MATHEMATICS OF COMPUTATION
Volume 71, Number 239, Pages 1075–1103
S 0025-5718(01)01364-3
Article electronically published on November 21, 2001

THE DYNAMICAL BEHAVIOR OF THE DISCONTINUOUS GALERKIN METHOD AND RELATED DIFFERENCE SCHEMES

DONALD J. ESTEP AND ANDREW M. STUART

Abstract. We study the dynamical behavior of the discontinuous Galerkin finite element method for initial value problems in ordinary differential equations. We make two different assumptions which guarantee that the continuous problem defines a dissipative dynamical system. We show that, under certain conditions, the discontinuous Galerkin approximation also defines a dissipative dynamical system and we study the approximation properties of the associated discrete dynamical system. We also study the behavior of difference schemes obtained by applying a quadrature formula to the integrals defining the discontinuous Galerkin approximation and construct two kinds of discrete finite element approximations that share the dissipativity properties of the original method.

1. Introduction

In this paper, we study the dynamical behavior of the discontinuous Galerkin (dG) finite element method for the initial value problem

$$\begin{cases} \dot{y} = f(y,t), & 0 < t \le T, \\ y(0) = y_0 \in \mathbf{R}^d, & d \ge 1, \end{cases} \tag{1.1}$$

as well as the autonomous counterpart, under various assumptions that guarantee some control over the long-time dynamical behavior of the system.

The use of finite elements to discretize time has a relatively long tradition in several areas of engineering such as neutron transport, analysis of multi-body structures, and optimal control and recently has garnered increased interest in a wide range of areas including conservation laws, fluid flow, and Hamilton-Jacobi equations. See the proceedings of the recent conference on dG methods in Cockburn, Karniadakis, and Shu [2] for various applications. Often the use of finite elements is motivated by their close relationship to variational analysis, which is the natural framework for understanding many sorts of physical models. The discontinuous Galerkin method that we study in this paper has found wide-spread application

Received by the editor May 24, 1999 and, in revised form, September 12, 2000.
2000 *Mathematics Subject Classification.* Primary 65L07.
Key words and phrases. Attractors, contractivity, discontinuous Galerkin method, dissipativity, dynamical system, existence, initial value problems, quadrature.
The research of the first author was partially supported by the National Science Foundation, DMS 9805748.
The research of the second author was partially supported by the Office of Naval Research under grant No. N00014-92-J-1876 and by the National Science Foundation under grant No. DMS-9201727.

to differential equations with a "dissipative" or "parabolic" nature in particular (see Estep [7], Estep, Larson, and Williams [8], and Johnson [14]), as well as being central to a new approach to computational error estimation (see Eriksson, Estep, Hansbo, and Johnson [6], Estep [7], and Estep, Larson, and Williams [8]).

Our aim is to investigate the underlying reasons that the dG method, as well as finite element methods obtained by applying quadrature to the dG method, is well-suited for two classes of "dissipative" problems. More precisely, we consider problems that are either *contractive* or *dissipative* in the senses defined in Stuart and Humphries [16], Chapter 2. Previous analyses of the dissipative properties of the dG method have been confined to strongly contractive problems, for which there are few practical applications other than the heat equation. The dissipative problems we consider here allow for much more interesting long-time behavior (while avoiding difficulties such as blow-up) and include several physically interesting models, such as the Cahn-Hilliard, Navier-Stokes, and Kuramoto-Sivashinsky equations. Another important class of "dissipative" problems, based on the existence of invariant regions, is considered in Estep, Larson, and Williams [8].

We carry out our analysis by considering the discrete approximation from the viewpoint of dynamical systems since this is the natural language for the analysis of the long time behavior of solutions of dissipative problems. We prove both approximation properties for invariant sets of the dynamical system, i.e., a convergence property as the time-step tends to zero, and the inheritance of global boundedness properties such as contractivity and dissipativity, i.e., stability properties, for fixed time-step.

The results we present in this paper are related to the analysis of Runge-Kutta difference schemes in Humphries and Stuart [13] and Hill [12, 11]. Indeed the *modified* dG methods we consider in this paper can often be interpreted as Runge-Kutta schemes, in which case the analysis in Humphries and Stuart [13] applies. However applying quadrature to a finite element method typically has a strong effect on the stability properties of the discretization and in fact can completely mask the properties of the underlying finite element method. Hence, thinking of difference schemes as finite element methods plus quadrature, the question of why the underlying dG method is well-suited for dissipative problems remains to be addressed. After establishing the reasons for two classes of problems, we then address the issue of understanding which kinds of quadrature rules yield schemes that inherit the dissipative characteristics of the dG method. Another feature of the analysis in this paper is that it is largely based on a variational framework. As mentioned, in contexts in which finite element methods are the natural choice, variational analysis is also natural.

In Section 2, we establish notation and describe the assumptions we make to guarantee that the semigroup generated by the autonomous version of (1.1) is contractive or dissipative. In Section 3, we introduce the discontinuous Galerkin methods and prove existence of solutions for all time under these contractivity or dissipativity assumptions. We also establish uniqueness in the contractive case. Section 4 contains finite time error analysis comparing the discrete time semigroup generated by the dG method with the semigroup for (1.1) in the autonomous case. We prove C^1 error estimates, enabling straightforward application of various results concerning approximation of invariant sets of dynamical systems such as those described in Chapters 6 and 7 of [16]. Section 5 is concerned with the preservation of dissipativity by the dG method. In practice, quadrature is used to evaluate

integrals defining the dG method and this has a strong effect on the dynamical properties of the resulting schemes. In Section 6, we extend our results to include the effect of quadrature on the method.

2. Continuous dynamical systems and dissipativity

In this section we establish the framework of analysis for the continuous system (1.1); the primary purpose is to give a point of reference for the corresponding analysis of the discrete system. Further discussion about the terminology and results we use may be found in [16], Chapter 2.

We begin by discussing the autonomous version of (1.1). In this case, (1.1) defines a *dynamical system* on an open set $\mathcal{U} \subset \mathbf{R}^d$ if for all $y_0 \in \mathcal{U}$ there exists a unique solution of (1.1) with $y(t) \in \mathcal{U}$ for all $t > 0$. The *evolution map* for the dynamical system is the map $S(t) : \mathcal{U} \to \mathcal{U}$ defined by $y(t) = S(t)y_0$. If $\mathcal{V} \subset \mathcal{U}$, then $S(t)\mathcal{V} := \{S(t)y_0, y_0 \in \mathcal{V}\}$. A dynamical system is *continuous with respect to initial data* or *continuous* if given any $y_0 \in \mathcal{U}$, any $T > 0$, and any $\epsilon > 0$, there is a $\delta(y_0, T, \epsilon)$ with $\|S(t)y_0 - S(t)v_0\| < \epsilon$, for all $0 \le t \le T$ and $z_0 \in \mathcal{U}$ with $\|z_0 - v_0\| < \delta$, where $\|\cdot\|$ denotes the Euclidean norm in $\mathbf{R}^d$. We use $(\cdot, \cdot)$ to denote the corresponding inner product. Assuming that f is locally Lipschitz continuous ensures that (1.1) is continuous with respect to initial data.

Dissipativity is defined by the action of the solution operator on bounded sets. A dynamical system is *dissipative* if there is a bounded set $\mathcal{B}$ with the property that for any bounded set $\mathcal{V} \subset \mathcal{U}$ there exists a $t^*(\mathcal{V})$ such that $S(t)\mathcal{V} \subset \mathcal{B}$ for $t > t^*$; $\mathcal{B}$ is called an *absorbing set*. The long time behavior of a dissipative dynamical system is determined by its behavior on any absorbing set. Since an absorbing set is bounded, the action of $S(t)$ on an absorbing set is more easily determined than its action on the whole domain. In particular, the attractor of an absorbing set is in fact a global attractor for the dynamical system.

We introduce concepts which enable us to define the attractor. For any $y_0 \in \mathcal{U}$, the *ω-limit set* of y_0, denoted $\omega(y_0)$, is defined

$$\omega(y_0) := \bigcap_{\tau \ge 0} \overline{\bigcup_{t \ge \tau} S(t)y_0}.$$

For a bounded set $\mathcal{V} \subset \mathcal{U}$, we define

$$\omega(\mathcal{V}) := \bigcap_{\tau \ge 0} \overline{\bigcup_{t \ge \tau} S(t)\mathcal{V}}.$$

The sets $\omega(y_0)$ and $\omega(\mathcal{V})$ are positively invariant under $S(t)$ and they are closed when the dynamical system is continuous.

Next we define the *distance between a point and a set*. Given $\mathcal{V} \subset \mathbf{R}^d$ and $x \in \mathbf{R}^d$,

$$\operatorname{dist}(x, \mathcal{V}) := \inf_{y \in \mathcal{V}} \|x - y\|.$$

For $\mathcal{V}$ and $\mathcal{W} \subset \mathbf{R}^d$,

$$\operatorname{dist}(\mathcal{V}, \mathcal{W}) := \sup_{x \in \mathcal{V}} \operatorname{dist}(x, \mathcal{W}).$$

Note that $\operatorname{dist}(\mathcal{V}, \mathcal{W}) \ne \operatorname{dist}(\mathcal{W}, \mathcal{V})$ and $\operatorname{dist}(\mathcal{V}, \mathcal{W}) = 0$ implies that $\mathcal{V} \subset \overline{\mathcal{W}}$. We define the ϵ neighborhood of $\mathcal{V}$ as

$$N(\mathcal{V}, \epsilon) := \{x : \operatorname{dist}(x, \mathcal{V}) < \epsilon\}.$$

For a continuous dynamical system, $\mathcal{A}$ *attracts* a set $\mathcal{V}$ under $S(t)$ if for any $\epsilon > 0$ there is a $t^*(\epsilon, \mathcal{A}, \mathcal{V}) \geq 0$ such that

$$S(t)\mathcal{V} \subset N(\mathcal{A}, \epsilon),$$

for all $t > t^*$. $\mathcal{A}$ is called a *local attractor* if it is a compact, invariant set that attracts an open neighborhood of itself. It is called a *global attractor* if it attracts all bounded subsets of $\mathcal{U}$. For a dissipative dynamical system, the global attractor is $\mathcal{A} = \omega(\mathcal{B})$, where $\mathcal{B}$ is any absorbing set. It follows that $\mathcal{A}$ is compact and invariant under $S(t)$.

When (1.1) is nonautonomous, we can no longer define a dynamical system in this fashion. However we can still consider the continuity and dissipativity properties of the solution operator, using some natural extensions of the definitions above. In this case, we abuse language by speaking of a dissipative system for example.

We now consider two sets of assumptions on f that guarantee that (1.1) defines a dissipative system. We use $(\ ,\ )$ to denote an inner product on vectors in $\mathbf{R}^d$ and $\|\ \|$ to denote the induced vector norm. First,

Assumption 2.1. There is a constant $c > 0$ such that

$$\big(f(u,t) - f(v,t), u - v\big) \leq -c\|u - v\|^2, \tag{2.1}$$

for all $u, v \in \mathbf{R}^d$ and all $t \geq 0$.

The following extension of Theorems 2.8.4 and 2.8.5 in [16] is straightforward.

Theorem 2.2. *Suppose that f is locally Lipschitz continuous and that Assumption 2.1 holds. Then, any two solutions u, v of (1.1) satisfy*

$$\|u(t) - v(t)\| \leq e^{-ct}\|u(0) - v(0)\|,$$

for all $t \geq 0$. The steady-state solutions define a closed, convex set $\mathcal{E}$ and

$$\lim_{t\to\infty} \operatorname{dist}(u(t), \mathcal{E}) = 0.$$

In particular if f is autonomous and there is a $\bar{u}$ with $f(\bar{u}) = 0$, then $\bar{u}$ is the unique equilibrium point and $\lim_{t\to\infty} u(t) = \bar{u}$ where u is any solution.

Remark 2.3. An f satisfying Assumption 2.1 is called *contractive* in the dynamical systems literature. (Unfortunately it is sometimes called *dissipative* in the numerical differential equations literature.)

The long time behavior of problems satisfying Assumption 2.1 is essentially that of linear decay (see [16] and Dekker and Verwer [3]) and has limited applicability. Therefore we are motivated to consider another class of problems. To define these we use the standard Sobolev spaces $W_p^q(\mathbf{R}^+)$.

