A stable discontinuous Galerkin-Petrov time discretization of higher order ^{*}

F. Schieweck [†]

September 2009

Abstract

We construct and analyze a discontinuous Galerkin-Petrov time discretization of a general evolution equation in a Hilbert space. The method is A-stable and exhibits an energy decreasing property. The approach consists in a continuous solution space and a discontinuous test space such that the time derivative of the discrete solution is contained in the test space. This is the key to get stability. We prove A-stability and optimal error estimates. Numerical results confirm the theoretical results.

Keywords: discontinuous finite elements, Galerkin-Petrov method, stability and error estimates

2000 Mathematics Subject Classification (MSC): 65M12, 65M15, 65M60

1 Introduction

A well-known approach to solve time dependent problems is the Galerkin method, see for instance the monograph [5]. In order to obtain a time marching process for this discretization a discontinuous Galerkin method is used. Typically, the discrete solution space consists of piecewise polynomials in time which are discontinuous at the boundary of the time intervals. Therefore, some jump terms appear in the discretization. However, there is no need to use a discontinuous discrete solution space. A Galerkin method leads also to a time marching process if only the test space consists of discontinuous piecewise polynomial functions. Then, the discrete solution and test spaces are different and the method called a *Galerkin-Petrov discretization*.

^{*}appears as Preprint No. 24/2009, Fakultät für Mathematik, Otto-von-Guericke Universität Magdeburg

[†]Institut für Analysis und Numerik, Otto-von-Guericke-Universität Magdeburg, Postfach 4120, D-39016 Magdeburg, Germany, URL: <http://www-ian.math.uni-magdeburg.de/home/schieweck>

In this paper, we study this approach which we call *discontinuous Galerkin-Petrov discretization* (dGP(k)-method). Here the discrete solution space consists of continuous piecewise polynomial functions of degree $k \geq 1$ and the discrete test space of discontinuous piecewise polynomial functions of degree $k - 1$. An advantage of this discretization is the *energy decreasing property* which was the main motivation for us to develop this method in the context of the numerical solution of the nonlinear Willmore flow problem [2]. If we want to minimize an energy functional $W(u)$ for functions u in a Hilbert space V we are looking at stationary points where the derivative of $W(\cdot)$ vanishes, i.e., $W'(u) = 0$. To compute such stationary points we can try to solve the gradient flow equation $d_t u(t) = -W'(u(t))$ and to find some stationary limits. Now, the variational discretization in the continuous case has the energy decreasing property that $W(u(t_2)) < W(u(t_1))$ for $t_2 > t_1$ unless $u(t_1)$ is a stationary point. This is proved by taking the time derivative of the solution as the test function. The key property of our discontinuous Galerkin-Petrov discretization is that also the time derivative of the discrete solution is contained in the discrete test space. Therefore, the energy decreasing property is maintained also for the discrete method. From the physical point of view, this can be regarded also as a kind of stability of the discrete method.

Looking in the literature we found that our approach has already been used by Aziz and Monk [1] for the linear heat equation. They have proved optimal error estimates as well as superconvergence results. However, there are some differences to our method. Whereas a Gauß-Legendre integration has been used in [1], we apply a $(k + 1)$ -point Gauß-Lobatto formula which decouples the resulting system, that has to be solved on each time interval, to a large extend. In general, the dGP(k)-method leads on each time interval to a coupled nonlinear system for k "unknowns" in a Hilbert space. One goal of this paper is to choose suitably the basis functions of the solution and test space and the numerical integration rule in order to uncouple this system as much as possible. For example, in the case of the dGP(2)-method, the system solver can be reduced to just a fixed point iteration in the second "unknown" where the first "unknown" is easily computed during the iteration by an explicit formula. Thus, we have computational costs that are not much more than for the Crank-Nicolson scheme but we get third order accuracy over the whole time interval and, due to superconvergence, even fourth order in the time mesh points. The dGP(1)-method reproduces exactly the Crank-Nicolson scheme.

Another new contribution of this paper is that we present the method in an abstract setting of a general evolution equation in a Hilbert space. In the analysis, we treat two important special cases - a linear parabolic equation in a Hilbert space and a d -dimensional nonlinear system of ordinary differential equations (ODE). For example, our results in the linear parabolic case include the time-dependent convection diffusion reaction problem. We

have proved an optimal error estimate of order k for the exact and the numerically integrated non-exact dGP(k)-method in a natural norm that contains also the time derivative of the solution. In the nonlinear Lipschitz continuous case, we have shown an optimal order of $k + 1$ in the maximum norm.

Moreover, we prove by some energy arguments that the dGP(k)-method is A-stable. Finally, we present some numerical results for a stiff linear 2×2 ODE system which show that the method is stable and that it has even superconvergence properties.

In the following, we denote by C a general constant which can have different values at different places whereas indexed constants have the same value at all positions. However, all these constants are independent of the local and global time step size, the length T of the global time interval and the continuous and discrete solution.

2 Evolution equation in a Hilbert space

Let V be a Hilbert space with the norm $\|\cdot\|_V$, V' its dual space and $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{V', V}$ the duality pairing. We assume that there is another Hilbert space H such that $V \subset H$ is a continuous embedding and V is dense in H . Instead of identifying H with H' we consider the linear continuous and invertible Riesz operator $M : H \rightarrow H'$ defined by

$$\langle Mu, v \rangle_{H', H} := (u, v)_H \quad \forall u, v \in H,$$

i.e., for a given $f \in H'$, there is a unique element $M^{-1}f \in H$ such that

$$(M^{-1}f, v)_H = \langle f, v \rangle_{H', H} \quad \forall v \in H. \quad (1)$$

Let $I = (0, T)$ be a time interval with a given positive final time T . For a given Banach space B with the norm $\|\cdot\|_B$ and a given time interval $J \subset \mathbb{R}$, we define the space $L^2(J, B)$ as

$$L^2(J, B) := \{u : J \rightarrow B : \|u\|_{L^2(J, B)} < \infty\} \quad \text{where} \quad \|u\|_{L^2(J, B)} := \left(\int_J \|u(t)\|_B^2 dt \right)^{1/2}$$

and the space $C(J, B)$ as the space of continuous functions $u : J \rightarrow B$ equipped with the norm

$$\|u\|_{C(J, B)} := \sup_{t \in J} \|u(t)\|_B.$$

Let $u \in L^2(I, V)$ be given such that $u(t) \in H$ for all $t \in I$. Then, we say that a function $d_t u \in L^2(I, V')$ is the *weak time derivative* of u if for all $\psi \in C^\infty([0, T])$ and all $v \in V$ it holds

$$\int_0^T \langle d_t u(t), v \rangle \psi(t) dt = \langle u(T), v \rangle \psi(T) - \langle u(0), v \rangle \psi(0) - \int_0^T (u(t), v)_H \psi'(t) dt.$$

For a time interval J and a Banach space B , we denote by $H^m(J, B)$ the space

$$H^m(J, B) := \{u : L^2(J, B) : \|u\|_{H^m(J, B)} < \infty\} \quad \text{where} \quad \|u\|_{L^2(J, B)} := \sum_{i=0}^m \|d_t^i u\|_{L^2(J, B)}$$

and define the semi-norm on $H^m(J, B)$ as

$$|u|_{H^m(J, B)} := \|d_t^m u\|_{L^2(J, B)}.$$

Now, we are ready to formulate our *continuous evolution equation in V'* . For a given sufficiently smooth function $F : \bar{I} \times V \rightarrow V'$ and a given initial value $u_0 \in V$, we consider the following problem : Find $u : \bar{I} \rightarrow V$ such that

$$\begin{aligned} d_t u(t) &= F(t, u(t)) & \text{for a.e. } t \in I, \\ u(0) &= u_0. \end{aligned} \tag{2}$$

The *weak formulation* of problem (2) reads: Find $u \in X$ such that $u(0) = u_0$ and

$$\int_0^T \langle d_t u(t), v(t) \rangle dt = \int_0^T \langle F(t, u(t)), v(t) \rangle dt \quad \forall v \in Y, \tag{3}$$

where X denotes the *solution space* (see e.g. [3, Sect. 6.1])

$$X := \{u \in L^2(I, V) : d_t u \in L^2(I, V')\} \quad \text{with} \quad \|u\|_X^2 := \|u\|_{L^2(I, V)}^2 + \|d_t u\|_{L^2(I, V')}^2 \tag{4}$$

and Y the *test space*

$$Y = L^2(I, V) \quad \text{with} \quad \|v\|_Y = \|v\|_{L^2(I, V)}. \tag{5}$$

We can split the solution u of problem (3) as

$$u(t) = u_0 + u^0(t) \quad \text{with} \quad u^0 \in X_0 := \{u \in X : u(0) = 0\}. \tag{6}$$

Since $d_t u = d_t u^0$, the problem (3) is equivalent to the u^0 -problem : Find $u^0 \in X_0$ such that

$$\int_0^T \langle d_t u^0(t), v(t) \rangle dt = \int_0^T \langle F(t, u_0 + u^0(t)), v(t) \rangle dt \quad \forall v \in Y, \tag{7}$$

In this paper, we assume that there exists a unique solution u of problem (3) which is sufficiently smooth with respect to t . In the following, we consider two important special cases of our general evolution equation.

2.1 Linear parabolic equation

Let $a : V \times V \rightarrow \mathbb{R}$ be a bilinear form on the Hilbert space V satisfying the assumptions:

$$a(v, v) \geq C_1 \|v\|_V^2 \quad \forall v \in V, \quad (8)$$

$$|a(u, v)| \leq C_2 \|u\|_V \|v\|_V \quad \forall u, v \in V, \quad (9)$$

with positive constants C_1, C_2 . Then, the bilinear form $a(\cdot, \cdot)$ defines a linear continuous operator $A : V \rightarrow V'$ in the following way. For any given $u \in V$, the image $Au \in V'$ is uniquely defined by

$$\langle Au, v \rangle = a(u, v) \quad \forall v \in V. \quad (10)$$

We assume that there is a subspace $D(A) \subset V$ such that $A : D(A) \rightarrow H'$ is a bijective linear continuous operator. As an example one can consider the time-dependent convection diffusion reaction problem in a domain $\Omega \subset \mathbb{R}^m$ where $V = H_0^1(\Omega)$, $H = L^2(\Omega)$, $V' = H^{-1}(\Omega)$ and

$$Au := -\varepsilon \Delta u + b \cdot \nabla u + cu \in H' \quad \forall u \in D(A) = H^2(\Omega) \cap H_0^1(\Omega)$$

where $\varepsilon \in \mathbb{R}$ is positive diffusion coefficient, $b \in (W^{1,\infty}(\Omega))^m$ a given convection field and $c \in L^\infty(\Omega)$ a reaction field satisfying $c(x) - \frac{1}{2} \operatorname{div} b(x) \geq c_0 > 0$ for all $x \in \Omega$.

