Proceedings of the Royal Society of Edinburgh, **146A**, 1265–1301, *2016*
DOI:10.1017/S0308210515000852

Exponentially accurate Hamiltonian embeddings of symplectic A-stable Runge–Kutta methods for Hamiltonian semilinear evolution equations

Claudia Wulff
Department of Mathematics, University of Surrey,
Guildford GU2 7XH, UK (c.wulff@surrey.ac.uk)

Marcel Oliver
School of Engineering and Science, Jacobs University,
28759 Bremen, Germany (oliver@member.ams.org)

(MS received 20 February 2015; accepted 28 July 2015)

We prove that a class of A-stable symplectic Runge–Kutta time semi-discretizations (including the Gauss–Legendre methods) applied to a class of semilinear Hamiltonian partial differential equations (PDEs) that are well posed on spaces of analytic functions with analytic initial data can be embedded into a modified Hamiltonian flow up to an exponentially small error. Consequently, such time semi-discretizations conserve the modified Hamiltonian up to an exponentially small error. The modified Hamiltonian is $O(h^p)$-close to the original energy, where p is the order of the method and h is the time-step size. Examples of such systems are the semilinear wave equation, and the nonlinear Schrödinger equation with analytic nonlinearity and periodic boundary conditions. Standard Hamiltonian interpolation results do not apply here because of the occurrence of unbounded operators in the construction of the modified vector field. This loss of regularity in the construction can be taken care of by projecting the PDE to a subspace in which the operators occurring in the evolution equation are bounded, and by coupling the number of excited modes and the number of terms in the expansion of the modified vector field with the step size. This way we obtain exponential estimates of the form $O(\exp(-c/h^{1/(1+q)}))$ with $c > 0$ and $q \geqslant 0$; for the semilinear wave equation, $q = 1$, and for the nonlinear Schrödinger equation, $q = 2$. We give an example which shows that analyticity of the initial data is necessary to obtain exponential estimates.

Keywords: Hamiltonian evolution equations; Runge–Kutta methods; symplectic discretizations; exponentially small error; approximate embedding of maps into flows

2010 *Mathematics subject classification:* Primary 65P10; 65M15; 65J08; 37K05

1. Introduction

Neishtadt [23] showed that a system of ordinary differential equations (ODEs) with a rapidly rotating phase of period ε can be transformed into a system of ODEs for the slow variables and a scalar ODE for the fast phase, both of which are, up to a small error, independent of the fast phase variable.

When the vector field is analytic in the slow coordinates the procedure can be carried out up to an exponentially small remainder of magnitude $O(\mathrm{e}^{-c/\varepsilon})$ for some $c > 0$. The result is proved by applying N near-identity transformations, each of which reduces the error by one order in ε, carefully estimating the remainders and concluding that the embedding is optimal when $N = O(1/\varepsilon)$. For rapidly forced Hamiltonian ODEs this implies approximate conservation of the Hamiltonian of the truncated slow system in the new coordinates with an error of order $O(\mathrm{e}^{-c/\varepsilon})$ over times of length $O(1)$ and, in particular, approximate conservation of the averaged Hamiltonian of the original slow system with error $O(\varepsilon)$ over exponentially long times provided that the trajectory of the system remains bounded.

An $O(\varepsilon)$-close-to-identity analytic map ψ^ε on a finite-dimensional space $\mathbb{R}^n$ is the time-ε map of a rapidly forced analytic vector field, where ε is the period of the forcing. Neishtadt's result therefore applies and shows that ψ^ε can be embedded into the flow of a system of autonomous ODEs up to an exponentially small error. When ψ^ε is symplectic, this flow is also symplectic. This proves that the iterates of ψ^ε approximately conserve the energy of this flow over exponentially long times so long as they remain bounded.

Benettin and Giorgilli [1] give an alternative proof of this embedding result by matching a Taylor expansion of a diffeomorphism ψ^ε with the formal power series expansion of the flow of an ε-dependent vector field $\tilde{f}(x)$, called a modified vector field [13,17], truncating at some order N, where, as before, the embedding is optimal when $N = O(1/\varepsilon)$.

In particular, if the diffeomorphism ψ^ε is a one-step discretization of order p with step size $h = \varepsilon$ of an ODE $\dot{x} = f(x)$, then ψ^ε can be approximately embedded into the flow of a modified vector field $\tilde{f}$ that satisfies $\|f - \tilde{f}\| = O(h^p)$. If the ODE is Hamiltonian with energy H and ψ^ε is symplectic, then the modified vector field is also Hamiltonian with energy $\tilde{H}$ and $|H - \tilde{H}| = O(h^p)$. This implies approximate conservation of the energy H of the original system by the symplectic time-stepping method ψ^ε over exponentially long times provided the numerical trajectory remains bounded. This strategy has been used to prove approximate energy conservation of many classes of symplectic numerical methods, in particular symplectic Runge–Kutta discretizations (see [13, 17, 29] and the references therein).

The question arises as to whether (and to what extent) these results extend to partial differential equations (PDEs). Here, the phase space is typically an infinite-dimensional Hilbert space, and the vector field contains unbounded operators, usually in the form of spatial derivatives. These unbounded operators propagate into the transformed vector field of Neishtadt [23] and into the formal series expansion of the modified vector field of Benettin and Giorgilli [1].

Note that the analytic difficulties persist when analysing full space-time discretizations of the problem. When discretizing space, unbounded operators turn into a sequence of bounded operators whose operator norms diverge as the spatial resolution increases. Consequently, the constant c in the exponential error estimate $O(\mathrm{e}^{-c/\varepsilon})$ tends to 0 with increasing spatial resolution, so that the approximate embedding result for general initial data without the requirement of high regularity fails unless ε is coupled to the spatial step size in a suitable way. This leads to severe restrictions on the time-step size. In the case of hyperbolic problems such as semilinear wave equations, the naive approach fails for step-size ratios

close to the Courant–Friedrichs–Lewy (CFL) limit, i.e. in the practically relevant regime.

Matthies and Scheel [21] consider semilinear Hamiltonian PDEs coupled to a high-frequency oscillator via a nonlinearity that is bounded on the underlying Hilbert space. They prove that Neishtadt's result of an approximate embedding into a flow where the slow variables are decoupled from the fast oscillator is still true, albeit with an error $O(\exp(-c/h^{1/(q+1)}))$ under the condition that the initial data are in a Gevrey class associated with the evolution equation. For the semilinear wave or nonlinear Schrödinger equation, this amounts to requiring real analyticity of the initial data. The positive constants q and c in the error estimate depend on the Gevrey class associated with the evolution equation. Matthies proves a similar result for rapidly forced parabolic PDEs [19] and an approximate embedding result for space-time discretizations of parabolic PDEs [20]. Matthies and Scheel [21] also provide an example showing that Gevrey-regular initial data are necessary. The conclusion is that exponential averaging still works in this context, but long-time approximate conservation of the averaged energy might fail because solutions of Hamiltonian evolution equations are not generally Gevrey regular over long times.

The aim of this paper is to prove a similar result for time semi-discretizations of PDEs. Note that, while formally a time discretization of a PDE can be embedded into the flow of a rapidly forced evolution equation via the construction of Fiedler and Scheurle [11, § 2], the rapidly forced nonlinear term in the interpolating evolution equation is not bounded, so the results of [21] do not apply.

Runge–Kutta time semi-discretizations of semilinear Hamiltonian PDEs are only well defined if the method is implicit; explicit or partially implicit Runge–Kutta time semi-discretizations such as partitioned Runge–Kutta methods, the simplest of which are the leapfrog and symplectic Euler schemes, cannot satisfy the CFL condition for any size of time step [14]. Thus, in this paper we consider a class of (implicit) symplectic A-stable Runge–Kutta methods that includes the Gauss–Legendre Runge–Kutta methods. The simplest of these methods is the implicit midpoint rule.

The analysis relies on our earlier work: in [24] we analysed the differentiability properties with respect to the initial value and time of the semi-flow of

$$\partial_t U = F(U) = AU + B(U) \tag{1.1}$$

on a scale of Hilbert spaces, and obtained analogous results for the time-h map of its corresponding A-stable Runge–Kutta time semi-discretization. In [25], we proved stability of the semi-flow and of the time-semi-discrete solution under spatial spectral Galerkin approximation.