Assumption 2.4. There exist a nonnegative function $\alpha \in W_1^1(\mathbf{R}^+) \cap W_\infty^1(\mathbf{R}^+)$ and a positive function $\beta \in W_\infty^1(\mathbf{R}^+)$ with

$$\lim_{t\to\infty} \int_0^t \beta(s)ds = \infty \tag{2.2}$$

and

$$\sup_{t\geq 0} \frac{\alpha(t)}{\beta(t)} = R^2 > 0, \tag{2.3}$$

such that

$$\big(f(v,t),v\big) \le \alpha(t) - \beta(t)\|v\|^2, \tag{2.4}$$

for all $v \in \mathbf{R}^d$ and $t \ge 0$.

This assumption means that f points "inwards" on balls in solution space of sufficiently large radius. Thus, the solution decays when it becomes very large and yet inside some fixed ball it may exhibit a variety of interesting dynamical behavior. This class of problems contains many important models; for example, the Cahn-Hilliard equation, the Navier-Stokes equation in two dimensions and the Kuramoto-Sivashinsky equation all satisfy infinite-dimensional analogs of Assumption 2.4 (see Temam [17]). For a discussion of the preservation of other sorts of invariant regions under discretization and some consequences for computational error estimation, see Estep, Larson, and Williams [8].

Theorem 2.5. *If f is locally Lipschitz continuous and Assumption 2.4 holds, then for any $\epsilon > 0$ there is a $t^*(y_0,\epsilon)$ such that for all $t > t^*$, $\|y(t)\| < R + \epsilon$. If f is autonomous, then* (1.1) *defines a dynamical system on $\mathbf{R}^d$.*

In other words, the system is dissipative and the open ball $\mathcal{B}(0, R+\epsilon)$ with center at 0 and radius $R+\epsilon$ is an absorbing set for any $\epsilon > 0$, and moreover in the autonomous case, (1.1) possesses a global attractor $\mathcal{A} = \omega\big(\mathcal{B}(0,R+\epsilon)\big)$.

Proof. The variational formulation of (1.1) reads: find $y \in C^1((0,T))$ such that

$$\begin{cases} \int_0^T (\dot{y},v)\,dt = \int_0^T (f(y,t),v)\,dt \quad \text{for all } v \in C^1((0,T)), \\ y(0) = y_0. \end{cases} \tag{2.5}$$

Taking $v = y$ in (2.5) and using (2.4), we get

$$\frac{1}{2}\frac{d}{dt}\|y(t)\|^2 = (y, f(y,t)) \le \alpha(t) - \beta(t)\|y(t)\|^2,$$

and setting $B(t) = \int_0^t \beta(s)\,ds$, we conclude that

$$\|y(t)\|^2 \le e^{-2B(t)}\|y(0)\|^2 + 2e^{-2B(t)}\int_0^t \alpha(s)e^{2B(s)}\,ds.$$

By (2.3),

$$0 \le 2\alpha(s)e^{2B(s)} \le 2R^2\beta(s)e^{2B(s)},$$

for $0 \le s \le t$. Hence

$$\|y(t)\|^2 \le R^2 + e^{-2B(t)}\|y(0)\|^2,$$

which shows that the system is dissipative with absorbing set $\mathcal{B}(0,R+\epsilon)$ for any $\epsilon > 0$. In the case that f is autonomous, since it is locally Lipschitz, the global bound on $\|y(t)\|$ implies that (1.1) defines a dynamical system on $\mathbf{R}^d$. □

Throughout the remainder of this paper the constant C denotes a constant independent of the mesh-spacing parameter k; its actual value may change from occurrence to occurrence.

3. Existence and uniqueness of the discontinuous Galerkin approximation

In this section, we define the discontinuous Galerkin finite element method for (1.1) and discuss the existence and uniqueness of the corresponding approximation. We discretize time as $t_0 = 0 < t_1 < t_2 < \cdots \to \infty$ with time intervals $I_m := (t_{m-1}, t_m]$ and time steps $k_m = t_m - t_{m-1}$. We assume that $\sup_m k_m < \infty$. Note this assumption is automatically satisfied if the step size is held constant. Most adaptive algorithms impose a maximum step-size, although they will typically allow the step-size to grow to this maximum near a stable equilibrium point.

The finite element space containing the piecewise polynomial approximate solution is defined as $\mathcal{V}^{(q)} = \left\{V : V\big|_{I_m} \in \mathcal{P}^{(q)}(I_m)\right\}$ where $\mathcal{P}^{(q)}(I_m)$ denotes the set of polynomials of degree q or less on I_m. A function in $\mathcal{V}^{(q)}$ has possibly two values at time nodes, so for $V \in \mathcal{V}^{(q)}$ we set $V_m^{\pm} = \lim_{s \to t_m^{\pm}} V(s)$ and $[V]_m = V_m^+ - V_m^-$.

Roughly speaking, the finite element approximation $Y \in \mathcal{V}^{(q)}$ satisfies the variational equation (2.5) for all test functions in $\mathcal{V}^{(q)}$. This has to be interpreted in the sense of distributions since Y is generally discontinuous. Recalling that the derivative at a point of discontinuity is an appropriately scaled delta function, Y solves the global problem

$$\sum_{m=1}^{n} \int_{I_m} (\dot{Y} - f(Y,t), X)\, dt + \sum_{m=1}^{n} ([Y]_{m-1}, X_{m-1}^+) = 0 \tag{3.1}$$

for all $X \in \mathcal{V}^{(q)}$ and $n = 1, 2, \cdots$, where $Y_0^- = y_0$. In practice, Y can be computed interval by interval since for $m = 1, 2, \cdots$, $Y \in \mathcal{P}^{(q)}(I_m)$ solves

$$\int_{I_m} (\dot{Y}, X)\, dt + (Y_{m-1}^+, X_{m-1}^+) = (Y_{m-1}^-, X_{m-1}^+) + \int_{I_m} (f(Y,t), X)\, dt, \tag{3.2}$$

for all $X \in \mathcal{P}^{(q)}(I_m)$. Thus, Y_{m-1}^- can be considered "initial data" for the computation on the m'th interval.

For $q = 0$, Y is a piecewise constant function whose value on I_m is given by

$$Y_m^- = Y_{m-1}^- + \int_{I_m} f(Y_m^-, t)\, dt.$$

If f is autonomous, then Y agrees at nodes with the values of the backward Euler difference scheme. For $q = 1$, Y is the piecewise linear function on I_m

$$Y\big|_{I_m} = \frac{(t - t_m)}{-k_m}\, Y_{m-1}^+ + \frac{(t - t_{m-1})}{k_m}\, Y_m^-, \tag{3.3}$$

with coefficients determined by

$$\begin{cases} Y_m^- = Y_{m-1}^- + \int_{I_m} f(Y(t), t)\, dt, \\ Y_m^- - Y_{m-1}^+ = 2 \int_{I_m} f(Y(t), t) \frac{(t - t_{m-1})}{k_m}\, dt. \end{cases}$$

The dG method using polynomials of degree q converges with order up to $q + 1$ at all points t while its nodal values from the left converge with order up to $2q + 1$. We refer the reader to Estep [7] and Eriksson, Estep, Hansbo, and Johnson [6] for more information.

The dG approximations are not equivalent to any standard difference scheme in general: for example, we can replace the integral defining the dG approximation

with $q = 0$ by $k_m f(Y_m^-, \tau_m)$ using the mean value theorem, where τ_m is a generally unknown point in (t_{m-1}, t_m). The resulting formula can be interpreted as a difference scheme for the nodal values $\{Y_m^-\}$ with the property that the nodes involved in the difference scheme depend on the nodal values of the approximation to y produced by the scheme. For higher order approximations, both the weights and the nodes depend on the approximation.

In the analysis of the approximation, we use the following *inverse estimates* that follow from standard results about norms on finite dimensional spaces (see Ciarlet [1]). In this result, we use the notation $\|V\|_{L^\infty(I_m)} = \sup_{t\in I_m} \|V(t)\|$.

Proposition 3.1. *For each integer $q \geq 0$ there is a constant $C = C(q)$ such that, for any $V \in \mathcal{V}^{(q)}$ and $m \geq 1$,*

$$k_m \|V\|^2_{L^\infty(I_m)} \leq C \int_{I_m} \|V\|^2 \, dt,$$

and

$$k_m \int_{I_m} \|V\|^2 \, dt \leq C \int_{I_m} (t - t_{m-1}) \|V\|^2 \, dt.$$

We start by discussing existence and uniqueness in the case that f is globally Lipschitz continuous. In this situation, we can specify an iterative process to produce the approximant. Given $a \in \mathbf{R}^d$, we define the map $\Phi_a : \mathcal{P}^{(q)}(I_m) \to \mathcal{P}^{(q)}(I_m)$ by $V = \Phi_a(U)$ if

$$\int_{I_m} (\dot{V}, X) \, dt + (V_{m-1}^+, X_{m-1}^+) = (a, X_{m-1}^+) + \int_{I_m} (f(U,t), X) \, dt, \tag{3.4}$$

for all $X \in \mathcal{P}^{(q)}(I_m)$. This is the dG approximation to the linear problem

$$\begin{cases} \dot{u} = f(U,t), & t_{m-1} < t \leq t_m, \\ u(t_{m-1}) = a, \end{cases}$$

and hence is well-defined for all q. It is easy to show that $Y = \Phi_{Y_{m-1}^-}(Y)$ if and only if Y satisfies the dG equations on I_m. We define the fixed point iteration

$$\begin{cases} Y^{(0)} = Y_{m-1}^-, \\ Y^{(i)} = \Phi_{Y_{m-1}^-}(Y^{(i-1)}), \end{cases} \tag{3.5}$$

and show the convergence in

Theorem 3.2. *Assume that $f(\cdot, t)$ is globally Lipschitz continuous with constant L independent of t and that $f(c, \cdot) \in C^0(\mathbf{R}^+) \cap L^\infty(\mathbf{R}^+)$ for some c. Then there is a constant $C = C(q)$ such that if $k_m < C/L$, the sequence given by* (3.5) *converges to the unique dG approximation on I_m.*

Note that the assumption on $f(c, \cdot)$ for some fixed c serves to give control of the nonhomogeneous part of f, which is not controlled by a Lipshitz assumption on the homogenous part of f. This condition is automatically satisfied for autonomous problems.

Remark 3.3. While we use this fixed point iteration to show the existence and uniqueness of the approximate solution, in practice a hybrid (quasi) Newton iteration would be used in order to try to avoid the time step restriction.

Proof. First we show that Φ_a maps a ball into itself. Given $a \in \mathbf{R}^d$, we choose R so that

$$\max\bigl\{\|a\|, \|c\|, \|f(c,\cdot)\|_{L^\infty(\mathbf{R}^+)}\bigr\} \le R.$$

For $U \in \mathcal{P}^{(q)}(I_m)$, we let $V = \Phi_a(U)$ and choose $X = V$ in (3.4) to get

$$\frac{1}{2}\int_{I_m} \frac{d}{dt}\|V\|^2\,dt + \|V_{m-1}^+\|^2 = (a, V_{m-1}^+) + \int_{I_m} (f(U,t), V)\,dt$$

or

$$\frac{1}{2}\|V_m^-\|^2 + \frac{1}{2}\|V_{m-1}^+\|^2 = (a, V_{m-1}^+) + \int_{I_m} (f(U,t), V)\,dt.$$

We conclude that

$$\|V_m^-\|^2 \le \|a\|^2 + 2\int_{I_m} \bigl|(f(U,t), V)\bigr|\,dt. \tag{3.6}$$

Next we choose $X = (t - t_{m-1})\dot{V}$ in (3.4) to get

$$\int_{I_m} (t - t_{m-1})\|\dot{V}\|^2\,dt = \int_{I_m} (t - t_{m-1})(f(U,t), \dot{V})\,dt.$$

Using Proposition 3.1, we conclude that

$$k_m\|\dot{V}\|^2_{L^\infty(I_m)} \le C\int_{I_m} \bigl|(f(U,t), \dot{V})\bigr|\,dt, \tag{3.7}$$

where $C = C(q)$. Since $\|V - c\|_{L^\infty(I_m)} \le \|V_m^- - c\| + k_m\|\dot{V}\|_{L^\infty(I_m)}$, (3.6) and (3.7) and a straightforward estimate yield

$$\|V - c\|_{L^\infty(I_m)} \le C\|a\| + \|c\| + Ck_m\|f(U(\cdot),\cdot)\|_{L^\infty(I_m)}.$$

By assumption, $\|f(U(\cdot),\cdot)\|_{L^\infty(I_m)} \le \|f(c,\cdot)\|_{L^\infty(I_m)} + L\|U - c\|_{L^\infty(I_m)}$. Hence,

$$\|V - c\|_{L^\infty(I_m)} \le C(1 + k_m)R + CLk_m\|U - c\|_{L^\infty(I_m)}.$$

Now assuming that $CLk_m < 1$, we define $\gamma = C(1 + k_m)/(1 - CLk_m)$. If U satisfies $\|U - c\|_{L^\infty(I_m)} \le \gamma R$, then $\|V - c\|_{L^\infty(I_m)} \le \gamma R$ as well. Hence, $\Phi_a : \mathcal{B}_\infty(c, \gamma R) \to \mathcal{B}_\infty(c, \gamma R)$.