Now, we consider the *linear parabolic equation* (2) with $F : \bar{I} \times V \rightarrow V'$ defined by

$$F(t, u) := f(t) - Au \quad (11)$$

where we assume $f \in L^2(\bar{I}, H')$ which implies $F(t, u) \in H'$ for all $(t, u) \in \bar{I} \times D(A)$. That means, for a given initial value $u_0 \in V$, we have to find a solution $u(t) = u_0 + u^0(t)$ where u^0 is the solution of the problem : Find $u^0 \in X_0$ such that

$$\int_0^T \langle d_t u^0(t), v(t) \rangle dt + \int_0^T \langle Au^0(t), v(t) \rangle dt = \int_0^T \langle f(t) - Au_0, v(t) \rangle dt \quad \forall v \in Y. \quad (12)$$

Applying Theorem 6.6 in [3], we conclude that there exists a unique solution $u^0 \in X_0$ of problem (12) which guarantees also the existence and uniqueness of a solution u of problem (2) in the special setting of this section.

2.2 Ordinary differential equation in $\mathbb{R}^d$

Here we have $V = H = \mathbb{R}^d$ equipped with the Euclidean norm $\|\cdot\|$ and $H' = V' = \mathbb{R}^d$ be identified with $H = V$. Then, the duality pairings $\langle \cdot, \cdot \rangle_{V', V} = \langle \cdot, \cdot \rangle_{H', H}$ are identical with the usual Euclidean scalar product and the Riesz operator $M : H \rightarrow H'$ corresponds to the identity matrix. For $F : \bar{I} \times V \rightarrow V$ we assume continuity and the following Lipschitz condition:

$$\|F(t, u) - F(t, v)\| \leq L \|u - v\| \quad \forall u, v \in V, t \in \bar{I}, \quad (13)$$

where L denotes the *global Lipschitz constant*. Then, it is well-known that the system (2) of ordinary differential equations has a unique solution $u \in C^1(I, \mathbb{R}^d)$.

3 Discretization by a Galerkin-Petrov method

We decompose the time interval I into N subintervals $I_n := (t_{n-1}, t_n)$, where $n \in \{1, \dots, N\}$ and

$$0 = t_0 < t_1 < \dots < t_{N-1} < t_N = T.$$

In our time discretization, we approximate the *continuous solution* $u(t)$ of problem (2) on each time interval I_n by a polynomial function:

$$u(t) \approx u_\tau(t) := \sum_{j=0}^k U_n^j \phi_{n,j}(t) \quad \forall t \in I_n, \quad (14)$$

where the "coefficients" U_n^j are elements of the Hilbert space V and the basis functions $\phi_{n,j} \in \mathbb{P}_k(I_n)$ are linearly independent elements of the standard space of polynomials on the interval I_n with a degree not larger than a given order k .

For a given time interval $J \subset \mathbb{R}$ and a Banach space B , we introduce the linear space of B -valued time polynomials with degree of at most k as

$$\mathbb{P}_k(J, B) := \left\{ u : J \rightarrow B : u(t) = \sum_{j=0}^k U^j t^j, \forall t \in J, U^j \in B, \forall j \right\}.$$

Now, the *discrete solution space* for the global approximation $u_\tau : I \rightarrow V$ is the space $X_\tau^k \subset X$ defined as

$$X_\tau^k := \{u \in C(I, V) : u|_{\bar{I}_n} \in \mathbb{P}_k(\bar{I}_n, V) \quad \forall n = 1, \dots, N\} \quad (15)$$

and the *discrete test space* is the space $Y_\tau^k \subset Y$ given by

$$Y_\tau^k := \{u \in L^2(I, V) : u|_{I_n} \in \mathbb{P}_{k-1}(I_n, V) \quad \forall n = 1, \dots, N\}. \quad (16)$$

The symbol τ denotes the *discretization parameter* which acts in the error estimates as the maximum time step size $\tau := \max_{1 \leq n \leq N} \tau_n$, where $\tau_n := t_n - t_{n-1}$ is the length of the n -th time interval I_n .

Let us denote by $X_{\tau,0}^k := X_\tau^k \cap X_0$ the subspace of X_τ^k with zero initial condition. Then, it is easy to see that the dimensions of the spaces $X_{\tau,0}^k$ and Y_τ^k coincide such that it makes sense to consider the following *discontinuous Galerkin-Petrov discretization* of order k for the weak problem (3) : Find $u_\tau \in u_0 + X_{\tau,0}^k$ such that

$$\int_0^T \langle d_t u_\tau(t), v_\tau(t) \rangle dt = \int_0^T \langle F(t, u_\tau(t)), v_\tau(t) \rangle dt \quad \forall v_\tau \in Y_\tau^k. \quad (17)$$

We will denote this discretization as the ”*exact* dGP(k)-*method*”. Since the discrete test space Y_τ^k is discontinuous, problem (17) can be solved in a time marching process. Therefore, we choose test functions $v_\tau(t) = v\psi_{n,i}(t)$ with an arbitrary $v \in V$ and a scalar function $\psi_{n,i} : I \rightarrow \mathbb{R}$ which is zero on $I \setminus \bar{I}_n$ and a polynomial $\psi_{n,i} \in \mathbb{P}_{k-1}(\bar{I}_n)$ on the time interval $\bar{I}_n = [t_{n-1}, t_n]$. Then, we obtain for each $i = 0, \dots, k-1$

$$\int_{I_n} \langle d_t u_\tau(t), v \rangle \psi_{n,i}(t) dt = \int_{I_n} \langle F(t, u_\tau(t)), v \rangle \psi_{n,i}(t) dt \quad \forall v \in V. \quad (18)$$

By the definition of the weak time derivative we get for u_τ represented by (14) the equation

$$\int_{I_n} \langle d_t u_\tau(t), v \rangle \psi_{n,i}(t) dt = \int_{I_n} \sum_{j=0}^k (U_n^j, v)_H \phi'_{n,j}(t) \psi_{n,i}(t) dt \quad \forall v \in V. \quad (19)$$

We define the basis functions $\phi_{n,j} \in \mathbb{P}_k(\bar{I}_n)$ of (14) via the reference transformation $\omega_n : \hat{I} \rightarrow \bar{I}_n$ where $\hat{I} := [-1, 1]$ and

$$t = \omega_n(\hat{t}) := \frac{t_{n-1} + t_n}{2} + \frac{\tau_n}{2} \hat{t} \in \bar{I}_n \quad \forall \hat{t} \in \hat{I}, \quad n = 1, \dots, N. \quad (20)$$

Let $\hat{\phi}_j \in \mathbb{P}_k(\hat{I})$, $j = 0, \dots, k$, be suitable basis functions satisfying the conditions

$$\hat{\phi}_j(-1) = \delta_{0,j}, \quad \hat{\phi}_j(1) = \delta_{k,j}, \quad (21)$$

where $\delta_{k,j}$ denotes the usual Kronecker symbol. Then, we define the basis functions on the original time interval $\bar{I}_n$ by

$$\phi_{n,j}(t) := \hat{\phi}_j(\hat{t}) \quad \text{with} \quad \hat{t} := \omega_n^{-1}(t) = \frac{2}{\tau_n} \left(t - \frac{t_n - t_{n-1}}{2} \right) \in \hat{I}. \quad (22)$$

Similarly, we define the test basis functions $\psi_{n,i}$ by suitable reference basis functions $\hat{\psi}_i \in \mathbb{P}_{k-1}(\hat{I})$, i.e.,

$$\psi_{n,i}(t) := \hat{\psi}_i(\omega_n^{-1}(t)) \quad \forall t \in \bar{I}_n, \quad i = 0, \dots, k-1. \quad (23)$$

By the property (21), the initial condition and the continuity (with respect to time) of the discrete solution $u_\tau : I \rightarrow V$ is equivalent to the conditions:

$$U_1^0 = u_0 \quad \text{and} \quad U_n^0 = U_{n-1}^k \quad \forall n > 2. \quad (24)$$

We transform the integrals in (18) to the reference interval $\hat{I}$ and obtain the following system of equations for the ”coefficients” $U_n^j \in V$, $j = 1, \dots, k$, in the ansatz (14) :

$$\sum_{j=0}^k \alpha_{i,j} (U_n^j, v)_H = \frac{\tau_n}{2} \int_{\hat{I}} \left\langle F \left(\omega_n(\hat{t}), \sum_{j=0}^k U_n^j \hat{\phi}_j(\hat{t}) \right), v \right\rangle \hat{\psi}_i(\hat{t}) d\hat{t} \quad \forall v \in V \quad (25)$$

where $i = 0, \dots, k-1$,

$$\alpha_{i,j} := \int_{\hat{I}} d\hat{t} \hat{\phi}_j(\hat{t}) \hat{\psi}_i(\hat{t}) d\hat{t},$$

and the "coefficient" $U_n^0 \in V$ is known. We approximate the integral on the right hand side of (25) by the $(k+1)$ -point *Gauß-Lobatto quadrature formula*:

$$\int_{\hat{I}} \left\langle F \left(\omega_n(\hat{t}), \sum_{j=0}^k U_n^j \hat{\phi}_j(\hat{t}) \right), v \right\rangle \hat{\psi}_i(\hat{t}) d\hat{t} \approx \sum_{\mu=0}^k \hat{w}_\mu \left\langle F \left(\omega_n(\hat{t}_\mu), \sum_{j=0}^k U_n^j \hat{\phi}_j(\hat{t}_\mu) \right), v \right\rangle \hat{\psi}_i(\hat{t}_\mu), \quad (26)$$

where $\hat{w}_\mu$ are the weights and $\hat{t}_\mu \in [-1, 1]$ are the integration points with $\hat{t}_0 = -1$ and $\hat{t}_k = 1$. Let us define the mapped Gauß-Lobatto points $t_{n,\mu} \in \bar{I}_n$ and the coefficients $\beta_{i,\mu}, \gamma_{j,\mu}$ by

$$t_{n,\mu} := \omega_n(\hat{t}_\mu), \quad \beta_{i,\mu} := \hat{w}_\mu \hat{\psi}_i(\hat{t}_\mu), \quad \gamma_{j,\mu} := \hat{\phi}_j(\hat{t}_\mu).$$

Then, the system (25) is equivalent to the following system of equations for the k unknown "coefficients" $U_n^j \in V, j = 1, \dots, k$,

$$\sum_{j=0}^k \alpha_{i,j} (U_n^j, v)_H = \frac{\tau_n}{2} \sum_{\mu=0}^k \beta_{i,\mu} \left\langle F \left(t_{n,\mu}, \sum_{j=0}^k \gamma_{j,\mu} U_n^j \right), v \right\rangle \quad \forall v \in V. \quad (27)$$

with the k "equations" $i = 0, \dots, k-1$ where $U_n^0 = U_{n-1}^k$ for $n > 1$ and $U_1^0 = u_0$.

Once we have solved this system we enter the next time interval and set the initial value of the new time interval I_{n+1} to $U_{n+1}^0 := U_n^k$. If the Gauß-Lobatto formula would be exact for the right hand side of (25) this time marching process would solve the global time discretization (17) exactly. Since in general there is an integration error we call the time marching process corresponding to (27) simply the "dGP(k)-method".