Our approach applies to a large class of semilinear Hamiltonian PDEs with analytic nonlinearities, including the semilinear wave equation and the nonlinear Schrödinger equation on the circle. Our main result, theorem 4.1, can be paraphrased as follows. If a semilinear Hamiltonian evolution equation with energy H is discretized by a symplectic A-stable Runge–Kutta method Ψ^h of order p, then there exists a modified Hamiltonian flow $\tilde{\Phi}$, defined for Gevrey-regular data, with a Hamiltonian $\tilde{H}$ that is $O(h^p)$ close to H, such that $\tilde{\Phi}^h$ interpolates Ψ^h with exponentially small error $O(\exp(-c_*/h^{1/(q+1)}))$, where c_* and q are positive constants. As a consequence, the modified energy $\tilde{H}$ is conserved by the symplectic integrator

Ψ^h with the same exponentially small error for Gevrey-regular initial data. This result is in a number of ways parallel to what has been proved for rapidly forced PDEs by Matthies and Scheel [21]: as in their work, the constants q and c_* depend on the Gevrey class associated with the evolution equation, and c_* also depends on the numerical scheme; moreover, as in [21], the result does not imply long-time approximate energy conservation for time semi-discretizations because both the solutions of the PDE and the numerical trajectories are typically not Gevrey regular over long times.

Let us mention some related work. Moore and Reich [22] derive a modified multi-symplectic PDE for a multi-symplectic discretization of the semilinear wave equation that is satisfied by the numerical solution with higher accuracy than the discretization error; further results in this direction are due to Islas and Schober [15]. Both papers derive higher-order modified equations, but leave it open as to whether these are well posed. In this paper, we can actually prove that the interpolating flow is well defined. In [26], it is shown that second-order finite-difference space semi-discretizations of analytic solutions of the semilinear wave equation approximately conserve a discrete momentum map up to an exponentially small error. Cano [2] considers symmetric–symplectic space-time discretizations of semilinear wave equations and constructs a finite-order modified Hamiltonian, assuming certain conjectures on the smoothness of the fully discrete system.

The approximate conservation of invariants by splitting methods for Hamiltonian PDEs has been studied extensively via normal form transformations or modulated Fourier expansions. Such results require less stringent regularity assumptions, but they are limited either to linear equations [5, 6] or to the weakly nonlinear regime, i.e. to small initial data of space-time discretizations near a homogeneous equilibrium [4, 9, 12], or they refer to modified numerical methods that dampen high oscillations [7, 9]. Note that (symplectic) Gauss–Legendre Runge–Kutta discretizations of linear Hamiltonian systems preserve energy exactly [13]. The results of [4, 8, 9, 12] give approximate conservation of actions and regularity of trajectories of splitting methods applied to semilinear wave and Schrödinger equations for small initial data over polynomially long times under non-resonance conditions. These results require initial values in high-order Sobolev spaces [4, 9, 12] or restrictive conditions on the coupling between space and time-step size [8]. In [7], exponentially accurate interpolations are constructed for modified splitting methods that dampen highly oscillatory motion.

Exponentially accurate estimates for PDEs, albeit without reference to a Hamiltonian structure, have also been obtained in the context of homogenization of linear elliptic problems for Gevrey-regular data [16], while a result for homogenization up to all orders for nonlinear elliptic PDEs can be found in [3].

The paper is structured as follows. In § 2 we define the precise class of semilinear Hamiltonian PDEs that we study. This class includes the semilinear wave equation and the nonlinear Schrödinger equation. In § 3, we introduce A-stable symplectic Runge–Kutta methods. These methods are well defined on Hilbert spaces when applied to a semilinear PDE of the class considered. In § 4, we present and prove our main result, theorem 4.1, on approximate Hamiltonian interpolation of the time-h map of such Runge–Kutta methods. Finally, in § 5, we give an example of a nonlinear Schrödinger equation in Fourier space, which shows that Gevrey-regular

initial data are necessary for an exponentially accurate embedding of a symplectic Runge–Kutta method into a flow.

2. Semilinear Hamiltonian PDEs

In this section, we describe the class of semilinear Hamiltonian systems on Hilbert spaces considered in this paper. We begin by reviewing the general functional setting for semilinear evolution equations from [27] and introduce Gevrey spaces. In § 2.2, we review results on the differentiability in time of the semi-flow from [24]. In § 2.3, we restrict to the Hamiltonian case and review a well-known integrability lemma in our Hilbert space setting. Section 2.4 introduces Hilbert spaces of analytic functions and superposition operators on these spaces. Finally, in §§ 2.5 and 2.6, respectively, we show how our main examples, the nonlinear Schrödinger equation and the semilinear wave equation, fit into this framework.

2.1. Semilinear evolution equations

We initially consider an abstract semilinear evolution equation of the form (1.1),

$$\partial_t U = F(U) = AU + B(U),$$

on a Hilbert space $\mathcal{Y}$. We assume the following.

(A) A is a normal operator on a Hilbert space $\mathcal{Y}$ that generates a $\mathcal{C}^0$-semigroup e^{tA}.

Recall that an operator A is normal if it is closed and $AA^* = A^*A$. For a definition of strongly continuous semigroups ($\mathcal{C}^0$-semigroups), see [27]. Assumption (A) implies that there is a constant $\omega > 0$ such that $\|\mathrm{e}^{tA}\| \leqslant \mathrm{e}^{t\omega}$ for all $t \geqslant 0$.

To formulate our assumptions on the nonlinearity B, we need some definitions. We write

$$\mathcal{B}_R^{\mathcal{X}}(U^0) = \{U \in \mathcal{X} \colon \|U - U^0\|_{\mathcal{X}} \leqslant R\}$$

to denote the closed ball of radius R in a Hilbert space $\mathcal{X}$ about $U^0 \in \mathcal{X}$. When no confusion about the space is possible, we may drop the superscript $\mathcal{X}$. Let $\mathcal{D} \subset \mathcal{Y}$ be open. For $\delta > 0$, let

$$\mathcal{D}^\delta = \bigcup_{U \in \mathcal{D}} \mathcal{B}_\delta^{\mathcal{Y}}(U).$$

Let $\mathcal{Y}^{\mathbb{C}} \equiv \mathcal{Y} + \mathrm{i}\mathcal{Y}$ denote the complexification of $\mathcal{Y}$. We define, for fixed $\delta > 0$,

$$\mathcal{D}^{\mathbb{C}} = \bigcup_{U \in \mathcal{D}} \mathcal{B}_\delta^{\mathcal{Y}^{\mathbb{C}}}(U).$$

Our assumption on B is then stated as follows.

(B0) There are some $\delta > 0$ and a bounded open set $\mathcal{D} \equiv \mathcal{D}_0$ such that $B \colon \mathcal{D}^{\mathbb{C}} \to \mathcal{Y}^{\mathbb{C}}$ is analytic with bound M_0.

Then, after casting (1.1) in its *mild formulation*

$$U(t) = \mathrm{e}^{tA}U^0 + \int_0^t \mathrm{e}^{(t-s)A} B(U(s))\,\mathrm{d}s, \tag{2.1}$$

we can apply the contraction mapping theorem with parameters to obtain well-posedness locally in time [27]. Let $U^0 \to \Phi^t(U^0)$ denote the flow of (2.1), i.e. $U(t) = \Phi^t(U^0) \in \mathcal{D}^{\mathbb{C}}$ satisfies (2.1) with $U(0) = U^0 \in \mathcal{D}^{\mathbb{C}}$. Then Φ^t is continuous in t and analytic in U^0.