Next we show that Φ_a is a contraction. Suppose that V and $\tilde{V}$ are two solutions of (3.4) corresponding to U and $\tilde{U}$. We subtract the respective equations and use the Lipschitz assumption on f and arguing using Proposition 3.1 as above to obtain

$$\|V - \tilde{V}\|^2_{L^\infty(I_m)} \le CLk_m\|U - \tilde{U}\|_{L^\infty(I_m)}\|V - \tilde{V}\|_{L^\infty(I_m)}.$$

Since $CLk_m < 1$, the map is contractive and we conclude that the iteration (3.5) converges to a unique fixed point of Φ_a in $\mathcal{B}_\infty(c, \gamma R)$. □

Before continuing with existence and uniqueness, we discuss how the dG method can be used to define a discrete time dynamical system when f is autonomous; the terminology is the same as in Chapter 1 of [16]. For fixed time steps $k_m \equiv k$, suppose that (3.2) can be solved for $Y \in \mathcal{P}^{(q)}(I_m)$ and that $Y(t) \in \mathcal{U}$ for each $t \in I_m$. Fix the time step $k > 0$; given $Y_{m-1}^- \in \mathcal{U}$, we define the map $S_k : \mathcal{U} \to \mathcal{U}$ by $S_k Y_{m-1}^- = Y_m^-$. Then for any initial value $y_0 \in \mathcal{U}$, the sequence $\{y_m\}_{m=0}^\infty$ defined by $y_m = S_k y_{m-1}$ is uniquely determined and S_k defines a discrete dynamical system. The definitions of continuity, dissipativity, ω-limit sets, and attractors from Section 2 can be extended in a straightforward way using the map S_k.

Thus, if we assume a constant step size and an autonomous problem, then the following holds.

Corollary 3.4. *Assume that f is autonomous and that a constant step size is used. Then, under the conditions of Theorem 3.2, the dG method defines a discrete dynamical system on $\mathbf{R}^d$ that is continuous with respect to initial data.*

Proof. We already have shown that the equation $\Psi_k(Y_{m-1}^-, Y_m^-) = 0$ defined through (3.4) and (3.5) has a unique solution so it remains to show that the resulting discrete dynamical system is continuous. We let Y and $\tilde{Y}$ denote dG approximations corresponding to Y_{m-1}^- and $\tilde{Y}_{m-1}^-$ respectively. We subtract the equations (3.2) for Y and $\tilde{Y}$ and argue as above using the Lipschitz assumption on f and Proposition 3.1 to get

$$\|Y - \tilde{Y}\|_{L^\infty(I_m)} \le C\|Y_{m-1}^- - \tilde{Y}_{m-1}^-\| + CLk_m\|Y - \tilde{Y}\|_{L^\infty(I_m)}.$$

Thus if $CLk_m < 1$, then

$$\|Y - \tilde{Y}\|_{L^\infty(I_m)} \le C\|Y_{m-1}^- - \tilde{Y}_{m-1}^-\|,$$

and continuity follows. □

We next consider the case when f is known only to be locally Lipschitz continuous. For a set $\mathcal{U} \subset \mathbf{R}^d$, we define

$$\mathcal{U}_\delta = \left\{ x \in \mathcal{U} : \inf_{y \in \mathbf{R}^d \setminus \mathcal{U}} \|x - y\| \ge \delta \right\},$$

and for a positive integer m, $U \in \mathcal{P}^{(q)}(I_m)$, and $\rho \in \mathbf{R}^+$, we define

$$\mathcal{B}_\infty(U, \rho) = \mathcal{B}_\infty(U, \rho, m) = \left\{ V \in \mathcal{P}^{(q)}(I_m) : \|V - U\|_{L^\infty(I_m)} < \rho \right\}.$$

In application, the interval in consideration will be clear. Note that when we write $\mathcal{B}_\infty(U, \rho) \subset \mathcal{U}$ we assume $U(t) \subset \mathcal{U}$ for all $t \in I_m$.

Theorem 3.5. *Suppose $\mathcal{U} \subset \mathbf{R}^d$ is a compact set and f is continuous on $\mathcal{U} \times \mathbf{R}^+$. There is a constant $C = C(q)$ such that if $k_m < C\delta/M$, where*

$$M = \sup_{z \in \mathcal{U}, t \ge 0} \|f(z, t)\|,$$

then for every $Y_{m-1}^- \in \mathcal{U}_\delta$, there is a solution Y of the dG equations with $Y \in \mathcal{B}_\infty(Y_{m-1}^-, \delta) \subset \mathcal{U}$. Moreover, if the iteration (3.5) converges, then it converges to Y. If in addition f is Lipschitz continuous on $\mathcal{U}$ with constant L uniformly in t and $k_m < C \min\{\delta/M, 1/L\}$, then the iteration (3.5) converges to the unique solution $Y \in \mathcal{B}_\infty(Y_{m-1}^-, \delta)$.

Theorem 3.5 is a local result in the sense that M and L may depend on the set $\mathcal{U}$. As the approximation of (1.1) advances, these values may grow with each step and thus the time steps would necessarily decrease with each step. One consequence is that this result alone cannot be used to show that the dG method defines a dynamical system even on a bounded set in $\mathbf{R}^d$ since the approximation may leave the set.

Proof. We define the constant function $W \in \mathcal{P}^{(q)}(I_m)$ by $W = Y^-_{m-1}$, so that $\dot{W} = 0$, $W^+_{m-1} = W^-_m$, and

$$\int_{I_m} (\dot{W}, X)\, dt + (W^+_{m-1}, X^+_{m-1}) = (Y^-_{m-1}, X^+_{m-1}), \tag{3.8}$$

for all $X \in \mathcal{P}^{(q)}(I_m)$. We claim that if $V = \Phi_{Y^-_{m-1}}(U)$ and $U \in \mathcal{B}_\infty(W, \delta)$, then $V \in \mathcal{B}_\infty(W, \delta)$. We subtract (3.8) from (3.4) to get

$$\int_{I_m} (\dot{V} - \dot{W}, X)\, dt + ((V - W)^+_{m-1}, X^+_{m-1}) = \int_{I_m} (f(U, t), X)\, dt,$$

for all $X \in \mathcal{P}^{(q)}(I_m)$. We now argue as in the proof of Theorem 3.2 and use the assumptions to obtain

$$\|V - W\|_{L^\infty(I_m)} \le Ck_m \|f(U(\cdot), \cdot)\|_{L^\infty(I_m)} < \delta.$$

$\mathcal{B}_\infty(W, \delta)$ is a closed, convex subset of the finite-dimensional space $\mathcal{P}^{(q)}(I_m)$ and f is continuous. Hence $\Phi_{Y^-_{m-1}}$ is a continuous map of $\mathcal{B}_\infty(W, \delta)$ into itself and the Brouwer fixed point theorem implies that there is a fixed point of $\Phi_{Y^-_{m-1}}$ in $\mathcal{B}_\infty(W, \delta)$.

Under the additional assumptions in the theorem, we want to show that $\Phi_{Y^-_{m-1}}$ is a contraction. But this is argued as in the proof of Theorem 3.2. □

The time step restrictions in Theorems 3.2 and 3.5 are rarely implemented in practice since this would severely curtail the ability to adapt the steps. We now consider the existence of approximation values when there is no step size restriction but the problem is dissipative.

For the next theorem, we use the notation

$$\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_m = \int_{I_m} (\cdot, \cdot)\, dt, \qquad \|\cdot\|_{L^2(I_m)} = \langle \cdot, \cdot \rangle_m^{1/2},$$

and

$$\mathcal{B}_2(U, \rho) = \mathcal{B}_2(U, \rho, m) = \big\{V \in \mathcal{P}^{(q)}(I_m) : \|V - U\|_{L^2(I_m)} < \rho\big\},$$

and the following proposition, proved in French and Jensen [10] and Temam [17],

Proposition 3.6. *Assume that $\Phi : \mathcal{P}^{(q)}(I_m) \to \mathcal{P}^{(q)}(I_m)$ is continuous and that there is an $R > 0$ such that $\langle \Phi(V), V \rangle < 0$ for all $V \in \partial\mathcal{B}_2(0, R) = \{V \in \mathcal{P}^{(q)}(I_m) : \|V\|_{L^2(I_m)} = R\}$. Then there is a $V^* \in \mathcal{B}_2(0, R)$ such that $\Phi(V^*) = 0$.*

Theorem 3.7. *Assume that f is locally Lipschitz continuous and $f(c, \cdot) \in C^0(\mathbf{R}^+) \cap L^\infty(\mathbf{R}^+)$ for some c. (a) If Assumption 2.1 holds, then for any $k_m > 0$, the dG formulas have a unique solution on I_m. (b) If Assumption 2.4 holds, then for any $k_m > 0$ there is an $r > 0$ such that the dG formulas have at least one solution in $\mathcal{B}_2(0, r)$ and all solutions must lie in $\mathcal{B}_2(0, r)$.*

Proof. We first prove part (b). We fix $a \in \mathbf{R}^d$ and for $U \in \mathcal{P}^{(q)}(I_m)$, we consider the linear functional $L_U : \mathcal{P}^{(q)}(I_m) \to \mathbf{R}$ defined by

$$L_U V = \int_{I_m} (\dot{U}, V) dt + (U^+_{m-1}, V^+_{m-1}) - (a, V^+_{m-1}) - \int_{I_m} (f(U, t), V)\, dt.$$

Using Proposition 3.1 and the assumptions on f, it is straightforward to show that L_U is bounded in the sense that

$$|L_U V| \le C\big(\|U\|_{L^2(I_m)}/k_m, ||f(c,\cdot)||_{L^\infty(I_m)}, \|a\|, \|U-c\|, L(U)\big)\,\|V\|_{L^2(I_m)},$$

for all $V \in \mathcal{P}^{(q)}(I_m)$, where $L(U)$ denotes the local Lipschitz constant of f. Thus, the Riesz representation theorem implies that there is a $\Phi(U) \in \mathcal{P}^{(q)}(I_m)$ such that $L_U V = \big\langle \Phi(U), V\big\rangle_m$ for all $V \in \mathcal{P}^{(q)}(I_m)$. The same argument used to show that L_U is bounded shows that $\Phi(U)$ depends continuously on U.

We now compute, using Assumption 2.4,

$$\begin{aligned}\big\langle \Phi(U), U\big\rangle_m &= \frac{1}{2}||U_m^-||^2 + \frac{1}{2}||U_{m-1}^+||^2 - (a, U_{m-1}^+) - \int_{I_m} (f(U,t), U)\,dt,\\ &\ge \frac{1}{2}||U_m^-||^2 - \frac{1}{2}||a||^2 + \int_{I_m} \beta(t)||U||^2 dt - \int_{I_m} \alpha(t)\,dt,\end{aligned}$$

or

$$\big\langle \Phi(U), U\big\rangle_m \ge \min_{I_m} \beta\,\|U\|^2_{L^2(I_m)} - \frac{1}{2}||a||^2 - \int_{I_m} \alpha(t)\,dt.$$

Hence for all $U \in \partial\mathcal{B}_2(0,r)$ with r sufficiently large, $\big\langle \Phi(U), U\big\rangle_m > 0$. By Proposition 3.6, $\Phi(Y) = 0$ for some $Y \in \mathcal{B}_2(0,r)$. But $L_Y V = 0$ for all $V \in \mathcal{P}^{(q)}(I_m)$ if and only if Y is a dG approximation. Note that this argument also shows that $\Phi(V) \ne 0$ for $V \in \mathcal{P}^{(q)}(I_m) \setminus \mathcal{B}_2(0,r)$.