In principle, we have to solve a coupled system for the $U_n^j \in V$ which could be very expensive. However, by a clever choice of the functions $\hat{\phi}_j$ and $\hat{\psi}_i$ it is possible to uncouple the system to a large extend. In the following, we will discuss this issue for the special methods dGP(1), dGP(2) and for the general method dGP(k), $k \geq 3$. In all cases, we choose the basis functions $\hat{\phi}_j \in \mathbb{P}_k(\hat{I})$ as the Lagrange basis functions with respect to the Gauß-Lobatto points $\hat{t}_\mu$, i.e.,

$$\hat{\phi}_j(\hat{t}_\mu) = \delta_{j,\mu} \quad \forall j, \mu \in \{0, \dots, k\}. \quad (28)$$

Then, the method (27) reduces to

$$\sum_{j=0}^k \alpha_{i,j} (U_n^j, v)_H = \frac{\tau_n}{2} \sum_{j=0}^k \beta_{i,j} \langle F(t_{n,j}, U_n^j), v \rangle \quad \forall v \in V, i = 0, \dots, k-1, \quad (29)$$

and by the choice of the test basis functions $\hat{\psi}_i \in \mathbb{P}_{k-1}(\hat{I})$ we try to get suitable values for the coefficients $\alpha_{i,j}$ and $\beta_{i,j}$. In the following, we will use the following abbreviation and

assumption:

$$F_n^j(U_n^j) := F(t_{n,j}, U_n^j) \in H' \quad \forall j = 0, \dots, k, n = 1, \dots, N. \quad (30)$$

3.1 The dGP(1) method

We use the 2-point Gauß-Lobatto formula (trapezoidal rule) with $\hat{w}_0 = \hat{w}_1 = 1$ and $\hat{t}_0 = -1$, $\hat{t}_1 = 1$. The only test function $\hat{\psi}_0$ is chosen as $\hat{\psi}_0(\hat{t}) = 1$. Then, we obtain

$$\alpha_{0,0} = -1, \quad \alpha_{0,1} = 1, \quad \beta_{0,0} = \beta_{0,1} = 1.$$

Using the notation $U^{n-1} := u_\tau(t_{n-1}) = U_n^0$ and $U^n := u_\tau(t_n) = U_n^1$, we obtain the following equation for the "unknown" $U^n \in V$:

$$(U^n, v)_H - (U^{n-1}, v)_H = \frac{\tau_n}{2} \{ \langle F(t_{n-1}, U^{n-1}) + F(t_n, U^n), v \rangle \} \quad (31)$$

for all $v \in V$ which is the well-known *Crank-Nicolson method*. In operator notation it can be written in the equivalent form:

$$U^n = U^{n-1} + \frac{\tau_n}{2} M^{-1} \{ F(t_{n-1}, U^{n-1}) + F(t_n, U^n) \}. \quad (32)$$

3.2 The dGP(2) method

We use the 3-point Gauß-Lobatto formula (Simpson rule) with $\hat{w}_0 = \hat{w}_2 = 1/3$, $\hat{w}_1 = 4/3$ and $\hat{t}_0 = -1$, $\hat{t}_1 = 0$, $\hat{t}_2 = 1$. For the test functions $\hat{\psi}_i \in \mathbb{P}_1(\hat{I})$, we choose

$$\hat{\psi}_0(\hat{t}) = -\frac{3}{4}\hat{t}, \quad \hat{\psi}_1(\hat{t}) = 1.$$

Then, we get

$$(\alpha_{i,j}) = \begin{pmatrix} -1/2 & 1 & -1/2 \\ -1 & 0 & 1 \end{pmatrix}, \quad (\beta_{i,j}) = \begin{pmatrix} 1/4 & 0 & -1/4 \\ 1/3 & 4/3 & 1/3 \end{pmatrix}$$

and, using (1) and the assumption (30), the system to compute the "unknowns" $U_n^1, U_n^2 \in V$ from the known $U_n^0 = U_{n-1}^2$ reads:

$$U_n^1 = \frac{1}{2}U_n^0 + \frac{1}{2}U_n^2 + \frac{\tau_n}{8}M^{-1} \{ F_n^0(U_n^0) - F_n^2(U_n^2) \} \quad (33)$$

$$U_n^2 = U_n^0 + \frac{\tau_n}{6}M^{-1} \{ F_n^0(U_n^0) + 4F_n^1(U_n^1) + F_n^2(U_n^2) \}. \quad (34)$$

Let us denote the value for U_n^1 computed from (33) and depending on U_n^2 by $U_n^1 = G_n^1(U_n^2)$ where $G_n^1 : V \rightarrow V$ in general is a nonlinear operator. We substitute this in the equation (34) and get, for the unknown $U_n^2 \in V$, the following fixed point equation :

$$U_n^2 = G_n^2(U_n^2) := U_n^0 + \frac{\tau_n}{6}M^{-1} \{ F_n^0(U_n^0) + 4F_n^1(G_n^1(U_n^2)) + F_n^2(U_n^2) \} \quad (35)$$

The mapping $G_n^2 : V \rightarrow V$ is a contraction if the time step size τ_n is sufficiently small.

Remark 1 We consider the linear case of Sect. 2.1 where $F(t, u) = f(t) - Au$ with a linear operator $A : V \rightarrow V'$ defined by a coercive and continuous bilinear form $a(\cdot, \cdot)$. Since we know that for the discrete solution $u_\tau \in X_\tau^k \subset X$ (its existence will be shown below) it holds $u_\tau \in C^0(\bar{I}, H)$ [3, Lemma 6.2], we get that the coefficients $U_n^j = u_\tau(t_{n,j})$, $j = 1, 2$, of the ansatz (14) satisfy $U_n^j \in H$. In the special case of finite dimensional discrete spaces $V \subset H$ with $\dim(V) = \dim(H)$, we have $V = H$. Thus, it is legal to build $AU_n^j \in V' = H'$ and the element $M^{-1}AU_n^j$ is contained in H . The Riesz operator M corresponds to the so-called mass-matrix. If we insert the linear expressions for $F_n^j(U) = F(t_{n,j}, U) = f_n^j - AU$ with $f_n^j := f(t_{n,j})$ into (35) and $U_n^1 = G_n^1(U_n^2)$ we get an implicit equation for U_n^2 :

$$\begin{aligned} \left\{ M + \frac{\tau_n}{2}A + \frac{\tau_n^2}{12}AM^{-1}A \right\} U_n^2 &= \left\{ M - \frac{\tau_n}{2}A + \frac{\tau_n^2}{12}AM^{-1}A \right\} U_n^0 \\ &+ \frac{\tau_n}{6} (f_n^0 + 4f_n^1 + f_n^2) - \frac{\tau_n^2}{12}AM^{-1} (f_n^0 - f_n^2). \end{aligned} \quad (36)$$

In the case of a moderate dimension of V , it is not a problem to build the operator $AM^{-1}A$ (it simply means matrix multiplication). However, if $\dim(V)$ is very large the generation of the operator $\tilde{A} = M + (\tau_n/2)A + (\tau_n^2/12)AM^{-1}A$ as a matrix would be very expensive. So it is useful to have an iterative process which solves (36) efficiently by using only the "matrix-vector product" $\tilde{A}U$ for a given $U \in H$ combined with a preconditioner being an approximative inverse of $\tilde{A}$. If A would be a differential operator on an infinite dimensional Hilbert space V some regularity problems can appear which we will not discuss here.

3.3 The dGP(k) method

We will present a general method for the polynomial degree $k \geq 2$ of the discrete solution space. We use the $(k+1)$ -point Gauß-Lobatto formula on the reference interval $[-1, 1]$ with the weights $\hat{w}_\mu$ and integration points $\hat{t}_\mu$, $\mu = 0, \dots, k$, where $\hat{t}_0 = -1$ and $\hat{t}_k = 1$. For the test function $\hat{\psi}_i \in \mathbb{P}_{k-1}(\hat{I})$, $i < k-1$, we require the k conditions

$$\hat{\psi}_i'(\hat{t}_j) = \frac{-1}{\hat{w}_{i+1}} \delta_{i+1,j} \quad \forall j \in \{1, \dots, k-1\}, \quad \forall i = 0, \dots, k-2. \quad (37)$$

and

$$\hat{\psi}_i(\hat{t}_{i+1}) = 0. \quad (38)$$

Then, we get from (37)

$$\alpha_{i,j} = \delta_{i+1,j} \quad \forall j \in \{1, \dots, k\}, \quad i \in \{0, \dots, k-2\}. \quad (39)$$

and from (38)

$$\beta_{i,i+1} = 0 \quad \forall i \in \{0, \dots, k-2\}. \quad (40)$$

Polynomials that satisfy the conditions (37) and (38) are

$$\hat{\psi}_i(\hat{t}) = \frac{-1}{\hat{w}_{i+1}} \int_{\hat{t}_{i+1}}^{\hat{t}} \prod_{\substack{\mu=1 \\ \mu \neq i+1}}^{k-1} \frac{s - \hat{t}_\mu}{\hat{t}_{i+1} - \hat{t}_\mu} ds \quad \forall i \in \{0, \dots, k-2\}.$$

For the last test function, we choose $\hat{\psi}_{k-1} \equiv 1$. Then, using (1) and the assumption (30), the resulting system to determine the k "unknowns" $U_n^j \in V$, $j = 1, \dots, k$, from the known $U_n^0 = U_{n-1}^k$ reads :

$$MU_n^{i+1} = -\alpha_{i,0}MU_n^0 - \alpha_{i,k}MU_n^k + \frac{\tau_n}{2} \sum_{\substack{j=0 \\ j \neq i+1}}^k \beta_{i,j}F_n^j(U_n^j), \quad i = 0, \dots, k-2, \quad (41)$$

$$MU_n^k = MU_n^0 + \frac{\tau_n}{2} \sum_{j=0}^k \hat{w}_j F_n^j(U_n^j). \quad (42)$$

We will solve this system by means of a defect correction iteration. Let us define the defect $d_i(\vec{U}_n)$ of the i -th equation for a given vector $\vec{U}_n := (U_n^j)_{j=1}^k \in V^k$ as

$$d_i(\vec{U}_n) = -MU_n^{i+1} - \alpha_{i,0}MU_n^0 - \alpha_{i,k}MU_n^k + \frac{\tau_n}{2} \sum_{\substack{j=0 \\ j \neq i+1}}^k \beta_{i,j}F_n^j(U_n^j)$$

for $i = 0, \dots, k-2$ and

$$d_k(\vec{U}_n) = -MU_n^k + MU_n^0 + \frac{\tau_n}{2} \sum_{j=0}^k \hat{w}_j F_n^j(U_n^j).$$

Now, for a given old approximate vector $\vec{U}_n$, we want to compute a correction vector $\vec{Z}_n = (Z_n^j)_{j=1}^k \in V^k$ such that $\vec{U}_n + \vec{Z}_n$ provides a better approximation. The most simple and stable initial approximation would be $\vec{U}_n = (U_n^j)_{j=1}^k$ with $U_n^j = U_n^0$ for all $j = 1, \dots, k$. To compute corrections Z_n^{i+1} for $i = 0, \dots, k-2$, we derive from (41)

$$M(U_n^{i+1} + Z_n^{i+1}) \approx -\alpha_{i,0}MU_n^0 - \alpha_{i,k}MU_n^k + \frac{\tau_n}{2} \sum_{\substack{j=0 \\ j \neq i+1}}^k \beta_{i,j}F_n^j(U_n^j)$$

and get the following equations for approximate corrections

$$Z_n^{i+1} = M^{-1}d_i(\vec{U}_n) \quad \forall i = 0, \dots, k-2. \quad (43)$$

For the correction Z_n^k , we derive from (42) and a linearization of $F_n^j(U_n^j + Z_n^j)$

$$M(U_n^k + Z_n^k) \approx MU_n^0 + \frac{\tau_n}{2} \sum_{j=0}^k \hat{w}_j \{F_n^j(U_n^j) + J_n^j Z_n^j\}$$

where $J_n^j := D_u F_n^j(U_n^j)$ denotes the derivative of $F_n^j(U)$ with respect to U . This leads to the following equation for an approximate correction Z_n^k depending on the corrections Z_n^j , $j = 1, \dots, k-1$, computed by (43) :

$$\left\{ M - \frac{\tau_n \hat{w}_k}{2} J_n^k \right\} Z_n^k = d_k(\vec{U}_n) + \frac{\tau_n}{2} \sum_{j=1}^{k-1} \hat{w}_j J_n^j Z_n^j. \quad (44)$$

Once all components of the correction vector $\vec{Z}_n$ have been computed, the old vector $\vec{U}_n$ is updated by $\vec{U}_n + \vec{Z}_n$. This defect correction process will be repeated until the norms of all defects $d_i(\vec{U}_n)$ are small enough.