For $m \in \mathbb{N}$, let $\mathbb{P}_m$ denote the sequence of spectral projectors of A onto the set $\mathcal{B}_m^{\mathbb{C}}(0) \cap \operatorname{spec} A$, set $\mathbb{P} \equiv \mathbb{P}_1$ and $\mathbb{Q} \equiv 1 - \mathbb{P}$. Assumption (A) implies that

$$\lim_{m\to\infty} \mathbb{P}_m U = U$$

for all $U \in \mathcal{Y}$, and that

$$\|A\mathbb{P}_m U\|_{\mathcal{Y}} \leqslant m\|\mathbb{P}_m U\|_{\mathcal{Y}} \tag{2.2}$$

for $m \in \mathbb{N}$. Let $q > 0$, $\tau \geqslant 0$ and $\ell \in \mathbb{N}_0$. Since A is normal, $|\mathbb{Q}A|^\ell \exp(\tau|\mathbb{Q}A|^{1/q})$ is a well-defined, generally unbounded and densely defined operator on $\mathcal{Y}$. We may thus introduce the abstract *Gevrey space*

$$\mathcal{Y}_{\tau,\ell,q} = D(|\mathbb{Q}A|^\ell \exp(\tau|\mathbb{Q}A|^{1/q})) \tag{2.3}$$

equipped with the inner product

$$\begin{aligned}\langle U_1, U_2\rangle_{\mathcal{Y}_{\tau,\ell,q}} &= \langle \mathbb{P}U_1, \mathbb{P}U_2\rangle_{\mathcal{Y}} \\ &\quad + \langle |\mathbb{Q}A|^\ell \exp(\tau|\mathbb{Q}A|^{1/q})\mathbb{Q}U_1, |\mathbb{Q}A|^\ell \exp(\tau|\mathbb{Q}A|^{1/q})\mathbb{Q}U_2\rangle_{\mathcal{Y}}. \end{aligned}\tag{2.4}$$

Gevrey-smooth functions $U \in \mathcal{Y}_{\tau,\ell,q}$ are exponentially well approximated by their Galerkin projections $\mathbb{P}_m U$. Indeed, setting $\mathbb{Q}_m = \mathrm{id} - \mathbb{P}_m$,

$$\|\mathbb{Q}_m U\|_{\mathcal{Y}} \leqslant m^{-\ell} \exp(-\tau m^{1/q})\|U\|_{\mathcal{Y}_{\tau,\ell,q}}. \tag{2.5}$$

Moreover, this definition of the norm ensures that

$$\|A\|_{\mathcal{Y}_{\tau,\ell+1,q}\to\mathcal{Y}_{\tau,\ell,q}} \leqslant 1 \quad \text{and} \quad \|U\|_{\mathcal{Y}_{\tau,\ell,q}} \leqslant \|U\|_{\mathcal{Y}_{\tau,\ell+1,q}} \tag{2.6}$$

for all $U \in \mathcal{Y}_{\tau,\ell+1,q}$. For convenience, we define $\mathcal{Y}_\ell \equiv \mathcal{Y}_{0,\ell,q}$. We can then state the following lemma, which will be needed later.

Lemma 2.1. *Let A satisfy (A). Then, for $\sigma > \tau$ and $p \in \mathbb{N}_0$,*

$$\|A^p\|_{\mathcal{Y}_{\sigma,\ell,q}\to\mathcal{Y}_{\tau,\ell,q}} \leqslant \max\left\{1, \left(\frac{pq}{\mathrm{e}(\sigma-\tau)}\right)^{pq}\right\}. \tag{2.7}$$

Proof. For $p = 0$, there is nothing to prove; hence, let $p > 0$. For fixed $U \in \mathcal{Y}_{\sigma,\ell,q}$,

$$\|A^p U\|^2_{\mathcal{Y}_{\tau,\ell,q}} = \|\mathbb{P}U\|^2_{\mathcal{Y}} + \||\mathbb{Q}A|^p \mathrm{e}^{(\tau-\sigma)|\mathbb{Q}A|^{1/q}} |\mathbb{Q}A|^\ell \mathrm{e}^{\sigma|\mathbb{Q}A|^{1/q}} \mathbb{Q}U\|^2_{\mathcal{Y}}. \tag{2.8}$$

The function $f(\lambda) = |\lambda|^p \mathrm{e}^{(\tau-\sigma)|\lambda|^{1/q}}$ is non-negative and has a global maximum at $\lambda_* = (pq/(\sigma-\tau))^q$. Replacing the corresponding term in (2.8) by its maximum value, we obtain (2.7). □

2.2. Differentiability of the semi-flow

To obtain a flow of the evolution equation that has higher-order time derivatives, as required in § 4, we need more specific assumptions on the regularity of B on the scale of Hilbert spaces defined above.

We use the following convention to denote derivatives. Given any Hilbert space $\mathcal{X}$, open set $\mathcal{D} \subset \mathcal{X}$ and map $Z\colon \mathcal{D} \to \mathbb{R}$, we write $\mathrm{D}Z(U)$ to denote the derivative of Z at $U \in \mathcal{D}$ as an element of $\mathcal{X}^*$, and denote by $\nabla Z(U)$ the canonical representation of $\mathrm{D}Z(U)$ by an element of $\mathcal{X}$. In other words, $\mathrm{D}Z(U)W = \langle \nabla Z(U), W\rangle$, where $\langle\cdot,\cdot\rangle$ denotes the inner product on $\mathcal{X}$.

For Hilbert spaces $\mathcal{X}$ and $\mathcal{Z}$, and $j \in \mathbb{N}_0$, we write $\mathcal{E}^j(\mathcal{Z},\mathcal{X})$ to denote the vector space of j-multilinear bounded mappings from $\mathcal{Z}$ to $\mathcal{X}$; we set $\mathcal{E}^j(\mathcal{X}) \equiv \mathcal{E}^j(\mathcal{X},\mathcal{X})$. Moreover, when $\mathcal{U} \subset \mathcal{X}$ is open and $k \in \mathbb{N}$, we write $\mathcal{C}^k_{\mathrm{b}}(\mathcal{U},\mathcal{Z})$ to denote the set of k-times continuously differentiable functions $F\colon \mathcal{U} \to \mathcal{Z}$ whose derivatives $\mathrm{D}^i F$ are bounded as maps from $\mathcal{U}$ to $\mathcal{E}^i(\mathcal{X},\mathcal{Z})$ and extend to the boundary of $\mathcal{U}$. When $\mathcal{U}$ is not open but has non-empty interior, we define $\mathcal{C}^k_{\mathrm{b}}(\mathcal{U},\mathcal{Z}) = \mathcal{C}^k_{\mathrm{b}}(\operatorname{int}\mathcal{U},\mathcal{Z})$, where $\operatorname{int}\mathcal{S}$ denotes the interior of a set $\mathcal{S}$.

Finally, for Hilbert spaces $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, and open subsets $\mathcal{U} \subset \mathcal{X}$, $\mathcal{V} \subset \mathcal{Y}$ and $\mathcal{W} \subset \mathcal{Z}$, we write

$$F \in \mathcal{C}^{(\underline{m},n)}_{\mathrm{b}}(\mathcal{U}\times\mathcal{V};\mathcal{W})$$

to denote a continuous, bounded function $F\colon \mathcal{U}\times\mathcal{V} \to \mathcal{W}$ whose partial Fréchet derivatives $\mathrm{D}^i_X \mathrm{D}^j_Y F(X,Y)$ exist, are bounded and are such that the maps

$$(X, Y, X_1, \dots, X_i) \mapsto \mathrm{D}^i_X \mathrm{D}^j_Y F(X,Y)(X_1,\dots,X_i)$$

are continuous from $\mathcal{U}\times\mathcal{V}\times\mathcal{X}^i$ into $\mathcal{E}^j(\mathcal{Y},\mathcal{Z})$ for $i = 0,\dots,m$ and $j = 0,\dots,n$, and extend continuously to the boundary. If $\mathcal{U}$ or $\mathcal{V}$ are not open but have non-empty interior, we again define $\mathcal{C}^{(\underline{m},n)}_{\mathrm{b}}(\mathcal{U}\times\mathcal{V};\mathcal{W}) = \mathcal{C}^{(\underline{m},n)}_{\mathrm{b}}(\operatorname{int}\mathcal{U}\times\operatorname{int}\mathcal{V};\mathcal{W})$.

Given $\delta > 0$ and a family of open sets $\mathcal{D}_\ell \subset \mathcal{Y}_\ell$ for $\ell = 0,\dots,L$ for $L \in \mathbb{N}$, we define the sets

$$\mathcal{D}^\delta_\ell = \bigcup_{U\in\mathcal{D}_\ell} \mathcal{B}^{\mathcal{Y}_\ell}_\delta(U) \tag{2.9}$$

analogously to the set $\mathcal{D}^\delta$ above.

We assume that the sets $\mathcal{D}_\ell$ are nested, i.e. $\mathcal{D}_{\ell+1} \subset \mathcal{D}_\ell$, for $\ell = 0,\dots,L-1$. Then, by construction, we also have $\mathcal{D}^\delta_{\ell+1} \subset \mathcal{D}^\delta_\ell$ for $\ell = 0,\dots,L-1$. For example, the family $\mathcal{D}_k = \operatorname{int}\mathcal{B}^{\mathcal{Y}_k}_R(U^0)$ is nested for every $U^0 \in \mathcal{Y}_L$ and $R > 0$.

We now make the following assumption on the nonlinearity of our semilinear evolution equation.