We give the proof of part (a) for $q = 1$. The proof for higher order q is similar but the notation is cumbersome. In this case, functions $U \in \mathcal{P}^{(q)}(I_m)$ are uniquely associated to vectors $\vec{U} \in \mathbf{R}^{2d}$ via

$$\vec{U} = \begin{pmatrix} U_{m-1}^+ \\ U_m^- \end{pmatrix}.$$

We again fix $a \in \mathbf{R}^d$. Given $\vec{U} \in \mathbf{R}^{2d}$, we define $G(\vec{U}) \in \mathbf{R}^{2d}$ by

$$G(\vec{U}) = \begin{pmatrix} \displaystyle\int_{I_m} \dot{U}\phi_1 dt - \int_{I_m} f(U,t)\phi_1\,dt \\ \displaystyle\int_{I_m} \dot{U}\phi_2 dt + U_{m-1}^+ - a - \int_{I_m} f(U,t)\phi_2\,dt \end{pmatrix},$$

where $\{\phi_1, \phi_2\}$ is the basis for $\mathcal{P}^{(q)}(I_m)$ used in (3.3). Note that $G(\vec{Y}) = 0$ if and only if Y is a dG approximation. We now show that G is uniformly monotone and continuous. Let $\vec{U}, \vec{V} \in \mathbf{R}^{2d}$. Then abusing notation with $(\cdot,\cdot)$ and $||\cdot||$,

$$\begin{aligned}&\big(G(\vec{U}) - G(\vec{V}), \vec{U} - \vec{V}\big)\\ &\quad= \int_{I_m} \big(\dot{U} - \dot{V}, U_m^- - V_m^-\big)\phi_1\,dt - \int_{I_m} \big(f(U,t) - f(V,t), U_m^- - V_m^-\big)\phi_1\,dt\\ &\qquad+ \int_{I_m} \big(\dot{U} - \dot{V}, U_{m-1}^+ - V_{m-1}^+\big)\phi_2\,dt + ||U_{m-1}^+ - V_{m-1}^+||^2\\ &\qquad- \int_{I_m} \big(f(U,t) - f(V,t), U_{m-1}^+ - V_{m-1}^+\big)\phi_2\,dt,\end{aligned}$$

or

$$\big(G(\vec{U}) - G(\vec{V}), \vec{U} - \vec{V}\big) = \int_{I_m} \big(\dot{U} - \dot{V}, U - V\big)\, dt + ||U_{m-1}^+ - V_{m-1}^+||^2 - \int_{I_m} \big(f(U,t) - f(V,t), U - V\big)\, dt.$$

Evaluating the first integral on the right and using Assumption 2.1 for the second gives

$$\big(G(\vec{U}) - G(\vec{V}), \vec{U} - \vec{V}\big) \geq \frac{1}{2}||\vec{U} - \vec{V}||^2.$$

The argument that shows that G is continuous is essentially the same line used to show that the map L_U above is bounded. By the uniform monotonicity theorem (see Ortega and Rheinboldt [15]), G is a homeomorphism of $\mathbf{R}^{2d}$ into itself and hence has a unique root Y. □

Note that if $\alpha \in L^\infty(I_m)$ and $\inf\, \beta > 1/2$, and r is chosen so that

$$r^2 > \frac{k_m \|\alpha\|_{L^\infty(I_m)}}{\inf\, \beta - 1/2},$$

then $\|Y_{m-1}^-\| < r$ implies that $\|Y\|_{L^2(I_m)} \leq r$. This follows because

$$\big\langle \Phi(U), U \big\rangle_m \geq \inf\, \beta\, r^2 - \frac{1}{2} r^2 - k_m ||\alpha||_{L^\infty(I_m)} > 0.$$

Recall that (3.4) and (3.5) implicitly define an equation $\Psi_k(Y_{m-1}^-, Y_m^-) = 0$. In the case that Y_m^- is not determined uniquely by Y_{m-1}^-, we can define a *generalized* evolution operator by allowing S_k^m to be a multi-valued map (see [16], Definition 1.1.7.) First for $a \in \mathbf{R}^d$,

$$S_k(a) = \big\{ b \in \mathbf{R}^d : \Psi_k(a,b) = 0 \big\}, \quad S_k(\mathcal{U}) = \bigcup_{a \in \mathcal{U}} S_k(a),$$

and now inductively

$$S_k^m(a) = S_k^1\big(S_k^{m-1}(a)\big), \quad S_k^m(\mathcal{U}) = \bigcup_{a \in \mathcal{U}} S_k^m(a).$$

If $\Psi_k(a,b)$ uniquely determines b for $a \in \mathbf{R}^d$, then S_k^m defines a discrete dynamical system. In general, S_k^m returns all the solutions of the dG equations for a given initial data and thus is a set-valued function on subsets of $\mathbf{R}^d$. We do not get a dynamical system in this case; however it is still makes sense to discuss the behavior of the trajectories that start from a given initial value. In particular, the definition of dissipativity can be extended in the obvious way to this case (see [16], Definition 1.8.3).

4. Local approximation properties of the discrete semigroup operator

In this section, we discuss the approximation properties of the discrete semigroup operator S_k defined implicitly by the dG method in Section 3 in the case that f is autonomous and uniform step sizes $k_m \equiv k$ are used. In this situation, recall, we can think of S_k as defining a discrete dynamical system involving the nodal values $\{Y_m^-\}$.

We define $dS_k(U)$ to be the Jacobian of S_kU with respect to U and, similarly, $dS(u,k)$ to be the Jacobian of $S(k)u$ with respect to u. The main result of this section is

Theorem 4.1. *Suppose that $f(\cdot)$ and $df(\cdot)$ are q times continuously differentiable with local Lipschitz constant $L = L(B)$ on a compact set B. For all $U, V \in B$ there are constants $C = C(B) > 0$ and $K = K(B) > 0$ such that, for all $0 < k < K$,*

$$\|S_kU - S(k)U\| \le Ck^{q+2}, \tag{4.1}$$

$$\|dS_k(U) - dS(U,k)\| \le Ck^{q+2}, \tag{4.2}$$

$$\|dS_k(U)\| \le 1 + Ck, \tag{4.3}$$

$$\|dS_k(U) - dS_k(V)\| \le Ck\|U - V\|. \tag{4.4}$$

The bounds (4.1)–(4.4) ensure that a wide variety of results concerning the approximation of the continuous-time semigroup $S(k)$ by its discrete-time counterpart S_k may be established. These results are described and proved in Chapters 6 and 7 of [16]. For example, stable and unstable manifolds, local phase portraits, periodic solutions and many other "hyperbolic" objects associated to $S(k)$ are approximated to $O(k^{q+1})$ by nearby objects associated to S_k. Furthermore, results about continuity of attractors may also be deduced, although 4.2)–(4.4) are not needed to prove these.

Proof. The proof is contained in Lemmas 4.2 – 4.5. To simplify notation in this section, we set $\|\cdot\|_\infty = \|\cdot\|_{L^\infty(0,k)}$, $\|\cdot\|_2 = \|\cdot\|_{L^2(0,k)}$, and $\mathcal{P}^{(q)} = \mathcal{P}^{(q)}(0,k)$.

The first pair of lemmas concern properties of the dG approximations of the initial value problem (1.1) over $(0,k)$ as well as the associated linearized variational problem. We recall that y solves

$$\begin{cases} \dot y = f(y), \quad 0 < t \le k, \\ y(0) = y_0. \end{cases} \tag{4.5}$$

The dG approximation $Y \in \mathcal{P}^{(q)}$ satisfies

$$\int_0^k \big(\dot Y - f(Y), V\big)\, dt + (Y(0) - y_0, V(0)) = 0 \tag{4.6}$$

for all $V \in \mathcal{P}^{(q)}$. The linearized variational problem associated to (4.5) reads

$$\begin{cases} \dot w = df(y)w, \quad 0 < t \le k, \\ w(0) = w_0, \end{cases} \tag{4.7}$$

while the corresponding dG approximation $W \in \mathcal{P}^{(q)}$ satisfies

$$\int_0^k \big(\dot W - df(Y)W, V\big)\, dt + (W(0) - w_0, V(0)) = 0 \tag{4.8}$$

for all $V \in \mathcal{P}^{(q)}$.

The first result is

Lemma 4.2. *Let $Y^{(i)}$ satisfy (4.6) with data $y_0^{(i)}$, $i = 1, 2$ respectively. Then there are constants $C > 0$ and $K = K(C) > 0$ such that*

$$\big\|Y^{(1)} - Y^{(2)}\big\|_\infty \le C\big\|Y^{(1)}(k) - Y^{(2)}(k)\big\|, \quad 0 < k < K, \tag{4.9}$$

and

$$\left\|Y^{(1)} - Y^{(2)}\right\|_\infty \le C\left\|y_0^{(1)} - y_0^{(2)}\right\|, \quad 0 < k < K. \tag{4.10}$$

Let W satisft (4.8). *Then there are constants $C > 0$ and $K = K(C) > 0$ such that*

$$\|W\|_\infty \le C\|W(k)\|, \quad 0 < k < K, \tag{4.11}$$

and

$$\|W(k)\| \le (1 + Ck)\|w_0\|, \quad 0 < k < K. \tag{4.12}$$

Proof. If we set $\Psi = Y^{(1)} - Y^{(2)}$ and $\psi = y_0^{(1)} - y_0^{(2)}$, then (4.6) implies

$$\int_0^k \left(\dot\Psi, V\right) dt + (\Psi(0) - \psi, V(0)) = \int_0^k \left(f(Y^{(1)}) - f(Y^{(2)}), V\right) dt. \tag{4.13}$$

Choosing $V = t\dot\Psi$ and using the inverse inequalities in Proposition 3.1 and the Lipschitz assumption on f, we obtain

$$k\left\|\dot\Psi\right\|_2^2 \le C\left\|t^{1/2}\dot\Psi\right\|_2^2 \le k\int_0^k L\|\Psi\|\left\|\dot\Psi\right\| dt$$

which leads to the conclusion that $\left\|\dot\Psi\right\|_2 \le L\|\Psi\|_2$. Since

$$\|P\|_\infty^2 \le \|P(k)\|^2 + 2\|P\|_2\|\dot P\|_2 \tag{4.14}$$

for every $P \in \mathcal{P}^{(q)}$, $\|\Psi\|_\infty^2 \le \|\Psi(k)\|^2 + \frac{Lk}{C}\|\Psi\|_\infty^2$ or

$$\|\Psi\|_\infty^2 \le \frac{C}{C - Lk}\|\Psi(k)\|^2,$$

which is (4.9). The analogous result (4.11) for W follows after choosing $V = t\dot W$ in (4.8).

To prove (4.10), we choose $V = \Psi(k)$ in (4.13) and estimate to get

$$\|\Psi(k)\|^2 \le \|\psi\|\|\Psi(k)\| + L\|\Psi(k)\|\int_0^k \|\Psi\| dt$$

or $\|\Psi(k)\| \le \|\psi\| + Lk\|\Psi\|_\infty$. By (4.9), it follows that $\|\Psi(k)\| \le (1 - CLk)^{-1}\|\psi\|$ and

$$\|\Psi\|_\infty \le \frac{C}{1 - CLk}\|\psi\|, \quad 0 < k < \frac{1}{CL}.$$

Finally, we take $V = W(k)$ in (4.8) and estimate to get

$$\|W(k)\|^2 \le \|w_0\|\|W(k)\| + L\|W(k)\|\int_0^k \|W\| dt$$

or $\|W(k)\| \le \|w_0\| + Lk\|W\|_\infty$. The bound (4.12) now follows from (4.11). □

The next result is

Lemma 4.3. *Let $W^{(i)}$ satisfy* (4.8) *where $Y^{(i)}$ satisfies* (4.6) *with data $y_0^{(i)}$ for $i = 1, 2$ respectively. Then there are constants $C > 0$ and $K = K(C) > 0$ such that*

$$\left\|W^{(1)}(k) - W^{(2)}(k)\right\| \le Ck\left\|y_0^{(1)} - y_0^{(2)}\right\|\|w_0\|.$$

Proof. With $\Phi = W^{(1)} - W^{(2)}$, (4.8) implies

$$\int_0^k \big(\dot{\Phi}, V\big)\,dt + \big(\Phi(0), V(0)\big) = \int_0^k \big(df(Y^{(1)})W^{(1)} - df(Y^{(2)})W^{(2)}, V\big)\,dt. \tag{4.15}$$

Choosing $V = \Phi(k)$ gives

$$\|\Phi(k)\| \le \int_0^k \big(\big\|df(Y^{(1)}) - df(Y^{(2)})\big\|\big\|W^{(1)}\big\| + \big\|df(Y^{(2)})\big\|\big\|W^{(1)} - W^{(2)}\big\|\big)\,dt$$

leading to

$$\|\Phi(k)\| \le Lk\big\|Y^{(1)} - Y^{(2)}\big\|_\infty \big\|W^{(1)}\big\|_\infty + Lk\|\Phi\|_\infty.$$