4 Energy decreasing property

Let us consider an *energy functional* $W : V \rightarrow \mathbb{R}^+$ which assigns to a given element u of a Hilbert space V an energy value $W(u) > 0$. If we are looking for a local minimizer $u^* \in V$ of $W(\cdot)$ we would like to solve the equation

$$\langle W'(u), v \rangle = 0 \quad \forall v \in V, \quad (45)$$

where we have assumed that $W(\cdot)$ is differentiable in the sense that for each $u \in V$ the following limit for $W'(u) \in V'$ exists

$$\langle W'(u), v \rangle := \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \{ W(u + \varepsilon v) - W(u) \} \quad \forall v \in V. \quad (46)$$

In order to find an approximation for a local minimizer $u^* \in V$, we can start with a suitable initial guess $u_0 \in V$ at a virtual time $t = 0$ and can solve approximately the following evolution equation: Find a function $u : [0, T] \rightarrow V$ such that

$$\begin{aligned} (d_t u(t), v) &= -\langle W'(u(t)), v \rangle \quad \forall v \in V, \quad t \in [0, T], \\ u(0) &= u_0, \end{aligned} \quad (47)$$

where $u(t)$ is changed in the direction of the negative gradient of $W(\cdot)$ at the actual "state" $u(t)$ and $T > 0$ is chosen large enough such that (hopefully) $d_t u(T) \approx 0_{V'}$. In this section, $(\cdot, \cdot)$ denotes the scalar product in the Hilbert space V and the connection to the notation in the previous sections is that we choose here $H = V$ and $F : [0, T] \times V \rightarrow V'$ as

$$\langle F(t, u), v \rangle := -\langle W'(u), v \rangle \quad \forall u, v \in V.$$

The weak formulation (with respect to time) of problem (47) is then: Find $u \in X$ such that $u(0) = u_0$ and

$$\int_0^T (d_t u(t), v(t)) dt = - \int_0^T \langle W'(u(t)), v(t) \rangle dt \quad \forall v \in Y. \quad (48)$$

We can show that, for the exact solution $u : [0, T] \rightarrow V$, the energy values $W(u(t))$ are monotonously decreasing for increasing time. Let t_{n-1} and t_n be two time levels with $0 \leq t_{n-1} < t_n \leq T$. We choose in (48) the test function $v(\cdot)$ such that

$$v(t) := d_t u(t) \quad \forall t \in [t_{n-1}, t_n] \quad \text{and} \quad v(t) := 0_V \quad \text{otherwise.}$$

For simplicity, we have assumed here that $d_t u(t) \in V$, i.e., we have identified the elements of V' and V . Now, we get from (48) the equation

$$\int_{t_{n-1}}^{t_n} \|d_t u(t)\|_V^2 dt = - \int_{t_{n-1}}^{t_n} \langle W'(u(t)), d_t u(t) \rangle dt = - \int_{t_{n-1}}^{t_n} d_t \{W(u(t))\} dt$$

which leads to

$$W(u(t_n)) = W(u(t_{n-1})) - \|d_t u\|_{L^2([t_{n-1}, t_n], V)}^2 < W(u(t_{n-1})) \quad (49)$$

as long as the time derivative $d_t u(\cdot)$ is not identically zero at the interval $[t_{n-1}, t_n]$. However, if $d_t u(\cdot)$ would be zero on $[t_{n-1}, t_n]$ we have found a stationary point of $W(\cdot)$, i.e., $u^* = u(t_n)$ would be a local minimizer of the energy functional.

An advantage of the exact dGP(k)-method is that the energy decreasing property (49) is maintained by this method also on the discrete level. This is easy to see since we can apply the same arguments as in the continuous case. The key ingredient of the dGP(k)-method is that the function $v_\tau = d_t u_\tau$ is always contained in the discrete test space Y_τ^k for all u_τ in the solution space X_τ^k . Thus, for our discrete solution $u_\tau \in X_\tau^k$, we get for each partition of the time interval $0 = t_0 < t_1 < \dots < t_N = T$ the energy decreasing property

$$W(u_\tau(t_n)) = W(u_\tau(t_{n-1})) - \|d_t u_\tau\|_{L^2([t_{n-1}, t_n], V)}^2 < W(u_\tau(t_{n-1})) \quad (50)$$

as long as the time derivative $d_t u_\tau(\cdot)$ is not identically zero at the interval $[t_{n-1}, t_n]$.

So, the exact dGP(k)-method could be a good candidate for optimization problems if the derivative of the functional $W(\cdot)$ is available.

5 A-stability of the dGP(k)-method

For a given $\lambda \in \mathbb{C}$ and an initial value $u_0 \in \mathbb{C}$, we consider the model problem: Find a function $u : [0, T] \rightarrow \mathbb{C}$ such that

$$\begin{aligned} d_t u(t) &= \lambda u(t) & \forall t \in [0, T], \\ u(0) &= u_0. \end{aligned} \quad (51)$$

To show that the dGP(k)-method is *A-stable*, we have to prove that for the simple situation of just one time step, i.e., for $t_1 = T = \tau$, it holds

$$|u_\tau(\tau)| < |u_0| \quad \forall \tau > 0, \quad \forall \lambda \in \{z \in \mathbb{C} : \operatorname{Re}(z) < 0\}, \quad \forall u_0 \in \mathbb{C} \setminus \{0\}. \quad (52)$$

Theorem 2 *The exact dGP(k)-method as well as the dGP(k)-method with numerical integration described in Section 3 are A-stable in the sense that (52) holds if the methods are applied to model problem (51).*

Proof. Let $\lambda \in \mathbb{C}$ be given with $\operatorname{Re}(\lambda) < 0$. We write the problem (51) in an equivalent way as an ODE-system in $\mathbb{R}^2$. To this end, we define for any complex function $u : [0, T] \rightarrow \mathbb{C}$ an associated vector function $\vec{u} : [0, T] \rightarrow \mathbb{R}^2$ by

$$\vec{u}(t) := \begin{pmatrix} \operatorname{Re}(u(t)) \\ \operatorname{Im}(u(t)) \end{pmatrix} \quad \forall t \in [0, T].$$

Then, the model problem (51) is equivalent to: Find a function $\vec{u} : [0, T] \rightarrow \mathbb{R}^2$ such that

$$\begin{aligned} d_t \vec{u}(t) &= A \vec{u}(t) \quad \forall t \in [0, T], \\ \vec{u}(0) &= \vec{u}_0, \end{aligned} \quad \text{where} \quad A := \begin{pmatrix} \operatorname{Re}(\lambda) & -\operatorname{Im}(\lambda) \\ \operatorname{Im}(\lambda) & \operatorname{Re}(\lambda) \end{pmatrix} \quad (53)$$

and $\vec{u}_0 := (\operatorname{Re}(u_0), \operatorname{Im}(u_0))^T$. The matrix A has the property

$$\xi^T A \xi = \xi^T A^T \xi = \operatorname{Re}(\lambda) \|\xi\|^2 \quad \forall \xi \in \mathbb{R}^2 \quad (54)$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^2$. This shows that A is regular and negative definite since $\operatorname{Re}(\lambda) < 0$. The application of the dGP(k)-method to problem (51) is equivalent to the application of the dGP(k)-method to problem (53) with the following setting. The Hilbert spaces $V = H = \mathbb{R}^2$ are equipped with the Euclidean norm $\|\cdot\|$. The dual spaces $H' = V'$ are identified with $\mathbb{R}^2$ in the sense that $\langle \cdot, \cdot \rangle_{V', V} = \langle \cdot, \cdot \rangle_{H', H} = (\cdot, \cdot)$ is the usual Euclidean scalar product. The Riesz operator $M : H \rightarrow H'$ corresponds to the identity matrix and the function $F : [0, T] \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is defined as $F(t, \vec{u}) := A \vec{u}$. Then, for the special case $t_1 = T = \tau$, the exact dGP(k)-method reads: Find $\vec{u}_\tau \in X_\tau^k$ such that $\vec{u}_\tau(0) = \vec{u}_0$ and

$$\int_0^\tau (d_t \vec{u}_\tau(t), \vec{v}_\tau(t)) dt = \int_0^\tau (A \vec{u}_\tau(t), \vec{v}_\tau(t)) dt \quad \forall \vec{v} \in Y_\tau^k. \quad (55)$$

Since the $(k+1)$ -point Gauß-Lobatto formula is exact for polynomials of degree less or equal to $2k-1$, the dGP(k)-method with numerical integration is equivalent to the exact dGP(k)-method such that equation (55) is also satisfied. The discrete spaces X_τ^k and Y_τ^k are chosen such that for each $\vec{u}_\tau \in X_\tau^k$ the function $\vec{v}_\tau : [0, T] \rightarrow \mathbb{R}^2$ defined by $\vec{v}_\tau(t) := (A^{-1})^T d_t \vec{u}_\tau(t)$ is contained in the test space Y_τ^k . Using this test function $\vec{v}_\tau$ in (55), leads to

$$\int_0^\tau (d_t \vec{u}_\tau(t), (A^{-1})^T d_t \vec{u}_\tau(t)) dt = \int_0^\tau (\vec{u}_\tau(t), d_t \vec{u}_\tau(t)) dt = \frac{1}{2} \int_0^\tau \frac{d}{dt} \|\vec{u}_\tau(t)\|^2 dt$$

which implies by (54)

$$\|\vec{u}_\tau(\tau)\|^2 - \|\vec{u}_0\|^2 = 2\operatorname{Re}(\lambda) \int_0^\tau \|A^{-1} d_t \vec{u}_\tau(t)\|^2 dt.$$