(B1) For $\delta > 0$ fixed as in (B0), there exist $K \in \mathbb{N}_0$, $N \in \mathbb{N}$ with $N > K+1$, and a nested sequence of $\mathcal{Y}_k$-bounded and open sets $\mathcal{D}_k$, such that $B \in \mathcal{C}^{N-k}_{\mathrm{b}}(\mathcal{D}^\delta_k,\mathcal{Y}_k)$ for $k = 0,\dots,K$.

We denote the bounds of the maps $B\colon \mathcal{D}^\delta_k \to \mathcal{Y}_k$ and their derivatives by constants M_k, M'_k, etc., for $k = 0,\dots,K$. In addition to the domains $\mathcal{D}_0,\dots,\mathcal{D}_K$ defined in (B1), we also need a domain $\mathcal{D}_{K+1}$ on the next highest scale rung $\mathcal{Y}_{K+1}$, which may be any $\mathcal{Y}_{K+1}$-bounded, open and nested subset of $\mathcal{D}_K$, and we define

$$R_{K+1} = \sup_{U\in\mathcal{D}^\delta_{K+1}} \|U\|_{\mathcal{Y}_{K+1}}. \tag{2.10}$$

We can then quote the following theorem on the uniform regularity of the flow [24, theorem 2.6 and remark 2.8].

THEOREM 2.2 (regularity of semi-flow). *Assume (A) and (B1). Then there exists $T_* > 0$ such that the semi-flow $(U,t) \mapsto \Phi^t(U)$ of (1.1) satisfies*

$$\Phi \in \bigcap_{\substack{j+k \leqslant N \\ \ell \leqslant k \leqslant K+1}} \mathcal{C}_{\mathrm{b}}^{(j,\ell)}(\mathcal{D}_{K+1} \times (0,T_*); \mathcal{Y}_{k-\ell}). \tag{2.11}$$

Moreover, Φ maps $\mathcal{D}_{K+1} \times [0,T_*]$ into $\mathcal{D}_K^\delta$. The bounds on Φ and T_* depend only on the bounds afforded by (B1), (2.10), on ω and on δ.

REMARK 2.3. In [24] we chose to shrink domains $\mathcal{D}$ in the range of the flow map to

$$\mathcal{D}^{-\delta} = \{U \in \mathcal{D} \colon \operatorname{dist}(U, \partial\mathcal{D}) \geqslant \delta\}$$

rather than work with extended domains $\mathcal{D}^\delta$ in the argument of the flow map as we do here. Since $\mathcal{D}^{\varepsilon-\delta} \subset (\mathcal{D}^\varepsilon)^{-\delta}$ for $\varepsilon > \delta > 0$, the formulation in [24] implies the version stated here; working with extended domains is more convenient for the purposes of this paper as the extension preserves star-shapedness, which is required in § 2.3.

REMARK 2.4. The precise form of assumption (B1) is motivated by the typical case where B is a superposition operator of a function $f \colon D \subset \mathbb{R}^d \to \mathbb{R}^m$ and $\mathcal{Y}_\ell$ is related to the standard Sobolev space $\mathcal{H}_\ell = \mathcal{H}_\ell(I;\mathbb{R}^d)$, where $I = [a,b] \subset \mathbb{R}$. Then, if f is $(N+1)$-times continuously differentiable on some open set $D \subset \mathbb{R}^d$, it is N-times differentiable as a map from the open set $\mathcal{D}$ of $\mathcal{H}_1$ to $\mathcal{H}_1$. We note that $u \in \mathcal{D}$ ensures that $u(x) \in D$ pointwise. Moreover, f is $(N-k)$-times differentiable from $\mathcal{D} \cap \mathcal{H}_k$ to $\mathcal{H}_k$ (see [24, theorem 2.12]; [26, remark 7.4]).

For the results of § 4, we need regularity of the flow on a space of Gevrey-regular functions as well. Hence, we assume the following.

(B2) There exist $\tau > 0$, $q > 0$, $L \geqslant 0$ and a $\mathcal{Y}_{\tau,L,q}$-bounded open set $\mathcal{D}_{\tau,L,q} \subset \mathcal{D}_{K+1}$ such that $B \in \mathcal{C}_{\mathrm{b}}^2(\mathcal{D}_{\tau,L,q}^\delta, \mathcal{Y}_{\tau,L,q})$.

We note that $\mathcal{Y}_{\tau,L,q} \subset \mathcal{Y}_{K+1}$ due to lemma 2.1. In the following, $\mathcal{D}_{\tau,L+1,q}$ refers to an arbitrary $\mathcal{Y}_{\tau,L+1,q}$-bounded open subset of $\mathcal{D}_{\tau,L,q}$. We note that under assumption (B2), theorem 2.2 applies with $\mathcal{Y}_{\tau,L,q}$ in place of $\mathcal{Y}$, with $\mathcal{D}_{\tau,L+1,q}$ in place of $\mathcal{D}_{K+1}$ and with $K = 0$.

2.3. Hamiltonian structures on Hilbert spaces

Our main result, theorem 4.1 below, requires that (1.1) is Hamiltonian, i.e. that there exist a symplectic structure operator $\mathbb{J}$ and a Hamiltonian $H \colon \mathcal{D} \to \mathbb{R}$ such that

$$\partial_t U = AU + B(U) = \mathbb{J}\nabla H(U). \tag{2.12}$$

In addition to (A) and (B0), we assume the following.

(H0) The symplectic structure operator $\mathbb{J}$ is a closed, skew-symmetric, densely defined and bijective linear operator on $\mathcal{Y}$.

(H1) A is skew-symmetric and $\mathbb{J}^{-1}A$ is bounded and self-adjoint on $\mathcal{Y}$.

(H2) For every $U \in \mathcal{D}^\delta$, the operator $\mathbb{J}^{-1}\mathrm{D}B(U)$ is self-adjoint on $\mathcal{Y}$.

(H3) $\mathcal{D}$ is star shaped.

Formally, the function $H\colon \mathcal{D}^\delta \to \mathbb{R}$ is an invariant of the motion.

Recall that a subset $\mathcal{S}$ of a linear space is star shaped if there exists $U_* \in \mathcal{S}$ such that for every $W \in \mathcal{S}$ the line segment U_*W is contained in $\mathcal{S}$. We then say that $\mathcal{S}$ is star shaped with respect to U_*. We remark that if $\mathcal{D}$ is star shaped with respect to $U^* \in \mathcal{D}$, then $\mathcal{D}^\delta$ is star shaped with respect to U^* as well. Moreover, by the closed graph theorem, (H0) implies that $\mathbb{J}$ is invertible with $\mathbb{J}^{-1} \in \mathcal{E}(\mathcal{Y})$. This implies, in particular, that $\mathbb{J}^{-1}\mathrm{D}B(U)$ is a bounded operator on $\mathcal{Y}$ for every $U \in \mathcal{D}^\delta$.

An operator A is skew if $A^* = -A$ and $D(A) = D(A^*)$. This implies that $\operatorname{spec}(A) \subset \mathrm{i}\mathbb{R}$ and that, by Stone's theorem, A generates a unitary $\mathcal{C}^0$-group on $\mathcal{Y}$ (see, for example, [28]). If $A = A_\mathrm{s} + A_\mathrm{b}$, where A_s is skew and A_b is bounded, we can redefine B as $B + A_\mathrm{b}$ and A as A_s to satisfy (H1). This situation is typical for semilinear wave equations (see § 2.5).

Further, by conditions (A), (H0), and (H1),

$$\mathbb{J}^{-1}A = (\mathbb{J}^{-1}A)^* = A^*(\mathbb{J}^{-1})^* = -A(-\mathbb{J}^{-1}) = A\mathbb{J}^{-1}. \tag{2.13}$$

Hence, A and $\mathbb{J}^{-1}$ commute, which also implies the following.

Lemma 2.5. *Assume (A), (H0) and (H1). Then $\mathbb{J}^{-1}\mathbb{P}_m = \mathbb{P}_m\mathbb{J}^{-1}$ for all $m \in \mathbb{N}_0$.*

Proof. By (2.13), the normal bounded operator $L = (A+1)^{-1}$ and $\mathbb{J}^{-1}$ commute, and so do $F(L)$ and $\mathbb{J}^{-1}$ for any polynomial F. Approximating the characteristic functions χ_Λ of measurable sets $\Lambda \subset \operatorname{spec}(L)$ by polynomials, we see that $\chi_\Lambda(L)$ and $\mathbb{J}^{-1}$ also commute [28]. With $\Lambda = \{\lambda \in \operatorname{spec}(L),\ \lambda^{-1} - 1 \in \mathcal{B}_m^{\mathbb{C}}(0)\}$, this implies that $\chi_\Lambda(L) = \mathbb{P}_m$ commutes with $\mathbb{J}^{-1}$. □

The existence of a Hamiltonian H is then guaranteed by the following integrability lemma.