Lemma 4.2 therefore implies there is a constant $C > 0$ such that

$$\|\Phi(k)\| \le Ck\big\|y_0^{(1)} - y_0^{(2)}\big\|\|w_0\| + Lk\|\Phi\|_\infty. \tag{4.16}$$

Now we choose $V = t\dot{\Phi}$ in (4.15) and use Proposition 3.1 and Lipschitz assumptions to find that

$$\begin{aligned} k\big\|\dot{\Phi}\big\|_2^2 &\le C\big\|t^{1/2}\dot{\Phi}\big\|_2^2 \le C\left\|\int_0^k t\left(df(Y^{(1)})W^{(1)} - df(Y^{(2)})W^{(2)}, \dot{\Phi}\right)dt\right\| \\ &\le CkL\big\|Y^{(1)} - Y^{(2)}\big\|_\infty \big\|W^{(1)}\big\|_\infty \int_0^k \big\|\dot{\Phi}\big\|\,dt + CkL\int_0^k \|\Phi\|\big\|\dot{\Phi}\big\|\,dt. \end{aligned}$$

Redefining the value of C as necessary, we find that

$$k\big\|\dot{\Phi}\big\|_2^2 \le Ck\big\|y_0^{(1)} - y_0^{(2)}\big\|\|w_0\| \left(\int_0^k 1\,dt\right)^{1/2} \big\|\dot{\Phi}\big\|_2 + CkL\|\Phi\|_2\big\|\dot{\Phi}\big\|_2$$

or

$$\big\|\dot{\Phi}\big\|_2 \le CL\|\Phi\|_2 + Ck^{1/2}\big\|y_0^{(1)} - y_0^{(2)}\big\|\|w_0\|.$$

Hence since $\|\Phi\|_\infty^2 \le \|\Phi(k)\|^2 + 2\{\|\Phi\|_2^2 + \|\dot{\Phi}\|_2^2\}$, and $\|\Phi\|_2^2 \le Ck\|\Phi\|_\infty^2$, we have

$$\|\Phi\|_\infty^2 \le \|\Phi(k)\|^2 + Ck\big\|y_0^{(1)} - y_0^{(2)}\big\|^2\|w_0\|^2 + Ck\|\Phi\|_\infty^2,$$

or

$$\|\Phi\|_\infty^2 \le C\|\Phi(k)\|^2 + Ck\big\|y_0^{(1)} - y_0^{(2)}\big\|^2\|w_0\|^2.$$

Combining this with (4.16) gives

$$\|\Phi\|_\infty^2 \le Ck\big\|y_0^{(1)} - y_0^{(2)}\big\|^2\|w_0\|^2$$

and the result follows from (4.16). □

To get at the approximation properties of S_k, we introduce the dG approximations corresponding to the linear equation $\dot{u} = f(y)$ where y denotes the exact solution of (1.1) that we wish to approximate. We define $Z \in \mathcal{P}^{(q)}$ to satisfy

$$\int_0^k \big(\dot{Z} - f(y), V\big)\,dt + (Z(0) - y_0, V(0)) = 0 \tag{4.17}$$

for all $V \in \mathcal{P}^{(q)}$ while $X \in \mathcal{P}^{(q)}$ satisfies

$$\int_0^k \big(\dot{X} - df(y)w, V\big)\,dt + (X(0) - w_0, V(0)) = 0 \tag{4.18}$$

for all $V \in \mathcal{P}^{(q)}$ where w solves (4.7). Standard analysis, using the assumed regularity on f, shows that there is a constant $C > 0$ such that

$$\|y - Z\|_\infty \le Ck^{q+1} \quad \text{and} \quad \|w - X\|_\infty \le C\|w_0\|k^{q+1}; \tag{4.19}$$

see Estep [7] and Eriksson, Estep, Hansbo, Johnson [6]. Moreover, choosing $V = e_j$ for $j = 1, ..., d$, where e_j is the standard j'th basis vector for $\mathbf{R}^d$, shows that

$$Z(k) = y_0 + \int_0^k f(y(t))\,dt, \quad X(k) = w_0 + \int_0^k df(y(t))w(t)dt.$$

Since y satisfies the same equation, we conclude that

$$Z(k) = y(k), \quad X(k) = w(k). \tag{4.20}$$

The next result is

Lemma 4.4. *There are constants $C > 0$ and $K = K(C) > 0$ such that*

$$\|y(k) - Y(k)\| + k\|y - Y\|_\infty \le Ck\|y - Z\|_\infty, \quad 0 < k < K.$$

Proof. We set $\Upsilon(t) = Y(t) - Z(t)$. Then (4.6) and (4.17) give

$$\int_0^k \left(\dot{\Upsilon}, V\right) dt + (\Upsilon(0), V(0)) = \int_0^k (f(Y) - f(y), V)\,dt. \tag{4.21}$$

Taking $V = \Upsilon(k)$ and estimating using the Lipschitz condition on f, we obtain

$$\|\Upsilon(k)\| \le Lk\|e\|_\infty \le Lk\|\mu\|_\infty + Lk\|\Upsilon\|_\infty, \tag{4.22}$$

where we write $e(t) = y(t) - Y(t) = \mu(t) - \Upsilon(t)$ with $\mu(t) = y(t) - Z(t)$.

Choosing $V = t\dot{\Upsilon}$ in (4.21) and using Proposition 3.1 and the Lipschitz condition as above gives

$$\|\dot{\Upsilon}\|_\infty^2 \le \frac{L^2}{C^2}\|e\|_\infty^2,$$

for some $C > 0$. But this implies that

$$\|\Upsilon\|_\infty \le \|\Upsilon(k)\| + k\|\dot{\Upsilon}\|_\infty \le \|\Upsilon(k)\| + \frac{Lk}{C}\|e\|_\infty,$$

or

$$\|\Upsilon\|_\infty^2 \le 2\|\Upsilon(k)\|^2 + \frac{2L^2k^2}{C^2}\|e\|_\infty^2. \tag{4.23}$$

But (4.22) and (4.23) imply there is a $C > 0$ such that

$$\|\Upsilon(k)\|^2 \le Ck^2\left(\|\mu\|_\infty^2 + \|\Upsilon\|_\infty^2\right)$$

and

$$\|\Upsilon\|_\infty^2 \le 2\|\Upsilon(k)\|^2 + Ck^2\left(\|\mu\|_\infty^2 + \|\Upsilon\|_\infty^2\right).$$

Eliminating $\|\Upsilon\|_\infty^2$ gives the result on $\|y(k) - Y(k)\|$, by use of (4.20). For the second, note that $\|\gamma\|_\infty \le C\|\gamma(k)\| = Ck\|y - Z\|$ and that $\|e\|_\infty \le \|\mu\|_\infty + \|\Upsilon\|_\infty$. □

The final result we need is

Lemma 4.5. *There exist constants $C > 0$ and $K = K(C) > 0$ such that*

$$\|w(k) - W(k)\|^2 \le Ck^2\left(\|y - Z\|_\infty^2\|w_0\|^2 + \|w - X\|_\infty^2\right).$$

Proof. We write $E(t) = w(t) - W(t) = \theta(t) - \Xi(t)$ where $\theta(t) = w(t) - X(t)$ and $\Xi(t) = W(t) - X(t)$. Then (4.8) and (4.18) imply that

$$\int_0^k \left(\dot{\Xi}, V\right) dt + (\Xi(0), V(0)) = \int_0^k (df(Y)W - df(y)w, V)\, dt. \tag{4.24}$$

Choosing $V = \Xi(k)$ in (4.24) and estimating as above using Proposition 3.1 and Lipschitz assumptions gives

$$\|\Xi(k)\|^2 \le \left\| \int_0^k ((df(Y) - df(y))W, \Xi(k)) + (df(y)(W - w), \Xi(k))\, dt \right\|$$

and so

$$\|\Xi(k)\| \le L \int_0^k \|e\|\|W\|\, dt + L \int_0^k \|E\|\, dt.$$

This means that

$$\|\Xi(k)\| \le Lk\|E\|_\infty + Lk\|e\|_\infty\|W\|_\infty. \tag{4.25}$$

Next choosing $V = t\dot{\Xi}$ in (4.24) gives

$$\int_0^k t\|\dot{\Xi}\|^2\, dt \le CLk^2\big(\|E\|_\infty + \|e\|_\infty\|W\|_\infty\big)\|\Xi\|_\infty$$

and therefore that, by Proposition 3.1,

$$\left\|\dot{\Xi}\right\|_\infty \le CL(\|E\|_\infty + \|e\|_\infty\|W\|_\infty).$$

This means that

$$\|\Xi\|_\infty^2 \le 2\|\Xi(k)\|^2 + Ck^2\big(\|E\|_\infty + \|e\|_\infty\|W\|_\infty\big)^2. \tag{4.26}$$

Together (4.25) and (4.26) imply there are constants $C, K > 0$ such that

$$\|\Xi(k)\|^2 \le Ck^2\left(\|\theta\|_\infty^2 + \|e\|_\infty^2\|W\|_\infty^2 + \|\Xi\|_\infty^2\right),$$

and

$$\|\Xi\|_\infty^2 \le C\left(\|\Xi(k)\|^2 + k^2\|\theta\|_\infty^2 + k^2\|e\|_\infty^2\|W\|_\infty^2\right),$$

for $0 < k < K$. Combining these we reach, by (4.20),

$$\|w(k) - W(k)\|^2 = \|\Xi(k)\|^2 \le Ck^2\left(\|\theta\|_\infty^2 + \|e\|_\infty^2\|W\|_\infty^2\right), \quad 0 < k < K, \tag{4.27}$$

for some $C, K > 0$.

Lemma 4.2 implies that there are constants $C, K > 0$ such that

$$\|W\|_\infty^2 \le C\|w_0\|^2, \quad 0 < k < K, \tag{4.28}$$

while Lemma 4.4 gives the necessary estimate on $\|e\|_\infty$. □

We now finish the proof of the theorem.

For (4.1), we note that $e(k) = S_kU - S(k)U$; thus (4.19) and Lemma 4.4 give the result.

For (4.2), we note that $E(k) = (dS_k(U) - dS(U,k))w_0$, so (4.19) and Lemma 4.5 show the result holds for the induced matrix norm.

For (4.3), we note that $W(k) = dS_k(U)w_0$ since $W(t) = \partial Y(t;U)/\partial U$. Thus the result follows from Lemma 4.2.

Finally for (4.4), since $W^{(i)}(k) = dS_k(U^{(i)})w_0$, $i = 1, 2$, the result follows from Lemma 4.3.

□

5. Preservation of dissipativity under Galerkin discretization

In this section, we show that the dG approximation inherits the dissipativity properties of the true solution operator under either Assumption 2.1 or 2.4. We deal with the contractive case first and show that the dG method automatically inherits the qualitative long-time behavior of the underlying ODE which is detailed in Section 1.

Theorem 5.1. *Assume that f is locally Lipschitz continuous and Assumption 2.1 holds and moreover there is a constant $k_0 > 0$ such that $k_n \geq k_0$ for $n > 0$. Then there is a constant $C = C(c,q) > 0$ such that any two solutions U, V of (3.1) satisfy*

$$\|(U-V)_n^-\| \leq e^{-Ct_n}\|(U-V)_0^-\|, \tag{5.1}$$

for $n \geq 0$ and

$$\lim_{n\to\infty} \operatorname{dist}(U_n^-, \mathcal{E}) = 0, \tag{5.2}$$

where $\mathcal{E}$ is the closed convex set of steady states. If f is autonomous and there is a $\bar{u}$ such that $f(\bar{u}) = 0$, then $\bar{u}$ is unique and $\lim_{n\to\infty} \operatorname{dist}(U_n^-, \bar{u}) = 0$.

Proof. We subtract the equation (3.2) for V from the equation (3.2) for U and choose $X = U - V$ to get

$$\frac{1}{2}||(U-V)_m^-||^2 + \frac{1}{2}||(U-V)_{m-1}^+||^2 = ((U-V)_{m-1}^-, (U-V)_{m-1}^+) + \int_{I_m} (f(U,t) - f(V,t), U-V)\,dt.$$

Now we estimate using Assumption 2.1 to get

$$\frac{1}{2}||(U-V)_m^-||^2 \leq \frac{1}{2}||(U-V)_{m-1}^-||^2 - c\int_{I_m} ||U-V||^2\,dt.$$

Using Proposition 3.1, we obtain

$$(1 + Ck_m)||(U-V)_m^-||^2 \leq ||(U-V)_{m-1}^-||^2$$

for some $C = C(c,q)$. We conclude that

$$||(U-V)_n^-||^2 \leq \prod_{m=1}^{n} \frac{1}{(1+Ck_m)}||(U-V)_0^-||^2 \tag{5.3}$$

and (5.1) follows immediately.