Suppose that the integral term is zero. Then we get, since A is regular, that $\|d_t \vec{u}_\tau(t)\| = 0$ for all $t \in [0, \tau]$. This implies $\vec{u}_\tau(t) = \vec{u}_0$ for all $t \in [0, \tau]$. If we take in (55) the time-constant test function $\vec{v}_\tau = \vec{u}_0 \in Y_\tau^k$ we obtain $\vec{u}_0^T A \vec{u}_0 = \operatorname{Re}(\lambda) \|\vec{u}_0\|^2 = 0$. This is a contradiction to the assumption that $u_0 \in \mathbb{C} \setminus \{0\}$, i.e. $\|\vec{u}_0\| \neq 0$. Therefore, the integral term multiplied by $2\operatorname{Re}(\lambda) < 0$ must be negative and we get that $\|\vec{u}_\tau(\tau)\| < \|\vec{u}_0\|$ which implies (52). $\square$

6 Analysis of the discrete Galerkin-Petrov method

6.1 Linear parabolic equation

We consider the special setting of Sect. 2.1. Instead of solving for the discrete solution $u_\tau \in u_0 + X_{\tau,0}^k$ we will solve only for the homogeneous part $u_\tau^0 := u_\tau - u_0 \in X_{\tau,0}^k$. Then, the exact dGP(k)-method (17) for u_τ^0 reads: Find $u_\tau^0 \in X_{\tau,0}^k$ such that

$$\int_0^T \langle d_t u_\tau^0(t), v_\tau(t) \rangle dt + \int_0^T \langle A u_\tau^0(t), v_\tau(t) \rangle dt = \int_0^T \langle f_0(t), v_\tau(t) \rangle dt \quad \forall v_\tau \in Y_\tau^k, \quad (56)$$

where $f_0 : [0, T] \rightarrow H'$ is defined as $f_0(t) := f(t) - A u_0$ with $u_0 \in D(A)$. To analyze the discrete problem (56) we write it in a more abstract form. Therefore, we introduce the bilinear form $B : X_0 \times Y \rightarrow \mathbb{R}$ and the linear form $\ell : Y \rightarrow \mathbb{R}$ as:

$$B(u, v) := \int_0^T \langle d_t u_\tau(t) + A u_\tau(t), v_\tau(t) \rangle dt, \quad (57)$$

$$\ell(v) := \int_0^T \langle f_0(t), v_\tau(t) \rangle dt. \quad (58)$$

Now, the continuous linear evolution problem (12) takes the form: Find $u^0 \in X_0$ such that

$$B(u^0, v) = \ell(v) \quad \forall v \in Y. \quad (59)$$

The exact dGP(k)-method (56) reads: Find $u_\tau^0 \in X_{\tau,0}^k$ such that

$$B(u_\tau^0, v_\tau) = \ell(v_\tau) \quad \forall v_\tau \in Y_\tau^k. \quad (60)$$

Next we will prove an inf-sup condition to get the existence, uniqueness and stability of the discrete solution of problem (60). Its proof is very close to that of Theorem 6.6 in [3].

Lemma 3 *Let $X_{\tau,0}^k = X_\tau^k \cap X_0 \subset X$ and $Y_\tau^k \subset Y$ be the discrete spaces defined in (15) and (16). Then, under the assumptions (8) and (9) on the bilinear form $a(\cdot, \cdot)$, it holds the inf-sup condition*

$$\inf_{u_\tau \in X_{\tau,0}^k \setminus \{0\}} \sup_{v_\tau \in Y_\tau^k \setminus \{0\}} \frac{B(u_\tau, v_\tau)}{\|u_\tau\|_X \|v_\tau\|_Y} \geq C_0 > 0. \quad (61)$$

Furthermore, the bilinear form $B(\cdot, \cdot)$ is continuous, i.e.,

$$|B(u, v)| \leq C_5 \|u\|_X \|v\|_Y \quad \forall u \in X, v \in Y. \quad (62)$$

The constants C_0 and C_5 are independent of τ and T .

Proof. Let $u_\tau \in X_{\tau,0}^k \setminus \{0\}$ be a given discrete function which has on the time interval I_n the representation

$$u_\tau(t) := \sum_{j=0}^k U_n^j \phi_{n,j}(t) \quad \forall t \in I_n,$$

Then, for the time derivative $d_t u_\tau$ restricted to I_n it holds

$$d_t u_\tau(t) := \sum_{j=0}^k U_n^j d_t \phi_{n,j}(t) \quad \forall t \in I_n.$$

This means by the construction of X_τ^k and Y_τ^k that $d_t u_\tau \in Y_\tau^k$. Due to the assumptions (8) and (9) on the bilinear form $a(\cdot, \cdot)$ which defines the operator $A : V \rightarrow V'$ we get by the Lax-Milgram Theorem that A is invertible and

$$\|A\| \leq C_2 \quad \text{and} \quad \|A^{-1}\| \leq 1/C_1. \quad (63)$$

Now, we choose for the given $u_\tau \in X_{\tau,0}^k$ the test function $v_\tau = A^{-1} d_t u_\tau + \mu u_\tau$ where μ is a positive constant defined later. For all $g \in V'$, it holds

$$\langle g, A^{-1} g \rangle = \langle A A^{-1} g, A^{-1} g \rangle \geq C_1 \|A^{-1} g\|_V^2 \geq \frac{C_1}{C_2^2} \|g\|_{V'}^2,$$

Thus, using $u_\tau(0) = 0$ and $\langle d_t u_\tau, u_\tau \rangle = 1/2 d_t (u_\tau, u_\tau)_H$ we get with $C_3 := C_1/C_2^2$

$$\begin{aligned} B(u_\tau, v_\tau) &= \int_0^T \langle d_t u_\tau(t) + A u_\tau(t), A^{-1} d_t u_\tau + \mu u_\tau \rangle dt \\ &\geq \frac{\mu}{2} \|u_\tau(T)\|_H^2 + \int_0^T \{ C_3 \|d_t u_\tau\|_{V'}^2 + \mu C_1 \|u_\tau\|_V^2 - \|A\| \|A^{-1}\| \|u_\tau\|_V \|d_t u_\tau\|_{V'} \} dt \\ &\geq \int_0^T \left\{ C_3 \|d_t u_\tau\|_{V'}^2 + \mu C_1 \|u_\tau\|_V^2 - \frac{C_2}{C_1} \|u_\tau\|_V \|d_t u_\tau\|_{V'} \right\} dt. \end{aligned}$$

By Young's inequality this implies

$$B(u_\tau, v_\tau) \geq \int_0^T \left\{ \frac{C_3}{2} \|d_t u_\tau\|_{V'}^2 + \left(\mu C_1 - \frac{C_2}{C_1} \varepsilon \right) \|u_\tau\|_V^2 \right\} dt$$

where $\varepsilon = 1/(2C_3)$. We choose $\mu = (C_3/2 + (C_2\varepsilon)/C_1)/C_1$ and get

$$B(u_\tau, v_\tau) \geq \frac{C_3}{2} \|u_\tau\|_X^2.$$

Finally, we estimate v_τ by

$$\|v_\tau\|_Y^2 = \|A^{-1}d_t u_\tau + \mu u_\tau\|_{L^2(I,V)}^2 \leq 2\|A^{-1}\|^2 \|d_t u_\tau\|_{L^2(I,V)}^2 + 2\mu^2 \|u_\tau\|_{L^2(I,V)}^2 \leq C_4^2 \|u_\tau\|_X^2,$$

where $C_4 := \sqrt{2} \max\{\|A^{-1}\|, \mu\}$, and obtain the inf-sup condition with $C_0 = C_3/(2C_4)$. The continuity of $B(\cdot, \cdot)$ is shown as follows:

$$\begin{aligned} |B(u, v)| &= \int_0^T \left| \langle d_t u(t) + Au(t), v(t) \rangle \right| dt \leq \int_0^T \|d_t u(t) + Au(t)\|_{V'} \|v(t)\|_V dt \\ &\leq \|d_t u + Au\|_{L^2(I,V')} \|v\|_{L^2(I,V)} \leq \{\|d_t u\|_{L^2(I,V')} + \|A\| \|u\|_{L^2(I,V)}\} \|v\|_Y \\ &\leq C_5 \|u\|_X \|v\|_Y \end{aligned}$$

with $C_5 = \sqrt{2} \max\{1, C_2\}$. $\square$

With the above lemma we easily obtain the existence, uniqueness and stability of the solution of the discrete evolution equation (60).

Theorem 4 *Let the assumptions (8) and (9) on the bilinear form $a(\cdot, \cdot)$ be satisfied. Then, for an arbitrary partition of the time interval $[0, T] = \bigcup_n \bar{I}_n$, there exists a unique solution $u_\tau^0 \in X_{\tau,0}^k$ of the discrete problem (60) of the exact dGP(k)-method and it holds the stability estimate*

$$\|u_\tau^0\|_X = \left(\int_0^T \|d_t u_\tau^0(t)\|_{V'}^2 + \|u_\tau^0(t)\|_V^2 dt \right)^{1/2} \leq \frac{1}{C_0} \|f_0\|_{L^2(I,V')}. \quad (64)$$

Proof. The discrete problem (60) satisfies the assumptions of the Banach-Nečas-Babuška-Theorem [3, Theorem 2.6] which guarantees the existence and uniqueness of the discrete solution u_τ^0 . Moreover, from (61) it follows that there exists a $v_\tau \in Y_\tau^k$ such that

$$C_0 \|u_\tau\|_X \|v_\tau\|_Y \leq B(u_\tau, v_\tau) = \int_0^T \langle f_0(t), v_\tau(t) \rangle dt \leq \|f_0\|_{L^2(I,V')} \|v_\tau\|_Y,$$

which implies (64). $\square$

In order to prove an error estimate for the discrete solution u_τ^0 of problem (60), we assume the following approximation property: There exists an interpolation operator $\Pi_\tau : X_0 \rightarrow X_{\tau,0}^k$ such that, for all sufficiently smooth $u \in X_0$ and all time intervals I_n , $n = 1, \dots, N$, it holds

$$\|u - \Pi_\tau u\|_{L^2(I_n,V)} \leq C \tau_n^{k+1} \|d_t^{k+1} u\|_{L^2(I_n,V)}, \quad (65)$$

$$\|d_t(u - \Pi_\tau u)\|_{L^2(I_n,V)} \leq C \tau_n^k \|d_t^{k+1} u\|_{L^2(I_n,V)}. \quad (66)$$

Now we are ready to prove an optimal error estimate for the solution $u_\tau^0 \in X_{\tau,0}^k$ of problem (60).

Theorem 5 *Let the solution $u^0 : [0, T] \rightarrow V$ of the continuous problem (59) be sufficiently smooth with respect to time such that the approximation properties (65) and (66) are fulfilled for $u = u^0$. Then, the solution $u_\tau^0 \in X_{\tau,0}^k$ of the discrete problem (60) of the exact dGP(k)-method satisfies the following error estimate*

$$\|u^0 - u_\tau^0\|_X \leq C \left\{ \sum_{n=1}^N \tau_n^{2k} \|d_t^{k+1} u^0\|_{L^2(I_n, V)}^2 \right\}^{1/2} \leq C \tau^k \|d_t^{k+1} u^0\|_{L^2(I, V)} \quad (67)$$

where the constants C are independent of τ_n , τ and T .