Lemma 2.6. *Assume (A), (B0) and (H0)–(H3) hold. Then there exists an analytic bounded Hamiltonian $H\colon \mathcal{D}^\delta \to \mathbb{R}$ for the evolution equation (1.1).*

Proof. We seek a Hamiltonian of the form

$$H(U) = \tfrac{1}{2}\langle U, \mathbb{J}^{-1}AU\rangle + V(U). \tag{2.14}$$

Due to (H1), the quadratic part of the Hamiltonian is well defined and possesses the properties claimed.

To proceed, we use that $\mathcal{D}^\delta$ is star shaped and fix U^0 such that $\mathcal{D}$ and $\mathcal{D}^\delta$ are star shaped with respect to U^0. We set

$$V(U) = \int_0^1 \langle \mathbb{J}^{-1}B(tU + (1-t)U^0), U - U^0\rangle \,\mathrm{d}t, \tag{2.15}$$

so that, for $W \in \mathcal{Y}$,

$$\begin{aligned}\langle \nabla V(U), W\rangle &= \int_0^1 \langle \mathbb{J}^{-1} B(tU + (1-t)U^0), W\rangle \,\mathrm{d}t \\ &\quad + t\int_0^1 \langle \mathbb{J}^{-1}\mathrm{D}B(tU + (1-t)U^0)W, U - U^0\rangle \,\mathrm{d}t \\ &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\langle t\mathbb{J}^{-1}B(tU + (1-t)U^0), W\rangle \,\mathrm{d}t, \end{aligned} \tag{2.16}$$

where the last equality is due to the self-adjointness of $\mathbb{J}^{-1}\mathrm{D}B(U)$. Then, by the fundamental theorem of calculus, $B(U) = \mathbb{J}\nabla V(U)$. Further, (2.15) shows that analyticity of B implies analyticity of V, and uniform bounds on B imply corresponding uniform bounds on V. □

For bounds on the modified Hamiltonian in § 4 we shall need (H3) on at least two scale rungs, so that, for simplicity, we assume the following.

(H4) Each $\mathcal{D}_k$ is star shaped for $k = 0, \ldots, K+1$.

In the next section we introduce concrete function spaces and superposition operators on these spaces in order to verify that our main examples, the nonlinear Schrödinger equation and the semilinear wave equation, fit into our abstract framework.

2.4. Spaces of analytic functions

We denote the Fourier coefficients of a function $u \in \mathcal{L}_2(\mathbb{S}^1;\mathbb{C}^d)$ on the circle $\mathbb{S}^1 \simeq \mathbb{R}/(2\pi\mathbb{Z})$ by $\hat{u}_k$, so that

$$u(x) = \frac{1}{\sqrt{2\pi}}\sum_{k\in\mathbb{Z}} \hat{u}_k \mathrm{e}^{\mathrm{i}kx}. \tag{2.17}$$

Let $\mathcal{G}_{\tau,\ell} \equiv \mathcal{G}_{\tau,\ell}(\mathbb{S}^1;\mathbb{C}^d)$ denote the Hilbert space of analytic functions $u \in \mathcal{L}_2(\mathbb{S}^1;\mathbb{C}^d)$ for which

$$\|u\|^2_{\mathcal{G}_{\tau,\ell}} \equiv \langle u, u\rangle_{\mathcal{G}_{\tau,\ell}} < \infty,$$

where the inner product is given by

$$\langle u, v\rangle_{\mathcal{G}_{\tau,\ell}} = \sum_{|k|\leqslant 1} \langle \hat{u}_k \hat{v}_k\rangle_{\mathbb{C}^d} + \sum_{|k|>1} k^{2\ell}\mathrm{e}^{2\tau|k|}\langle \hat{u}_k\hat{v}_k\rangle_{\mathbb{C}^d}. \tag{2.18}$$

It can be shown that $\mathcal{G}_{\tau,\ell}(\mathbb{S}^1;\mathbb{C}^d)$ contains all real analytic functions whose radius of analyticity is at least τ. In particular, functions in $\mathcal{G}_{\tau,\ell}(\mathbb{S}^1;\mathbb{C}^d)$ can be differentiated infinitely often. This follows from lemma 2.1 with $\mathcal{Y} = \mathcal{L}_2(\mathbb{S}^1;\mathbb{C}^d)$ and $A = \partial_x$. We write $\mathcal{H}_\ell \equiv \mathcal{G}_{0,\ell}$ to denote the usual Sobolev space of functions whose weak derivatives up to order ℓ are square integrable.

The additional index ℓ in $\mathcal{G}_{\tau,\ell}$ is important because of the following.

Lemma 2.7 (Ferrari and Titi [10, lemma 1]). *The space $\mathcal{G}_{\tau,\ell}(\mathbb{S}^1;\mathbb{C})$ is a topological algebra for every $\tau \geqslant 0$ and $\ell > \frac{1}{2}$. Specifically, there exists a constant $c = c(\ell)$ such that for every $u, v \in \mathcal{G}_{\tau,\ell}(\mathbb{S}^1;\mathbb{C})$ the product $uv \in \mathcal{G}_{\tau,\ell}(\mathbb{S}^1;\mathbb{C})$ with*

$$\|uv\|_{\mathcal{G}_{\tau,\ell}(\mathbb{S}^1;\mathbb{C})} \leqslant c\|u\|_{\mathcal{G}_{\tau,\ell}(\mathbb{S}^1;\mathbb{C})}\|v\|_{\mathcal{G}_{\tau,\ell}(\mathbb{S}^1;\mathbb{C})}. \tag{2.19}$$

To treat general nonlinear potentials, we need to consider superposition operators $f\colon \mathcal{G}_{\tau,\ell} \to \mathcal{G}_{\tau,\ell}$ of analytic functions. The following lemma is a minor adaptation of results proved in [10, 19].

LEMMA 2.8. *If $f\colon \mathbb{C}^d \to \mathbb{C}^d$ is entire, then f is also entire as a function from $\mathcal{G}_{\tau,\ell}(\mathbb{S}^1;\mathbb{C}^d)$ to itself for every $\tau \geqslant 0$ and $\ell > \frac{1}{2}$. If f is analytic on $\mathcal{B}_r(u_0) \subset \mathbb{C}^d$, where $u_0 \in \mathbb{C}^d$, then f is analytic from $\mathcal{D}_{\tau,\ell} \equiv \mathcal{B}_R(u_0) \subset \mathcal{G}_{\tau,\ell}(\mathbb{S}^1;\mathbb{C}^d)$ to $\mathcal{G}_{\tau,\ell}(\mathbb{S}^1;\mathbb{C}^d)$ for any $R < r/c$, where $c = c(\ell)$ is the constant from lemma 2.7. Moreover, $f \in \mathcal{C}^2_{\mathrm{b}}(\mathcal{D}_{\tau,\ell};\mathcal{G}_{\tau,\ell})$.*

Proof. We prove the result for $d = 1$; it generalizes to $d > 1$ if multi-indices are used. Let f be entire and let

$$f(z) = \sum_{n=0}^{\infty} a_n (z - u_0)^n \tag{2.20}$$

be the Taylor series of f around $u_0 \in \mathbb{C}$. Let $\phi\colon \mathbb{R} \to \mathbb{R}$ be its majorization

$$\phi(s) = \sum_{n=0}^{\infty} |a_n| s^n.$$

By applying the algebra inequality (2.19) to each term of the power series expansion (2.20) of $f(u)$, we see that the series converges for every $u \in \mathcal{G}_{\tau,\ell}$ provided $\tau \geqslant 0$ and $\ell > \frac{1}{2}$, and that

$$\|f(u)\|_{\mathcal{G}_{\tau,\ell}} \leqslant c^{-1}\phi(c\|u - u_0\|_{\mathcal{G}_{\tau,\ell}}) + |a_0|(\sqrt{2\pi} - c^{-1}), \tag{2.21}$$

where c is as in lemma 2.7 (see [10]). In other words, f is entire on $\mathcal{G}_{\tau,\ell}(\mathbb{S}^1)$.

