_{n,i}(t_{n,l}) = \delta_{i,l}$  for  $i, l = 1, \dots, r$  are defined similarly; cf. [15,37] for details. Now we transform all the time integrals in (2.16), (2.17) to the reference interval  $\hat{I}$ . By a subsequent application of the  $r$ -point Gaussian quadrature formula with weights  $\hat{\omega}_i$  and quadrature nodes  $\hat{t}_i$  on  $\hat{I}$  as well as the further notation

$$\hat{\alpha}_{ij} := \hat{\omega}_i \cdot \frac{d}{d\hat{t}} \hat{\varphi}_j(\hat{t}_i) \quad \text{and} \quad \hat{\beta}_{ij} := \hat{\omega}_i \cdot \delta_{i,j}$$



for  $i = 1, \dots, r, j = 0, \dots, r$  (cf. [15,36,46]), we obtain the following system of variational problems for the coefficient functions  $U_n^j \in W$  and  $Q_n^j \in V$  of the representation (2.18): For  $n = 1, \dots, N$  and  $j = 1, \dots, r$  find coefficient functions  $\{U_n^j, Q_n^j\} \in W \times V$  such that

$$\sum_{j=0}^r \hat{\alpha}_{ij} \langle U_n^j, w \rangle + \tau_n \hat{\beta}_{ii} \langle \nabla \cdot Q_n^i, w \rangle = \tau_n \hat{\beta}_{ii} \langle f(t_{n,i}), w \rangle, \tag{2.19}$$

$$\langle D^{-1} Q_n^i, v \rangle - \langle U_n^i, \nabla \cdot v \rangle = 0, \tag{2.20}$$

for  $i = 1, \dots, r$  and all  $\{w, v\} \in W \times V$ , and where due to continuity in time  $U_n^0 = u_{\tau|I_{n-1}}(t_{n-1}), Q_n^0 = q_{\tau|I_{n-1}}(t_{n-1})$  for  $n \geq 2$  and  $U_n^0 := u_0, Q_n^0 := -D\nabla u_0$  for  $n = 1$ .

*Remark 2.2* In the numerical scheme (2.19), (2.20), the flux coefficient functions  $Q_n^j$ , for  $j = 1, \dots, r$ , arise only in the  $r$  Gaussian quadrature points  $t_{n,1}, \dots, t_{n,r} \in (t_{n-1}, t_n)$  of the subinterval  $I_n$ . Nevertheless, the coefficient functions  $Q_n^0$  for  $n \geq 1$ , are needed for the unique determination of the semidiscrete flux function  $q_{\tau} \in \mathcal{X}^r(V)$  and an explicit evaluation of  $q_{\tau|I_n}$  by the representation (2.18) in other time points of  $I_n$  than in the Gaussian quadrature nodes. The fact that the coefficient functions  $Q_n^0$  do not arise in (2.19), (2.20) is due to the definition of the Lagrange basis functions  $\varphi_{n,j}$  in (2.18) and the fact that the time derivative of the flux variable  $q$  does not arise in the model equations.

For the derivation of (2.19), (2.20) from (2.16), (2.17) we tacitly replaced the integrand  $f$  on the right-hand side by its Lagrange interpolate  $\Pi_r f \in \mathbb{P}_r(I_n; L^2(\Omega))$  defined by

$$\Pi_r f(t)|_{I_n} = \sum_{j=0}^r f(t_{n,j})\varphi_{n,j}(t) \quad \text{for } t \in I_n. \tag{2.21}$$

We note that the constants  $\hat{\beta}_{ii}$  are satisfying the following property.

**Lemma 2.3** [Coefficient property (C)] *There exist constants  $\beta_m, \beta_M \in \mathbb{R}$  such that*

$$0 < \beta_m \leq \hat{\beta}_{ii} \leq \beta_M < \infty, \quad \text{for } i = 1, \dots, r, \tag{2.22}$$

*is satisfied. The constants do not depend on the time step size, but only on the number  $r$  of involved Gaussian quadrature points.*

*Proof* Indeed, the coefficients  $\beta_{ii} = \hat{w}_i$  are the Gauss–Legendre quadrature weights

$$\tilde{w}_i = \int_{-1}^1 \prod_{\substack{j=1 \\ j \neq i}}^r \left( \frac{x - x_j}{x_i - x_j} \right)^2 dx = \frac{1}{(1 - x_i^2)(P_r'(x_i))^2}, \quad i = 1, \dots, r, \tag{2.23}$$

scaled to the interval  $[0, 1]$ , i.e.  $\hat{\omega}_i = \tilde{\omega}_i/2$ . In (2.23),  $P_r$  denotes the Legendre polynomial of degree  $r$  and  $x_i$ , for  $i = 1, \dots, r$ , are its roots, cf. e.g., [41, p. 436]. Since the sum of the weights  $\tilde{\omega}_i$  equals to two and the weights are all strictly positive, we immediately conclude that an upper bound for  $\hat{\omega}_i$  is given by one. On the other hand, we know that  $|P'_r(x)| \leq r(r + 1)/2$  for any  $x \in [-1, 1]$ ; cf. [16, p. 73]. This gives us the lower bound  $\hat{\omega}_i \geq 2/(r(r + 1))^2$ .  $\square$

Below, we will also need the following auxiliary results.

**Lemma 2.4** *Let  $F(t, \mathbf{x}) = \sum_{i=0}^r F_n^i(\mathbf{x})\varphi_{n,i}(t)$ , for  $t \in I_n$ , with coefficient functions  $F_n^i \in W$  for  $i = 0, \dots, r$ . Then it holds that*

$$\sum_{i=1}^r \sum_{j=0}^r \hat{\alpha}_{ij} \langle F_n^j, F_n^i \rangle = \int_{t_{n-1}}^{t_n} \langle \partial_t F, F \rangle dt = \frac{1}{2} \|F(t_n)\|^2 - \frac{1}{2} \|F(t_{n-1})\|^2 \tag{2.24}$$

and

$$\|F\|_{L^2(I_n; W)}^2 \leq c \tau_n \sum_{j=0}^r \|F_n^j\|_W^2, \tag{2.25}$$

for some  $c > 0$  independent of  $\tau_n$ . An analogous results holds for coefficients  $F_n^i \in V$ .

*Proof* Using the properties of the basis functions  $\varphi_i$  and that the  $r$ -point Gaussian quadrature formula is exact for polynomials of maximum degree  $2r - 1$  there holds that

$$\begin{aligned} \int_{t_{n-1}}^{t_n} \langle \partial_t F, F \rangle dt &= \int_{t_{n-1}}^{t_n} \int_{\Omega} \sum_{j=0}^r \varphi'_{n,j}(t) F_n^j(\mathbf{x}) \sum_{i=0}^r \varphi_{n,i}(t) F_n^i(\mathbf{x}) d\mathbf{x} dt \\ &= \sum_{j=0}^r \sum_{i=0}^r \int_0^1 \frac{d}{d\hat{t}} \hat{\varphi}_j(\hat{t}) \cdot \hat{\varphi}_i(\hat{t}) d\hat{t} \langle F_n^i, F_n^j \rangle \\ &= \sum_{j=0}^r \sum_{i=1}^r \hat{\omega}_i \hat{\varphi}'_j(\hat{t}_i) \langle F_n^i, F_n^j \rangle = \sum_{i=1}^r \sum_{j=0}^r \hat{\alpha}_{ij} \langle F_n^j, F_n^i \rangle. \end{aligned}$$

The second of the equalities in (2.24) follows immediately from the first one. It remains to prove (2.25). It holds that

$$\|F\|_{L^2(I_n; W)}^2 \leq (r + 1) \sum_{i=0}^r \int_{t_{n-1}}^{t_n} \varphi_{n,i}^2(t) dt \|F_n^i\|^2 \leq c(r + 1) \sum_{i=0}^r \tau_n \|F_n^i\|^2,$$

with  $c$  independent of  $\tau_n$ . Here we used that  $\int_{t_{n-1}}^{t_n} \varphi_{n,i}^2(t) dt \leq c \tau_n$ ; cf. [35, p. 1790].  $\square$

### 2.4 Discretization in space by the mixed finite element method

Now, we present the fully discrete approximation scheme that is obtained by discretizing (2.19), (2.20) with respect to their spatial variables. For this we choose a pair of finite element spaces  $W_h \subset W$  and  $V_h \subset V$  satisfying the inf–sup stability condition; cf. [17,21]. Here, we denote by  $\mathcal{T}_h = \{K\}$  a finite element decomposition of mesh size  $h$  of the polyhedral domain  $\bar{\Omega}$  into closed subsets  $K$ , quadrilaterals in two space dimensions and hexahedrons in three space dimensions. Since the software library `deal.ii` [8] that we use for our implementation of the schemes allows only quadrilateral and hexahedral elements, we restrict ourselves to these types of elements in the following. Triangular and tetrahedral elements can be treated in an analogous way. In our calculations (cf. Sect. 5) we use the Raviart–Thomas element on quadrilateral meshes for two space dimensions. For an application in three dimensions based on the Raviart–Thomas–Nédélec element we refer to [15,37].

The construction of the discrete function spaces  $W_h$  and  $V_h$  on quadrilateral and hexahedral finite elements is done by a transformation  $\mathcal{T}_K: \hat{K} \rightarrow K$  of the reference element  $\hat{K} = [0, 1]^d$ , with  $d = 2$  or  $d = 3$ , to the element  $K$  through a diffeomorphism  $\mathcal{T}_K$  for all  $K \in \mathcal{T}_h$ . We sketch this briefly for  $d = 2$ ; cf. [21,37] for  $d = 3$ . For this, let

$$\hat{Q}^{p_1, p_2} := \left\{ \hat{p} : [0, 1]^2 \rightarrow \mathbb{R} \mid \hat{p}(\hat{x}) = \sum_{i=0}^{p_1} \sum_{j=0}^{p_2} p_{i,j} x_1^i x_2^j, \quad p_{i,j} \in \mathbb{R} \right\}.$$

We then define the discrete subspaces  $W_h^p \subset W$  and  $V_h^p \subset V$  by

$$W_h = W_h^p := \left\{ w \in W \mid w_K \circ \mathcal{T}_K^{-1} \in \hat{Q}^{p,p}, \text{ for } K \in \mathcal{T}_h \right\}, \tag{2.26}$$

$$V_h = V_h^p := \left\{ v \in V \mid v_K \circ \mathcal{T}_K^{-1} \in \hat{Q}^{p+1,p} \times \hat{Q}^{p,p+1}, \text{ for } K \in \mathcal{T}_h \right\}. \tag{2.27}$$

The fully discrete continuous Galerkin–Petrov and MFE approximation scheme, referred to as cGP( $r$ )–MFEM( $p$ ), then defines fully discrete solutions  $u_{\tau,h} \in \mathcal{X}^r(W_h)$  and  $q_{\tau,h} \in \mathcal{X}^r(V_h)$  that are represented in terms of basis functions in time by

$$u_{\tau,h}|_{\bar{I}_n}(t) = \sum_{j=0}^r U_{n,h}^j \varphi_{n,j}(t) \quad \text{and} \quad q_{\tau,h}|_{\bar{I}_n}(t) = \sum_{j=0}^r Q_{n,h}^j \varphi_{n,j}(t), \quad \text{for } t \in I_n,$$

with coefficient functions  $U_{n,h}^j \in W_h$  and  $Q_{n,h}^j \in V_h$  for  $j = 0, \dots, r$ . The coefficient functions are obtained by solving the variational problem (2.19), (2.20) in the discrete subspaces  $W_h \subset W$  and  $V_h \subset V$ : For  $n = 1, \dots, N$  and  $j = 1, \dots, r$  find coefficient functions  $\{U_{n,h}^j, Q_{n,h}^j\} \in W_h \times V_h$  such that

$$\sum_{j=0}^r \hat{\alpha}_{ij} \langle U_{n,h}^j, w_h \rangle + \tau_n \hat{\beta}_{ii} \langle \nabla \cdot Q_{n,h}^i, w_h \rangle = \tau_n \hat{\beta}_{ii} \langle f(t_{n,i}), w_h \rangle, \tag{2.28}$$

$$\langle \mathbf{D}^{-1} \mathbf{Q}_{n,h}^i, \mathbf{v}_h \rangle - \langle U_{n,h}^i, \nabla \cdot \mathbf{v}_h \rangle = 0 \tag{2.29}$$

for  $i = 1, \dots, r$  and all  $\{w_h, \mathbf{v}_h\} \in W_h \times \mathbf{V}_h$ , where  $U_{n,h}^0 \in W_h$  and  $\mathbf{Q}_{n,h}^0 \in \mathbf{V}_h$  are defined by means of the continuity constraint (cf. Remark 2.2), i.e.

$$\begin{aligned} U_{n,h}^0 &:= \sum_{j=0}^r U_{n-1,h}^j \varphi_{n-1,j}(t_{n-1}) \text{ if } n \geq 2, & U_{n,h}^0 &:= P_h u_0 \text{ if } n = 1, \\ \mathbf{Q}_{n,h}^0 &:= \sum_{j=0}^r \mathbf{Q}_{n-1,h}^j \varphi_{n-1,j}(t_{n-1}) \text{ if } n \geq 2, & \mathbf{Q}_{n,h}^0 &:= \mathbf{P}_h(-\mathbf{D}\nabla u_0) \text{ if } n = 1, \end{aligned} \tag{2.30}$$

with  $P_h: L^2(\Omega) \mapsto W_h$  and  $\mathbf{P}_h: L^2(\Omega) \mapsto \mathbf{V}_h$  denoting the  $L^2$  projections onto  $W_h$  and  $\mathbf{V}_h$ , respectively.

For the derivation of the algebraic formulation of the fully discrete variational problem (2.28), (2.29) we also refer to [15,36]. In [15,36], the iterative solution of the arising linear systems and the construction of an efficient preconditioner is further addressed. For solving the algebraic counterpart of Eqs. (2.28), (2.29) we do not apply an additional hybridization technique as it was done, for instance, in [11,12,14] and the references therein. We solve the algebraic system by using a Schur complement technique. In [36] the efficiency of the proposed iterative solver along with an adapted preconditioning technique is analyzed numerically. In [15,36], the approximation properties of some families of space-time discretization schemes, including the cGP( $r$ )–MFEM( $p$ ) approach, in terms of convergence rates and their robustness are studied by numerous numerical experiments. Test cases in three space dimensions and with heterogeneous and strongly anisotropic material properties are also included.

### 3 Existence and uniqueness of the semidiscrete approximation and error estimates

In this subsection we prove the existence and uniqueness of solutions to the semidiscrete approximation scheme that is defined by (2.14), (2.15) and its numerically integrated counterpart (2.19), (2.20), respectively. The time discretization error is also studied in this section. The spatial discretization error is analyzed in Sect. 4.

#### 3.1 Existence and uniqueness of the semidiscrete approximation

**Theorem 3.1** (Uniqueness of solutions) *Let the assumptions of Sect. 2.2 about  $\Omega$ ,  $u_0$  and  $f$  be satisfied. Then the solution  $\{u_\tau, \mathbf{q}_\tau\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(\mathbf{V})$  of the semidiscrete problem (2.14), (2.15) is unique.*

*Proof* Suppose that  $\{u_{\tau,1}, \mathbf{q}_{\tau,1}\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(\mathbf{V})$  and  $\{u_{\tau,2}, \mathbf{q}_{\tau,2}\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(\mathbf{V})$ , respectively, satisfy the semidiscrete problem (2.14), (2.15) and let  $u_\tau := u_{\tau,1} - u_{\tau,2}$  and  $\mathbf{q}_\tau := \mathbf{q}_{\tau,1} - \mathbf{q}_{\tau,2}$ . Then, the tuple  $\{u_\tau, \mathbf{q}_\tau\}$  satisfies (2.14), (2.15) with

$f \equiv 0$ . Choosing the test function  $w_\tau = A^{-1}\partial_t u_\tau \in \mathcal{Y}^{r-1}(D(A)) \subset \mathcal{Y}^{r-1}(W)$  in (2.14) yields that

$$\int_0^T \langle \partial_t u_\tau, A^{-1}\partial_t u_\tau \rangle dt + \int_0^T \langle \nabla \cdot \mathbf{q}_\tau, A^{-1}\partial_t u_\tau \rangle dt = 0. \tag{3.1}$$

From estimate (2.5) we get that

$$\int_0^T \langle \partial_t u_\tau, A^{-1}\partial_t u_\tau \rangle dt \geq \frac{\alpha}{\beta^2} \int_0^T \|\partial_t u_\tau\|_{H^{-1}(\Omega)}^2 dt. \tag{3.2}$$

Using integration by parts in the second of the integrals in (3.1) and recalling that  $A^{-1}\partial_t u_\tau \in D(A) \subset H_0^1(\Omega)$ , Eq. (3.1) along with (3.2) yields that

$$0 \geq \frac{\alpha}{\beta^2} \int_0^T \|\partial_t u_\tau\|_{H^{-1}(\Omega)}^2 dt - \int_0^T \langle \mathbf{q}_\tau, \nabla A^{-1}\partial_t u_\tau \rangle dt. \tag{3.3}$$

Next, by choosing the test function  $\mathbf{v}_\tau = \mathbf{D}\nabla A^{-1}\partial_t u_\tau \in \mathcal{Y}^{r-1}(V)$  in Eq. (2.15) we find that

$$\int_0^T \langle \mathbf{D}^{-1}\mathbf{q}_\tau, \mathbf{D}\nabla A^{-1}\partial_t u_\tau \rangle dt - \int_0^T \langle u_\tau, \nabla \cdot (\mathbf{D}\nabla A^{-1}\partial_t u_\tau) \rangle dt = 0. \tag{3.4}$$

Since

$$\nabla \cdot (\mathbf{D}\nabla A^{-1}\partial_t u_\tau) = -AA^{-1}\partial_t u_\tau = -\partial_t u_\tau$$

and  $\mathbf{D} = \mathbf{D}^\top$  by assumption, it follows from (3.4) that

$$\int_0^T \langle \mathbf{q}_\tau, \nabla A^{-1}\partial_t u_\tau \rangle dt + \frac{1}{2} \int_0^T \frac{d}{dt} \|u_\tau\|^2 dt = 0.$$

Since  $u_\tau(0) = u_{\tau,1}(0) - u_{\tau,2}(0) = 0$  it follows that

$$\int_0^T \langle \mathbf{q}_\tau, \nabla A^{-1}\partial_t u_\tau \rangle dt + \frac{1}{2} \|u_\tau(T)\|^2 = 0. \tag{3.5}$$

Combing relations (3.3) and (3.5) shows that

$$0 \geq c \int_0^T \|\partial_t u_\tau\|_{H^{-1}(\Omega)}^2 dt + \frac{1}{2} \|u_\tau(T)\|^2 \tag{3.6}$$

and, therefore,  $u_\tau = 0$ . This implies the uniqueness of solutions  $u_\tau$  to (2.14), (2.15).