Proof. We apply the approximation properties (65) and (66) to $u = u^0$ and get

$$\|u^0 - \Pi_\tau u^0\|_X \leq C \left\{ \sum_{n=1}^N \tau_n^{2k} \|d_t^{k+1} u^0\|_{L^2(I_n, V)}^2 \right\}^{1/2} \leq C \tau^k \|d_t^{k+1} u^0\|_{L^2(I, V)}. \quad (68)$$

For the discrete function $w_\tau := u_\tau^0 - \Pi_\tau u^0$, there exists, due to the inf-sup condition (61), a function $v_\tau \in Y_\tau^k$ such that

$$C_0 \|w_\tau\|_X \|v_\tau\|_Y \leq B(w_\tau, v_\tau) = B(u^0 - \Pi_\tau u^0, v_\tau) \leq C_5 \|u^0 - \Pi_\tau u^0\|_X \|v_\tau\|_Y$$

where the Galerkin orthogonality has been exploited. Using (68) we get

$$\|w_\tau\|_X \leq C \left\{ \sum_{n=1}^N \tau_n^{2k} \|d_t^{k+1} u^0\|_{L^2(I_n, V)}^2 \right\}^{1/2} \leq C \tau^k \|d_t^{k+1} u^0\|_{L^2(I, V)}$$

which proves (67) by the triangle inequality. $\square$

In the following, we will derive an analogous estimate for the (non-exact) dGP(k)-method based on the time marching process given in (29). We use the notation of Section 3. The basis representation of the corresponding global discrete solution $u_\tau^0 \in X_{\tau,0}^k$ reads:

$$u_\tau^0(t) = \sum_{n=1}^N \sum_{j=0}^k U_n^j \phi_{n,j}(t) \quad \forall t \in [0, T] \quad (69)$$

where each local basis function $\phi_{n,j}(t)$ is defined to be zero outside of $\bar{I}_n = [t_{n-1}, t_n]$. For the function $f_0 : [0, T] \rightarrow H'$, we define the piecewise polynomial interpolate $\pi_k f_0 : [0, T] \rightarrow H'$ such that

$$\pi_k f_0(t) = \sum_{j=0}^k f_0(t_{n,j}) \phi_{n,j}(t) \quad \forall t \in \bar{I}_n. \quad (70)$$

Since the $(k+1)$ -Gauß-Lobatto quadrature rule is exact for polynomials of degree $2k-1$ we get for each test basis function $\psi_{n,i} \in \mathbb{P}_{k-1}(I_n)$ that

$$\int_{I_n} \langle \pi_k f_0(t), v \rangle \psi_{n,i}(t) dt = \frac{\tau_n}{2} \sum_{j=0}^k \beta_{i,j} \langle f_0(t_{n,j}), v \rangle \quad \forall v \in H.$$

Moreover, we have for all $i = 0, \dots, k-1$:

$$\int_{I_n} \langle d_t u_\tau^0(t) + Au_\tau^0(t), v \rangle \psi_{n,i}(t) dt = \sum_{j=0}^k \left\{ \alpha_{i,j} (U_n^j, v)_H + \frac{\tau_n}{2} \beta_{i,j} \langle AU_n^j, v \rangle \right\} \quad \forall v \in V.$$

Thus, the time marching process that computes the coefficients U_n^j of the representation (69) is equivalent to the following variational formulation : Find $u_\tau^0 \in X_{\tau,0}^k$ such that

$$B(u_\tau^0, v_\tau) = \int_0^T \langle \pi_k f_0(t), v_\tau(t) \rangle dt \quad \forall v_\tau \in Y_\tau^k. \quad (71)$$

We assume that the function $f : [0, T] \rightarrow H'$ is sufficiently smooth with respect to time such that for each time interval I_n it holds

$$\|f - \pi_k f\|_{L^2(I_n, H')} = \|f_0 - \pi_k f_0\|_{L^2(I_n, H')} \leq C \tau_n^{k+1} \|d_t^{k+1} f\|_{L^2(I_n, H')}. \quad (72)$$

Note, that the time-dependent functions $f(t)$ and $f_0(t) = f(t) - Au_0$ differ only by the time-constant $Au_0 \in H'$.

Theorem 6 *Let the solution $u^0 : [0, T] \rightarrow V$ of the continuous problem (59) be sufficiently smooth with respect to time such that the approximation properties (65) and (66) are fulfilled for $u = u^0$. Assume that $u_0 \in D(A)$, i.e. $Au_0 \in H'$, and that the function $f : [0, T] \rightarrow H'$ is sufficiently smooth with respect to time such that the approximation property (72) is satisfied. Then, the solution $u_\tau^0 \in X_{\tau,0}^k$ of the discrete problem (71) of the dGP(k)-method exists uniquely and satisfies the following error estimate*

$$\begin{aligned} \|u^0 - u_\tau^0\|_X &\leq C \left(\sum_{n=1}^N \tau_n^{2k} \left\{ \|d_t^{k+1} u^0\|_{L^2(I_n, V)}^2 + \tau_n^2 \|d_t^{k+1} f\|_{L^2(I_n, H')}^2 \right\} \right)^{1/2} \\ &\leq C \tau^k \left\{ \|d_t^{k+1} u^0\|_{L^2(I, V)} + \tau \|d_t^{k+1} f\|_{L^2(I, H')} \right\} \end{aligned} \quad (73)$$

where the constants C are independent of τ_n , τ and T .

Proof. Existence and uniqueness of the solution $u_\tau^0 \in X_{\tau,0}^k$ easily follow from Theorem 4 applied to the right hand side $\pi_k f_0$ instead of f_0 . The norm $\|\pi_k f_0\|_{L^2(I, V')}$ is of course bounded if $Au_0 \in H'$ and f is sufficiently smooth.

For the discrete function $w_\tau := \Pi_\tau u^0 - u_\tau^0$, there exists, due to the inf-sup condition (61), a function $v_\tau \in Y_\tau^k$ such that

$$\begin{aligned} C_0 \|w_\tau\|_X \|v_\tau\|_Y &\leq B(w_\tau, v_\tau) = B(u^0, v_\tau) - B(u_\tau^0, v_\tau) - B(u^0 - \Pi_\tau u^0, v_\tau) \\ &= \int_0^T \langle f_0(t) - \pi_k f_0(t), v_\tau(t) \rangle dt - B(u^0 - \Pi_\tau u^0, v_\tau) \\ &\leq \|f_0 - \pi_k f_0\|_{L^2(I, H')} \|v_\tau\|_{L^2(I, H)} + C_5 \|u^0 - \Pi_\tau u^0\|_X \|v_\tau\|_Y \\ &\leq C \left(\sum_{n=1}^N \left\{ \|f_0 - \pi_k f_0\|_{L^2(I_n, H')}^2 \right\} + \|u^0 - \Pi_\tau u^0\|_{L^2(I_n, X)}^2 \right)^{1/2} \|v_\tau\|_Y \end{aligned}$$

where the inequality $\|v_\tau(t)\|_H \leq C\|v_\tau(t)\|_V$ due to the continuous embedding $V \subset H$ has been exploited. By means of the approximation properties (72) and (68) we get

$$\|w_\tau\|_X \leq C \left(\sum_{n=1}^N \tau_n^{2k} \left\{ \|d_t^{k+1} u^0\|_{L^2(I_n, V)}^2 + \tau_n^2 \|d_t^{k+1} f\|_{L^2(I_n, H')}^2 \right\} \right)^{1/2}$$

which implies (73) by the triangle inequality. $\square$

Corollary 7 *According to the solutions $u^0 \in X_0$ and $u_\tau^0 \in X_{\tau,0}^k$ of the continuous and discrete homogeneous problems (59) and (60) or (71), let $u(t) = u_0 + u^0(t)$ and $u_\tau(t) = u_0 + u_\tau^0(t)$ denote the corresponding solutions of the original problems (3) and (17) with a non-homogeneous initial value $u_0 \in D(A) \subset V$. Then, we have*

$$u^0(t) - u_\tau^0(t) = u(t) - u_\tau(t) \quad \text{and} \quad d_t^{k+1} u^0(t) = d_t^{k+1} u(t) \quad \forall t \in [0, T]$$

and the error estimates of Theorems 5 and 6 hold analogously also for the solutions u and u_τ with the non-homogeneous initial value.

Remark 8 *The localized error bounds in Theorems 5 and 6 with the sum of error parts over the time intervals I_n show that it makes sense to choose a small time step size τ_n for an interval I_n on which the local norms of derivatives of the exact solution u or the right hand side f are large. So it is useful to have a local indicator that can detect a large local norm $\|d_t^{k+1} u\|_{L^2(I_n, V)}$.*

In order to prove a better L^2 -error bound we need the following local Poincaré-type inequality.

Lemma 9 *Let $I_n = (t_{n-1}, t_n)$ be a time interval and $u \in H^1(I_n, V)$. Then, it holds the inequality*

$$\|u\|_{L^2(I_n, V)}^2 \leq 2\tau_n \|u(t_{n-1})\|_V^2 + \tau_n^2 \|d_t u\|_{L^2(I_n, V)}^2. \quad (74)$$

Proof. The assumption $u \in H^1(I_n, V)$ implies $u \in C([t_{n-1}, t_n], V)$, see e.g. [4, Theorem 2 in §5.9], such that the "point values" $u(t) \in V$ are well-defined for all $t \in [t_{n-1}, t_n]$ and that it holds

$$u(t) = u(t_{n-1}) + \int_{t_{n-1}}^t d_t u(s) ds \quad \text{in } V.$$

This gives

$$\begin{aligned} \|u(t)\|_V &\leq \|u(t_{n-1})\|_V + \int_{t_{n-1}}^t \|d_t u(s)\|_V ds \\ &\leq \|u(t_{n-1})\|_V + \left(\int_{t_{n-1}}^t 1 ds \right)^{1/2} \left(\int_{t_{n-1}}^t \|d_t u(s)\|_V^2 ds \right)^{1/2} \end{aligned}$$

for all $t \in [t_{n-1}, t_n]$ and therefore

$$\|u(t)\|_V^2 \leq 2\|u(t_{n-1})\|_V^2 + 2(t - t_{n-1})\|d_t u\|_{L^2(I_n, V)}^2.$$

Integration over the interval I_n yields the inequality (74). $\square$

We want to apply a duality argument to prove an L^2 error estimate. Therefore, we need the following regularity assumption for the dual problem. Let $A' : V \rightarrow V'$ be the dual operator of A defined by

$$\langle A'u, v \rangle := a'(u, v) := a(v, u) \quad \forall u, v \in V$$

where the bilinear form $a' : V \times V \rightarrow \mathbb{R}$ satisfies the analogous assumptions to (8) and (9). Consider the problem: Find $w \in X_0$ such that

$$\int_0^T \langle d_t w(t), v(t) \rangle dt + \int_0^T \langle A'w(t), v(t) \rangle dt = \int_0^T (r(t), v(t))_H dt \quad \forall v \in Y. \quad (75)$$

By Theorem 6.6 in [3] we know that this problem has a unique solution $w \in X_0$ for each right hand side $r \in L^2(I, H)$. Now, our regularity assumption for the dual problem (75) is that, for each right hand side $r \in L^2(I, H)$, the solution $w \in X_0$ of (75) satisfies $w \in H^1(I, V)$ and the a priori estimate

$$\|d_t w\|_{L^2(I, V)} \leq C_a \|r\|_{L^2(I, H)} \quad (76)$$

where the constant C_a is independent of r and T . An example where this assumption is satisfied is at least the case of a linear ODE system where $V = H = V' = \mathbb{R}^d$.