Given $R > 0$, we know that if $U_0^- \in \{u \in \mathbf{R}^d : \operatorname{dist}(u, \mathcal{E}) \leq R\}$, then there is a $\bar{u} \in \mathcal{E}$ with $||U_0^- - \bar{u}|| \leq R$. By (5.3), we know that $||U_n^- - \bar{u}|| \leq R$ for all $n \geq 0$. Now choose $r : 0 < r < R$. By (5.3), there is an $n^*(r, C) \geq 0$ such that $||U_{n^*}^- - \bar{u}|| \leq r$. Moreover repeating the argument with $R = r$, we see that $||U_n^- - \bar{u}|| \leq r$ for $n > n^*$. Since r is arbitrary, (5.2) follows. □

We need to amend Assumption 2.4 slightly for the discrete case. On each interval I_m, we set $\hat{\beta}_m = \sup_{I_m} \beta$ and $\check{\beta}_m = \inf_{I_m} \beta$.

Assumption 5.2. For the positive function $\beta \in W_\infty^1(\mathbf{R}^+)$ given in Assumption 2.4 there are constants $\rho > 0$ and $K > 0$ such that, if $k_m \in (0, K)$ for all m,

$$\sup_{0\leq m\leq\infty} \frac{\hat{\beta}_m}{\check{\beta}_m} \leq 1 + \rho. \tag{5.4}$$

Note that Assumption 5.2 is trivially true in the important case when β is a constant.

The following result shows that the ball $\mathcal{B}(0,\tilde{R})$ is absorbing for the *nodal* values of the dG approximation so in that sense the dG method inherits the consequences of this form of dissipativity. However, we do *not* show that $Y(t)$ enters $\mathcal{B}(0,\tilde{R})$ for $t \geq t_m$ with $m \geq n^*$. All we deduce in this direction is that there is a constant C such that

$$\|Y\|^2_{L^\infty(I_m)} \leq \frac{C}{k_m \check{\beta}_m}\|Y^-_{m-1}\|^2 + CR^2.$$

Theorem 5.3. *Assume that f is locally Lipschitz continuous and that Assumptions 2.4 and 5.2 hold. Assume also that the sequence of time steps $\{k_m\}$ have the property that $\sum_m k_m = \infty$. Then there are constants $C = C(q) > 0$ and $\tilde{R} > 0$ such that for all $\epsilon > 0$ and $y_0 \in \mathbf{R}^d$ there exists an $n^* = n^*(y_0,\epsilon)$ so that the nodal values of any solution $Y(t)$ from (3.1) satisfy $Y^-_m \in \mathcal{B}(0,\tilde{R})$, for all $m \geq n^*$, provided $k_m \in (0,K)$ for all m.*

Proof. We choose $V = Y$ in (3.2) to get

$$\frac{1}{2}\|Y^-_m\|^2 + \frac{1}{2}\|Y^+_{m-1}\|^2 = \left(Y^-_{m-1}, Y^+_{m-1}\right) + \int_{I_m} (f(Y,t), Y)\, dt$$

which implies that

$$\|Y^-_m\|^2 \leq \|Y^-_{m-1}\|^2 + 2\int_{I_m} \left(\alpha(t) - \beta(t)\|Y(t)\|^2\right) dt. \tag{5.5}$$

Therefore either

$$\|Y^-_m\|^2 \leq \|Y^-_{m-1}\|^2 - 2k_m\epsilon\check{\beta}_m \tag{5.6}$$

or

$$\int_{I_m} \beta(t)\|Y(t)\|^2\, dt \leq \int_{I_m} \alpha(t)\, dt + k_m\epsilon\check{\beta}_m$$

which implies that

$$\check{\beta}_m \int_{I_m} \|Y(t)\|^2\, dt \leq R^2 k_m \hat{\beta}_m + k_m\epsilon\check{\beta}_m. \tag{5.7}$$

In the second case, Proposition 3.1 and the assumptions above thus imply that there is a constant $C = C(q)$ such that

$$\|Y^-_m\|^2 \leq C(R^2(1+\rho) + \epsilon). \tag{5.8}$$

Notice that this dichotomy implies that $\mathcal{B}(0,\tilde{R})$ is positively invariant; i.e., if $Y^-_{m-1} \in \mathcal{B}(0,\tilde{R})$, then either (5.6) or (5.8) yield $Y^-_m \in \mathcal{B}(0,\tilde{R})$. It remains to show that all nodal values enter $\mathcal{B}(0,\tilde{R})$ after a finite number of steps.

Assume the contrary, so that (5.8) does not hold for any m. Then (5.6) implies that

$$\|Y^-_m\|^2 \leq \|Y^-_{m-1}\|^2 - 2k_m\epsilon\check{\beta}_m, \quad 0 \leq m.$$

If we show that

$$\lim_{M\to\infty} \sum_{m=1}^{M} k_m\check{\beta}_m = \infty,$$

then we have a contradiction and the proof is complete.

By (2.2), for any $\delta > 0$ there is an $L > 0$ such that

$$\int_0^L \beta(t)\,dt > \frac{1}{\delta}. \tag{5.9}$$

Given L, we choose the smallest M such that $\sum_{m=1}^{M} k_m \geq L$, which is possible by the assumptions of the theorem. We increase L if necessary so that $L = \sum_{m=1}^{M} k_m$, noting that (5.9) still holds. Then

$$\int_0^L \beta(t)\,dt - \sum_{m=1}^{M} \check{\beta}_m k_m \leq \sum_{m=1}^{M} \bigl(\hat{\beta}_m - \check{\beta}_m\bigr) k_m$$

or

$$\int_0^L \beta(t)\,dt \leq \sum_{m=1}^{M} \check{\beta}_m k_m + \sum_{m=1}^{M} \check{\beta}_m \left(\frac{\hat{\beta}_m}{\check{\beta}_m} - 1\right) k_m.$$

By (5.4)

$$\int_0^L \beta(t)\,dt \leq (1+\rho) \sum_{m=1}^{M} \check{\beta}_m k_m$$

and so, by (5.9),

$$\sum_{m=1}^{M} \check{\beta}_m k_m \geq \frac{1}{\delta(1+\rho)}.$$

The desired conclusion follows since δ is arbitrary. □

Recall from Section 3 that we do not prove that there is global uniqueness of the solution – only existence; hence this theorem does not mean that the dG method defines a dissipative dynamical system under these assumptions. However using the local existence Theorem 3.5 under the assumption of constant step sizes, we can show:

Corollary 5.4. *Let the assumptions of Theorem 5.3 hold. Suppose that the step sizes are constant $k_m \equiv k$ and the iteration* (3.5) *is used to solve the dG equations. Let $\mathcal{N}$ be an open neighborhood of $\mathcal{B}\bigl(0,\tilde{R}\bigr)$. Then there is a $K > 0$ such that if $k < K$, the nodal values of the dG method define a discrete dynamical system on $\mathcal{B}\bigl(0,\tilde{R}\bigr)$ in the sense that if $y_0 \in \mathcal{B}\bigl(0,\tilde{R}\bigr)$, then $Y_m^- \in \mathcal{B}\bigl(0,\tilde{R}\bigr)$, for all m and $Y(t) \in \mathcal{N}$ for all $t \geq 0$; this dynamical system is continuous with respect to initial data.*

Since $\mathcal{B}\bigl(0,\tilde{R}\bigr)$ is bounded, the dynamical system defined by the dG method is automatically dissipative and has the global attractor $\mathcal{A}_k = \omega\bigl(\mathcal{B}\bigl(0,\tilde{R}\bigr)\bigr)$. Outside $\mathcal{B}\bigl(0,\tilde{R}\bigr)$, there can be multiple solutions. However for n^* large enough, the nodal values of all solutions enter $\mathcal{B}\bigl(0,\tilde{R}\bigr)$ for $m > n^*$. Thereafter under the time step restriction, there is a unique solution associated to each entry point.

Proof. We know that $\mathcal{B}\bigl(0,\tilde{R}\bigr)$ is absorbing and invariant for the nodal values of the dG method. If we choose $\mathcal{U}_\delta = \mathcal{B}\bigl(0,\tilde{R}\bigr)$ and $\mathcal{U} = \mathcal{N}$ and if $Y_{m-1}^- \in \mathcal{B}\bigl(0,\tilde{R}\bigr)$, then Theorem 3.5 implies that there is a $K\left(\mathcal{B}\bigl(0,\tilde{R}\bigr),\mathcal{N}\right) > 0$ such that if $k_m < K$, then the iteration (3.5) converges to a unique solution $Y(t) \in \mathcal{N}$ for $t \in I_m$. Since

$Y_m^- \in \mathcal{B}(0,\tilde{R})$ the dG method defines a discrete dynamical system on $\mathcal{B}(0,\tilde{R})$. Continuity follows by arguing as in the proof of Theorem 3.2. □

Finally, we show that the dG method can be used to get some information on the long time behavior of the continuous solution by showing that the numerical attractor is upper semicontinuous with respect to the continuous attractor at $k = 0$.

Corollary 5.5. *Let the assumptions of Theorem 5.3 hold and assume that* (1.1) *is autonomous so that* (1.1) *has a global attractor* $\mathcal{A}$. *Then there is a* $K > 0$ *such that the dynamical system constructed in Corollary 5.4 has an attractor* $\mathcal{A}_k$ *for all* $k < K$. *Furthermore,*

$$\lim_{k\to 0} \operatorname{dist}(\mathcal{A}_k, \mathcal{A}) = 0.$$

Proof. We know that $\mathcal{B}(0,\tilde{R})$ is an absorbing set for the nodal values of the dG method and furthermore the dG method defines a continuous dynamical system on the invariant set $\mathcal{B}(0,\tilde{R})$. Therefore, $\mathcal{A}_k = \omega(\mathcal{B}(0,\tilde{R}))$. The proof now follows from Theorem 7.6.3 in [16]. □

6. Preservation of dissipativity under the use of quadrature

We now discuss the dynamical behavior of approximate solutions derived by applying a quadrature formula to the second integral appearing in (3.2). It is already known that the use of quadrature can affect the accuracy of approximation (see Delfour, Hager, and Trochu [5] and Delfour and Dubeau [4] for an extensive discussion of the possible choices of quadrature, the connection between the resulting approximations and standard difference schemes, the A-stability of the schemes, and a priori convergence results). See also Eriksson, Estep, Hansbo, and Johnson [6]. In this section, we show that the choice of quadrature can also have a strong effect on the dissipativity properties of the resulting numerical solution and show how certain choices of quadrature preserve dissipativity properties.

In this section, we abuse notation to let Y denote any approximate solution obtained from the Galerkin equations either by exact integration or by the application of a quadrature formula.

To illustrate the effect of quadrature on the long time behavior of the approximate solutions, we recall two simple examples considered in Stuart and Humphries [16], Chapter 5. For $q = 0$, the dG approximation Y is given by

$$Y_m^- = Y_{m-1}^- + \int_{I_m} f(Y(t),t)dt.$$

In the next two examples, we replace

$$\int_{I_m} f(Y(t),t) \to f(Y_{m-1}^-, t_{m-1})k_m$$

to produce the approximation $\tilde{Y}$. This method, equivalent to the forward Euler scheme, converges with the same order as the dG method. Of course using extrapolation, as in deriving this scheme, when a problem is dissipative is perhaps not wise, but we make this choice to illustrate the fact that the choice of quadrature is critical in determining whether or not a scheme is dissipative.

Example 6.1. Consider $\dot{y} = -\beta y$, β a positive real constant. This is a dissipative problem and every solution converges to 0 as $t \to \infty$. In the case of constant time step k, the scheme above gives

$$\tilde{Y}_m^- = (1 - k\beta)^m \tilde{Y}_0^-,$$

so that it is dissipative if

$$k < \frac{2}{\beta},$$

but it is unbounded otherwise. Thus preserving dissipativity requires an upper bound on the step size.

Example 6.2. Next consider $\dot{y} = -y^3$, $y(0) = y_0$. Since $-s^3 \cdot s \leq 1 - s^2$ for all real s, this problem is dissipative. This scheme has the property that

$$|\tilde{Y}_0| < \sqrt{\frac{2}{k}} \Rightarrow \lim_{m\to\infty} |\tilde{Y}_m| = 0,$$

while

$$|\tilde{Y}_0| > \sqrt{\frac{2}{k}} \Rightarrow \lim_{m\to\infty} |\tilde{Y}_m| = \infty.$$

Hence there is no step size condition that makes the discrete scheme dissipative. As we have seen, this contrasts with the behavior of the dG approximation.