To show the uniqueness of solutions  $\mathbf{q}_\tau$  of (2.14), (2.15), we choose the test function  $\mathbf{v}_\tau = \partial_t \mathbf{q}_\tau \in \mathcal{Y}^{r-1}(V)$ . Recalling that  $u_\tau = 0$  by means of the uniqueness result (3.6) we obtain from Eq. (2.15) that

$$\int_0^T \langle \mathbf{D}^{-1} \mathbf{q}_\tau, \partial_t \mathbf{q}_\tau \rangle dt = 0.$$

From  $\langle \mathbf{D}^{-1} \mathbf{q}_\tau, \partial_t \mathbf{q}_\tau \rangle = \frac{1}{2} \frac{d}{dt} \langle \mathbf{D}^{-1} \mathbf{q}_\tau, \mathbf{q}_\tau \rangle$  and  $\mathbf{q}_\tau(0) = \mathbf{0}$  we conclude that

$$0 = \frac{1}{2} \|\mathbf{D}^{-1/2} \mathbf{q}_\tau(T)\|^2. \tag{3.7}$$

Next, we choose  $\mathbf{v}_\tau = \partial_t^2 \mathbf{q}_\tau \in \mathcal{Y}^{r-2}(\mathbf{V})$ ,  $\mathcal{Y}^{r-2}(\mathbf{V}) \subset \mathcal{Y}^{r-1}(\mathbf{V})$  by definition, in (2.15), recall that  $u_\tau = 0$  and use that

$$\frac{d}{dt} \langle \mathbf{D}^{-1} \mathbf{q}_\tau, \partial_t \mathbf{q}_\tau \rangle = \langle \mathbf{D}^{-1} \partial_t \mathbf{q}_\tau, \partial_t \mathbf{q}_\tau \rangle + \langle \mathbf{D}^{-1} \mathbf{q}_\tau, \partial_t^2 \mathbf{q}_\tau \rangle.$$

Together, this implies that

$$0 = \int_0^T \frac{d}{dt} \langle \mathbf{D}^{-1} \mathbf{q}_\tau, \partial_t \mathbf{q}_\tau \rangle dt - \int_0^T \|\mathbf{D}^{-1/2} \partial_t \mathbf{q}_\tau\|^2 dt. \tag{3.8}$$

Since  $\mathbf{q}_\tau(0) = \mathbf{q}_{\tau,1}(0) - \mathbf{q}_{\tau,2}(0) = \mathbf{0}$  and, further,  $\mathbf{q}_\tau(T) = \mathbf{0}$  by means of (3.7), it follows from Eq. (3.8) along with property (3.7) that  $\mathbf{q}_\tau = 0$ . The uniqueness of solutions to the variational problem (2.14), (2.15) is thus proved.  $\square$

**Theorem 3.2** (Existence of solutions) *Let the assumptions of Sect. 2.2 about  $\Omega$ ,  $u_0$ ,  $\mathbf{D}$  and  $f$  be satisfied. Then the semidiscrete problem (2.14), (2.15) admits a solution  $\{u_\tau, \mathbf{q}_\tau\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(\mathbf{V})$ .*

*Proof* To prove existence of solutions to problem (2.14), (2.15), we will use an equivalent conformal formulation, see [43] for a similar approach.

Find  $\tilde{u}_\tau \in X^r(H_0^1(\Omega))$  such that  $\tilde{u}_\tau(0) = u_0$  and

$$\int_0^T \langle \partial_t \tilde{u}_\tau, w_\tau \rangle dt + \int_0^T a(\tilde{u}_\tau, w_\tau) dt = \int_0^T \langle f, w_\tau \rangle dt \tag{3.9}$$

for all  $w_\tau \in Y^{r-1}(H_0^1(\Omega))$ .

The existence and uniqueness of the semidiscrete approximation satisfying (3.9) can be established. This is shown in the ‘‘Appendix’’ of this work. Then we define

$$u_\tau := \tilde{u}_\tau \quad \text{and} \quad \mathbf{q}_\tau := -\mathbf{D}\nabla \tilde{u}_\tau. \tag{3.10}$$

Obviously, it holds that  $u_\tau \in X^r(W)$  since  $H_0^1(\Omega) \subset W$ . Further, we have that  $\partial_t \tilde{u}_\tau \in L^2(I; H_0^1(\Omega))$  since on each of the subintervals  $I_n$ ,  $n = 1, \dots, N$  the function  $\tilde{u}_\tau \in X^r(H_0^1(\Omega))$  admits the representation

$$u_\tau|_{I_n}(t) = \sum_{j=0}^r U_n^j \varphi_{n,j}(t), \quad \text{for } t \in I_n,$$

with coefficients  $U_n^j \in H_0^1(\Omega)$  and polynomial basis functions  $\varphi_{n,j} \in \mathbb{P}_r(I_n; \mathbb{R})$ .

Next, we prove that  $\mathbf{q}_\tau \in X^r(\mathbf{V})$ . Under the assumption of Sect. 2.2 that  $f \in L^2(I; W)$  it follows that

$$\begin{aligned} \int_0^T \langle -\mathbf{q}_\tau, \nabla w_\tau \rangle dt &= \int_0^T \langle \mathbf{D}\nabla u_\tau, \nabla w_\tau \rangle dt \\ &= \int_0^T \langle f - \partial_t u_\tau, w_\tau \rangle dt =: \int_0^T \langle \tilde{f}, w_\tau \rangle dt \end{aligned}$$

for all  $w_\tau \in Y^{r-1}(C_0^\infty(\Omega))$  with  $\tilde{f} \in L^2(I; L^2(\Omega))$ . Thus, we have that

$$\int_0^T \langle -\mathbf{q}_\tau, \nabla w_\tau \rangle dt = \int_0^T \langle \tilde{f}, w_\tau \rangle dt.$$

Consequently, it holds that (cf. [17, p. 18, Eq. (3.38)])

$$\int_0^T \langle \nabla \cdot \mathbf{q}_\tau, w_\tau \rangle dt = \int_0^T \langle \tilde{f}, w_\tau \rangle dt$$

for all  $w_\tau \in Y^{r-1}(C_0^\infty(\Omega))$  in the sense of distributions. Since  $\tilde{f} \in L^2(I; L^2(\Omega))$ , it follows that  $\nabla \cdot \mathbf{q}_\tau \in L^2(I; L^2(\Omega))$  and, therefore, that  $\mathbf{q}_\tau \in L^2(I; \mathbf{V})$  is fulfilled. Finally, from the expansion in terms of polynomial basis functions

$$\mathbf{q}_\tau(t) = - \sum_{j=0}^r \mathbf{D}\nabla U_n^j \varphi_{n,j}(t), \tag{3.11}$$

we conclude that  $\mathbf{q}_\tau \in C([0, T]; \mathbf{V})$ .

Equation (3.9) then directly implies that the functions  $u_\tau$  and  $\mathbf{q}_\tau$  defined in (3.10) satisfy the first equation of the variational problem (2.14), (2.15). The second equation of the system (2.14), (2.15) then follows from the representation (3.11) of the variable  $\mathbf{q}_\tau$  by testing the identity (3.11) with some function  $\mathbf{v}_\tau \in \mathcal{Y}^{r-1}(\mathbf{V})$  and applying the divergence theorem of Gauss. Hence, the assertion of the theorem is proved.  $\square$

As a corollary of the previous two theorems proving the existence of a unique solution to the semidiscrete problem (2.14), (2.15) we obtain an inf-sup stability condition within our space-time framework. This result will play a fundamental role in our error analyses. For this we need some further notation. Let  $\{u_\tau, \mathbf{q}_\tau\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(\mathbf{V})$  denote the solution of the semidiscrete problem (2.14), (2.15). We split  $u_\tau$  as

$$u_\tau(t) = u_0 + u_\tau^0(t) \quad \text{with} \quad u_\tau^0 \in \mathcal{X}_0^r(W). \tag{3.12}$$

In terms of the tuple  $\{u_\tau^0, \mathbf{q}_\tau\}$  of unknowns we recast the existence and uniqueness result of Theorems 3.1 and 3.2 in the following form.

**Corollary 3.3** *Let the assumptions of Sect. 2.2 about  $\Omega, u_0, \mathbf{D}$  and  $f$  be satisfied. Let  $\{u_\tau, \mathbf{q}_\tau\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(\mathbf{V})$  be the unique solution of the semidiscrete problem (2.14), (2.15) according to Theorems 3.1 and 3.2. Then, the tuple  $\{u_\tau^0, \mathbf{q}_\tau\} \in \mathcal{X}_0^r(W) \times \mathcal{X}^r(\mathbf{V})$  with  $u_\tau^0$  being defined in (3.12) is the unique solution of the following variational problem: Find  $\{u_\tau^0, \mathbf{q}_\tau\} \in \mathcal{X}_0^r(W) \times \mathcal{X}^r(\mathbf{V})$  such that*

$$\int_0^T \langle \partial_t u_\tau^0, w_\tau \rangle dt + \int_0^T \langle \nabla \cdot \mathbf{q}_\tau, w_\tau \rangle dt = \int_0^T \langle f, w_\tau \rangle dt, \tag{3.13}$$

$$\int_0^T \langle \mathbf{D}^{-1} \mathbf{q}_\tau, \mathbf{v}_\tau \rangle dt - \int_0^T \langle u_\tau^0, \nabla \cdot \mathbf{v}_\tau \rangle dt = \int_0^T \langle u_0, \nabla \cdot \mathbf{v}_\tau \rangle dt \tag{3.14}$$

for all  $w_\tau \in \mathcal{Y}^{r-1}(W)$  and  $\mathbf{v} \in \mathcal{Y}^{r-1}(\mathbf{V})$ .

As a corollary we get the following inf–sup stability condition.

**Corollary 3.4** *Let the assumptions of Sect. 2.2 about  $\Omega, u_0, \mathbf{D}$  and  $f$  be satisfied. Then, there exists a constant  $\gamma > 0$  such that*

$$\inf_{\{u_\tau^0, \mathbf{q}_\tau\} \in \mathcal{W} \setminus \{\mathbf{0}\}} \sup_{\{w_\tau, \mathbf{v}_\tau\} \in \mathcal{V} \setminus \{\mathbf{0}\}} \frac{a_\tau(\{u_\tau^0, \mathbf{q}_\tau\}, \{w_\tau, \mathbf{v}_\tau\})}{\|\{u_\tau^0, \mathbf{q}_\tau\}\|_{\mathcal{W}} \|\{w_\tau, \mathbf{v}_\tau\}\|_{\mathcal{V}}} \geq \gamma > 0. \tag{3.15}$$

*Proof* The discrete problem (3.13), (3.14) satisfies the assumptions of the Banach–Nečas–Babuška theorem [24, p. 85]. Since the discrete problem (3.13), (3.14) is well-posed according to Corollary 3.3, the Banach–Nečas–Babuška theorem implies the inf–sup stability condition (3.15). □

### 3.2 Estimates for the error between the continuous and the semidiscrete solution

Now we shall show error estimates for the exact form (2.14), (2.15) of the cGP( $r$ ) approach applied to the mixed formulation (2.10), (2.11) of our parabolic model problem.

For this we assume that the following approximation property are satisfied. There exist interpolation operators  $I_\tau: H_0^1(I, W) \mapsto \mathcal{X}_0^r(W), \mathbf{J}_\tau: L^2(I; \mathbf{V}) \mapsto \mathcal{X}^r(\mathbf{V})$  such that for sufficiently smooth functions  $u \in H^1(I; W)$  and  $\mathbf{q} \in L(I; \mathbf{V})$  and all time intervals  $I_n$ , for  $n = 1, \dots, N$ , it holds that

$$\|u - I_\tau u\|_{L^2(I_n; W)} \leq c \tau_n^{r+1} \|\partial_t^{r+1} u\|_{L^2(I_n; W)}, \tag{3.16}$$

$$\|\partial_t(u - I_\tau u)\|_{L^2(I_n; W)} \leq c \tau_n^r \|\partial_t^{r+1} u\|_{L^2(I_n; W)}, \tag{3.17}$$

$$\|\mathbf{q} - \mathbf{J}_\tau \mathbf{q}\|_{L^2(I_n; \mathbf{V})} \leq c \tau_n^{r+1} \|\partial_t^{r+1} \mathbf{q}\|_{L^2(I_n; \mathbf{V})} \tag{3.18}$$

with some constant  $c$  independent of  $\tau_n$  and  $\tau$ . The existence of such approximations is obviously ensured, for instance, by using Lagrange interpolation [48].

We get the following error estimates in the natural norm of the time discretization.



**Theorem 3.5** [Space-time error estimate for exact form of cGP( $r$ )] *Let the assumptions of Sect. 2.2 about  $\Omega$ ,  $u_0$ ,  $\mathbf{D}$  and  $f$  be satisfied. Let  $\{u, \mathbf{q}\} \in H^1(I; W) \times L^2(I; V)$  denote the unique solution of the mixed problem (2.10), (2.11) that is supposed to be sufficiently regular. Then the solution  $\{u_\tau, \mathbf{q}_\tau\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(V)$  of the semidiscrete problem (2.14), (2.15) satisfies the error estimate*

$$\begin{aligned} \|\{u - u_\tau, \mathbf{q} - \mathbf{q}_\tau\}\|_{\mathcal{W}} &\leq c \left\{ \sum_{n=1}^N \tau_n^{2r} \left( \|\partial_t^{r+1} u\|_{L^2(I_n; W)}^2 + \|\partial_t^{r+1} \mathbf{q}\|_{L^2(I_n; V)}^2 \right) \right\}^{1/2} \\ &\leq c\tau^r \left( \|\partial_t^{r+1} u\|_{L^2(I; W)} + \|\partial_t^{r+1} \mathbf{q}\|_{L^2(I; V)} \right), \end{aligned}$$

where the constant  $c$  is independent of  $\tau_n$ ,  $\tau$  and  $T$ .

*Proof* By splitting

$$u(t) = u_0 + u^0(t) \quad \text{with} \quad u^0 \in H_0^1(I; W) \tag{3.19}$$

and recalling the semidiscrete counterpart (3.12), we get that

$$u(t) - u_\tau(t) = u^0(t) - u_\tau^0(t), \quad \partial_t^r u(t) = \partial_t^r u^0(t)$$

for almost every  $t \in (0, T)$ , such that it is sufficient to derive the asserted error bounds of the theorem for  $u^0 - u_\tau^0$  instead of estimating  $u - u_\tau$ . This will be done in the following.

By (3.16)–(3.18) it holds that

$$\begin{aligned} &\|\{u^0 - I_\tau u^0, \mathbf{q} - \mathbf{J}_\tau \mathbf{q}\}\|_{\mathcal{W}} \\ &\leq c \left\{ \sum_{n=1}^N \tau_n^{2r} \left( \|\partial_t^{r+1} u^0\|_{L^2(I_n; W)}^2 + \|\partial_t^{r+1} \mathbf{q}\|_{L^2(I_n; V)}^2 \right) \right\}^{1/2} \\ &\leq c\tau^r \left( \|\partial_t^{r+1} u^0\|_{L^2(I; W)} + \|\partial_t^{r+1} \mathbf{q}\|_{L^2(I; V)} \right). \end{aligned} \tag{3.20}$$

For the discrete functions  $w_\tau := u_\tau^0 - I_\tau u^0 \in \mathcal{X}_0^r(W)$ ,  $\mathbf{v}_\tau = \mathbf{q}_\tau - \mathbf{J}_\tau \mathbf{q} \in \mathcal{X}^r(V)$  there exist, due to the inf–sup stability condition (3.15), functions  $\varphi_\tau \in \mathcal{X}_0^r(W)$ ,  $\psi_\tau \in \mathcal{X}^r(V)$  such that

$$\begin{aligned} \gamma \|\{w_\tau, \mathbf{v}_\tau\}\|_{\mathcal{W}} \|\{\varphi_\tau, \psi_\tau\}\|_{\mathcal{V}} &\leq a_\tau (\{w_\tau, \mathbf{v}_\tau\}, \{\varphi_\tau, \psi_\tau\}) \\ &= a_\tau (\{u^0 - I_\tau u^0, \mathbf{q} - \mathbf{J}_\tau \mathbf{q}\}, \{\varphi_\tau, \psi_\tau\}) \\ &\leq c \|\{u^0 - I_\tau u^0, \mathbf{q} - \mathbf{J}_\tau \mathbf{q}\}\|_{\mathcal{W}} \|\{\varphi_\tau, \psi_\tau\}\|_{\mathcal{V}}, \end{aligned} \tag{3.21}$$

where the Galerkin orthogonalities

$$\int_0^T \langle \partial_t (u_\tau^0 - u^0), w_\tau \rangle dt + \int_0^T \langle \nabla \cdot (\mathbf{q}_\tau - \mathbf{q}), w_\tau \rangle dt = 0,$$

$$\int_0^T \langle \mathbf{D}^{-1}(\mathbf{q}_\tau - \mathbf{q}), \mathbf{v}_\tau \rangle dt - \int_0^T \langle u_\tau^0 - u^0, \nabla \cdot \mathbf{v}_\tau \rangle dt = 0$$

have been used. From (3.21) along with (3.20), we find that

$$\begin{aligned} \|\{u_\tau^0 - I_\tau u^0, \mathbf{q}_\tau - \mathbf{J}_\tau \mathbf{q}\}\|_{\mathcal{W}} &\leq c\gamma^{-1} \|\{u^0 - I_\tau u^0, \mathbf{q} - \mathbf{J}_\tau \mathbf{q}\}\|_{\mathcal{W}} \\ &\leq c\gamma^{-1} \tau^r \left( \|\partial_t^{r+1} u\|_{L^2(I; W)} + \|\partial_t^{r+1} \mathbf{q}\|_{L^2(I; V)} \right). \end{aligned} \tag{3.22}$$

From inequality (3.22) along with the interpolation error estimate (3.20) we conclude the assertion of the theorem by means of the triangle inequality.  $\square$

Theorem 3.5 yields an error estimate with respect to the natural space-time norm of the discretization scheme. The estimate is sharp with respect to the contribution of  $\|\partial_t(u - u_\tau)\|_{L^2(I; W)}$  to the overall norm (2.12). However, the estimate is suboptimal with respect to  $\|u - u_\tau\|_{L^2(I; W)}$ . In the following theorem, we sharpen our analysis by providing an optimal order error estimate also for  $\|u - u_\tau\|_{L^2(I; W)}$ . This is done by a duality argument. For this, the following additional regularity assumption is needed.