When f has only a finite radius of analyticity, we argue as follows. Assume that $|f(z)| \leqslant M$ on $\mathcal{B}^{\mathbb{C}}_r(u_0)$. Then, by Cauchy's estimate,

$$|a_n| \leqslant \frac{M}{r^n}.$$

Consequently, the majorant ϕ is bounded on any $\mathcal{B}^{\mathbb{C}}_\rho(0)$ with $\rho < r$ with uniform bound

$$\mu = \frac{M}{1 - \rho/r}.$$

Due to (2.21), the superposition operator f is then analytic and bounded by

$$M_{\mathrm{spp}} = \frac{\mu}{c} + |a_0|(\sqrt{2\pi} - c^{-1})$$

as a map from a ball $\mathcal{D}_{\tau,\ell} = \mathcal{B}_R(u_0)$ of radius $R = \rho/c$ around $u_0 \in \mathcal{G}_{\tau,\ell}$ into $\mathcal{G}_{\tau,\ell}$; similarly, we see that $f \in \mathcal{C}^2_{\mathrm{b}}(\mathcal{D}_{\tau,\ell};\mathcal{G}_{\tau,\ell})$. □

2.5. Functional setting for the semilinear wave equation

Consider the semilinear wave equation

$$\partial_{tt} u = \partial_{xx} u - V'(u) \tag{2.22}$$

on $\mathbb{S}^1$. Its Hamiltonian can be written

$$H(u,v) = \int_{\mathbb{S}^1} [\tfrac{1}{2}v^2 + \tfrac{1}{2}(\partial_x u)^2 + V(u)]\,\mathrm{d}x,$$

where $v = \partial_t u$. We write $U = (u,v)^{\mathrm{T}}$ and set $\mathcal{Y} \equiv \mathcal{H}_1(\mathbb{S}^1;\mathbb{R}) \times \mathcal{L}_2(\mathbb{S}^1;\mathbb{R})$ so that the Hamiltonian is well defined on $\mathcal{Y}$. For $U = (u,v) \in \mathcal{Y}$ let $\mathbb{P}_0 U = (\mathsf{p}_0 u, \mathsf{p}_0 v)$ where, for $u \in \mathcal{L}_2(\mathbb{S}^1;\mathbb{R})$, we define $\mathsf{p}_0 u = \hat{u}_0$ and let $\mathbb{Q}_0 = \mathrm{id} - \mathbb{P}_0$. Setting

$$\tilde{A} = \begin{pmatrix} 0 & \mathrm{id} \\ \partial_x^2 & 0 \end{pmatrix},$$

we then define

$$A = \mathbb{Q}_0 \tilde{A}, \qquad B(U) = \begin{pmatrix} 0 \\ -V'(u) \end{pmatrix} + \mathbb{P}_0 \tilde{A} U, \tag{2.23}$$

and the symplectic structure operator $\mathbb{J}$ via

$$\langle \mathbb{J}^{-1} U_1, U_2 \rangle_{\mathcal{Y}} = \int_{\mathbb{S}^1} (u_1 v_2 - u_2 v_1)\,\mathrm{d}x \tag{2.24}$$

for all $U_1 = (u_1, v_1)^{\mathrm{T}}, U_2 = (u_2, v_2)^{\mathrm{T}} \in \mathcal{Y}$.

Since the Laplacian is diagonal in the Fourier representation (2.17) with eigenvalues $-k^2$ for $k \in \mathbb{Z}$, the eigenvalue problem for A separates into 2×2 eigenvalue problems on each Fourier mode, and $\operatorname{spec} A = \mathrm{i}\mathbb{Z}\backslash\{0\}$. Clearly, A is skew-symmetric on $\mathcal{Y}$ if $\mathcal{H}_1 = \mathcal{G}_{0,1}$ is endowed with the inner product (2.18). Note that $\mathbb{P}_0 \tilde{A}$ has a Jordan block and is hence included with the nonlinearity B. Thus, with

$$\mathcal{Y}_{\tau,\ell,q} = \mathcal{G}_{\tau,\ell+1}(\mathbb{S}^1;\mathbb{R}) \times \mathcal{G}_{\tau,\ell}(\mathbb{S}^1;\mathbb{R}),$$

$B(U)$ from (2.23) satisfies (B2) with $q = 1$.

The symplectic structure operator $\mathbb{J}$ defined by (2.24) is an unbounded operator on $\mathcal{Y}_{\tau,\ell,1}$ with domain $\mathcal{Y}_{\tau,\ell+1,1}$. It is possible, though not necessary for anything which follows, to compute $\mathbb{J}^{-1}$ explicitly. Namely, (2.24) reads

$$\langle \mathbb{J}^{-1} U_1, U_2 \rangle_{\mathcal{Y}} = \langle (\mathbb{J}^{-1} U_1)_u, u_2 \rangle_{\mathcal{H}_1} + \langle (\mathbb{J}^{-1} U_1)_v, v_2 \rangle_{\mathcal{L}_2} = \int_{\mathbb{S}^1} (u_1 v_2 - u_2 v_1)\,\mathrm{d}x. \tag{2.25}$$

The definition of the inner product (2.18) implies

$$\langle (\mathbb{J}^{-1} U_1)_u, u_2 \rangle_{\mathcal{H}_1} = \langle (\mathsf{p}_0 - \partial_x^2)(\mathbb{J}^{-1} U_1)_u, u_2 \rangle_{\mathcal{L}_2},$$

so that (2.25) splits into

$$(\mathsf{p}_0 - \partial_x^2)(\mathbb{J}^{-1} U_1)_u = -v_1 \quad \text{and} \quad (\mathbb{J}^{-1} U_1)_v = u_1.$$

We conclude that

$$\mathbb{J}^{-1} = \begin{pmatrix} 0 & -(\mathsf{p}_0 - \partial_x^2)^{-1} \\ 1 & 0 \end{pmatrix}.$$

If the potential $V\colon D \to \mathbb{R}$ is analytic on an open set $D \subset \mathbb{R}$, then, by lemma 2.8, B is analytic from $\mathcal{B}_R^{\mathcal{Y}_{\tau,\ell,1}}(U^0)$ to $\mathcal{Y}_{\tau,\ell,1}$ for any $\tau, \ell \geqslant 0$ and $U^0 = (u^0, v^0) \in \mathcal{Y}_{\tau,\ell,1}$ with u^0 independent of x, provided $\mathcal{B}_r^{\mathbb{R}}(u^0) \subset D$ with $R < r/c(\ell)$ as in lemma 2.8. In this setting, all the above assumptions are satisfied.

2.6. Functional setting for the nonlinear Schrödinger equation

Consider the nonlinear Schrödinger equation

$$\mathrm{i}\partial_t u = -\partial_{xx} u + \partial_{\bar{u}} V(u, \bar{u}) \tag{2.26}$$

on the circle $\mathbb{S}^1$, where $V(u, \bar{u})$ is analytic in $\operatorname{Re} u$ and $\operatorname{Im} u$. Setting $U \equiv u$, we can write

$$A = \mathrm{i}\partial_x^2, \quad B(U) = -\mathrm{i}\partial_{\bar{u}} V(u, \bar{u}). \tag{2.27}$$

Similar to (2.24), we have

$$\langle \mathbb{J}^{-1} u_1, u_2 \rangle_{\mathcal{Y}} = \int \operatorname{Re}(\mathrm{i} u_1 \bar{u}_2)\, \mathrm{d}x \tag{2.28}$$

with $\mathcal{Y} = \mathcal{H}_1(\mathbb{S}^1, \mathbb{C})$. Therefore, as for the semilinear wave equation in § 2.5, we see that $\mathbb{J}^{-1} \colon \mathcal{H}_2 \to \mathcal{L}_2$ and that (H0) and (H1) hold. The first term of the Hamiltonian

$$H(U) = \frac{1}{2} \int_{\mathbb{S}^1} (|\partial_x u|^2 + V(u, \bar{u}))\, \mathrm{d}x \tag{2.29}$$

is then well defined for all $u \in \mathcal{Y}$. The Laplacian is diagonal in the Fourier representation (2.17) with eigenvalues $-k^2$. Hence, $\operatorname{spec} A = \{-\mathrm{i}k^2 \colon k \in \mathbb{Z}\}$ so that A generates a unitary group on $\mathcal{L}_2(\mathbb{S}^1; \mathbb{C})$ and, more generally, on every $\mathcal{G}_{\tau,\ell}$ with $\ell \in \mathbb{N}_0$ and $\tau \geqslant 0$.