As mentioned, there are many possibilities for the choice of quadrature. We concentrate on quadrature formulas that satisfy two criteria:

- The formulas give one-step difference schemes using points contained in the interval $(t_{m-1}, t_m]$, $m = 1, 2, \ldots$. It is possible to define multi-step schemes using the dG formulation (see Estep and Larsson [9] for examples). However, the consequences on the stability of the approximation are rather severe.
- The resulting schemes reproduce the dG approximant when applied to linear, autonomous problems. This ensures that the formulas inherit at least the linear stability and superconvergence properties of the dG method.

The goal is to derive schemes that preserve the dynamical properties of the dG method on dissipative problems. The fundamental reason that the dG method is dissipative under Assumptions 2.1 or Assumptions 2.4, 5.2 is that the Galerkin formulation is based on the inner product $\langle \cdot, \cdot \rangle$ and the associated norm $||\cdot||_2$ which allows "energy" arguments to be used to carry over the dissipative properties of the solution naturally to the dG approximation. If $u(t)$ and $v(t)$ are two continuous functions taking I_m to $\mathbf{R}^d$, then we can apply a quadrature rule to $\int_{I_m} (u, v)\, dt$ to get a "discrete" inner product. We will see that if the quadrature formula has the property that this discrete inner product determines a semi-norm, then the resulting difference scheme preserves dissipativity.

We consider two possible quadratures.

Quadrature Assumption 6.3. In this quadrature, we replace f by a suitable interpolant $P_m f \in \mathcal{P}^{(q)}(I_m)$ so that (3.2) becomes

(6.1)
$$\int_{I_m} (\dot{Y}, X)\, dt + (Y_{m-1}^+, X_{m-1}^+) = (Y_{m-1}^-, X_{m-1}^+) + \int_{I_m} (P_m f(Y(t), t), X(t))\, dt,$$

for all $X \in \mathcal{P}^{(q)}(I_m)$, where the integral on the right is computed exactly. We call this an *internal* quadrature formula. We assume that the interpolation basis is a set of orthogonal polynomials. Choose $q'_m \geq q \in \mathbf{Z}$ and let $\{\phi_{m,i}\}_{i=0}^{q'_m}$ be a basis for $\mathcal{P}^{q'_m}(I_m)$ that is orthogonal with respect to $\langle \cdot, \cdot \rangle$ and is associated to the nodes $\{\tau_{m,i}\}_{i=0}^{q'_m}$, i.e.,

$$\phi_{m,i}(\tau_{m,j}) = \begin{cases} 0, & i \neq j, \\ 1, & i = j. \end{cases}$$

We consider the interpolant

$$P_m g = \sum_{i=0}^{q'_m} g(\tau_{m,i}) \phi_{m,i}(t),$$

for g continuous.

Quadrature Assumption 6.4. In the second quadrature formula, we replace the entire integral term involving f in (3.2) with a quadrature formula to get

(6.2)
$$\int_{I_m} (\dot{Y}, X)\, dt + (Y_{m-1}^+, X_{m-1}^+) = (Y_{m-1}^-, X_{m-1}^+) + Q_m^{q'_m}\big((f(Y(t),t), X(t)), k_m\big),$$

for all $X \in \mathcal{P}^{(q)}(I_m)$. We call this an *external* quadrature. We assume that

$$Q_m^{q'_m}\big(g(t), k_m\big) = \sum_{i=0}^{q'_m} \omega_{m,i} g(\tau_{m,i}),$$

where the nodes $\{\tau_{m,i}\}_{i=0}^{q'_m}$ are in I_m, the weights $\{\omega_{m,i}\}_{i=0}^{q'_m}$ are positive and $q'_m \in \mathbf{Z}^+$ is chosen large enough so that the formula has order of precision at least $2q$.

Note that we abuse notation by using q'_m in different ways for internal and external quadratures; the meaning will always be clear from the context.

Example 6.5. When $q = 0$, internal and external quadratures yield the same set of formulas for $q'_m = 0$, namely,

$$\int_{I_m} (f(Y(t),t), V(t)) dt \to (f(Y_m^-, \tau_{m,0}), V_m^-) k_m,$$

for some $t_{m-1} \leq \tau_{m,0} \leq t_m$. For autonomous problems this yields the backward Euler method.

Example 6.6. For an internal quadrature with $q = 1$, we choose

$$\tau_{m,0} = t_{m-1} + \frac{k_m}{3} \text{ and } \tau_{m,1} = t_m$$

and use the basis

$$\left\{ \frac{(t - \tau_{m,0})}{(\tau_{m,1} - \tau_{m,0})}, \frac{(t - \tau_{m,1})}{(\tau_{m,1} - \tau_{m,0})} \right\}.$$

These are known as Radau points and the resulting scheme has order of precision three. This formula is equivalent to a standard Runge-Kutta scheme that can be

written

$$\begin{cases} \tilde{Y}_{m,0} = Y_{m-1} + \frac{5}{12}k_m f(\tilde{Y}_{m,0}, t_{m-1} + k_m/3) - \frac{1}{12}k_m f(\tilde{Y}_{m,1}, t_m), \\ \tilde{Y}_{m,1} = Y_{m-1} + \frac{3}{4}k_m f(\tilde{Y}_{m,0}, t_{m-1} + k_m/3) + \frac{1}{4}k_m f(\tilde{Y}_{m,1}, t_m), \\ Y_m = Y_{m-1} + \frac{3}{4}k_m f(\tilde{Y}_{m,0}, t_{m-1} + k_m/3) + \frac{1}{4}k_m f(\tilde{Y}_{m,1}, t_m). \end{cases}$$

Example 6.7. For an external quadrature with $q = 1$, we choose the two point Gauss rule with order of precision three,

$$\tau_{m,0} = t_{m-1} + \frac{k_m}{2} - \frac{k_m}{2\sqrt{3}}, \quad \omega_{m,0} = \frac{1}{2}$$

and

$$\tau_{m,1} = t_{m-1} + \frac{k_m}{2} + \frac{k_m}{2\sqrt{3}}, \quad \omega_{m,1} = \frac{1}{2}.$$

The associated Runge-Kutta scheme can be written

$$\begin{cases} \tilde{Y}_{m,0} = Y_{m-1} + \frac{1}{3}k_m f(\tilde{Y}_{m,0}, t_{m-1} + \frac{k_m}{2} - \frac{k_m}{2\sqrt{3}}) \\ \qquad\qquad + \frac{(1-\sqrt{3})}{6}k_m f(\tilde{Y}_{m,1}, t_{m-1} + \frac{k_m}{2} + \frac{k_m}{2\sqrt{3}}), \\ \tilde{Y}_{m,1} = Y_{m-1} + \frac{(1+\sqrt{3})}{6}k_m f(\tilde{Y}_{m,0}, t_{m-1} + \frac{k_m}{2} - \frac{k_m}{2\sqrt{3}}) \\ \qquad\qquad + \frac{1}{3}k_m f(\tilde{Y}_{m,1}, t_{m-1} + \frac{k_m}{2} + \frac{k_m}{2\sqrt{3}}), \\ Y_m = Y_{m-1} + \frac{1}{2}k_m f(\tilde{Y}_{m,0}, t_{m-1} + \frac{k_m}{2} - \frac{k_m}{2\sqrt{3}}) \\ \qquad\qquad + \frac{1}{2}k_m f(\tilde{Y}_{m,1}, t_{m-1} + \frac{k_m}{2} + \frac{k_m}{2\sqrt{3}}). \end{cases}$$

It is straightforward to show that these last three examples produce nonconfluent, B-N stable Runge-Kutta schemes (see Dekker and Verwer [3] for the definition of such schemes) and hence, the approximations are algebraically stable.

There is a close relationship between certain internal and external quadratures.

Proposition 6.8. *Internal and external quadrature formulas satisfying Quadrature Assumptions 6.3 and 6.4 respectively that are based on interpolation using orthogonal polynomials with the same nodes yield the same difference scheme.*

Proof. Assume that $f(t)$ is continuous on I_m. Since

$$X(t) = \sum_{i=0}^{q'_m} X(\tau_{m,i})\phi_{m,i}(t),$$

orthogonality implies that

$$\int_{I_m} (P_m f(t), X(t))\, dt = \sum_{i=0}^{q'_m} (f(\tau_{m,i}), X(\tau_{m,i})) \int_{I_m} \phi_{m,i}^2\, dt.$$

But using orthogonality once more, we get

$$\begin{aligned} \sum_{i=0}^{q'_m} (f(\tau_{m,i}), X(\tau_{m,i})) \int_{I_m} \phi_{m,i}^2\, dt &= \int_{I_m} \Big(\sum_{i=0}^{q'_m} (f(\tau_{m,i}), X(\tau_{m,i}))\phi_{m,i}\, \phi_{m,i}\Big)\, dt \\ &= \int_{I_m} \Big(\sum_{i=0}^{q'_m} (f(\tau_{m,i}), X(\tau_{m,i}))\phi_{m,i}\Big)\Big(\sum_{j=0}^{q'_m} \phi_{m,j}\Big)\, dt. \end{aligned}$$

Since $\sum_{j=0}^{q'_m} \phi_{m,j} \equiv 1$ by interpolation, we conclude that

$$\begin{aligned}\int_{I_m} (P_m f(t), X(t))\, dt &= \sum_{i=0}^{q'_m} (f(\tau_{m,i}), X(\tau_{m,i})) \int_{I_m} \phi_{m,i}\, dt \\ &= Q_m^{q'_m}\big((f(t), X(t)), k_m\big)\end{aligned}$$

for all $X \in \mathcal{P}^{(q)}(I_m)$. □

Not all external quadrature formulae satisfying Quadrature Assumption 6.4 are equivalent to an internal orthogonal quadrature formula satisfying Quadrature Assumption 6.3 as a somewhat lengthy calculation on the following example illustrates.

Example 6.9. Consider Simpson's rule with nodes and weights

$$\tau_{m,0} = t_{m-1}, \quad \omega_{m,0} = \frac{k_m}{6}, \quad \tau_{m,1} = t_{m-1} + k_m/2, \quad \omega_{m,1} = \frac{2k_m}{3},$$
$$\tau_{m,2} = t_m, \quad \omega_{m,2} = \frac{k_m}{6}.$$

This is equivalent to the three-stage Runge-Kutta scheme

$$\begin{cases} \tilde{Y}_{m,0} = Y_{m-1} + \frac{1}{6}k_m f(\tilde{Y}_{m,0}, t_{m-1}) - \frac{1}{6}k_m f(\tilde{Y}_{m,2}, t_m), \\ \tilde{Y}_{m,1} = Y_{m-1} + \frac{1}{6}k_m f(\tilde{Y}_{m,0}, t_{m-1}) + \frac{1}{3}k_m f(\tilde{Y}_{m,1}, t_{m-1} + k_m/2), \\ \tilde{Y}_{m,2} = Y_{m-1} + \frac{1}{6}k_m f(\tilde{Y}_{m,0}, t_{m-1}) + \frac{2}{3}k_m f(\tilde{Y}_{m,1}, t_{m-1} + k_m/2) \\ \qquad\qquad + \frac{1}{6}k_m f(\tilde{Y}_{m,2}, t_m), \\ Y_m = Y_{m-1} + \frac{1}{6}k_m f(\tilde{Y}_{m,0}, t_{m-1}) + \frac{2}{3}k_m f(\tilde{Y}_{m,1}, t_{m-1} + k_m/2) \\ \qquad\qquad + \frac{1}{6}k_m f(\tilde{Y}_{m,2}, t_m). \end{cases}$$

This is also a nonconfluent, B-N stable, hence algebraically stable, method. Note that the Jacobian of this method is more expensive to evaluate than for the previous two examples.

The assumptions and definitions of the previous sections carry over to these schemes in the obvious way.

Theorem 6.10. *Suppose that f is locally Lipschitz continuous, $f(c, \cdot) \in C^0(\mathbf{R}^+) \cap L^\infty(\mathbf{R}^+)$ for some c, and furthermore satisfies the assumptions of the theorems in Sections 3 and 5 while either an internal or external quadrature formula satisfying Quadrature Assumption 6.3 or 6.4 respectively is used to compute the approximation. Then the conclusions of the theorems in these sections hold for the resulting scheme.*

The results in Section 4 also extend to fully discrete discontinuous Galerkin schemes when an a priori accuracy result is known. Proving such error bounds is beyond the scope of this paper (see Delfour, Hager, and Trochu [5] and Delfour and Dubeau [4] for some results). Indeed, the results in Stuart and Humphries [16] apply to such fully discrete methods when they coincide with standard Runge-Kutta methods, as do the examples in this section.