**Regularity condition ( $\mathbf{R}_{\text{mix}}$ )** *Suppose that  $g \in L^2(I; W)$ . The variational problem, find  $z \in H^1(I; W)$  and  $\mathbf{p} \in L^2(I; V)$  with  $z(T) = 0$  such that*

$$\int_0^T (\langle -\partial_t z, w \rangle + \langle \nabla \cdot \mathbf{p}, w \rangle) dt = \int_0^T \langle g, w \rangle dt, \tag{3.23}$$

$$\int_0^T (\langle \mathbf{D}^{-1} \mathbf{p}, \mathbf{v} \rangle - \langle z, \nabla \cdot \mathbf{v} \rangle) dt = 0 \tag{3.24}$$

for all  $w \in L^2(T, 0; W)$ ,  $\mathbf{v} \in L^2(T, 0; V)$  admits a unique solution  $\{z, \mathbf{p}\} \in H^1(I; L^2(\Omega)) \times L^2(I; V)$  with the improved regularity  $\mathbf{p} \in H^1(I; V)$  and the a priori estimate

$$\|\partial_t \mathbf{p}\|_{L^2(I; V)} \leq c \|g\|_{L^2(I; W)}. \tag{3.25}$$

Formally, the corresponding strong form of (3.23), (3.24) is given by

$$-\partial_t z + \nabla \cdot \mathbf{p} = g, \quad \mathbf{D}^{-1} \mathbf{p} + \nabla z = 0 \quad \text{in } \Omega \times (0, T), \tag{3.26}$$

with  $z(T) = 0$  and homogeneous Dirichlet boundary conditions, that is obtained by rewriting the dual problem associated with (2.6)–(2.8),

$$-\partial_t z - \nabla \cdot (\mathbf{D} \nabla z) = g \quad \text{in } \Omega \times (0, T), \quad z(T) = 0 \quad \text{in } \Omega, \quad z = 0 \quad \text{on } \partial\Omega \times (0, T), \tag{3.27}$$

as a system of first order equations. Defining the transformation  $\tilde{z}(t) := z(T - t)$  and  $\tilde{g}(t) := z(T - t)$  we recast (3.27) as a forward parabolic problem in  $\tilde{z}$ ,

$$\partial_t \tilde{z} - \nabla \cdot (\mathbf{D} \nabla \tilde{z}) = \tilde{g} \quad \text{in } \Omega \times I, \quad \tilde{z}(0) = 0 \quad \text{in } \Omega, \quad \tilde{z} = 0 \quad \text{on } \partial\Omega \times (0, T),$$

such that standard existence and stability estimates can be applied; cf. [26, p. 382, Theorem 5]. Then, defining the variable  $\mathbf{p}$  by means of the second of the identities in

(3.26), the thus obtained tuple  $\{z, \mathbf{p}\}$  satisfies the variational problem (3.23), (3.24). Moreover, for  $g \in L^2(I; W)$ , from [26, p. 382, Theorem 5] we get the a priori estimate

$$\|\partial_t z\|_{L^2(I; W)} + \|\mathbf{p}\|_{L^2(I; V)} \leq c \|g\|_{L^2(I; W)}. \tag{3.28}$$

For this we note that  $\mathbf{p} \in L^2(I; V)$  can be shown by using the arguments of the proof of Theorem 3.2. The a priori estimate of the vector variable  $\mathbf{p}$  in (3.28) is then a direct consequence of the variational equation (3.23).

*Remark 3.6* A regularity condition similar to  $(R_{\text{mix}})$  is also used in [46, p. 48, Eq. (6.16)] to prove the optimal order convergence of a variational time discretization of second order parabolic problems in the non-mixed formulation. Currently, it remains an open problem how this limiting condition can be avoided in the theoretical analysis. The techniques that were developed recently in [25] might be helpful. However, in our numerical convergence studies of Sect. 5 the optimal convergence rate that is proved in Theorem 3.8 under the condition  $(R_{\text{mix}})$  is nicely observed.

Below we also need the following auxiliary lemma.

**Lemma 3.7** *Let  $I_0: H^1(I, W) \mapsto \mathcal{Y}^0(W)$  and  $\mathbf{J}_0: H^1(I, V) \mapsto \mathcal{Y}^0(V)$  be interpolation operators that are defined on each subinterval  $I_n$  by means of*

$$I_0 u(t) := u(t_{n-1}) \quad \text{and} \quad \mathbf{J}_0 \mathbf{v}(t) := \mathbf{v}(t_{n-1}) \quad \text{for all } t \in I_n.$$

Then it holds that

$$\|z - I_0 z\|_{L^2(I_n, W)} \leq \tau_n \|\partial_t z\|_{L^2(I_n; W)}, \tag{3.29}$$

$$\|\mathbf{p} - \mathbf{J}_0 \mathbf{p}\|_{L^2(I_n; V)} \leq \tau_n \|\partial_t \mathbf{p}\|_{L^2(I_n; V)}. \tag{3.30}$$

*Proof* The assertions directly follow from [46, Lemma 6.2]; cf. also [46, p. 49].  $\square$

**Theorem 3.8** [ $L^2$  Error estimate for the exact form of cGP( $r$ )] *Let the assumptions of Sect. 2.2 about  $\Omega, u_0, \mathbf{D}$  and  $f$  be satisfied. Further, suppose that the regularity condition  $(R_{\text{mix}})$  holds. Let  $\{u, \mathbf{q}\} \in H^1(I; W) \times L^2(I; V)$  denote the unique solution of the mixed problem (2.10), (2.11) that is supposed to be sufficiently regular. Then the solution  $\{u_\tau, \mathbf{q}_\tau\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(V)$  of the semidiscrete problem (2.14), (2.15) satisfies the error estimate*

$$\begin{aligned} \|u - u_\tau\|_{L^2(I; W)} &\leq c \tau \left\{ \sum_{n=1}^N \tau_n^{2r} \left( \|\partial_t^{r+1} u\|_{L^2(I_n; W)}^2 + \|\partial_t^{r+1} \mathbf{q}\|_{L^2(I_n; V)}^2 \right) \right\}^{1/2} \\ &\leq c \tau^{r+1} \left( \|\partial_t^{r+1} u\|_{L^2(I; W)} + \|\partial_t^{r+1} \mathbf{q}\|_{L^2(I; V)} \right). \end{aligned}$$

*Proof* We put  $e_u := u^0 - u_\tau^0 \in L^2(I; W)$  and  $\mathbf{e}_q := \mathbf{q} - \mathbf{q}_\tau \in L^2(I; V)$  with the splitting (3.19) and (3.12) of the scalar variable and its semidiscrete approximation, respectively. Further, let  $\{z, \mathbf{p}\} \in H^1(0, T; W) \cap C([0, T]; W) \times L^2(0, T; V)$  with

$z(T) = 0$  denote the unique solution of (3.23), (3.24) with right-hand side function  $g = e_u$ .

Firstly, recalling that  $z(T) = 0$  and  $e_u(0) = 0$  by definition, we get that

$$\int_0^T \langle -\partial_t z, e_u \rangle dt = -z(T)e_u(T) + z(0)e_u(0) + \int_0^T \langle \partial_t e_u, z \rangle dt = \int_0^T \langle \partial_t e_u, z \rangle dt. \tag{3.31}$$

Choosing the test function  $w = e_u$  in (3.23) and using (3.31), we find that

$$\begin{aligned} \int_0^T \|e_u\|^2 dt &= \int_0^T (\langle \partial_t e_u, z \rangle + \langle \nabla \cdot \mathbf{p}, e_u \rangle) dt \\ &= \int_0^T (\langle \partial_t e_u, z \rangle + \langle \nabla \cdot \mathbf{e}_q, z \rangle) dt - \int_0^T (\langle \nabla \cdot \mathbf{e}_q, z \rangle - \langle \nabla \cdot \mathbf{p}, e_u \rangle) dt. \end{aligned} \tag{3.32}$$

Choosing the test function  $\mathbf{v} = \mathbf{e}_q$  in (3.24) and recalling that the matrix  $\mathbf{D}$  is symmetric by assumption, we conclude that

$$\int_0^T \langle \nabla \cdot \mathbf{e}_q, z \rangle dt = \int_0^T \langle \mathbf{D}^{-1} \mathbf{p}, \mathbf{e}_q \rangle dt = \int_0^T \langle \mathbf{D}^{-1} \mathbf{e}_q, \mathbf{p} \rangle dt. \tag{3.33}$$

From (3.32) and (3.33) it then follows that

$$\int_0^T \|e_u\|^2 dt = \int_0^T (\langle \partial_t e_u, z \rangle + \langle \nabla \cdot \mathbf{e}_q, z \rangle) dt - \int_0^T (\langle \mathbf{D}^{-1} \mathbf{e}_q, \mathbf{p} \rangle - \langle e_u, \nabla \cdot \mathbf{p} \rangle) dt. \tag{3.34}$$

Secondly, by Galerkin orthogonality we find that

$$\int_0^T (\langle \partial_t e_u, w_\tau \rangle + \langle \nabla \cdot \mathbf{e}_q, w_\tau \rangle) dt = 0, \tag{3.35}$$

$$\int_0^T (\langle \mathbf{D}^{-1} \mathbf{e}_q, \mathbf{v}_\tau \rangle - \langle e_u, \nabla \cdot \mathbf{v}_\tau \rangle) dt = 0 \tag{3.36}$$

for all  $w_\tau \in \mathcal{Y}^{r-1}(W)$  and  $\mathbf{v}_\tau \in \mathcal{Y}^{r-1}(V)$ . Choosing  $w_\tau = I_0 z$  in (3.35), it follows that

$$\int_0^T (\langle \partial_t e_u, I_0 z \rangle + \langle \nabla \cdot \mathbf{e}_q, I_0 z \rangle) dt = 0. \tag{3.37}$$

Further, choosing  $\mathbf{v}_\tau = \mathbf{J}_0 \mathbf{p}$  in (3.36) yields that

$$\int_0^T (\langle \mathbf{D}^{-1} \mathbf{e}_q, \mathbf{J}_0 \mathbf{p} \rangle - \langle e_u, \nabla \cdot \mathbf{J}_0 \mathbf{p}_\tau \rangle) dt = 0. \tag{3.38}$$

Thirdly, combining (3.34) with (3.37) and (3.38), and then using the Cauchy–Schwarz inequality as well as the interpolation error estimates (3.29) and (3.30) yields that

$$\begin{aligned}
 \|e_u\|_{L^2(I;W)}^2 &= \int_0^T (\langle \partial_t e_u, z - I_0 z \rangle + \langle \nabla \cdot \mathbf{e}_q, z - I_0 z \rangle) dt \\
 &\quad - \int_0^T (\langle \mathbf{D}^{-1} \mathbf{e}_q, \mathbf{p} - \mathbf{J}_0 \mathbf{p} \rangle - \langle e_u, \nabla \cdot (\mathbf{p} - \mathbf{J}_0 \mathbf{p}) \rangle) dt. \\
 &\leq (\|\partial_t e_u\|_{L^2(I;W)} + \|\mathbf{e}_q\|_{L^2(I;V)}) \|z - I_0 z\|_{L^2(I;W)} \\
 &\quad + (\|\mathbf{D}^{-1}\|_2 \|\mathbf{e}_q\|_{L^2(I;V)} + \|e_u\|_{L^2(I;W)}) \|\mathbf{p} - \mathbf{J}_0 \mathbf{p}\|_{L^2(I;V)} \\
 &\leq \tau (\theta_M \|\partial_t e_u\|_{L^2(I;W)} + \|\mathbf{e}_q\|_{L^2(I;V)}) \|\partial_t z\|_{L^2(I;W)} \\
 &\quad + c\tau (\|\mathbf{e}_q\|_{L^2(I;V)} + \|e_u\|_{L^2(I;W)}) \|\partial_t \mathbf{p}\|_{L^2(I;V)}.
 \end{aligned}$$

Applying the a priori estimate (3.28) and the additional regularity assumption (3.25) with  $g = e_u$  as well as using the error estimate of Theorem 3.5, we then find that

$$\begin{aligned}
 \|e_u\|_{L^2(I;W)} &\leq c\tau \left\{ \sum_{n=1}^N \tau_n^{2r} \left( \|\partial_t^{r+1} u\|_{L^2(I_n;W)}^2 + \|\partial_t^{r+1} \mathbf{q}\|_{L^2(I_n;V)}^2 \right) \right\}^{1/2} \\
 &\leq c\tau^{r+1} \left( \|\partial_t^{r+1} u\|_{L^2(I;W)} + \|\partial_t^{r+1} \mathbf{q}\|_{L^2(I;V)} \right).
 \end{aligned}$$

This proves the assertion of the theorem. □

Next we derive an error estimate for the non-exact form (2.19), (2.20) of the cGP( $r$ ) method. The difference of the non-exact form of cGP( $r$ ) to (2.14), (2.15) comes through the numerically integrated right-hand side term in (2.19). Firstly, we ensure the existence and uniqueness of the solution to the non-exact form of cGP( $r$ ).

**Theorem 3.9** (Existence and uniqueness) *Let the assumptions of Sect. 2.2 about  $\Omega, u_0, \mathbf{D}$  and  $f$  be satisfied. Then the non-exact form (2.19), (2.20) of the semidiscrete problem admits a unique solution  $\{U_n^j, \mathbf{Q}_n^j\} \in W \times V$  for  $j = 1, \dots, r$  and  $n = 1, \dots, N$  defining semidiscrete approximations  $\{u_\tau, \mathbf{q}_\tau\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(V)$  by means of the expansions (2.18) and the initial condition  $u_\tau(0) = u_0$ .*

*Proof* By the definition of the Lagrange interpolation operator  $\Pi_r$  given in (2.21) and the representations (2.18) of  $u_\tau$  and  $\mathbf{q}_\tau$  in terms of basis functions we recast the non-exact form (2.19), (2.20) of the semidiscrete problem in the equivalent form

$$\int_0^T \langle \partial_t u_\tau, w_\tau \rangle dt + \int_0^T \langle \nabla \cdot \mathbf{q}_\tau, w_\tau \rangle dt = \int_0^T \langle \Pi_r f, w_\tau \rangle dt, \tag{3.39}$$

$$\int_0^T \langle \mathbf{D}^{-1} \mathbf{q}_\tau, \mathbf{v}_\tau \rangle dt - \int_0^T \langle u_\tau, \nabla \cdot \mathbf{v}_\tau \rangle dt = 0 \tag{3.40}$$

for all  $w_\tau \in \mathcal{Y}^{r-1}(W)$  and  $\mathbf{v} \in \mathcal{Y}^{r-1}(V)$  with the initial condition  $u_\tau(0) = u_0$ .

Existence and uniqueness of the solution  $\{u_\tau, \mathbf{q}_\tau\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(V)$  to the system (3.39), (3.40) then follows as in Theorems 3.1 and 3.2 with  $\Pi_r f$  replacing  $f$  in the arguments of the proofs. □

Next, we present the corresponding a priori error estimate.

**Theorem 3.10** *Let the assumptions of Sect. 2.2 about  $\Omega, u_0, \mathbf{D}$  and  $f$  be satisfied. Suppose that  $f$  is sufficiently regular with respect to the time variable. Let  $\{u, \mathbf{q}\} \in H^1(I; W) \times L^2(I; V)$  denote the unique solution of the mixed problem (2.10), (2.11) that is supposed to be sufficiently regular. Then the solution  $\{u_\tau, \mathbf{q}_\tau\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(V)$  of the non-exact semidiscrete problem (2.14), (2.15) satisfies the error estimate*

$$\begin{aligned} \|\{u - u_\tau, \mathbf{q} - \mathbf{q}_\tau\}\|_{\mathcal{W}} &\leq c \left( \sum_{n=1}^N \tau_n^{2r} \left\{ \|\partial_t^{r+1} u\|_{L^2(I_n; W)}^2 + \tau_n^2 \|\partial_t^{r+1} \mathbf{q}\|_{L^2(I_n; V)}^2 \right. \right. \\ &\quad \left. \left. + \tau_n^2 \|\partial_t^{r+1} f\|_{L^2(I_n; W)}^2 \right\} \right)^{1/2} \\ &\leq c \tau^r \left\{ \|\partial_t^{r+1} u\|_{L^2(I; W)} + \tau \|\partial_t^{r+1} \mathbf{q}\|_{L^2(I; W)} + \tau \|\partial_t^{r+1} f\|_{L^2(I; W)} \right\}. \end{aligned}$$

where the constants  $c$  is independent of  $\tau_n, \tau$  and  $T$ .