To continue, we identify $\mathbb{R}^2 \simeq \mathbb{C}$ so that $\mathcal{Y} = \mathcal{H}_1(\mathbb{S}^1, \mathbb{R}^2)$. If the potential $V \colon D \subset \mathbb{R}^2 \to \mathbb{R}$ is analytic as a function of $(q, p) \equiv (\operatorname{Re} u, \operatorname{Im} u)$, then, by lemma 2.8, the nonlinearity $B(U)$ defined in (2.27) is analytic as a map from a ball in $\mathcal{G}_{\tau,\ell}(\mathbb{S}^1; \mathbb{R}^2)$ to itself for every $\tau \geqslant 0$ and $\ell > \frac{1}{2}$. The construction of the domain hierarchy works as in § 2.5, so that (B0)–(B2) hold with $\mathcal{Y}_{\tau,\ell,q} = \mathcal{G}_{\tau,2\ell+1}(\mathbb{S}^1; \mathbb{R}^2)$, where $\ell \in \mathbb{N}_0$ and $q = 2$.

We remark that if we were to write out the nonlinear Schrödinger equation in real coordinates with $U = (\operatorname{Re} u, \operatorname{Im} u)$, the structure operator $\mathbb{J}$ would be the canonical symplectic matrix on $\mathbb{R}^2$ for the space $\mathcal{L}_2(\mathbb{S}^1; \mathbb{R}^2)$.

An example of a nonlinear Schrödinger equation in Fourier space that is defined on a more complicated set than a ball can be found in remark 5.1.

3. A-stable Runge–Kutta methods on Hilbert spaces

In this section, we introduce a class of A-stable Runge–Kutta methods that are well defined when applied to the semilinear PDE (1.1) under assumptions (A) and (B0), and review some regularity and convergence results for those methods from [24]. In most of this section, we need not assume that (1.1) is Hamiltonian.

Applying an s-stage Runge–Kutta method of the form (3.1) to the semilinear evolution equation (1.1), we obtain

$$W = U^0 \mathbb{1} + h\mathsf{a} F(W), \tag{3.1 a}$$

$$\Psi^h(U^0) = U^0 + h\mathsf{b}^{\mathrm{T}} F(W). \tag{3.1 b}$$

For $U \in \mathcal{Y}$, we write

$$\mathbb{1}U = \begin{pmatrix} U \\ \vdots \\ U \end{pmatrix} \in \mathcal{Y}^s, \qquad W = \begin{pmatrix} W^1 \\ \vdots \\ W^s \end{pmatrix}, \qquad B(W) = \begin{pmatrix} B(W^1) \\ \vdots \\ B(W^s) \end{pmatrix},$$

where $W^1, \dots, W^s$ are the stages of the Runge–Kutta method,

$$(\mathsf{a}W)^i = \sum_{j=1}^{s} \mathsf{a}_{ij} W^j, \qquad \mathsf{b}^{\mathrm{T}} W = \sum_{j=1}^{s} \mathsf{b}_j W^j,$$

and A acts diagonally on the stages, i.e. $(AW)^i = AW^i$ for $i = 1, \dots, s$. In the following, we denote s copies of $\mathcal{Y}$ by $\mathcal{Y}^s$ endowed with norm

$$\|W\|_{\mathcal{Y}^s} = \max_{j=1,\dots,s} \|W^j\|_{\mathcal{Y}}.$$

The scheme (3.1) can be written in a more suitable form, required later, namely

$$W = \Pi(W; U, h) \equiv (\mathrm{id} - h\mathsf{a}A)^{-1}(\mathbb{1}U + h\mathsf{a}B(W)) \tag{3.2a}$$

and

$$\Psi^h(U) = \mathsf{S}(hA)U + h\mathsf{b}^{\mathrm{T}}(\mathrm{id} - h\mathsf{a}A)^{-1}B(W(U,h)), \tag{3.2b}$$

where S is the so-called *stability function*

$$\mathsf{S}(z) = 1 + z\mathsf{b}^{\mathrm{T}}(\mathrm{id} - z\mathsf{a})^{-1}\mathbb{1}. \tag{3.3}$$

We now make a number of assumptions on the method and its interaction with the linear operator A. First, we assume that the method is A-stable. Setting $\mathbb{C}^- = \{z \in \mathbb{C} \colon \operatorname{Re} z \leqslant 0\}$, the conditions are as follows.

(RK1) The stability function (3.3) is bounded with $|\mathsf{S}(z)| \leqslant 1$ for all $z \in \mathbb{C}^-$.

(RK2) The $s \times s$ matrices $\mathrm{id} - z\mathsf{a}$ are invertible for all $z \in \mathbb{C}^-$.

We also require two further conditions.

(RK3) The matrix a is invertible.

(RK4) The method is symplectic.

Recall that the flow map of a Hamiltonian system is a symplectic map, i.e. Φ^t satisfies

$$(\mathrm{D}_U \Phi^t(U))^{\mathrm{T}} \mathbb{J}^{-1} \mathrm{D}_U \Phi^t(U) = \mathbb{J}^{-1} \tag{3.4}$$

for all U and t for which this relation is well defined [18]. A numerical one-step method is called symplectic if, when applied to a Hamiltonian system, its time-h map Ψ^h is symplectic.

It is known that a Runge–Kutta method of the form (3.1) is symplectic when its coefficients satisfy

$$\mathsf{b}_i \mathsf{a}_{ij} + \mathsf{b}_j \mathsf{a}_{ji} - \mathsf{b}_i \mathsf{b}_j = 0$$

for $i, j = 1, \dots, s$ (see, for example, [30]). The simplest example of a symplectic Runge–Kutta method is the *implicit midpoint rule*, given by

$$\Psi^h(U) = U + hF\left(\frac{U + \Psi^h(U)}{2}\right),$$

which, equivalently, can be written in the form of a general Runge–Kutta scheme (3.1) with $s = 1$, $\mathsf{a}_{11} = \frac{1}{2}$, and $\mathsf{b}_1 = 1$. This is an example of a Gauss–Legendre Runge–Kutta method; Gauss–Legendre Runge–Kutta methods satisfy conditions (RK1)–(RK4); see [24, lemma 3.6] for conditions (RK1)–(RK3) and [30], for example, for (RK4).

In the following, we also need to refer to a set of key estimates on the linear operators that appear in the formulation (3.2 *a*), (3.2 *b*) of the Runge–Kutta method, namely

$$\|(\mathrm{id} - h\mathsf{a}A)^{-1}\|_{\mathcal{Y}^s \to \mathcal{Y}^s} \leqslant \Lambda, \tag{3.5 a}$$

$$\|h\mathsf{a}A(\mathrm{id} - h\mathsf{a}A)^{-1}\|_{\mathcal{Y}^s \to \mathcal{Y}^s} \leqslant 1 + \Lambda, \tag{3.5 b}$$

$$\|\mathsf{S}(hA)\|_{\mathcal{Y}^s \to \mathcal{Y}^s} \leqslant 1 + \sigma h \tag{3.5 c}$$

for all $h \in [0, h_*]$ and constants $\Lambda \geqslant 1$ and $\sigma \geqslant 0$. These estimates naturally hold true on each rung of our hierarchy of spaces. For proofs, see [24, § 3.2].

A-stable Runge–Kutta methods have the remarkable property that their time-h map is of the same regularity class as the flow of the evolution equation stated in theorem 2.2. We state this result as an abbreviated version of [24, theorem 3.15 and remark 3.17].

THEOREM 3.1 (regularity of numerical method). *Assume (A), (B1) and (RK1)–(RK3). Then there exists $h_* > 0$ such that the components W^j of the stage vector $W(U, h)$ and numerical method $\Psi(U, h) = \Psi^h(U)$ are of class (2.11), with T_* there replaced by h_* here. Moreover, Ψ and W^j map into $\mathcal{D}^\delta_K$. The bounds on W, Ψ and h_* only depend on the bounds afforded by (B1) and (2.10), on the coefficients of the method, on the constants afforded by (3.5) and on δ.*

REMARK 3.2. Even when $B\colon \mathcal{Y} \to \mathcal{Y}$ is analytic, the numerical time-h map $\Psi^h(U)$ is generally not analytic in h unless A is bounded. Take, for example, the linear Schrödinger equation, i.e. (2.26) with $B \equiv 0$, discretized by the implicit midpoint rule. Then

$$h \mapsto \mathsf{S}(hA)e_k = (\mathrm{id} + \tfrac{1}{2}hA)(\mathrm{id} - \tfrac{1}{2}hA)^{-1}e_k = \frac{1 + \frac{1}{2}hk^2\mathrm{i}}{1 - \frac{1}{2}hk^2\mathrm{i}}e_k$$

has radius of analyticity $2/k$, where e_k is the kth Galerkin mode of A as described in § 2.6. Therefore, if the Fourier expansion of U does not terminate finitely, then $\Psi^h(U) = \mathsf{S}(hA)U$ cannot be analytic in h. This argument applies to any A-stable Runge–Kutta method: since $|\mathsf{S}(z)| \leqslant 1$ for all $z \in \mathrm{i}\mathbb{R}$ by assumption (RK1), the stability function is a rational polynomial $\mathsf{S}(z) = \mathsf{P}(z)/\mathsf{Q}(z)$ with $\deg \mathsf{Q} \geqslant \deg \mathsf{P}$. Hence, $\deg \mathsf{Q} \geqslant 1$ so that $\mathsf{S}(z)$ has at least one pole z_0. The radius of analyticity of $\mathsf{S}(z)$ around 0 is $r_0 = |z_0|$, so that $h \mapsto \mathsf{S}(hA)e_k$ cannot be analytic outside a ball around $h = 0$ of radius r_0/k. As for the implicit midpoint rule, this implies that $\Psi^h(U)$ is not analytic in h unless the Fourier expansion of U is finite.