Theorem 10 *Let the solution $u^0 : [0, T] \rightarrow V$ of the continuous problem (59) be sufficiently smooth with respect to time such that the approximation properties (65) and (66) are fulfilled for $u = u^0$. Assume that the regularity assumption (76) for the dual problem is satisfied. Then, the solution $u_\tau^0 \in X_{\tau, 0}^k$ of the discrete problem (60) of the exact dGP(k)-method satisfies the following error estimate*

$$\|u^0 - u_\tau^0\|_{L^2(I, H)} \leq C\tau \left\{ \sum_{n=1}^N \tau_n^{2k} \|d_t^{k+1} u^0\|_{L^2(I_n, V)}^2 \right\}^{1/2} \leq C\tau^{k+1} \|d_t^{k+1} u^0\|_{L^2(I, V)} \quad (77)$$

where the constants C are independent of τ_n , τ , T and u^0 .

Proof. For the error $e := u^0 - u_\tau^0 \in L^2(I, V)$ we define the continuous dual problem: Find a function $z : [0, T] \rightarrow V$ such that $z(T) = 0$ and

$$-d_t z(t) + A'z(t) = e(t) \quad \forall t \in (0, T).$$

For the function $w(t) := z(T-t)$ we get the problem: Find $w : [0, T] \rightarrow V$ such that $w(0) = 0$ and

$$d_t w(t) + A' w(t) = e(t) \quad \forall t \in (0, T).$$

The weak formulation of this problem is just problem (75) with $r = e$. If we use the regularity assumption (76) and transform back to z we get

$$\|d_t z\|_{L^2(I, V)} \leq C_a \|e\|_{L^2(I, H)} \quad (78)$$

and the equation

$$-\int_0^T \langle d_t z(t), v(t) \rangle dt + \int_0^T \langle A' z(t), v(t) \rangle dt = \int_0^T (e(t), v(t))_H dt \quad \forall v \in Y.$$

Thus, in particular, we have $z \in C(\bar{I}, V)$ since $z \in H^1(I, V)$. Let $\Pi_0 : H^1(I, V) \rightarrow Y_\tau^k$ be an interpolation operator which is defined on each time interval I_n as

$$\Pi_0 z(t) := z(t_{n-1}) \quad \forall t \in I_n.$$

If we apply the inequality (74) to $u = z - \Pi_0 z$ we obtain

$$\|z - \Pi_0 z\|_{L^2(I_n, V)}^2 \leq \tau_n^2 \|d_t z\|_{L^2(I_n, V)}^2.$$

In the above variational equation for z , we choose $v = e$, apply partial integration to the first integral term, use the fact that $e(0) = 0$ and the Galerkin orthogonality of the dGP(k)-method and get

$$\begin{aligned} \|e\|_{L^2(I, H)}^2 &= \int_0^T \langle d_t e(t), z(t) \rangle dt + \int_0^T a(e(t), z(t)) dt \\ &= \int_0^T \langle d_t e(t), z(t) - \Pi_0 z(t) \rangle dt + \int_0^T a(e(t), z(t) - \Pi_0 z(t)) dt \\ &\leq \|d_t e\|_{L^2(I, V')} \|z - \Pi_0 z\|_{L^2(I, V)} + C_2 \|e\|_{L(I, V)} \|z - \Pi_0 z\|_{L^2(I, V)} \\ &\leq \{ \|d_t e\|_{L^2(I, V')} + C_2 \|e\|_{L(I, V)} \} \tau \|d_t z\|_{L^2(I, V)} \\ &\leq C \|e\|_X C_a \tau \|e\|_{L^2(I, H)}. \end{aligned}$$

We apply the error estimate of Theorem 5 and obtain the estimate (77). $\square$

Note, that the statement of Corollary 7 holds in an analogous sense.

6.2 Ordinary differential equation in $\mathbb{R}^d$

We consider the setting of Sect. 2.2. Although $V = \mathbb{R}^d$ we will write the norm $\|\cdot\|_{C(I, V)}$ instead of $\|\cdot\|_{C(I, \mathbb{R}^d)}$ because this stronger suggests that the results can be generalized to the case of an infinite dimensional separable Hilbert space V with an inner product $(\cdot, \cdot)$ and a

norm $\|\cdot\|$. The Lagrange type interpolation operator $\Pi_k : C(I, V) \rightarrow X_\tau^k$, defined on each time interval I_n by

$$\Pi_k u(t) := \sum_{j=0}^k u(t_{n,j}) \phi_{n,j}(t) \quad \forall t \in I_n,$$

satisfies for all time intervals $I_n = [t_{n-1}, t_n]$, $n = 1, \dots, N$, the approximation property

$$\|u - \Pi_k u\|_{C(I_n, V)} \leq \frac{1}{(k+1)!} \tau_n^{k+1} \|d_t^{k+1} u\|_{C(I_n, V)} \quad \forall u \in C^{k+1}(I_n, V). \quad (79)$$

For the error analysis, we need a special interpolation operator $R_k : C^1(I, V) \rightarrow X_\tau^k$ with the properties

$$\int_{I_n} (d_t R_k u(t), v_\tau(t)) dt = \int_{I_n} (d_t u(t), v_\tau(t)) dt \quad \forall v_\tau \in Y_\tau^k, \quad \forall I_n = [t_{n-1}, t_n] \quad (80)$$

and

$$\|u - R_k u\|_{C(I_n, V)} \leq C_R \tau_n^{k+1} \|d_t^{k+1} u\|_{C(I_n, V)} \quad \forall u \in C^{k+1}(I_n, V), \quad \forall n = 1, \dots, N. \quad (81)$$

The continuous interpolate $R_k u : I \rightarrow V$ is defined on each time interval I_n by the continuity conditions

$$R_k u(t_{n-1}) = u(t_{n-1}) \quad \text{and} \quad R_k u(t_n) = u(t_n) \quad (82)$$

and the $k-1$ equations

$$\int_{I_n} (R_k u(t), v \theta_{n,i}(t)) dt = \int_{I_n} (u(t), v \theta_{n,i}(t)) dt \quad \forall v \in V, \quad \forall i = 1, \dots, k-1 \quad (83)$$

which disappear in the case $k=1$. The functions $\theta_{n,i}$ are basis functions of the polynomial space $\mathbb{P}_{k-2}(I_n)$ defined via the reference transformation (20) from the basis functions $\hat{\theta}_i \in \mathbb{P}_{k-2}(\hat{I})$ satisfying

$$\hat{\theta}_i(\hat{t}_j) = \frac{1}{\hat{w}_j} \delta_{i,j} \quad \forall i, j \in \{1, \dots, k-1\}$$

where $\hat{t}_j$ and $\hat{w}_j$ denote the points and weights of the $(k+1)$ -Gauß-Lobatto quadrature rule on $[-1, 1]$. We show that the conditions (82) and (83) guarantee the property (80). It is sufficient to show it for test functions $v_\tau(t) = v\psi(t)$ where $v \in V$ is arbitrary and $\psi \in \mathbb{P}_{k-1}(I_n)$. Then, the fact that $\psi' \in \mathbb{P}_{k-2}(I_n)$ implies together with (82) and (83) that

$$\begin{aligned} \int_{I_n} (d_t R_k u(t), v\psi(t)) dt &= - \int_{I_n} (R_k u(t), v\psi'(t)) dt + (R_k u(t_n), v\psi(t_n)) \\ &\quad - (R_k u(t_{n-1}), v\psi(t_{n-1})) \\ &= - \int_{I_n} (u(t), v\psi'(t)) dt + (u(t_n), v\psi(t_n)) - (u(t_{n-1}), v\psi(t_{n-1})) \\ &= \int_{I_n} (d_t u(t), v\psi(t)) dt. \end{aligned}$$

In order to determine $R_k u$, we have to determine by the conditions (82) and (83) the "coefficients" U_n^j , $j = 0, \dots, k$, in the ansatz

$$R_k u(t) := \sum_{j=0}^k U_n^j \phi_{n,j}(t) \quad \forall t \in I_n.$$

If we use the functions $\phi_{n,j}$ of Section 3.3 we obtain $U_n^0 = u(t_{n-1})$, $U_n^k = u(t_n)$ and

$$U_n^i = -\hat{w}_0 \hat{\theta}_i(-1) U_n^0 - \hat{w}_k \hat{\theta}_i(1) U_n^k + \frac{2}{\tau_n} \int_{I_n} u(t) \theta_{n,i}(t) dt \quad (84)$$

for all $i = 1, \dots, k-1$. This yields the estimate

$$\sum_{j=0}^k \|U_n^j\| \leq C_6 \|u\|_{C(I_n, V)}$$

with

$$C_6 = 2 + \sum_{i=1}^{k-1} (|\hat{w}_0| + |\hat{w}_k| + 2) \sup_{\hat{t} \in [-1, 1]} |\hat{\theta}_i(\hat{t})|.$$

where the sum term is defined to be zero in the case $k = 1$. We conclude for all $u \in C(I, V)$

$$\|R_k u\|_{C(I_n, V)} \leq C_7 \|u\|_{C(I_n, V)} \quad \text{with} \quad C_7 := C_6 \max_{0 \leq j \leq k} \sup_{\hat{t} \in [-1, 1]} |\hat{\phi}_j(\hat{t})|.$$

Using the fact that $R_k u = u$ on I_n for all $u \in \mathbb{P}_k(I_n, V)$ and (79) we get for all $n \geq 1$

$$\|u - R_k u\|_{C(I_n, V)} \leq \|u - \Pi_k u\|_{C(I_n, V)} + \|R_k(u - \Pi_k u)\|_{C(I_n, V)} \leq (1 + C_7) \|u - \Pi_k u\|_{C(I_n, V)}$$

which leads to the estimate (81) with $C_R = (1 + C_7)/(k+1)!$.

Theorem 11 *Let the solution $u : I = [0, T] \rightarrow V$ of the ODE system (2) be smooth enough such that $d_t^{k+1} u \in C(I, V)$. Assume that the function $F : I \times V \rightarrow V$ is Lipschitz-continuous with the constant L (see (13)) and that the maximum time step size τ satisfies the condition $L\tau \leq \delta_0$ where the constant $\delta_0 \leq \frac{1}{2}$ is sufficiently small but independent of L , T , τ and u . Then, there exists a unique solution $u_\tau \in X_\tau^k$ of the discrete problem (17) of the exact dGP(k)-method and for each time interval I_n , $n = 1, \dots, N$, it holds the error estimate*

$$\|u - u_\tau\|_{C(I_n, V)} \leq C \max_{1 \leq \mu \leq n} \tau_\mu^{k+1} \|d_t^{k+1} u\|_{C(I_\mu, V)} \quad (85)$$

where $C = 2C_R e^{2Lt_{n-1}}$.

Proof. In order to show the existence and uniqueness of the discrete solution, we have to show that the equations (25) for the coefficients U_n^j in the ansatz (14) have a unique solution.