Since the proof of Theorem 3.10 follows from the proof of Theorem 3.5 by a standard estimate of the interpolation error, we skip it here. For the sake of completeness we summarize the proof in the ‘‘Appendix’’ of this work.

### 4 Existence and uniqueness of the fully discrete approximation and error estimates

In the first subsection of Sect. 4 we prove the existence and uniqueness of solutions to the fully discrete approximation scheme (2.28), (2.29). Then, in Sect. 4.2 we establish an estimate for the error between the non-exact form of the semidiscrete approximation defined by Eqs. (2.19), (2.20) and the fully discrete solution given by Eqs. (2.28), (2.29). Finally, in Sect. 4.3 we combine the error estimates of the temporal discretization that are derived in Sect. 3 with the error estimates of Sect. 4.2 to get the desired error estimates.

#### 4.1 Existence and uniqueness of the fully discrete approximation

Firstly we prove the existence and uniqueness of solutions to the fully discrete cGP( $r$ )–MFEM( $p$ ) scheme (2.28), (2.29). For this we need the following lemma (cf. [44, p. 302]).

**Lemma 4.1** *For given  $w_h \in W_h$  there exists a function  $\mathbf{v}_h \in V_h$  satisfying*

$$\nabla \cdot \mathbf{v}_h = w_h \quad \text{and} \quad \|\mathbf{v}_h\| \leq c \|w_h\|$$

for some constant  $c > 0$  depending on  $\Omega$  and the space dimension  $d$  but not on  $w_h$  or the mesh size  $h$ .

**Theorem 4.2** (Existence and uniqueness of solutions) *Let the assumptions of Sect. 2.2 about  $\Omega$ ,  $u_0$ ,  $\mathbf{D}$  and  $f$  be satisfied. Then the fully discrete problem (2.28), (2.29) admits a unique solution  $\{u_{\tau,h}, \mathbf{q}_{\tau,h}\} \in \mathcal{X}^r(W_h) \times \mathcal{X}^r(V_h)$ .*

*Proof* Since the fully discrete problem (2.28), (2.29) is finite dimensional and linear, it is sufficient to show the uniqueness of the solution. The existence is then a direct consequence. Assume that there exist two pairs of solutions  $\{u_{\tau,h}^k, \mathbf{q}_{\tau,h}^k\} \in \mathcal{X}^r(W_h) \times \mathcal{X}^r(V_h)$ , for  $k = 1, 2$ , that are represented in terms of basis functions by

$$u_{\tau,h}^k(t)|_{I_n} = \sum_{j=0}^r U_{n,h}^{j,k} \varphi_{n,j}(t) \quad \text{and} \quad \mathbf{q}_{\tau,h}^k(t)|_{I_n} = \sum_{j=0}^r \mathbf{Q}_{n,h}^{j,k} \varphi_{n,j}(t), \quad \text{for } k = 1, 2,$$

and  $t \in I_n$  with coefficient functions  $U_{n,h}^{j,k} \in W_h$  and  $\mathbf{Q}_{n,h}^{j,k} \in V_h$ . The continuity constraint that is imposed by the definition of  $\mathcal{X}^r(W_h)$  and  $\mathcal{X}^r(V_h)$ , respectively, directly implies that  $U_{n,h}^{0,1} = U_{n,h}^{0,2}$  and  $\mathbf{Q}_{n,h}^{0,1} = \mathbf{Q}_{n,h}^{0,2}$ . Further, the pairs  $\{u_{\tau,h}^k(t), \mathbf{q}_{\tau,h}^k(t)\}$ , for  $k = 1, 2$ , both satisfy the discrete equations (2.28), (2.29). Therefore, it follows that

$$\sum_{j=0}^r \hat{\alpha}_{ij} \langle U_{n,h}^{j,1} - U_{n,h}^{j,2}, w_h \rangle + \tau_n \hat{\beta}_{ii} \langle \nabla \cdot (\mathbf{Q}_{n,h}^{i,1} - \mathbf{Q}_{n,h}^{i,2}), w_h \rangle = 0, \quad (4.1)$$

$$\langle \mathbf{D}^{-1}(\mathbf{Q}_{n,h}^{i,1} - \mathbf{Q}_{n,h}^{i,2}), \mathbf{v}_h \rangle - \langle U_{n,h}^{i,1} - U_{n,h}^{i,2}, \nabla \cdot \mathbf{v}_h \rangle = 0 \quad (4.2)$$

for  $i = 1, \dots, r$  and all  $\{w_h, \mathbf{v}_h\} \in W_h \times V_h$ . Now, by subtracting the equations (4.1) and (4.2) from each other and choosing the test functions  $w_h = U_{n,h}^{i,1} - U_{n,h}^{i,2}$  and  $\mathbf{v}_h = \tau_n \hat{\beta}_{ii} (\mathbf{Q}_{n,h}^{i,1} - \mathbf{Q}_{n,h}^{i,2})$ , for  $i = 1, \dots, r$  in (4.1) and (4.2), respectively, we get that

$$\sum_{j=0}^r \hat{\alpha}_{ij} \langle U_{n,h}^{j,1} - U_{n,h}^{j,2}, U_{n,h}^{i,1} - U_{n,h}^{i,2} \rangle + \tau_n \hat{\beta}_{ii} \langle \mathbf{D}^{-1}(\mathbf{Q}_{n,h}^{i,1} - \mathbf{Q}_{n,h}^{i,2}), \mathbf{Q}_{n,h}^{i,1} - \mathbf{Q}_{n,h}^{i,2} \rangle = 0, \quad (4.3)$$

for  $i = 1, \dots, r$ . Summing up Eq. (4.3) from  $i = 1$  to  $i = r$ , using Lemma 2.4 and recalling that  $U_{n,h}^{0,1} = U_{n,h}^{0,2}$  then implies that

$$\frac{1}{2} \|u_{\tau,h}^1(t_n) - u_{\tau,h}^2(t_n)\|^2 + \sum_{i=1}^r \tau_n \hat{\beta}_{ii} \langle \mathbf{D}^{-1}(\mathbf{Q}_{n,h}^{i,1} - \mathbf{Q}_{n,h}^{i,2}), \mathbf{Q}_{n,h}^{i,1} - \mathbf{Q}_{n,h}^{i,2} \rangle = 0. \quad (4.4)$$

The symmetric matrix  $\mathbf{D}^{-1}$  is positive definite by assumption (2.2) and  $\hat{\beta}_{ii} > 0$  under the coefficient property (C); cf. Lemma 2.3. Therefore, Eq. (4.4) immediately implies that  $\mathbf{Q}_{n,h}^{i,1} = \mathbf{Q}_{n,h}^{i,2}$  for  $i = 1, \dots, r$ . By Lemma 4.1 there exists some  $\mathbf{v}_h \in V_h$  such that  $\nabla \cdot \mathbf{v}_h = U_{n,h}^{i,1} - U_{n,h}^{i,2}$ . Using this  $\mathbf{v}_h$  as test function in (4.2) and noting that the first term in (4.2) now vanishes, we obtain that  $U_{n,h}^{i,1} = U_{n,h}^{i,2}$ , for  $i = 1, \dots, r$ . This implies the uniqueness of the solution to the fully discrete problem (2.28), (2.29) and proves the assertion of the theorem.  $\square$

### 4.2 Estimates for the error between the semidiscrete and the fully discrete solution

In this subsection we derive estimates for the error between the semidiscrete approximation defined by Eqs. (2.14), (2.15) and the fully discrete solution given by Eqs. (2.28), (2.29). For this we use the following projection operators (cf. [5,17] and [40, p. 237]) defined in  $W$  and  $V$ , respectively, by

$$P_h: W \rightarrow W_h, \quad \langle P_h w - w, w_h \rangle = 0 \tag{4.5}$$

for all  $w_h \in W_h$  and

$$\Pi_h: V \rightarrow V_h, \quad \langle \nabla \cdot (\Pi_h \mathbf{v} - \mathbf{v}), w_h \rangle = 0, \tag{4.6}$$

$$P_h: V \rightarrow V_h \quad \langle P_h \mathbf{v} - \mathbf{v}, \mathbf{v}_h \rangle = 0, \tag{4.7}$$

for all  $w_h \in W_h$  and  $\mathbf{v}_h \in V_h$ , respectively. We point out that  $\Pi_h$  is firstly defined on  $\mathbf{H}^1(\Omega)$  and then extended to  $V$  by following [40, p. 237]. For these operators and the family of Raviart–Thomas elements on quadrilateral elements for the two-dimensional case and the class of Raviart–Thomas–Nédélec elements in three space dimensions there holds that

$$\|w - P_h w\| \leq ch^{p+1} \|w\|_{p+1}, \tag{4.8}$$

$$\|\mathbf{v} - \Pi_h \mathbf{v}\| \leq ch^{p+1} \|\mathbf{v}\|_{p+1}, \quad \|\nabla \cdot (\mathbf{v} - \Pi_h \mathbf{v})\| \leq ch^{p+1} \|\nabla \cdot \mathbf{v}\|_{p+1}, \tag{4.9}$$

$$\|\mathbf{v} - P_h \mathbf{v}\| \leq ch^{p+1} \|\mathbf{v}\|_{p+1}, \quad \|\nabla \cdot (\mathbf{v} - P_h \mathbf{v})\| \leq ch^{p+1} \|\nabla \cdot \mathbf{v}\|_{p+1}, \tag{4.10}$$

for any  $w \in H^{p+1}(\Omega)$  and  $\mathbf{v} \in \mathbf{H}^{p+1}(\Omega)$ ,  $\nabla \cdot \mathbf{v} \in H^{p+1}(\Omega)$ , respectively.

For the error between the semidiscrete solution and fully discrete we use the notation

$$E_u(t) = u_\tau(t) - u_{\tau,h}(t), \quad E_q(t) = \mathbf{q}_\tau(t) - \mathbf{q}_{\tau,h}(t), \\ E_{u,n}^i = E_u(t_{n,i}), \quad E_{q,n}^i = E_q(t_{n,i})$$

for  $t \in I$  and  $n \in \{1, \dots, N\}$ ,  $i \in \{0, \dots, r\}$ . Representing the semidiscrete and fully discrete solution in terms of basis functions [cf. (2.18)] there holds that

$$E_u(t) = \sum_{i=0}^r E_{u,n}^i \varphi_{n,i}(t) \quad \text{and} \quad E_q(t) = \sum_{i=0}^r E_{q,n}^i \varphi_{n,i}(t), \quad \text{for } t \in I_n.$$

Next, we prove two preliminary lemmas.

**Lemma 4.3** *Let the assumptions of Sect. 2.2 about  $\Omega$ ,  $u_0$ ,  $\mathbf{D}$  and  $f$  be satisfied. Let the semidiscrete approximation  $\{u_\tau, \mathbf{q}_\tau\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(V)$  be defined by (2.18)–(2.20). Further, let  $\{u_{\tau,h}, \mathbf{q}_{\tau,h}\} \in \mathcal{X}^r(W_h) \times \mathcal{X}^r(V_h)$  be the unique solution of the fully discrete problem (2.28), (2.29). Then, for any  $K = 1, \dots, N$  it holds that*



$$\begin{aligned} & \|E_u(t_K)\|^2 + \sum_{n=1}^K \sum_{i=1}^r \tau_n \|E_{u,n}^i\|^2 + \sum_{n=1}^K \sum_{i=1}^r \tau_n \|E_{q,n}^i\|^2 \\ & \leq \|u_\tau(t_K) - P_h u_\tau(t_K)\|^2 + c \sum_{n=1}^K \sum_{i=1}^r \tau_n \left( \|U_n^i - P_h U_n^i\|^2 + \|\mathcal{Q}_n^i - \Pi_h \mathcal{Q}_n^i\|^2 \right) \end{aligned} \tag{4.11}$$

with some constant  $c > 0$  not depending on the discretization parameters  $h$  and  $\tau$ .

*Proof* By subtracting (2.28), (2.29) from (2.19), (2.20), respectively, it follows that

$$\sum_{j=0}^r \hat{\alpha}_{ij} \langle U_n^j - U_{n,h}^j, w_h \rangle + \tau_n \hat{\beta}_{ii} \langle \nabla \cdot (\mathcal{Q}_n^i - \mathcal{Q}_{n,h}^i), w_h \rangle = 0, \tag{4.12}$$

$$\langle \mathcal{D}^{-1}(\mathcal{Q}_n^i - \mathcal{Q}_{n,h}^i), \mathbf{v}_h \rangle - \langle U_n^i - U_{n,h}^i, \nabla \cdot \mathbf{v}_h \rangle = 0 \tag{4.13}$$

for  $i = 1, \dots, r$  and all  $\{w_h, \mathbf{v}_h\} \in W_h \times V_h$ . For any  $i = 1, \dots, r$  we choose the test functions  $w_h = P_h U_n^i - U_{n,h}^i \in W_h$  and  $\mathbf{v}_h = \tau_n \hat{\beta}_{ii} (\Pi_h \mathcal{Q}_n^i - \mathcal{Q}_{n,h}^i) \in V_h$  in (4.12) and (4.13), respectively. By adding the thus obtained equations, using the properties of the projection projectors  $P_h$  and  $\Pi_h$  defined in (4.5) and (4.6), respectively, and summing up from  $i = 1$  to  $r$  we get that

$$\begin{aligned} & \sum_{i=1}^r \sum_{j=0}^r \hat{\alpha}_{ij} \langle P_h U_n^j - U_{n,h}^j, P_h U_n^i - U_{n,h}^i \rangle \\ & + \sum_{i=1}^r \tau_n \hat{\beta}_{ii} \langle \mathcal{D}^{-1}(\mathcal{Q}_n^i - \mathcal{Q}_{n,h}^i), \Pi_h \mathcal{Q}_n^i - \mathcal{Q}_{n,h}^i \rangle = 0. \end{aligned} \tag{4.14}$$

We note that due to Lemma 2.4, the first term in (4.14) can be rewritten as

$$\begin{aligned} & \sum_{i=1}^r \sum_{j=0}^r \hat{\alpha}_{ij} \langle P_h U_n^j - U_{n,h}^j, P_h U_n^i - U_{n,h}^i \rangle \\ & = \frac{1}{2} \|P_h E_{u,n}(t_n)\|^2 - \frac{1}{2} \|P_h E_{u,n-1}(t_{n-1})\|^2. \end{aligned}$$

Along with some further algebraic manipulations we then conclude from (4.14) that

$$\begin{aligned} & \frac{1}{2} \|P_h E_{u,n}(t_n)\|^2 - \frac{1}{2} \|P_h E_{u,n-1}(t_{n-1})\|^2 \\ & + \sum_{i=1}^r \tau_n \hat{\beta}_{ii} \left\langle \mathcal{D}^{-1}(\mathcal{Q}_n^i - \mathcal{Q}_{n,h}^i), \mathcal{Q}_n^i - \mathcal{Q}_{n,h}^i \right\rangle \\ & = \sum_{i=1}^r \tau_n \hat{\beta}_{ii} \left\langle \mathcal{D}^{-1}(\mathcal{Q}_n^i - \mathcal{Q}_{n,h}^i), \mathcal{Q}_n^i - \Pi_h \mathcal{Q}_n^i \right\rangle. \end{aligned} \tag{4.15}$$

Recalling assumption (2.2) about  $\mathbf{D}$  and property (C) in (2.22) of the coefficients  $\hat{\beta}_{ii}$  and we obtain from Eq. (4.15) by applying Cauchy–Young’s inequality that

$$\begin{aligned} & \|P_h E_u(t_n)\|^2 - \|P_h E_u(t_{n-1})\|^2 + \sum_{i=1}^r \tau_n \beta_m \theta_m \| \mathbf{Q}_n^i - \mathbf{Q}_{n,h}^i \|^2 \\ & \leq \frac{\beta_M^2}{\beta_m \theta_m} \sum_{i=1}^r \tau_n \| \mathbf{Q}_n^i - \Pi_h \mathbf{Q}_n^i \|^2. \end{aligned} \tag{4.16}$$

Summing up inequality (4.16) from  $n = 1$  to  $K$  and noting that  $P_h E_u(t_0) = 0$  then shows that

$$\begin{aligned} & \|P_h E_u(t_K)\|^2 + \sum_{n=1}^K \sum_{i=1}^r \tau_n \beta_m \| \mathbf{Q}_n^i - \mathbf{Q}_{n,h}^i \|^2 \\ & \leq \frac{\beta_M^2}{\beta_m \theta_m} \sum_{n=1}^K \sum_{i=1}^r \tau_n \| \mathbf{Q}_n^i - \Pi_h \mathbf{Q}_n^i \|^2 \end{aligned} \tag{4.17}$$

for any  $K \in \mathbb{N}$  with  $K \leq N$ . By using now Lemma 4.1, there exists for any  $i \in \{1, \dots, r\}$  a  $\mathbf{v}_h \in \mathbf{V}_h$  such that  $\nabla \cdot \mathbf{v}_h = P_h E_{u,n}^i$  and  $\|\mathbf{v}_h\| \leq c \|P_h E_{u,n}^i\|$ . By testing (4.13) with this  $\mathbf{v}_h$ , we get by using the Cauchy–Schwarz inequality along with assumption (2.2) about  $\mathbf{D}$  that

$$\|P_h E_{u,n}^i\| \leq c \theta_M \| \mathbf{Q}_n^i - \mathbf{Q}_{n,h}^i \|, \tag{4.18}$$

for  $n = 1, \dots, N, i = 1, \dots, r$ .