We conjecture that the Quadrature Assumptions 6.3 or 6.4 imply that the resulting scheme is an algebraically stable Runge-Kutta scheme. A lengthy direct computation shows that, in general for $q = 1$, interpolants using the nodes $\{t_{m-1} + \gamma k_m, t_m - \gamma k_m\}$, $0 \le \gamma < 1/2$, give nonalgebraically stable Runge-Kutta

schemes unless $\gamma = \frac{1}{2} \pm \frac{1}{2\sqrt{3}}$, i.e., unless the nodes agree with the Gauss rule. Similarly, interpolants using the nodes $\{t_{m-1} + \gamma k_m, t_m\}$, $0 \leq \gamma < 1$, give algebraically stable Runge-Kutta schemes if and only if $\gamma = 1/3$, i.e., the nodes agree with the corresponding Radau rule.

Proof. The proofs of the results for fully discrete dG schemes follow the original proofs closely and we only discuss some of the technical differences due to the use of quadrature.

Proof of the analog of Theorem 3.2. For an internal quadrature, we define Φ_a by $V = \Phi_a(U)$ if

$$\int_{I_m} (\dot{V}, X)\, dt + (V_{m-1}^+, X_{m-1}^+) = (a, X_{m-1}^+) + \int_{I_m} (P_m f(U,t), X)\, dt,$$

for all $X \in \mathcal{P}^{(q)}(I_m)$. This is a well-defined map and any fixed point of $\Phi_{Y_{m-1}}^+$ is a discrete dG approximation. To show that Φ_a maps a ball into itself, we estimate as in the proof of Theorem 3.2 with obvious modifications; for example we obtain, instead of (3.6),

$$||V||_{L^\infty(I_m)} \leq C||a|| + Ck_m \sum_{i=0}^{q_m'} ||f(U(\tau_{m,i}), \tau_{m,i})||.$$

In the case of external quadratures, we define Φ_a by $V = \Phi_a(U)$ if

$$\int_{I_m} (\dot{V}, X)\, dt + (V_{m-1}^+, X_{m-1}^+) = (a, X_{m-1}^+) + Q_m^{q_m'}\big((f(U,t), X), k_m\big),$$

for all $X \in \mathcal{P}^{(q)}(I_m)$. Now estimating, we find that

$$||V||_{L^\infty(I_m)} \leq C||a|| + C \sum_{i=0}^{q_m'} \omega_{m,i} ||f(U(\tau_{m,i}), \tau_{m,i})||.$$

Since $\sum_{i=0}^{q_m'} \omega_{m,i} = k_m$ by construction, the proof now proceeds as in the original case.

The proofs that Φ_a is contractive and continuous are modified in a similar fashion. □

Proof of the analog of Theorem 3.5. The proofs are modified just as above. □

Proof of the analog of Theorem 3.7. For part (a), we define G using an internal quadrature in the obvious fashion and then to estimate $\big(G(\vec{U}) - G(\vec{V}), \vec{U} - \vec{V}\big)$ we compute

$$\begin{aligned}
&\int_{I_m} (P_m f(U,t) - P_m f(V,t), U - V)\, dt \\
&\quad = \sum_{i=0}^{q_m'} \big(f(U(\tau_{m,i}), \tau_{m,i}) - f(V(\tau_{m,i}), \tau_{m,i}), U(\tau_{m,i}) - V(\tau_{m,i})\big) \int_{I_m} \phi_{m,i}^2\, dt \\
&\quad \leq -c \sum_{i=0}^{q_m'} ||U(\tau_{m,i}) - V(\tau_{m,i})||^2 \int_{I_m} \phi_{m,i}^2\, dt = -c \int_{I_m} ||U - V||^2\, dt.
\end{aligned}$$

The proof now proceeds as in the original proof. In the case of an external quadrature, we make a similar estimate and use the fact the weights are positive to reach the conclusion.

To show (b) for an internal quadrature, we define L_U using the quadrature in the obvious way. In showing that L_U is bounded, the only difference encountered is estimating

$$\begin{aligned}\int_{I_m} ||P_m f(U,t)||^2\,dt &= \int_{I_m}\left(\sum_{i=0}^{q'_m} f(U(\tau_{m,i}),\tau_{m,i})\phi_{m,i}, \sum_{i=0}^{q'_m} f(U(\tau_{m,i}),\tau_{m,i})\phi_{m,i}\right)dt\\ &= \sum_{i=0}^{q'_m} ||f(U(\tau_{m,i}),\tau_{m,i})||^2 \int_{I_m}\phi_{m,i}^2\,dt\\ &\le \sum_{i=0}^{q'_m} ||f(0,\tau_{m,i})||^2 \int_{I_m}\phi_{m,i}^2\,dt + L(U)\sum_{i=0}^{q'_m}||U(\tau_{m,i})||^2\int_{I_m}\phi_{m,i}^2\,dt\\ &\le ||f(0,\cdot)||^2_{L^\infty(I_m)}k_m + L(U)||U||^2_{L^2(I_m)},\end{aligned}$$

We continue as in the original proof. When we compute $\langle \Phi(U), U\rangle_m$, we find that

$$\begin{aligned}\langle \Phi(U),U\rangle_m &\ge \frac{1}{2}||U_m^-||^2 - \frac{1}{2}||a||^2 - \int_{I_m}\left(\sum_{i=0}^{q'_m} f(U(\tau_{m,i}),\tau_{m,i})\phi_{m,i}\,, U\right)dt\\ &= \frac{1}{2}||U_m^-||^2 - \frac{1}{2}||a||^2 - \sum_{i=0}^{q'_m}\Big(f(U(\tau_{m,i}),\tau_{m,i}),U(\tau_{m,i})\Big)\int_{I_m}\phi_{m,i}^2\,dt\\ &\ge \frac{1}{2}||U_m^-||^2 - \frac{1}{2}||a||^2 + \sum_{i=0}^{q'_m}\beta(\tau_{m,i})||U(\tau_{m,i})||^2\int_{I_m}\phi_{m,i}^2\,dt\\ &\qquad - \sum_{i=0}^{q'_m}\alpha(\tau_{m,i})\int_{I_m}\phi_{m,i}^2\,dt\\ &\ge \frac{1}{2}||U_m^-||^2 - \frac{1}{2}||a||^2 + \check{\beta}_m\int_{I_m}||U||^2dt - ||\alpha||_{L^\infty(I_m)}k_m,\end{aligned}$$

and the proof proceeds. Note that orthogonality is a key ingredient in these estimates. In the case of external quadratures, we make similar estimates but use instead the assumption that the quadrature is exact of degree at least $2q$. □

Proof of the analog of Theorem 5.1. The proofs are modified just as for (a) above. □

Proof of the analog of Theorem 5.3. For an internal quadrature, we find, instead of (5.5),

$$\begin{aligned}||Y_m^-||^2 &\le ||Y_{m-1}^-||^2\\ &\quad + 2\sum_{i=0}^{q'_m}\alpha(\tau_{m,i})\int_{I_m}\phi_{m,i}^2\,dt - 2\sum_{i=0}^{q'_m}\beta(\tau_{m,i})||Y(\tau_{m,i})||^2\int_{I_m}\phi_{m,i}^2\,dt.\end{aligned}$$

Now either (5.6) holds or

$$\sum_{i=0}^{q'_m}\beta(\tau_{m,i})||Y(\tau_{m,i})||^2\int_{I_m}\phi_{m,i}^2\,dt \le \sum_{i=0}^{q'_m}\alpha(\tau_{m,i})\int_{I_m}\phi_{m,i}^2\,dt + k_m\epsilon\check{\beta}_m$$

which yields (5.7). The proof now follows as in the original case.

The proof for an external quadrature is very similar. □

This completes the proof of Theorem 6.10. □

Note that we did not extend the corollaries to include the case of quadrature because this would require proving C^1 approximation properties. However we believe that this can be achieved without any difficulty.

Note that not all likely-looking quadratures satisfy the assumptions of Theorem 6.10.

Example 6.11. Consider the (nonorthogonal) internal interpolant

$$P_m g = g(t_{m-1}^+)\frac{(t-t_m)}{-k_m} + g(t_m^-)\frac{(t-t_{m-1})}{k_m}.$$

This is can be written as

$$\begin{cases} \tilde{Y}_{m,0} = Y_{m-1} + \frac{1}{6}k_m f(\tilde{Y}_{m,0}, t_{m-1}) - \frac{1}{6}k_m f(\tilde{Y}_{m,1}, t_m), \\ \tilde{Y}_{m,1} = Y_{m-1} + \frac{1}{2}k_m f(\tilde{Y}_{m,0}, t_{m-1}) + \frac{1}{2}k_m f(\tilde{Y}_{m,1}, t_m), \\ Y_m = Y_{m-1} + \frac{1}{2}k_m f(\tilde{Y}_{m,0}, t_{m-1}) + \frac{1}{2}k_m f(\tilde{Y}_{m,1}, t_m). \end{cases}$$

This Runge-Kutta scheme is not algebraically stable and so cannot be B-N stable either (see Dekker and Verwer [3]).

References

1. P. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, New York, 1978. MR **58:**25001
2. B. Cockburn, G. E. Karniadakis, and C.-W. Shu (eds.), *Discontinuous Galerkin Methods: Theory, Computation and Applications*, Lecture Notes in Computational Science and Engineering, vol. 11, Springer-Verlag, New York, 2000.
3. K. Dekker and J. Verwer, *Stability of Runge-Kutta methods for stiff nonlinear differential equations*, North-Holland, New York, 1984. MR **86g:**65003
4. M. Delfour and F. Dubeau, *Discontinuous polynomial approximations in the theory of one-step, hybrid and multistep methods for nonlinear ordinary differential equations*, Math. Comp. **47** (1986), 169–189. MR **87h:**65134
5. M. Delfour, W. Hager, and F. Trochu, *Discontinuous Galerkin methods for ordinary differential equations*, Math. Comp. **36** (1981), 455–472. MR **82b:**65066
6. K. Eriksson, D. Estep, P. Hansbo, and C. Johnson, *Computational Differential Equations*, Cambridge University Press, New York, 1996. MR **97m:**65006
7. D. Estep, *A posteriori error bounds and global error control for approximations of ordinary differential equations*, SIAM J. Numer. Anal. **32** (1995), 1–48. MR **96i:**65049
8. D. Estep, M. Larson, and R. Williams, *Estimating the error of numerical solutions of systems of nonlinear reaction-diffusion equations*, Memoirs of the A.M.S. **146** (2000), 1–109. MR **2000m:**65103
9. D. Estep and S. Larsson, *The discontinuous Galerkin method for semilinear parabolic problems*, RAIRO Model. Math. Anal. Numer. **27** (1993), 35–54. MR **94b:**65131
10. D. French and S. Jensen, *Global dynamics of a discontinuous Galerkin approximation to a class of reaction-diffusion equations*, Appl. Numer. Math. **18** (1995), 473–487. MR **96m:**35155
11. A. Hill, *Dissipativity of Runge-Kutta methods in Hilbert spaces*, BIT **37** (1997), 37–42. MR **97h:**65096
12. ———, *Global dissipativity for A-stable methods*, SIAM J. Num. Anal. **34** (1997), 119–142. MR **98b:**65091
13. A. Humphries and A. M. Stuart, *Runge-Kutta methods for dissipative and gradient dynamical systems*, SIAM J. Numer. Anal. **31** (1994), 1452–1485. MR **95m:**65116
14. C. Johnson, *Error estimates and adaptive time step control for a class of one step methods for stiff ordinary differential equations*, SIAM J. Numer. Anal. **25** (1988), 908–926. MR **90a:**65160
15. J. Ortega and W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970. MR **42:**8686

16. A. M. Stuart and A. Humphries, *Dynamical Systems and Numerical Analysis*, Cambridge University Press, Cambridge, 1996. MR **97g:**65009
17. R. Temam, *Infinite Dimensional Dynamical Systems in Mechanics and Physics*, Springer-Verlag, New York, 1988. MR **89m:**58056

Department of Mathematics, Colorado State University, Fort Collins, Colorado 80523
E-mail address: estep@math.colostate.edu

Mathematics Institute, University of Warwick, Coventry CV4 7AL, England
E-mail address: stuart@maths.warwick.ac.uk