Using the test functions $\hat{\psi}_i \in \mathbb{P}_{k-1}(\hat{I})$ as in Section 3.3, we have the property (39) which leads in our context to the following system for the k "unknowns" U_n^j with $j = 1, \dots, k$:

$$\begin{aligned} U_n^{i+1} + \alpha_{i,k} U_n^k &= -\alpha_{i,0} U_n^0 + \frac{\tau_n}{2} \int_{\hat{I}} F \left(\omega_n(\hat{t}), \sum_{j=0}^k U_n^j \hat{\phi}_j(\hat{t}) \right) \hat{\psi}_i(\hat{t}) d\hat{t} \quad \forall 0 \leq i \leq k-2, \\ U_n^k &= U_n^0 + \frac{\tau_n}{2} \int_{\hat{I}} F \left(\omega_n(\hat{t}), \sum_{j=0}^k U_n^j \hat{\phi}_j(\hat{t}) \right) \hat{\psi}_{k-1}(\hat{t}) d\hat{t}. \end{aligned}$$

If we multiply the second equation by $\alpha_{i,k}$ and subtract it from the first equation we get for all $i = 1, \dots, k$ a system of the form

$$U_n^i = \alpha_i U_n^0 + \frac{\tau_n}{2} \int_{\hat{I}} F \left(\omega_n(\hat{t}), \sum_{j=0}^k U_n^j \hat{\phi}_j(\hat{t}) \right) \tilde{\psi}_i(\hat{t}) d\hat{t}$$

where $\tilde{\psi}_i(\hat{t}) := \hat{\psi}_{i-1}(\hat{t}) - \alpha_{i-1,k} \hat{\psi}_{k-1}(\hat{t})$ and $\alpha_i := -\alpha_{i-1,0} - \alpha_{i-1,k}$ if $i = 1, \dots, k-1$ and $\tilde{\psi}_k(\hat{t}) := \hat{\psi}_{k-1}(\hat{t})$, $\alpha_k := 1$. We introduce the vectors $\vec{U}_n := (U_n^i)_{i=1}^k \in V^k$ and $\vec{U}_n^0 := (\alpha_i U_n^0)_{i=1}^k \in V^k$ and the notations

$$u_\tau(t, \vec{U}_n) := \sum_{j=0}^k U_n^j \phi_{n,j}(t) \quad \forall t \in I_n, \quad U_n^0 := u(t_{n-1}) \quad \text{fixed},$$

and $\vec{F}_n(\vec{U}_n) := (F_n^i(\vec{U}_n))_{i=1}^k \in V^k$ with

$$F_n^i(\vec{U}_n) := \int_{I_n} F \left(t, u_\tau(t, \vec{U}_n) \right) \tilde{\psi}_i(\omega_n^{-1}(t)) dt.$$

Then, we can rewrite our system as the fixed point iteration: Find $\vec{U}_n \in V^k$ such that

$$\vec{U}_n = \vec{\Phi}(\vec{U}_n) := \vec{U}_n^0 + \vec{F}_n(\vec{U}_n).$$

To show that $\vec{\Phi} : V^k \rightarrow V^k$ is a contraction we define on V^k the norm

$$\|\vec{U}_n\|_{V^k} := \max_{1 \leq i \leq k} \|U_n^i\| \quad \forall \vec{U}_n := (U_n^i)_{i=1}^k \in V^k.$$

Then, we have for all $i = 1, \dots, k$ and all $\vec{U}_n^1, \vec{U}_n^2 \in V^k$

$$\begin{aligned} \|F_n^i(\vec{U}_n^1) - F_n^i(\vec{U}_n^2)\| &\leq C_8 L \tau_n \max_{t \in I_n} \|u_\tau(t, \vec{U}_n^1) - u_\tau(t, \vec{U}_n^2)\| \\ &\leq C_9 L \tau_n \|\vec{U}_n^1 - \vec{U}_n^2\|_{V^k} \end{aligned}$$

where

$$C_8 = \max_{1 \leq i \leq k} \max_{\hat{t} \in \hat{I}} |\tilde{\psi}_i(\hat{t})| \quad \text{and} \quad C_9 = C_8 \sum_{j=1}^k \max_{\hat{t} \in \hat{I}} |\hat{\phi}_j(\hat{t})|.$$

This implies

$$\|\vec{\Phi}(\vec{U}_n^1) - \vec{\Phi}(\vec{U}_n^2)\|_{V^k} \leq C_9 L \tau_n \|\vec{U}_n^1 - \vec{U}_n^2\|_{V^k} \quad \forall \vec{U}_n^1, \vec{U}_n^2 \in V^k$$

which says that $\vec{\Phi} : V^k \rightarrow V^k$ is a contraction provided that $L\tau_n \leq \delta_0$ with $\delta_0 < 1/C_9$. Thus, we get the existence and uniqueness of the discrete solution from Banach's fixed point theorem under the assumption $L\tau \leq \delta_0$.

Now, we will prove the estimate (85). We split the error $e(t) := u(t) - u_\tau(t)$ into $e = \eta + w_\tau$ where

$$\eta(t) := u(t) - R_k u(t) \quad \text{and} \quad w_\tau(t) := R_k u(t) - u_\tau(t)$$

and introduce the abbreviations

$$e_n := \|e\|_{C(I_n, V)}, \quad \eta_n := \|\eta\|_{C(I_n, V)} \quad \text{and} \quad w_n := \|w_\tau\|_{C(I_n, V)}.$$

Let $v_\tau \in Y_\tau^k$ be a function such that $v_\tau(t) = 0_V$ for all $t \in I \setminus I_n$. Then, the property (80) of the operator R_k implies

$$\begin{aligned} \int_{I_n} (d_t w_\tau(t), v_\tau(t)) dt &= \int_{I_n} (F(t, u(t)) - F(t, u_\tau(t)), v_\tau(t)) dt \\ &\leq L \int_{I_n} \|e(t)\| \|v_\tau(t)\| dt \\ &\leq L\tau_n^{1/2} e_n \|v_\tau\|_{L^2(I_n, V)}. \end{aligned} \tag{86}$$

For each $t \in I_n = [t_{n-1}, t_n]$, it holds

$$w_\tau(t) = w_\tau(t_{n-1}) + \int_{t_{n-1}}^t d_t w_\tau(s) ds.$$

Thus, it follows

$$\|w_\tau(t)\| \leq \|w_\tau(t_{n-1})\| + \left(\int_{t_{n-1}}^t 1 ds \right)^{1/2} \|d_t w_\tau\|_{L^2(I_n, V)}$$

which gives

$$w_n \leq \|w_\tau(t_{n-1})\| + \tau_n^{1/2} \|d_t w_\tau\|_{L^2(I_n, V)}.$$

We choose in (86) the test function $v_\tau = d_t w_\tau|_{I_n} \in \mathbb{P}_{k-1}(I_n, V)$ and get

$$\|d_t w_\tau\|_{L^2(I_n, V)} \leq L\tau_n^{1/2} e_n. \tag{87}$$

Therefore, using $e_n \leq w_n + \eta_n$ we obtain

$$e_n \leq \|w_\tau(t_{n-1})\| + L\tau_n e_n + \eta_n$$

With $L\tau_n \leq 1/2$ we conclude

$$e_n \leq 2\|w_\tau(t_{n-1})\| + 2\eta_n.$$

Using the fact that $w_\tau(0) = 0_V$ and (87) for arbitrary n , we get for all $n \geq 2$

$$\|w_\tau(t_{n-1})\| \leq \int_0^{t_{n-1}} \|d_t w_\tau\| dt \leq \sum_{\mu=1}^{n-1} \tau_\mu^{1/2} \|d_t w_\tau\|_{L^2(I_\mu, V)} \leq \sum_{\mu=1}^{n-1} L \tau_\mu e_\mu$$

and finally the inequality

$$e_n \leq \sum_{\mu=1}^{n-1} 2L \tau_\mu e_\mu + 2 \max_{1 \leq \mu \leq n} \eta_\mu$$

By means of the discrete Gronwall inequality we obtain

$$e_n \leq 2 \exp \left\{ \sum_{\mu=1}^{n-1} 2L \tau_\mu \right\} \max_{1 \leq \mu \leq n} \eta_\mu \leq 2C_R e^{2Lt_{n-1}} \max_{1 \leq \mu \leq n} \left\{ \tau_\mu^{k+1} \|d_t^{k+1} u\|_{C(I_\mu, V)} \right\}.$$

□

7 Numerical results

As a test example we consider the following linear ODE system: Find $u : [0, 1] \rightarrow \mathbb{R}^2$ such that

$$\begin{aligned} d_t u(t) &= A u(t) \quad \forall t \in [0, 1], \\ u(0) &= u_0, \end{aligned} \quad \text{where } A = \begin{pmatrix} -0.5 & 0.1 \\ 10 & -50 \end{pmatrix}, \quad u_0 = \begin{pmatrix} 2 \\ 1.6 \end{pmatrix}. \quad (88)$$

The eigenvalues of A are $\lambda_1 \approx -0.4798$ and $\lambda_2 \approx -50.0202$ such that the ODE system is relatively stiff. We apply the dGP(2)-method (which is the exact one here) with an equidistant time step size τ . Figure 1 shows the graphs of the two components of the exact solution $u(t)$ and of the discrete solution $u_\tau(t)$ for $\tau = 1/10$ and $\tau = 2/10$. Even for these relatively coarse step sizes, the discrete solution is rather accurate and also stable without any oscillations. In Table 1, the error norms

$$\|u - u_\tau\|_\infty := \max_{1 \leq n \leq N} \|u(t_n) - u_\tau(t_n)\|$$

with the Euclidean norm $\|\cdot\|$ and $\tau = 1/N$ are presented. We see that the dGP(2)-method is superconvergent of order 4 at the points t_n .

Acknowledgment

I would like to thank Klaus Deckelnick for the fruitful discussions leading me to the development of the presented method in the context of the numerical computation of Willmore flow [2].

Figure 1: Components of the exact and discrete $dGP(2)$ solution for $\tau = 1/10$ (left) and $\tau = 1/20$ (right).

τ	$\ u - u_\tau\ _\infty$	EOC
1/10	1.170 e-01	
1/20	1.875 e-02	2.6413
1/40	1.592 e-03	3.5583
1/80	9.301 e-05	4.0968
1/160	5.858 e-06	3.9891
1/320	3.645 e-07	4.0063

Table 1: Error norms $\|u - u_\tau\|_\infty$ for the test problem with known analytical solution and experimental orders of convergence (EOC).

References

- [1] A. K. Aziz and Peter Monk. Continuous finite elements in space and time for the heat equation. *Math. Comp.*, 52(186):255–274, 1989.
- [2] K. Deckelnick and F. Schieweck. Error analysis for the approximation of axisymmetric willmore flow by C^1 -finite elements. Preprint 23/09, Otto-von-Guericke Universität Magdeburg, Fakultät für Mathematik, 2009.
- [3] Alexandre Ern and Jean-Luc Guermond. *Theory and practice of finite elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.
- [4] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.
- [5] Vidar Thomée. *Galerkin finite element methods for parabolic problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006.