Combining (4.17) with (4.18) it follows that

$$\begin{aligned} & \|P_h E_u(t_K)\|^2 + \sum_{n=1}^K \sum_{i=1}^r \tau_n \beta_m \| \mathbf{Q}_n^i - \mathbf{Q}_{n,h}^i \|^2 + \sum_{n=1}^K \sum_{i=1}^r \tau_n \|P_h E_{u,n}^i\|^2 \\ & \leq c \sum_{n=1}^K \sum_{i=1}^r \tau_n \| \mathbf{Q}_n^i - \Pi_h \mathbf{Q}_n^i \|^2. \end{aligned} \tag{4.19}$$

By  $\mathbf{E}_{q,n}^i = \mathbf{Q}_n^i - \mathbf{Q}_{n,h}^i$  and the triangle inequality relation (4.19) implies that

$$\begin{aligned} & \|E_u(t_K)\|^2 + \sum_{n=1}^K \sum_{i=1}^r \tau_n \beta_m \| \mathbf{E}_{q,n}^i \|^2 + \sum_{n=1}^K \sum_{i=1}^r \tau_n \|E_{u,n}^i\|^2 \\ & \leq c \sum_{n=1}^K \sum_{i=1}^r \tau_n \| \mathbf{Q}_n^i - \Pi_h \mathbf{Q}_n^i \|^2 + c \|u_\tau(t_k) - P_h u_\tau(t_k)\|^2 \\ & \quad + c \sum_{n=1}^K \sum_{i=1}^r \tau_n \|E_{u,n}^i - P_h E_{u,n}^i\|^2. \end{aligned} \tag{4.20}$$

Observing that  $E_{u,n}^i - P_h E_{u,n}^i = U_n^i - P_h U_n^i$ , inequality (4.20) proves (4.11).  $\square$

In the second lemma we restrict ourselves to the case that  $\mathbf{D} = d\mathbf{I}$  with some  $d > 0$  is satisfied. An extension of the provided estimates to more general matrices  $\mathbf{D}(\mathbf{x})$  still remains an open problem.

**Lemma 4.4** *Let the assumptions of Sect. 2.2 about  $\Omega, u_0$  and  $f$  be satisfied and  $\mathbf{D} = d\mathbf{I}$  with some  $d > 0$ . Let the semidiscrete approximation  $\{u_\tau, \mathbf{q}_\tau\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(\mathbf{V})$  be defined by (2.18)–(2.20). Further, let  $\{u_{\tau,h}, \mathbf{q}_{\tau,h}\} \in \mathcal{X}^r(W_h) \times \mathcal{X}^r(\mathbf{V}_h)$  be the unique solution of the fully discrete problem (2.28), (2.29). Then, for any  $K = 1, \dots, N$  it holds that*

$$\begin{aligned} & \sum_{n=1}^K \tau_n \sum_{i=1}^r \hat{\beta}_{ii} \|\nabla \cdot \mathbf{\Pi}_h \mathbf{E}_{q,n}^i\|^2 + \|\mathbf{P}_h \mathbf{E}_q(t_K)\|^2 \\ & \leq \sum_{n=1}^K \sum_{i=1}^r \tau_n \hat{\beta}_{ii} \|\nabla \cdot (\mathbf{P}_h - \mathbf{\Pi}_h) \mathbf{Q}_n^i\|^2. \end{aligned} \tag{4.21}$$

*Proof* Introducing the projectors into the error equations (4.12)–(4.13) yields that

$$\sum_{j=0}^r \hat{\alpha}_{ij} \langle P_h E_{u,n}^j, w_h \rangle + \tau_n \hat{\beta}_{ii} \langle \nabla \cdot \mathbf{\Pi}_h \mathbf{E}_{q,n}^i, w_h \rangle = 0, \tag{4.22}$$

$$\langle \mathbf{P}_h \mathbf{E}_{q,n}^i, \mathbf{v}_h \rangle - \langle P_h E_{u,n}^i, \nabla \cdot \mathbf{v}_h \rangle = 0 \tag{4.23}$$

for  $n = 1, \dots, N, i = 1, \dots, r$  and all  $\{w_h, \mathbf{v}_h\} \in W_h \times \mathbf{V}_h$ . Observing that for any  $n \geq 2$  the quantities  $\mathbf{E}_{q,n}^0$  and  $E_{u,n}^0$  are linear combinations of  $\mathbf{E}_{q,n-1}^i$  and  $E_{u,n-1}^i$ , for  $i = 0, \dots, r$ , respectively, and that  $P_h E_{u,1}^0 = 0$  and  $\mathbf{P}_h \mathbf{E}_{q,1}^0 = \mathbf{0}$  by definition of  $\{U_1^0, \mathbf{Q}_1^0\}$  and  $\{U_{n,h}^1, \mathbf{Q}_{n,h}^1\}$ , it follows that Eq. (4.23) is also satisfied for  $i = 0$  and any  $n \geq 1$ . Using this, we obtain by multiplying (4.23) with  $\hat{\alpha}_{ji}$  and summing up the resulting identity from  $i = 0$  to  $r$  that

$$\left\langle \sum_{j=0}^r \hat{\alpha}_{ij} \mathbf{P}_h \mathbf{E}_{q,n}^j, \mathbf{v}_h \right\rangle - \left\langle \sum_{j=0}^r \hat{\alpha}_{ij} P_h E_{u,n}^j, \nabla \cdot \mathbf{v}_h \right\rangle = 0 \tag{4.24}$$

for any  $\mathbf{v}_h \in \mathbf{V}_h$ . We note that we changed the notation for the indices. By testing now (4.22) with  $w_h = \sum_{j=0}^r \hat{\alpha}_{ij} P_h E_{u,n}^j \in W_h$  and (4.24) with  $\mathbf{v}_h = \tau_n \hat{\beta}_{ii} \mathbf{P}_h \mathbf{E}_{q,n}^i \in \mathbf{V}_h$ , we get by summing the resulting equations and using the inequalities of Cauchy–Schwarz and Cauchy–Young that

$$\begin{aligned} & \left\| \sum_{j=0}^r \hat{\alpha}_{ij} P_h E_{u,n}^j \right\|^2 + \tau_n \hat{\beta}_{ii} \left\langle \sum_{j=0}^r \hat{\alpha}_{ij} \mathbf{P}_h \mathbf{E}_{q,n}^j, \mathbf{P}_h \mathbf{E}_{q,n}^i \right\rangle \\ & = \tau_n \hat{\beta}_{ii} \left\langle \sum_{j=0}^r \hat{\alpha}_{ij} P_h E_{u,n}^j, \nabla \cdot (\mathbf{P}_h - \mathbf{\Pi}_h) \mathbf{E}_{q,n}^i \right\rangle \end{aligned}$$

$$\leq \frac{1}{2} \left\| \sum_{j=0}^r \hat{\alpha}_{ij} P_h E_{u,n}^j \right\|^2 + \frac{1}{2} \tau_n^2 \hat{\beta}_{ii}^2 \|\nabla \cdot (\mathbf{P}_h - \mathbf{\Pi}_h) \mathbf{E}_{q,n}^i\|^2$$

for  $n = 1, \dots, N$  and  $i = 1, \dots, r$ . The inequality above further simplifies to

$$\begin{aligned} & \left\| \sum_{j=0}^r \hat{\alpha}_{ij} P_h E_{u,n}^j \right\|^2 + 2\tau_n \hat{\beta}_{ii} \left\langle \sum_{j=0}^r \hat{\alpha}_{ij} P_h \mathbf{E}_{q,n}^j, P_h \mathbf{E}_{q,n}^i \right\rangle \\ & \leq \tau_n^2 \hat{\beta}_{ii}^2 \|\nabla \cdot (\mathbf{P}_h - \mathbf{\Pi}_h) \mathbf{E}_{q,n}^i\|^2, \end{aligned} \tag{4.25}$$

for  $n = 1, \dots, N, i = 1, \dots, r$ . Dividing (4.25) by  $\tau_n \hat{\beta}_{ii}$  (note that  $\hat{\beta}_{ii} > 0$  for all  $i = 1, \dots, r$ ), summing up the resulting inequality from  $i = 1, \dots, r$  and using Lemma 2.4 gives that

$$\begin{aligned} & \sum_{i=1}^r \frac{1}{\tau_n \hat{\beta}_{ii}} \left\| \sum_{j=0}^r \hat{\alpha}_{ij} P_h E_{u,n}^j \right\|^2 + \|P_h \mathbf{E}_q(t_n)\|^2 \leq \|P_h \mathbf{E}_q(t_{n-1})\|^2 \\ & + \sum_{i=1}^r \tau_n \hat{\beta}_{ii} \|\nabla \cdot (\mathbf{P}_h - \mathbf{\Pi}_h) \mathbf{E}_{q,n}^i\|^2 \end{aligned} \tag{4.26}$$

for  $n = 1, \dots, N$ . By summing up (4.26) from  $n = 1, \dots, K$  and noting that  $P_h \mathbf{E}_q(t_0) = P_h \mathbf{E}_{q,1}^0 = \mathbf{0}$  for the choices of the semidiscrete and fully discrete coefficient functions  $\mathbf{Q}_1^0$  and  $\mathbf{Q}_{1,h}^0$  [cf. their definition below (2.19), (2.20) and Eq. (2.30)] we get that

$$\begin{aligned} & \sum_{n=1}^K \sum_{i=1}^r \frac{1}{\tau_n \hat{\beta}_{ii}} \left\| \sum_{j=0}^r \hat{\alpha}_{ij} P_h E_{u,n}^j \right\|^2 + \|P_h \mathbf{E}_q(t_K)\|^2 \\ & \leq \sum_{n=1}^K \sum_{i=1}^r \tau_n \hat{\beta}_{ii} \|\nabla \cdot (\mathbf{P}_h - \mathbf{\Pi}_h) \mathbf{E}_{q,n}^i\|^2. \end{aligned} \tag{4.27}$$

We now estimate the divergence of the flux. By testing (4.22) with  $w_h = \nabla \cdot \mathbf{\Pi}_h \mathbf{E}_{q,n}^i \in W_h$ , and using the inequalities of Cauchy–Schwarz and Cauchy–Young ( $\hat{\beta}_{ii} > 0$  for all  $i = 1, \dots, r$ ) we get that

$$\begin{aligned} \tau_n \hat{\beta}_{ii} \|\nabla \cdot \mathbf{\Pi}_h \mathbf{E}_{q,n}^i\|^2 & = - \left\langle \sum_{j=0}^r \hat{\alpha}_{ij} P_h E_{u,n}^j, \nabla \cdot \mathbf{\Pi}_h \mathbf{E}_{q,n}^i \right\rangle \\ & \leq \frac{1}{2\tau_n \hat{\beta}_{ii}} \left\| \sum_{j=0}^r \hat{\alpha}_{ij} P_h E_{u,n}^j \right\|^2 + \frac{\tau_n \hat{\beta}_{ii}}{2} \|\nabla \cdot \mathbf{\Pi}_h \mathbf{E}_{q,n}^i\|^2 \end{aligned}$$

for  $n = 1, \dots, N, i = 1, \dots, r$ . Summing up the previous inequality from  $n = 1, \dots, K$  as well as from  $i = 1, \dots, r$ , using (4.27) along with  $(\mathbf{P}_h - \mathbf{\Pi}_h) \mathbf{E}_{q,n}^i = (\mathbf{P}_h - \mathbf{\Pi}_h) \mathbf{Q}_n^i$  by definition of the projectors  $\mathbf{P}_h$  and  $\mathbf{\Pi}_h$  we obtain that

$$\sum_{n=1}^K \sum_{i=1}^r \tau_n \hat{\beta}_{ii} \|\nabla \cdot \Pi_h \mathbf{E}_{\mathbf{q},n}^i\|^2 \leq \sum_{n=1}^K \sum_{i=1}^r \tau_n \hat{\beta}_{ii} \|\nabla \cdot (\mathbf{P}_h - \Pi_h) \mathbf{Q}_n^i\|^2,$$

which proves the assertions of the lemma. □

Now we combine the inequalities of the previous lemmas to estimate the error between the semidiscrete and the fully discrete solutions in the norms of  $L^2(I; W)$  and  $L^2(I; \mathbf{V})$ .

**Theorem 4.5** *Let the assumptions of Sect. 2.2 about  $\Omega, u_0, \mathbf{D}$  and  $f$  be satisfied. Let the sufficiently regular semidiscrete approximation  $\{u_\tau, \mathbf{q}_\tau\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(\mathbf{V})$  be defined by (2.18)–(2.20). Further, let  $\{u_{\tau,h}, \mathbf{q}_{\tau,h}\} \in \mathcal{X}^r(W_h) \times \mathcal{X}^r(\mathbf{V}_h)$  be the solution of the fully discrete problem (2.28), (2.29). For the scalar variable  $u_\tau$  it holds that*

$$\|u_\tau - u_{\tau,h}\|_{L^2(I;W)} \leq ch^{p+1}. \tag{4.28}$$

For the vectorial variable  $\mathbf{q}_\tau$  it holds that

$$\left( \sum_{n=1}^N \tau_n \sum_{i=1}^r \|\mathbf{q}_\tau(t_{n,i}) - \mathbf{q}_{\tau,h}(t_{n,i})\|^2 \right)^{1/2} \leq ch^{p+1}. \tag{4.29}$$

Further, for  $\mathbf{D}(\mathbf{x}) = d\mathbf{I}$ , for some  $d > 0$ , it holds that

$$\|\mathbf{q}_\tau - \mathbf{q}_{\tau,h}\|_{L^2(I;L^2(\Omega))} \leq ch^{p+1} \tag{4.30}$$

and

$$\left( \sum_{n=1}^N \tau_n \sum_{i=1}^r \|\mathbf{q}_\tau(t_{n,i}) - \mathbf{q}_{\tau,h}(t_{n,i})\|_{\mathbf{V}}^2 \right)^{1/2} \leq ch^{p+1}. \tag{4.31}$$

The constant  $c$  does not depend on the discretization parameters  $h$  and  $\tau$ .

*Proof* By using (2.25) and recalling that  $E_{u,n}^0 = E_u(t_{n-1})$  we find that

$$\|u_\tau - u_{\tau,h}\|_{L^2(I;W)}^2 \leq c \left( \sum_{n=1}^N \sum_{i=1}^r \tau_n \|E_{u,n}^i\|^2 + \sum_{n=1}^N \tau_n \|E_u(t_{n-1})\|^2 \right). \tag{4.32}$$

By inequality (4.11) we get for the first term on the right-hand side of (4.32) that

$$\begin{aligned} \sum_{n=1}^N \sum_{i=1}^r \tau_n \|E_{u,n}^i\|^2 + \sum_{n=1}^N \sum_{i=1}^r \tau_n \|E_{q,n}^i\|^2 &\leq c \left( \sum_{n=1}^N \sum_{i=1}^r \tau_n \|U_n^i - P_h U_n^i\|^2 \right. \\ &\left. + \tau_N \|u_\tau(t_N) - P_h u_\tau(t_N)\|^2 + \sum_{n=1}^N \sum_{i=1}^r \tau_n \|Q_n^i - \Pi_h Q_n^i\|^2 \right). \end{aligned} \tag{4.33}$$

Using (4.11) again, we find for the second term on the right-hand side of (4.32) that

$$\begin{aligned} \|E_u(t_K)\|^2 &\leq \|u_\tau(t_K) - P_h u_\tau(t_K)\|^2 \\ &+ c \sum_{n=1}^K \sum_{i=1}^r \left( \|U_n^i - P_h U_n^i\|^2 + \|Q_n^i - \Pi_h Q_n^i\|^2 \right) \end{aligned} \tag{4.34}$$

for  $K = 1, \dots, N$ . Combining now (4.32) with (4.33) and (4.34) and using the approximation properties (4.9)–(4.10) of the projection operators we then get that

$$\begin{aligned} \|u_\tau - u_{\tau,h}\|_{L^2(I;W)}^2 &\leq c \left( \sum_{n=1}^N \tau_n h^{2(p+1)} \sum_{i=0}^r \|U_n^i\|_{p+1}^2 \right. \\ &\left. + h^{2(p+1)} \max_{K=0,\dots,N} \|u_\tau(t_K)\|_{p+1}^2 + \sum_{n=1}^N \tau_n h^{2(p+1)} \sum_{i=0}^r \|Q_n^i\|_{p+1}^2 \right), \end{aligned} \tag{4.35}$$

where the arising constant does not depend on the discretization parameters  $h$  and  $\tau$ . The result (4.28) directly follows from (4.35) under the assumption of the theorem of sufficiently regular coefficient functions  $\{U_n^j, Q_n^j\} \in W \times V$ . From (4.33) along with (4.34) and (4.9)–(4.10) we further conclude that

$$\sum_{n=1}^N \tau_n \sum_{i=1}^r \|E_q(t_{n,i})\|^2 \leq ch^{2(p+1)}. \tag{4.36}$$

This proves (4.29).

By using (2.25), recalling that  $E_{q,n}^0 = E_q(t_{n-1})$  and applying the boundedness of the  $L^2$  projection operator  $P_h$  we find that

$$\begin{aligned} \|q_\tau - q_{\tau,h}\|_{L^2(I;L^2(\Omega))}^2 &\leq c \left( \|q_\tau - P_h q_\tau\|_{L^2(I;L^2(\Omega))}^2 + \|P_h q_\tau - q_{\tau,h}\|_{L^2(I;L^2(\Omega))}^2 \right) \\ &\leq c \left( \sum_{n=1}^N \sum_{i=1}^r \tau_n \|E_{q,n}^i\|^2 + \sum_{n=1}^N \tau_n \|P_h E_q(t_{n-1})\|^2 \right. \\ &\left. + \|q_\tau - P_h q_\tau\|_{L^2(I;L^2(\Omega))}^2 \right). \end{aligned} \tag{4.37}$$

For  $D(x) = dI$  the second term on the right-hand side of (4.37) can be bounded from above by means of the inequality (4.21) along with the observation that  $(P_h - \Pi_h)E_{q,n}^i = (P_h - \Pi_h)Q_n^i$  by definition of the projectors  $P_h$  and  $\Pi_h$ . Recalling further the boundedness of  $\beta_{ii}$  (cf. Lemma 2.3) we conclude that

$$\|P_h E_q(t_K)\|^2 \leq c \sum_{n=1}^K \sum_{i=1}^r \tau_n \|(P_h - \Pi_h)Q_n^i\|^2 \tag{4.38}$$

for  $K = 1, \dots, N$ . Finally, combining (4.37) with (4.33) and (4.38) and using the approximation properties (4.9)–(4.10) of the projection operators we then get that

$$\begin{aligned} & \|q_\tau - q_{\tau,h}\|_{L^2(I; L^2(\Omega))}^2 \\ & \leq c \left( \sum_{n=1}^N \tau_n h^{2(p+1)} \sum_{i=0}^r \|U_n^i\|_{p+1}^2 + h^{2(p+1)} \max_{K=0, \dots, N} \|u_\tau(t_K)\|_{p+1}^2 \right. \\ & \quad \left. + \sum_{n=1}^N \tau_n h^{2(p+1)} \sum_{i=0}^r \|Q_n^i\|_{p+1}^2 + h^{2(p+1)} \|q_\tau\|_{L^2(I; H^{p+1}(\Omega))}^2 \right), \end{aligned} \tag{4.39}$$

where the arising constant does not depend on the discretization parameters  $h$  and  $\tau$ . The result (4.30) directly follows from (4.39) under the assumption of sufficiently regular coefficient functions  $\{U_n^j, Q_n^j\} \in W \times V$ .

To estimate the divergence part of the error in (4.31), we use that by definition of the projection operators it holds that

$$\begin{aligned} & \sum_{n=1}^N \tau_n \sum_{i=1}^r \|\nabla \cdot E_q(t_{n,i})\|^2 \\ & \leq \sum_{n=1}^N \tau_n \sum_{i=1}^r \|\nabla \cdot (Q_n^i - \Pi_h Q_n^i)\|^2 + \sum_{n=1}^N \tau_n \sum_{i=1}^r \|\nabla \cdot \Pi_h E_{q,n}^i\|^2. \end{aligned} \tag{4.40}$$

The assertion (4.31) then follows from (4.40) combined with (4.21) and the approximation properties (4.9)–(4.10). □

We remark that the inequalities (4.30) and (4.31) provide an error control for the spatial discretization in the Gaussian quadrature points or temporal degrees of freedom of the subintervals  $I_n$  with respect to the norm of  $L^2(\Omega)$  and  $V$ , respectively. For an error control with respect to the norm of  $L^2(I; V)$  or  $L^2(I; V)$  a further estimate of  $E_{q,n}^0$  is required which remains an open problem.

### 4.3 Error estimates for the error between the continuous and the fully discrete solution

In this section we combine the results of Theorems 3.8 and 3.10 with the estimates of Theorem 4.5 to prove the convergence of the fully discrete scheme.

**Theorem 4.6** *Let the assumptions of Sect. 2.2 about  $\Omega$ ,  $u_0$ ,  $\mathbf{D}$  and  $f$  be satisfied. Let  $\{u, \mathbf{q}\} \in H^1(I; W) \times L^2(I; V)$  denote the unique solution of (2.10), (2.11) that is supposed to be sufficiently regular. Further, let  $\{u_{\tau, h}, \mathbf{q}_{\tau, h}\} \in \mathcal{X}^r(W_h) \times \mathcal{X}^r(V_h)$  be the uniquely defined solution of the fully discrete problem (2.28), (2.29), respectively. Suppose that the semidiscrete problem (2.19), (2.20) admits a sufficiently regular solution  $\{u_{\tau}, \mathbf{q}_{\tau}\} \in \mathcal{X}^r(W) \times \mathcal{X}^r(V)$ . Then, there holds that*

$$\|u - u_{\tau, h}\|_{L^2(I; L^2(\Omega))} \leq c(\tau^r + h^{p+1}). \quad (4.41)$$

For homogeneous diffusion coefficients  $\mathbf{D}(\mathbf{x}) = d\mathbf{I}$ , with some constant  $d > 0$ , there holds that

$$\|\mathbf{q} - \mathbf{q}_{\tau, h}\|_{L^2(I; L^2(\Omega))} \leq c(\tau^r + h^{p+1}). \quad (4.42)$$

Under the regularity condition ( $R_{\text{mix}}$ ) given in (3.25) and for interpolated right-hand side functions (2.21) there holds that

$$\|u - u_{\tau, h}\|_{L^2(I; L^2(\Omega))} \leq c(\tau^{r+1} + h^{p+1}). \quad (4.43)$$

The constant  $c$  in (4.41)–(4.43), respectively, does not depend on the discretization parameters  $h$  and  $\tau$ .

*Proof* By using the triangle inequality, Theorems 3.10 and 4.5 it follows that

$$\begin{aligned} \|u - u_{\tau, h}\|_{L^2(I; L^2(\Omega))}^2 &\leq 2\|u - u_{\tau}\|_{L^2(I; L^2(\Omega))}^2 + 2\|u_{\tau} - u_{\tau, h}\|_{L^2(I; L^2(\Omega))}^2 \\ &\leq c\left(\tau^{2r} + h^{2(p+1)}\right), \end{aligned}$$

where sufficient regularity of the continuous and semidiscrete solution with appropriate upper bounds for the solutions (cf. Theorems 3.10 and 4.5) is assumed. The inequality (4.42) is obtained similarly. The estimate (4.43) can be concluded in the same way by using now the result of Theorem 3.8 instead of Theorem 3.10.  $\square$

**Remark 4.7** • The error estimate (4.43) is optimal in time and space. The assumption of an interpolated right-hand side function (2.21) can still be dropped even though this is not explicitly done in this work. It requires to estimate the error between the exact form of  $c\text{GP}(r)$  defined in (2.14), (2.15) and the fully discrete solution. In this case the arguments used to prove Theorem 4.5 have to be augmented by an estimate of the interpolation error for the right-hand side function, similarly to the proof of Theorem 3.10.

- The error estimate (4.42) is suboptimal in time. It remains an open problem to analyze if the estimates can still be sharpened to order  $r + 1$ . In our numerical study presented in Sect. 5 convergence of order  $r + 1$  will be observed for the temporal



discretization of the scalar and the flux variable. Moreover, this is even observed in the (spatially) stronger norm of  $L^2(0, T; V)$  instead of  $L^2(0, T; L^2(\Omega))$  for the flux variable.

### 5 Numerical studies

In this section we present numerical studies in order to illustrate the error estimate given in Theorem 4.6 for the fully discrete scheme (2.28), (2.29) combining a variational time discretization with the MFEM. Moreover, we analyze the robustness of the convergence behaviour with respect to random perturbations of the meshes. Thereby we mimic mesh distributions of applications that are of practical interest. Additional convergence studies for variational space-time discretizations of the proposed type as well as for discontinuous time discretizations can be found in [15,37] for parabolic problems and in [36,37] for variational space-time discretizations of wave equations. In [15,37] the efficient iterative solution of the resulting algebraic system of Eqs. (2.28), (2.29) along with the construction of appropriate preconditioning techniques is carefully addressed. In the literature, further computational studies of variational time discretization schemes are presented also for different kind of flow and transport problems in, e.g., [1–3,30–32,38,46].

In order to determine the space-time convergence behavior we consider in our numerical study the cGP(2)–MFEM(2) approach. That is (2.14)–(2.15) with  $r = 2$  combined with the mixed finite element method MFEM(2) based on the choice  $p = 2$  in the definition (2.26) and (2.27) of the tuple of MFE spaces. We prescribe the solution

$$u_E(\mathbf{x}, t) := \sin(\omega t) \sin(\pi x_1) \sin(\pi x_2), \quad \text{in } \Omega \times (0, T),$$

with  $\Omega = (0, 1)^2$ ,  $\omega = 10\pi$  of problem (2.6)–(2.8). The corresponding flux function is then given by  $\mathbf{q}_E = -\mathbf{D}\nabla u_E$  for  $\mathbf{D} = \mathbf{I}$ . We choose the final time  $T = 1$ . On the coarsest level (level 0) the temporal mesh is uniformly refined into  $N = 10$  time subintervals and the corresponding spatial mesh consists of a single cell. In the following, we use the abbreviation

$$e_u^{cGP(2)}(t) := u_E(t) - u_{\tau,h}(t) \quad \text{and} \quad e_q^{cGP(2)}(t) := \mathbf{q}_E(t) - \mathbf{q}_{\tau,h}(t),$$

where we denote by  $u_{\tau,h}$  and by  $\mathbf{q}_{\tau,h}$  the fully discrete cGP(2)–MFEM(2) approximation of the primal variable and the flux field. The discretization errors for  $e_u^{cGP(2)}$  are measured in the  $L^2(I; L^2(\Omega))$ -norm and for  $e_q^{cGP(2)}$  in the  $L^2(I; V)$ -norm. As usual, the integral over the spatial domain  $\Omega$  and the integral over the time domain  $I = (0, T)$  in the error norms are evaluated elementwise in space and time by appropriate quadrature rules of sufficiently high order of accuracy.

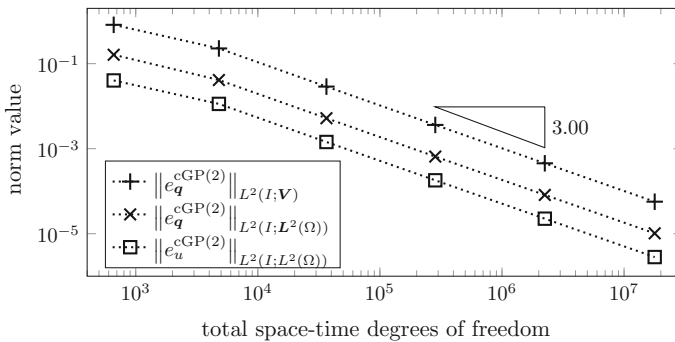
#### 5.1 Uniform meshes

To determine the experimental orders of convergence the space-time mesh is refined uniformly by a factor of two in each of the space dimensions and in the time dimension.

The characteristic mesh numbers are provided in Table 1. We summarize the calculated errors and their experimental order of convergence (EOC) for the proposed space-time discretization in Table 2 and further illustrate them in Fig. 1. The numerical results confirm the expected third order rate of convergence established in Theorem 4.6 (cf. also Remark 4.7) for the discretization in the space-time domain with polynomial order  $r = 2$  and  $p = 2$ , respectively, in the definition of the underlying finite element spaces. We note that the theoretically optimal order of convergence in time and space is obtained for the primal and the flux variable. Thus, the estimate (4.42) might be suboptimal with respect to the time discretization; cf. Remark 4.7. The estimate (4.43) is nicely confirmed by the presented numerical results. Further, we note that the optimal rate of convergence is obtained for the spatial discretization of the flux field in the norm of  $\mathbf{V}$ . In this point the family of Raviart–Thomas pairs of mixed finite elements is superior to the family of Brezzi–Douglas–Marini pairs of mixed finite elements (cf. [17]) for that the optimal order of convergence of the flux variable can be obtained only in the norm of  $L^2(\Omega)$ .

### 5.2 Distorted meshes

In the second part of the numerical convergence studies we approximate the same analytical solution as before but we use spatial meshes with randomly distorted interior



**Fig. 1** Calculated errors and experimental orders of convergence in space-time for cGP(2)–MFEM(2)

**Table 1** Space-time mesh with number of time subintervals  $N$ , global time discretization parameter  $\tau_n$ , number of cells  $|\mathcal{T}_h|$ , global space discretization parameter  $h$  and degrees of freedom  $N_{\text{DoF}}$  per degree of freedom in time

Level	$N$	$\tau_n$	$ \mathcal{T}_h $	$h$	$N_{\text{DoF}}$
0	10	1.000e-01	1	1.4142e-00	33
1	20	5.000e-02	4	7.0711e-01	120
2	40	2.500e-02	16	3.5355e-01	456
3	80	1.250e-02	64	1.7678e-01	1776
4	160	6.250e-03	256	8.8388e-02	7008
5	320	3.125e-03	1024	4.4194e-02	27,840

vertices. Precisely, in each of the computations we start on a coarse mesh and do uniform refinement steps by halvening the spatial mesh width. On the thus obtained finest mesh each of the interior vertices is distorted by a randomly chosen vector. The magnitude of the distortion vector is chosen randomly up to a given factor of relative length to the corresponding edge length. The characteristic numbers of the refinement levels are summarized in Table 3. The resulting distorted meshes are illustrated in Fig. 2 for the refinement level 3. The temporal mesh is chosen in the same way as in the first numerical experiment; cf. Table 1.

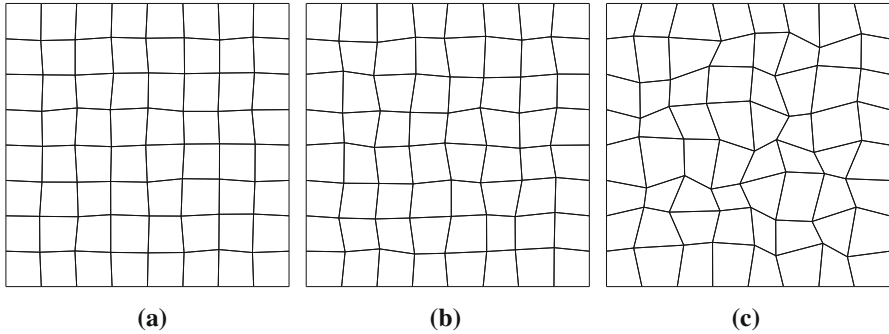
We summarize the calculated errors and the corresponding experimental order of convergence (EOC) for the proposed space-time discretization on the distorted spatial meshes in Table 4 for the scalar-valued primal variable and in Tables 5 and 6 for the vector-valued flux variable and further illustrate them in Fig. 3. Tables 5 and 6 differ by the norms in that the errors of the flux approximation are measured. In Tables 4 and 5 the expected order of convergence in space and time, for the primal variable measured in the norm of  $L^2(I; L^2(\Omega))$  and for the flux variable measured in the norm of  $L^2(I; L^2(\Omega))$ , is largely confirmed even for the strongly perturbed meshes with a distortion factor of 25%. This nicely demonstrates the robustness of the numerical scheme. Solely in Table 6 a slight reduction of the experimental order of convergence is observed depending on the degree of mesh perturbation. On the randomly distorted meshes the quasi uniformity condition, that is typically assumed about the finite element meshes in the numerical analyses, deteriorates successively. We conjec-

**Table 2** Norm values and corresponding experimental orders of convergence in space-time for cGP(2)–MFEM(2) on the refinement levels as given in Table 1

Level	$\ e_u^{cGP(2)}\ _{L^2(I; L^2(\Omega))}$	EOC	$\ e_q^{cGP(2)}\ _{L^2(I; V)}$	EOC
0	4.0298e-02	–	8.2000e-01	–
1	1.1316e-02	1.83	2.2827e-01	1.84
2	1.4371e-03	2.98	2.8876e-02	2.98
3	1.8037e-04	2.99	3.6208e-03	3.00
4	2.2569e-05	3.00	4.5295e-04	3.00
5	2.8219e-06	3.00	5.6631e-05	3.00

**Table 3** Distorted spatial mesh:  $h_{\max}$  largest cell diameter and  $h_{\text{red}}$  cell diameter reduction factor for 0%, 5%, 10% and 25% random vertex movement

Level	0%		5%		10%		25%	
	$h_{\max}$	$h_{\text{red}}$	$h_{\max}$	$h_{\text{red}}$	$h_{\max}$	$h_{\text{red}}$	$h_{\max}$	$h_{\text{red}}$
0	1.4142	–	1.4142	–	1.4142	–	1.4142	–
1	0.7071	2.00	0.7320	1.93	0.7570	1.87	0.8127	1.70
2	0.3536	2.00	0.3673	1.99	0.3814	1.98	0.4694	1.96
3	0.1768	2.00	0.1885	1.95	0.2002	1.91	0.2393	1.81
4	0.0884	2.00	0.0946	1.99	0.1008	1.99	0.1195	1.97
5	0.0442	2.00	0.0473	2.00	0.0504	2.00	0.0598	2.00



**Fig. 2** Distorted spatial meshes for 5% (a), 10% (b) and 25% (c) random vertex movement for refinement level 3

**Table 4** Calculated errors and corresponding experimental order of convergence for  $\|e_u^{cGP(2)}\|_{L^2(I;L^2(\Omega))}$  on distorted meshes given in Table 3

Level	5%	EOC	10%	EOC	25%	EOC
0	4.0298e-02	–	4.0298e-02	–	4.0298e-02	–
1	1.1353e-02	1.83	1.1463e-02	1.81	1.2155e-02	1.73
2	1.4381e-03	2.98	1.4783e-03	2.95	1.8690e-03	2.70
3	1.8290e-04	2.97	1.9117e-04	2.95	2.5232e-04	2.89
4	2.3024e-05	2.99	2.4402e-05	2.97	3.4463e-05	2.87
5	2.8787e-06	3.00	3.0483e-06	3.00	4.2870e-06	3.01

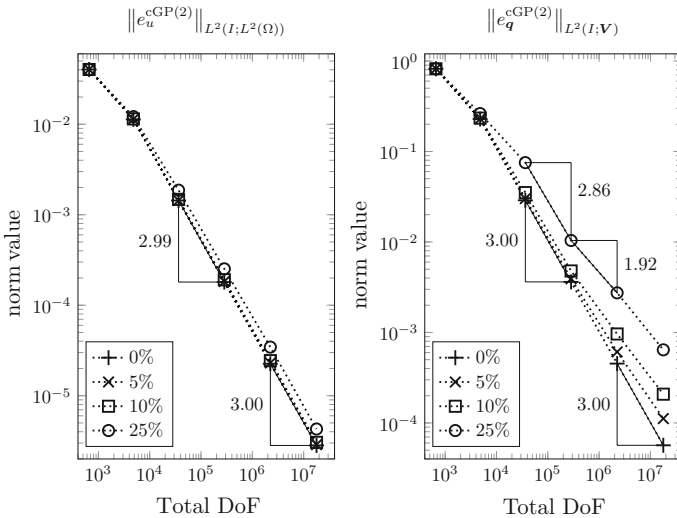
**Table 5** Calculated errors and corresponding experimental order of convergence for  $\|e_q^{cGP(2)}\|_{L^2(I;L^2(\Omega))}$  on distorted meshes given in Table 3

Level	5%	EOC	10%	EOC	25%	EOC
0	1.6206e-01	–	1.6206e-01	–	1.6206e-01	–
1	4.1287e-02	1.97	4.1460e-02	1.97	4.2889e-02	1.92
2	5.1846e-03	2.99	5.2156e-03	2.99	5.8069e-03	2.88
3	6.5995e-04	2.97	6.7916e-04	2.94	8.1938e-04	2.83
4	8.2527e-05	3.00	8.5013e-05	3.00	1.0344e-04	2.99
5	1.0310e-05	3.00	1.0628e-05	3.00	1.2982e-05	2.99

ture that this impacts the convergence behavior in the stronger  $L^2(I; V)$  norm more severely than in the  $L^2(I; L^2(\Omega))$  norm. The higher sensitivity of the derivatives in the  $L^2(I; V)$  norm with respect to the mesh perturbations seems to be quite natural. Nevertheless, we note that even though a strong random mesh perturbation is applied, a robust convergence behavior is still ensured and optimal order of convergence in the  $L^2(I; L^2(\Omega))$  norm is obtained. Finally, we note that the space-time convergence studies on the distorted spatial meshes were done with exactly the same numerical solver settings as for the above-given studies on uniform meshes.

**Table 6** Calculated errors and corresponding experimental order of convergence for  $\|e_q^{cGP(2)}\|_{L^2(I;V)}$  on distorted meshes given in Table 3

Level	5%	EOC	10%	EOC	25%	EOC
0	8.2000e-01	–	8.2000e-01	–	8.2000e-01	–
1	2.2964e-01	1.84	2.3376e-01	1.81	2.6241e-01	1.64
2	3.0172e-02	2.93	3.5136e-02	2.73	7.5321e-02	1.80
3	3.9244e-03	2.94	4.7838e-03	2.88	1.0391e-02	2.86
4	6.0990e-04	2.69	9.6250e-04	2.31	2.7408e-03	1.92
5	1.1175e-04	2.45	2.0797e-04	2.21	6.4287e-04	2.09



**Fig. 3** Calculated errors and corresponding experimental order of convergence on distorted meshes given in Table 3

### 6 Conclusions

In this work a numerical analysis of a family of variational space approximation schemes that combine continuous finite elements in time with the MFEM in space was presented for a parabolic prototype model of flow in porous media. The existence and uniqueness of the temporally semidiscrete and the fully discrete approximations were proved. Error estimates with explicit rates of convergence, including an optimal order error estimate, in natural norms of the scheme were established. The error estimates were illustrated and confirmed by numerical convergence studies. We believe that our analyses and techniques can be extended and applied to more sophisticated flow and transport processes in porous media or to incompressible viscous free flow. This will be our work for the future.

**Acknowledgements** This work was supported by the German Academic Exchange Service (DAAD) under the Grant IDs 56435737 and 57238185 and by the Research Council of Norway under the Grants DAADppp225267 and DAADppp255715. The authors wish to thank the anonymous reviewers for their help to improve the presentation of this paper.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## A Supplementary proofs

In the sequel we introduce a variational semidiscretization in time of the weak formulation of the second order problem (2.6)–(2.8), i.e. without rewriting Eq. (2.6) as a first order system of equations as it is done in Sect. 2.3. Then we prove the existence and uniqueness of solutions to the resulting semidiscrete variational problem. This results is used to establish the existence of the semidiscrete approximation in mixed form defined by the variational problem (2.14), (2.15) in Sect. 3. Here, we present a different technique of proof than in [46] since one of the arguments that is used [46, Lemma 6.1] does not hold in the applied form from our point of view. Thereby, we aim to keep our work self-contained. Further, we summarize the proof of Theorem 3.10.

### A.1 Variational time discretization of the second order problem

In the following we use the notation that is introduced in Sects. 2.1 and 2.3, respectively. Moreover we use the splitting [cf. Eq. (3.12)]

$$u_\tau(t) = u_0 + u_\tau^0(t) \quad \text{with} \quad u_\tau^0 \in \mathcal{X}_0^r(H_0^1(\Omega)).$$

Further, we put

$$f_0(t) = f(t) - Au_0$$

for  $u_0 \in H_0^1(\Omega)$  such that by definition  $Au_0 \in H^{-1}(\Omega)$ , cf. Sect. 2.1. Under the additional regularity condition that

$$D(A) = H^2(\Omega) \cap H_0^1(\Omega)$$

it even holds that  $f_0 \in L^2(I; L^2(\Omega))$ . The semidiscrete variational approximation of the system (2.6)–(2.8) is now defined by: Find  $u_\tau^0 \in \mathcal{X}_0^r(H_0^1(\Omega))$  such that

$$\int_0^T \langle \partial_t u_\tau^0, w_\tau \rangle dt + \int_0^T a(u_\tau^0, w_\tau) dt = \int_0^T \langle f_0, w_\tau \rangle dt \quad (\text{A.1})$$

for all  $w_\tau \in Y^{r-1}(H_0^1(\Omega))$ .

Firstly, we show the uniqueness of solutions to (A.1). In the sequel we denote by  $\varphi_{n,j} = \varphi_{n,j}(t)$  for  $j = 0, \dots, r$  the Lagrange basis functions in  $I_n = (t_{n-1}, t_n]$  with respect to  $r + 1$  quadrature points  $t_{n,l}, l = 0, \dots, r$ . Here, we choose the Gauss–Lobatto quadrature rule that is exact for polynomials of maximum degree  $2r - 1$ . In particular, for the quadrature nodes in  $\overline{I_n}$  it holds that  $t_{n,0} = t_{n-1}$  and  $t_{n,r} = t_n$ . Then, any function  $u_\tau^0 \in X_0^r(H_0^1(\Omega))$  and its time derivative admit the representation

$$u_\tau^0(t) = \sum_{j=0}^r U_n^j \varphi_{n,j}(t), \quad \partial_t u_\tau^0(t) = \sum_{j=0}^r U_n^j \varphi'_{n,j}(t) \tag{A.2}$$

for all  $t \in I_n$  with coefficient functions  $U_n^j \in H_0^1(\Omega)$  for  $j = 0, \dots, r$ .

**Theorem A.1** [Uniqueness of solutions to (A.1)] *Let the assumptions of Sect. 2.2 about  $\Omega, u_0$  and  $f$  be satisfied. Then the solution  $u_\tau^0 \in \mathcal{X}_0^r(H_0^1(\Omega))$  of the semidiscrete problem (A.1) is unique.*

*Proof* Let  $u_{\tau,1}^0, u_{\tau,2}^0 \in \mathcal{X}_0^r(H_0^1(\Omega))$  denote two solutions of the semidiscrete variational problem (A.1). We put  $u_\tau^0(t) := u_{\tau,1}^0 - u_{\tau,2}^0$ . We choose the test function  $w_\tau := A^{-1} \partial_t u_\tau^0 + \mu \partial_t u_\tau^0$  for some fixed parameter  $\mu \geq 0$ . By means of (A.2), it holds that  $w_\tau \in \mathcal{Y}^{r-1}(H_0^1(\Omega))$ . For this choice of  $w_\tau$  it follows that

$$I := \int_0^T \langle \partial_t u_\tau^0(t), A^{-1} \partial_t u_\tau^0 + \mu \partial_t u_\tau^0 \rangle dt + \int_0^T \langle Au_\tau^0(t), A^{-1} \partial_t u_\tau^0 + \mu \partial_t u_\tau^0 \rangle dt = 0. \tag{A.3}$$

By the symmetry of  $a(\cdot, \cdot)$  we have that  $a(u_\tau^0, \partial_t u_\tau^0) = \frac{1}{2} \frac{d}{dt} a(u_\tau^0, u_\tau^0)$ . Further, we have that  $\langle u_\tau^0, \partial_t u_\tau^0 \rangle = \frac{1}{2} \frac{d}{dt} \|u_\tau^0\|_{L^2(\Omega)}^2$ . Recalling (2.3)–(2.5) and noting that  $u_\tau^0(0) = 0$  and  $\partial_t u_\tau^0 \in \mathcal{Y}^{r-1}(H_0^1(\Omega))$ , it follows from (A.3) that

$$\begin{aligned} 0 = I &= \int_0^T \langle \partial_t u_\tau^0(t), A^{-1} \partial_t u_\tau^0 \rangle dt + \int_0^T \langle \partial_t u_\tau^0(t), \mu \partial_t u_\tau^0 \rangle dt \\ &+ \int_0^T \langle Au_\tau(t), A^{-1} \partial_t u_\tau^0 \rangle dt + \int_0^T \langle Au_\tau^0(t), \mu \partial_t u_\tau^0 \rangle dt \\ &\geq c \int_0^T \|\partial_t u_\tau^0\|_{H^{-1}(\Omega)}^2 dt + \mu \int_0^T \|\partial_t u_\tau^0\|_{L^2(\Omega)}^2 dt \\ &+ \int_0^T \frac{1}{2} \frac{d}{dt} \|u_\tau^0\|_{L^2(\Omega)}^2 dt + \mu \int_0^T \frac{1}{2} \frac{d}{dt} a(u_\tau^0, u_\tau^0) dt \\ &\geq c \int_0^T \|\partial_t u_\tau^0\|_{H^{-1}(\Omega)}^2 dt + \mu \int_0^T \|\partial_t u_\tau^0\|_{L^2(\Omega)}^2 dt \\ &+ \frac{1}{2} \|u_\tau^0(T)\|_{L^2(\Omega)}^2 + \frac{\alpha\mu}{2} \|u_\tau^0(T)\|_{H_0^1(\Omega)}^2. \end{aligned}$$

This implies that  $u_\tau^0 = 0$  and, consequently, that  $u_{\tau,1}^0 = u_{\tau,2}^0$ . The uniqueness of solutions to (A.1) is thus established.  $\square$

We remark that testing Eq. (A.1) with  $v_\tau = A^{-1}\partial_t u_\tau^0$  or  $v_\tau = \partial_\tau u_\tau^0$  would already be sufficient for proving the uniqueness result. Further, the symmetry of  $a(\cdot, \cdot)$  is essential in the previous proof. A generalization of the arguments to problems with nonsymmetric bilinear forms, for instance to convection–diffusion equations, still remains an open problem.

The existence of a solution to the semidiscrete problem (A.1) follows from the uniqueness of the solutions. Using the eigenspaces of  $A$ , problem (A.1) can be reduced to a set of finite dimensional problems, for each of which obviously uniqueness implies existence. For this we recall the following result from [26, Appendix D.6].

**Lemma A.2** *Let  $H$  be a separable Hilbert space, and suppose that  $S : H \mapsto H$  is a compact and symmetric operator. Then there exists a countable orthonormal basis of  $H$  consisting of eigenfunctions of  $S$ .*

**Theorem A.3** [Existence of solutions to (A.1)] *Let the assumptions of Sect. 2.2 about  $\Omega$ ,  $u_0$  and  $f$  as the be satisfied. Then the semidiscrete problem (A.1) admits a solution  $u_\tau^0 \in \mathcal{X}_0^r(H_0^1(\Omega))$ .*

*Proof* The operator  $S := A^{-1} : L^2(\Omega) \mapsto L^2(\Omega)$  with  $A$  being defined in (2.1) is a bounded, linear compact operator mapping  $L^2(\Omega)$  into itself. By means of Lemma A.2 there exists a set of appropriately scaled eigenfunctions  $\{w_k\}_{k=1}^\infty \subset L^2(\Omega)$  with  $w_k \in H_0^1(\Omega)$  such that  $\{w_k\}_{k=1}^\infty$  is an orthogonal basis of  $H_0^1(\Omega)$  and an orthonormal basis of  $L^2(\Omega)$ .

In terms of these eigenfunctions  $\{w_k\}_{k=1}^\infty \subset H_0^1(\Omega)$  the solution  $u_\tau^0$  of problem (A.1) can be represented as

$$u_\tau(x, t) = \sum_{j=0}^r U_n^{(j)}(x)\varphi_n^{(j)}(t) = \sum_{j=0}^r \sum_{k=1}^\infty d_{n,k}^{(j)} w_k(x)\varphi_n^{(j)}(t), \quad \text{for } t \in \bar{I}_n$$

with coefficients  $d_{n,k}^{(j)} \in \mathbb{R}$  for  $k = 1, \dots, \infty$  and each  $j = 0, \dots, r$  and  $n = 0, \dots, N$ . We choose test functions  $v_\tau \in \mathcal{Y}^{r-1}(H_0^1(\Omega))$  being defined by

$$v_\tau = \begin{cases} w_k \psi_n^{(i)}, & \text{for } t \in \bar{I}_n, \\ 0, & \text{for } t \in I \setminus \bar{I}_n \end{cases}$$

for  $i = 1, \dots, r, k = 1, \dots, \infty$  and  $n = 1, \dots, N$ . Then, for each  $k = 1, \dots, \infty$ , we get the finite dimensional problem

$$\begin{aligned} & \sum_{j=0}^r d_{n,k}^{(j)} \underbrace{\int_{I_n} d_t \varphi_n^{(j)}(t) \psi_n^{(i)}(t) dt}_{:=\alpha_{ij}} + \sum_{j=0}^r d_{n,k}^{(j)} \underbrace{a(w_k, w_k)}_{:=\gamma_k} \underbrace{\int_{I_n} \varphi_n^{(j)}(t) \psi_n^{(i)}(t) dt}_{:=\beta_{ij}} \\ & = \underbrace{\int_0^T \langle f_0(t), w_k \rangle \psi_n^{(i)}(t) dt}_{:=b_{k,i}} \end{aligned} \tag{A.4}$$



for  $i = 1, \dots, r$  and  $n = 1, \dots, N$ . Due to the continuity of functions  $u_\tau \in \mathcal{X}_0^r(H_0^1(\Omega))$  and the choice of the Gauss–Lobatto quadrature rule it holds that

$$U_n^{(0)} = U_{n-1}^{(r)} \quad \text{or} \quad d_{n,k}^{(0)} = d_{n-1,k}^{(r)},$$

respectively. Therefore, we recast the finite dimensional problem (A.4) as

$$\sum_{j=1}^r (\alpha_{ij} + \gamma_k \beta_{ij}) d_{n,k}^{(j)} = b_{k,i} + (\alpha_{i0} + \gamma_k \beta_{i0}) d_{n,k}^{(0)} \tag{A.5}$$

for  $i = 1, \dots, r$  and each  $k = 1, \dots, \infty$ . For the finite dimensional problem (A.5) the uniqueness of the solution established in Theorem A.1 then implies the existence of a solution.  $\square$

Similarly to Corollary 3.4, the existence and uniqueness of the semidiscrete solution implies that an inf–sup stability condition in the underlying space–time framework is satisfied [24, p. 85, Thm. 2.6].

**Corollary A.4** *Let the assumptions of Sect. 2.2 about  $\Omega$ ,  $u_0$ ,  $\mathbf{D}$  and  $f$  be satisfied. Let  $u_\tau^0 \in \mathcal{X}_0^r(H_0^1(\Omega))$  be the unique solution of the semidiscrete problem (A.1) according to Theorems A.1 and A.3. Then, there exists a constant  $c > 0$  such that*

$$\inf_{u_\tau \in \mathcal{X}_0^r(H_0^1(\Omega)) \setminus \{0\}} \sup_{v_\tau \in \mathcal{Y}^{r-1}(H_0^1(\Omega)) \setminus \{0\}} \frac{B(u_\tau, v_\tau)}{\|u_\tau\|_{\mathcal{X}} \|v_\tau\|_{\mathcal{Y}}} \geq c > 0$$

with

$$\begin{aligned} B(u_\tau, v_\tau) &= \int_0^T \langle \partial_t u_\tau + Au_\tau, v_\tau \rangle dt, \\ \|u_\tau\|_{\mathcal{X}} &= \left( \|u_\tau\|_{L^2(I, H_0^1(\Omega))}^2 + \|\partial_t u_\tau\|_{L^2(I, H^{-1}(\Omega))}^2 \right)^{1/2}, \\ \|v_\tau\|_{\mathcal{Y}} &= \|u_\tau\|_{L^2(I, H_0^1(\Omega))}. \end{aligned} \tag{A.6}$$

*Remark A.5* By the arguments of [46, Thm. 6.2] the inf–sup stability condition implies an error estimate for the semidiscretization (A.1) where the error is measured in the corresponding natural norm (A.6) of the scheme.

**A.2 Proof of Theorem 3.10**

*Proof* We let  $w_\tau := I_\tau u - u_\tau$ ,  $\mathbf{v}_\tau = \mathbf{J}_\tau \mathbf{q} - \mathbf{q}_\tau$  with the interpolation operators  $I_\tau$  and  $\mathbf{J}_\tau$  of Sect. 3.2. Using the inf–sup stability condition along with problem (2.10), (2.11) and the non-exact semidiscrete problem (3.39), (3.40), applying the inequality of Cauchy–Schwarz and the continuity (2.13) of  $a_\tau(\cdot, \cdot)$  we conclude that

$$\alpha \|\{w_\tau, \mathbf{v}_\tau\}\|_{\mathcal{V}} \|\{\varphi_\tau, \boldsymbol{\psi}_\tau\}\|_{\mathcal{V}} \leq a_\tau(\{w_\tau, \mathbf{v}_\tau\}, \{\varphi_\tau, \boldsymbol{\psi}_\tau\})$$

$$\begin{aligned}
 &= a_\tau \left( \{u^0, \mathbf{q}\}, \{\varphi_\tau, \boldsymbol{\psi}_\tau\} \right) - a_\tau \left( \{u_\tau^0, \mathbf{q}_\tau\}, \{\varphi_\tau, \boldsymbol{\psi}_\tau\} \right) \\
 &\quad - a_\tau \left( \{u^0 - I_\tau u_\tau^0, \mathbf{q} - \mathbf{J}_\tau \mathbf{q}\}, \{\varphi_\tau, \boldsymbol{\psi}_\tau\} \right) \\
 &= \int_0^T \langle f(t) - \Pi_r f(t), \varphi_\tau \rangle dt - a_\tau \left( \{u^0 - I_\tau u_\tau^0, \mathbf{q} - \mathbf{J}_\tau \mathbf{q}\}, \{\varphi_\tau, \boldsymbol{\psi}_\tau\} \right) \\
 &\leq \|f - \Pi_r f\|_{L^2(I; W)} \|\{\varphi_\tau, \boldsymbol{\psi}_\tau\}\|_{\mathcal{V}} + c \|\{u^0 - I_\tau u_\tau^0, \mathbf{q} - \mathbf{J}_\tau \mathbf{q}\}\|_{\mathcal{W}} \|\{\varphi_\tau, \boldsymbol{\psi}_\tau\}\|_{\mathcal{V}} \\
 &\leq \left( \sum_{n=1}^N \left\{ \|f - \Pi_r f\|_{L^2(I; W)}^2 + c \|u^0 - I_\tau u_\tau^0\|_{L^2(I_n; W)}^2 \right. \right. \\
 &\quad \left. \left. + c \|\partial_t(u^0 - I_\tau u_\tau^0)\|_{L^2(I_n; W)}^2 + c \|\mathbf{q} - \mathbf{J}_\tau \mathbf{q}\|_{L^2(I_n; V)}^2 \right\} \right)^{1/2} \cdot \|\{\varphi_\tau, \boldsymbol{\psi}_\tau\}\|_{\mathcal{V}}.
 \end{aligned}$$

By means of the approximation properties (3.16) to (3.18) we then get that

$$\begin{aligned}
 \|\{w_\tau, \mathbf{v}_\tau\}\|_{\mathcal{W}} &\leq c \left( \sum_{n=1}^N \tau_n^{2r} \left\{ \|\partial_t^{r+1} u\|_{L^2(I_n; W)}^2 \right. \right. \\
 &\quad \left. \left. + \tau_n^2 \|\partial_t^{r+1} \mathbf{q}\|_{L^2(I_n; V)}^2 + \tau_n^2 \|\partial_t^{r+1} f\|_{L^2(I_n; W)}^2 \right\} \right)^{1/2}.
 \end{aligned}$$

Combining this estimate with the triangle inequality yields the assertion of Theorem 3.10. □

## B Summary of notation

### Function spaces and norms

$I, \Omega$	Time and space domain, $I = (0, T]$
$L^2(\Omega), H^p(\Omega), H_0^1(\Omega), W^{1,\infty}(\Omega)$	Standard Sobolev spaces
$H^{-1}(\Omega)$	Dual space of $H_0^1(\Omega)$
$\ \cdot\ , \ \cdot\ _p, \langle \cdot, \cdot \rangle$	Norms in $L^2(\Omega), H^p(\Omega)$ , inner product in $L^2(\Omega)$
$W, V$	$W = L^2(\Omega), V = \mathbf{H}(\text{div}; \Omega)$
$\ \cdot\ _V$	$\ \cdot\ _V = (\ \cdot\ ^2 + \ \nabla \cdot (\cdot)\ ^2)^{1/2}$
$C(\bar{I}; X), L^2(I; X), H^1(I; X)$	Bogner spaces with values in the Banach space $X$
$D, u_0, f$	Data of the problem
$u, \mathbf{q}$	Solution and flux of the continuous problem
$a(u, v)$	Bilinear form $a(u, v) = \langle \mathbf{D}u, v \rangle$

**Time discretization**

$\mathcal{X}^r(W), \mathcal{X}^r(V)$	Trial spaces of semidiscretization in time
$\mathcal{X}_0^r(W)$	Trial space of semidiscretization with $u_\tau(0) = 0$
$\mathcal{Y}^{r-1}(W), \mathcal{Y}^r(V)$	Test spaces of semidiscretization in time
$\mathbb{P}_r(J; X)$	$\mathbb{P}_r(J; X) = \{p(t) = \sum_{j=0}^r \xi_n^j t^j \mid \xi_n^j \in X\}$
$\mathcal{W}, \mathcal{V}$	$\mathcal{W} = X_0^r(W) \times X^r(V), \mathcal{V} = Y^{r-1}(W) \times Y^{r-1}(V)$
$I_n, \tau_n$	Subinterval $I_n = (t_{n-1}, t_n], \tau_n = t_n - t_{n-1}$
$t_{n,0}$	$t_{n,0} = t_{n-1}$
$t_{n,1}, \dots, t_{n,r}$	$r$ -point Gaussian quadrature nodes
$\varphi_{n,j}(t)$	Lagrange polynomial on $I_n$ w.r.t $t_{n,0}, \dots, t_{n,r}$
$\psi_{n,i}(t)$	Lagrange polynomial on $I_n$ w.r.t $t_{n,1}, \dots, t_{n,r}$
$u_\tau$	$u_\tau _{\bar{I}_n}(t) = \sum_{j=0}^r U_n^j \varphi_{n,j}(t)$
$q_\tau$	Semidiscrete solution
$\hat{I}$	$q_\tau _{\bar{I}_n}(t) = \sum_{j=0}^r Q_n^j \varphi_{n,j}(t)$
$\hat{t}_1, \dots, \hat{t}_r$	Semidiscrete flux
$\hat{\omega}_1, \dots, \hat{\omega}_r$	$\hat{I} = [0, 1]$ reference interval
$\hat{\varphi}_j(\hat{t})$	Gaussian quadrature nodes on $\hat{I}$
$\hat{\alpha}_{ij}, \hat{\beta}_{ij}$	Gaussian quadrature weights on $\hat{I}$
$\Pi_\tau$	Transformed Lagrange polynomial on $\hat{I}$
$I_\tau$	$\hat{\alpha}_{ij} = \hat{\omega}_i, \hat{\beta}_{ij} = \hat{\omega}_i \delta_{i,j}$ with Kronecker symbol $\delta_{i,j}$
$J_\tau$	From transformation of time integrals to $\hat{I}$ and Application of quadrature on $\hat{I}$
$a_\tau(\{\cdot, \cdot\}, \{\cdot, \cdot\})$	Temporal Lagrange interpolant
$u^0(t), u_\tau^0(t)$	W.r.t. to $t_{n,0}, \dots, t_{n,r}$
$z, p$	Temporal interpolation operator for variable $u$
$I_0$	Temporal interpolation operator for variable $q$
$J_0$	Space-time bilinear form
	$u^0(t) = u(t) - u_0, u_\tau^0(t) = u_\tau(t) - u_0$
	Solution of dual problem
	Temporal interpolation operator for $z$
	Temporal interpolation operator for $p$

**Space discretization and error analysis**

$W_h, V_h$	Inf-sup stable pair of finite element spaces
$\mathcal{X}^r(W_h), \mathcal{X}^r(V_h)$	Raviart–Thomas(–Nédélec) elements
$\mathcal{Y}^{r-1}(W_h), \mathcal{Y}^{r-1}(V_h)$	Trial spaces of space-time discretization
$u_{\tau,h}$	Test spaces of space-time discretization
$q_{\tau,h}$	$u_{\tau,h}(t) _{I_n} = \sum_{j=0}^r U_{n,h}^j \varphi_{n,j}(t)$
$P_h, P_h$	Fully discrete solution
$\Pi_h$	$q_{\tau,h}(t) _{I_n} = \sum_{j=0}^r Q_{n,h}^j \varphi_{n,j}(t)$
	Fully discrete flux
	$L^2$ projection onto $W_h$ and $V_h$
	Projection onto $V_h$ :
	$\langle \nabla \cdot (\Pi_h v - v), w_h \rangle = 0$ for all $w_h \in W_h$

$E_u(t)$	$E_u(t) = u_\tau(t) - u_{\tau,h}(t)$ $E_u(t) = \sum_{j=0}^r E_{u,n}^j \varphi_{n,j}(t)$ Error between semidiscrete and fully discrete Approximation of scalar variable
$E_q(t)$	$E_q(t) = q_\tau(t) - q_{\tau,h}(t)$ $E_q(t) = \sum_{j=0}^r E_{q,n}^j \varphi_{n,j}(t)$ Error between semidiscrete and fully discrete Approximation of flux variable
$E_{u,n}^i$	$E_{u,n}^i = E_u(t_{n,i})$ : error in node $t_{n,i}$ , $i = 0, \dots, r$
$E_{q,n}^i$	$E_{q,n}^i = E_q(t_{n,i})$ : error in node $t_{n,i}$ , $i = 0, \dots, r$

## References

1. Ahmed, N., John, V.: Adaptive time step control for higher order variational time discretizations applied to convection–diffusion–reaction equations. *Comput. Methods Appl. Mech. Eng.* **285**, 83–101 (2015)
2. Ahmed, N., Matthies, G.: Numerical studies of Galerkin-type time-discretizations applied to transient convection–diffusion–reaction equations. *World Acad. Sci. Eng. Technol.* **66**, 586–593 (2012)
3. Ahmed, N., Matthies, G.: Numerical studies of variational-type time-discretization techniques for transient Oseen problem. In: *Proceedings of Contributed Papers and Posters*, pp. 404–415. Slovak University of Technology, Faculty of Civil Engineering, Department of Mathematics and Descriptive Geometry, Bratislava. ISBN 978-80-227-3742-5/pbk (2012)
4. Andreev, R.: Space-time discretization of the heat equation. *Numer. Algorithm* **67**, 713–731 (2014)
5. Arbogast, T., Wheeler, M.F., Zhang, N.Y.: A nonlinear mixed finite element method for a degenerate parabolic equation arising in flow in porous media. *SIAM J. Numer. Anal.* **33**, 1669–1687 (1996)
6. Aziz, A.K., Monk, P.: Continuous finite elements in space and time for the heat equation. *Math. Comput.* **52**, 255–274 (1989)
7. Bangerth, W., Geiger, M., Rannacher, R.: Adaptive Galerkin finite element methods for the wave equation. *Comput. Methods Appl. Math.* **10**(1), 3–48 (2010)
8. Bangerth, W., Heister, T., Kanschat, G.: *deal.II Differential Equations Analysis Library*, Technical reference. <http://www.dealii.org> (2014)
9. Bangerth, W., Rannacher, R.: *Adaptive Finite Element Methods for Differential Equations*. Birkhäuser, Basel (2003)
10. Basting, S., Weller, S.: Efficient preconditioning of variational time discretization methods for parabolic partial differential equations. *Math. Model. Numer. Anal.* **49**(2), 331–347 (2015)
11. Bause, M.: Higher and lowest order mixed finite element methods for subsurface flow problems with solutions of weak regularity. *Adv. Water Res.* **31**, 370–382 (2008)
12. Bause, M., Brunner, F., Knabner, P., Radu, F.: An improved optimal order mixed finite element method for semilinear transport problems. In: Cangiani, A., et al. (eds.) *Numerical Mathematics and Advanced Applications 2011*. pp. 247–256 Springer, Berlin (2013)
13. Bause, M., Hoffmann, J., Knabner, P.: First-order convergence of multi point flux approximation on triangular grids and comparison with mixed finite element methods. *Numer. Math.* **116**, 1–29 (2010)
14. Bause, M., Knabner, P.: Computation of variably saturated subsurface flow by adaptive mixed hybrid finite element methods. *Adv. Water Res.* **27**, 565–581 (2004)
15. Bause, M., Köcher, U.: Variational time discretization for mixed finite element approximations of nonstationary diffusion problems. *J. Comput. Appl. Math.* **289**, 208–224 (2015)
16. Bernardi, C., Maday, Y.: *Approximations spectrales des problèmes aux limites elliptique*. Springer, Paris (1992)
17. Brezzi, F., Fortin, M.: *Mixed and Hybrid Finite Element Methods*. Springer, New York (1991)
18. Brunner, F., Radu, F.A., Bause, M., Knabner, P.: Optimal order convergence of a modified BDM<sub>1</sub> mixed finite element scheme for reactive transport in porous media. *Adv. Water Res.* **35**, 163–171 (2012)
19. Celia, M.A., Bouloutas, E.T., Zarba, R.L.: A general mass-conservative numerical solution for the unsaturated flow equation. *Water Resour. Res.* **26**(7), 1483–1496 (1990)

20. Česenek, J., Feistauer, M.: Theory of the space-time discontinuous Galerkin method for nonstationary parabolic problems with nonlinear convection and diffusion. *SIAM J. Numer. Anal.* **50**(3), 1181–1206 (2012)
21. Chen, Z.: *Finite Element Methods and Their Applications*. Springer, Berlin (2010)
22. Chen, Z., Huan, G., Ma, Y.: *Computational Methods for Multiphase Flows in Porous Media*. SIAM, Philadelphia (2005)
23. Cristina, M., Squeff, J.: Superconvergence of mixed finite element methods for parabolic equations. *RAIRO Modél. Anal. Numér.* **21**, 327–352 (1987)
24. Ern, A., Guermond, J.L.: *Theory and Practice of Finite Elements*. Springer, Berlin (2010)
25. Ern, A., Schieweck, F.: Discontinuous Galerkin method in time combined with an stabilized finite element method in space for linear first-order PDEs. *Math. Comput.* **85**, 2099–2129 (2016)
26. Evans, L.C.: *Partial Differential Equations*. American Mathematical Society, Providence (2010)
27. Farthing, M.W., Kees, C.E., Miller, C.T.: Mixed finite element methods and higher-order temporal approximations. *Adv. Water Resour.* **25**, 85–101 (2002)
28. Grisvard, P.: *Elliptic Problems in Nonsmooth Domains*. Pitman, Boston (1985)
29. Helmig, R.: *Multiphase Flow and Transport Processes in the Subsurface: A Contribution to the Modeling of Hydrosystems*. Springer, Berlin (1997)
30. Hussain, S., Schieweck, F., Turek, S.: Higher order Galerkin time discretization for nonstationary incompressible flow. In: Cangiani, A., et al. (eds.) *Numerical Mathematics and Advanced Applications 2011*, pp. 509–517. Springer, Berlin (2013)
31. Hussain, S., Schieweck, F., Turek, S.: A note on accurate and efficient higher order Galerkin time stepping schemes for nonstationary Stokes equations. *Open Numer. Methods J.* **4**, 35–45 (2012)
32. Hussain, S., Schieweck, F., Turek, S.: Higher order Galerkin time discretizations and fast multigrid solvers for the heat equation. *J. Numer. Math.* **19**(1), 41–61 (2011)
33. Johnson, C., Thomée, V.: Error estimates for some mixed finite element methods for parabolic problems. *RAIRO Modél. Anal. Numér.* **15**, 71–78 (1981)
34. Karakashin, O., Makridakis, C.: Convergence of a continuous Galerkin method with mesh modification for nonlinear wave equations. *Math. Comput.* **74**, 85–102 (2004)
35. Karakashin, O., Makridakis, C.: A space-time finite element method for the nonlinear Schrödinger equation: the continuous Galerkin method. *SIAM J. Numer. Anal.* **36**, 1779–1807 (1999)
36. Köcher, U., Bause, M.: Variational space-time methods for the wave equation. *J. Sci. Comput.* **61**, 424–453 (2014)
37. Köcher, U.: *Variational Space-Time Methods for the Elastic Wave Equation and the Diffusion Equation*, Ph.D. Thesis, Helmut Schmidt University Hamburg, submitted, Hamburg (2015)
38. Matthies, G., Schieweck, F.: Higher Order Variational Time Discretizations for Nonlinear Systems of Ordinary Differential equations. Preprint No. 23/2011 Otto von Guericke Universität Magdeburg, pp. 1–30 (2011)
39. Maz'ya, V., Nazarov, S., Plamenevskij, B.: *Asymptotic Theory of Elliptic Boundary Value Problems in Singularly Perturbed Domains*. Birkhäuser, Basel (2000)
40. Quarteroni, A., Valli, A.: *Numerical Approximation of Partial Differential Equations*. Springer, Heidelberg (1994)
41. Quarteroni, A., Sacco, R., Saleri, F.: *Numerical Mathematics*. Springer, Berlin (2007)
42. Radu, F.A., Pop, I.S., Attinger, S.: Analysis of an Euler implicit-mixed finite element scheme for reactive solute transport in porous media. *Numer. Methods Partial Differ. Equ.* **26**, 320–344 (2010)
43. Radu, F.A., Pop, I.S., Knabner, P.: Order of convergence estimates for an Euler implicit, mixed finite element discretization of Richards' equation. *SIAM J. Numer. Anal.* **42**, 1452–1478 (2004)
44. Radu, F.A., Pop, I.S., Knabner, P.: Error estimates for a mixed finite element discretization of some degenerate parabolic equations. *Numer. Math.* **109**, 285–311 (2008)
45. Radu, F.A., Suci, N., Hoffmann, J., Vogel, A., Kolditz, O., Park, C.-H., Attinger, S.: Accuracy of numerical simulations of contaminant transport in heterogeneous aquifers: a comparative study. *Adv. Water Res.* **34**, 47–61 (2011)
46. Schieweck, F.: A-stable discontinuous Galerkin–Petrov time discretization of higher order. *J. Numer. Math.* **18**(1), 25–57 (2010)
47. Sudirham, J.J., van der Vegt, J.J.W., van Damme, R.M.J.: Space-time discontinuous Galerkin method for advection–diffusion problems on time-dependent domains. *Appl. Numer. Math.* **56**, 1491–1518 (2006)

48. Thomeé, V.: Galerkin Finite Element Methods for Parabolic Problems. Springer, Berlin (2006)
49. Woodward, C.S., Dawson, C.N.: Analysis of expanded mixed finite element methods for a nonlinear parabolic equation modeling flow into variably saturated porous media. *SIAM J. Numer. Anal.* **37**, 701–724 (2000)