Thus, while differentiability in h can be obtained by stepping down on a scale of Hilbert spaces, analyticity can only be obtained by projecting onto a subspace on which the vector field is bounded. This will become necessary in § 4.3, where analyticity is essential for obtaining exponential error estimates.

4. Exponentially accurate Hamiltonian embeddings of time semi-discretizations

We are now ready to state and prove our main result on approximate embeddings of symplectic time discretizations of Hamiltonian evolution equations into flows.

4.1. Statement of the main result

In the following, we write $j = \lfloor r \rfloor$ to denote the largest integer $j \leqslant r$, and $j = \lceil r \rceil$ to denote the smallest integer $j \geqslant r$.

THEOREM 4.1 (main theorem). *Assume that the semilinear Hamiltonian evolution equation (2.12) with energy (2.14) satisfies (A), (B0)–(B2) and (H0)–(H4). Apply a symplectic Runge–Kutta method of order $p \geqslant 1$ and step size h which satisfies (RK1)–(RK4) to (2.12). Assume further that*

$$K + 1 \geqslant P \equiv \left\lceil \frac{p(q+1)^2}{q} + q \right\rceil + 1$$

with K from (B1) and q from (B2). Then there exists $h_* > 0$ and a modified energy $\tilde{H} \colon \mathcal{D}_1^{\delta/2} \times [0, h_*] \to \mathbb{R}$ which is analytic in U for each $h \in [0, h_*]$ and satisfies

$$\sup_{U \in \mathcal{D}_P} |\tilde{H}(U,h) - H(U,h)| = O(h^p) \tag{4.1 a}$$

such that $\mathbb{J}\nabla\tilde{H}$ generates a modified flow $\tilde{\Phi} \colon \mathcal{D}_1 \times [0, h_*] \to \mathcal{D}_1^{\delta/4}$. The numerical method is approximately embedded into the modified flow with an exponentially small error in the sense that there is $c_* > 0$ such that

$$\sup_{U \in \mathcal{D}_{\tau, L+1, q}} \|\Psi^h(U) - \tilde{\Phi}^h(U)\|_{\mathcal{Y}_1} \leqslant c_{\tilde{\Phi}} \exp(-c_* h^{-1/(1+q)}). \tag{4.1 b}$$

The modified energy is also approximately conserved by the numerical method:

$$\sup_{U \in \mathcal{D}_{\tau, L+1, q}} |\tilde{H}(\Psi^h(U), h) - \tilde{H}(U, h)| \leqslant c_{\tilde{H}} \exp(-c_* h^{-1/(1+q)}). \tag{4.1 c}$$

For the semilinear wave equation, $q = 1$ (see § 2.5), so the exponents in (4.1 *b*) and (4.1 *c*) scale like $h^{-1/2}$. For the nonlinear Schrödinger equations, $q = 2$ (see § 2.6), so the exponents in (4.1 *b*) and (4.1 *c*) scale like $h^{-1/3}$. Note that, due to (B2), $\mathcal{D}_{\tau, L+1, q} \subset \mathcal{D}_{K+1} \subset \mathcal{D}_P$, so the supremum in (4.1 *a*) can be taken, in particular, over $\mathcal{D}_{\tau, L+1, q}$.

REMARK 4.2. For ODEs, estimate (4.1 *c*) holds with $q = 0$ and implies approximate conservation of the energy H over exponentially long times so long as the numerical trajectory remains bounded. For PDEs this conclusion does not hold, because solutions of semilinear PDEs and their discretizations are not generally Gevrey regular over long times, while Gevrey regularity is needed for the embedding estimate (4.1 *b*) (see § 5).

REMARK 4.3. When the evolution equation (2.12) is linear, e.g. a linear wave equation or linear Schrödinger equation, its Hamiltonian is conserved exactly, since symplectic Runge–Kutta methods conserve quadratic invariants [13].

REMARK 4.4. Our result implies conservation of the modified energy with exponentially small error over finite times under slightly stronger conditions. We assume that there is a triple of Gevrey spaces as follows.

(B3) There are $\tau > 0$, $q > 0$, $L \geqslant 0$, and a sequence of nested $\mathcal{Y}_{\tau,L+k,q}$-bounded and open sets $\mathcal{D}_{\tau,L+k,q}$ such that $B \in \mathcal{C}^{3-k}_{\mathrm{b}}(\mathcal{D}^{\delta}_{\tau,L+k,q}, \mathcal{Y}_{\tau,L+k,q})$ for $k = 0, 1, 2$.

Let $\mathcal{D}_{\tau,L+3,q} \subset \mathcal{D}_{\tau,L+2,q}$ be an open and bounded subset of $\mathcal{Y}_{\tau,L+3,q}$. Fix $T > 0$ and $\delta > \varepsilon > 0$. Then for any U^0 with

$$\{\Phi^t(U^0) \colon t \in [0,T]\} \subset \mathcal{D}_{\tau,L+3,q} \tag{4.2}$$

and for any $h \in [0, h_*]$, the convergence theorem [24, theorem 3.20] with $p \equiv 1$, with $\mathcal{Y}_{\tau,L+1,q}$ in place of $\mathcal{Y}$, ensures that there is $h_* > 0$ such that, for any $h \in (0, h_*]$, the discrete trajectory $U^j = (\Psi^h)^j(U^0)$ is $O(h)$-close to the flow in the $\mathcal{Y}_{\tau,L+1,q}$ norm, and hence satisfies $U^j \in \mathcal{D}^{\varepsilon}_{\tau,L+1,q}$ so long as $0 \leqslant j \leqslant \lfloor T/h \rfloor$ and $h > 0$ is sufficiently small. Theorem 4.1 with $\mathcal{D}_{\tau,L+1,q}$ replaced by $\mathcal{D}^{\varepsilon}_{\tau,L+1,q}$, $\mathcal{D}_j$ replaced by $\mathcal{D}^{\varepsilon}_j$ and δ reduced to $\delta - \varepsilon$ then implies approximate conservation of the modified energy $\tilde{H}$ with an error

$$h^{-1} O(\exp(-c_* h^{-1/(1+q)})) = O(\exp(-\beta h^{-1/(1+q)}))$$

for any $\beta \in (0, c_*)$ with order constants uniform over all U^0 satisfying (4.2).

The remainder of the section is devoted to the proof of theorem 4.1, where claims (4.1 *a*)–(4.1 *c*) correspond to lemmas 4.18, 4.9 and 4.12, respectively.

Lemma 4.9 generalizes the well-known embedding result for ODEs, stated as theorem 4.7, to the Hilbert space setting. Theorem 4.7 is not directly applicable to PDEs because the formal expansion in h of the numerical method contains powers of the unbounded operator A. We thus resort to the following construction, which is also used in [21]: in § 4.2, we truncate the evolution equation (2.12) to the subspace $\mathbb{P}_m\mathcal{Y}$. Then, in § 4.3, we obtain an embedding result on this subspace and choose an optimal cut-off m as a function of h to obtain the embedding result in the Hilbert space setting. Finally, in § 4.4, we prove estimate (4.1 *a*).

4.2. Galerkin truncation

For given $m \in \mathbb{N}$, we define a truncated Hamiltonian evolution equation by restricting the Hamiltonian phase space to the subspace $\mathbb{P}_m\mathcal{Y}$. Since $\nabla H|_{\mathbb{P}_m\mathcal{Y}} = \mathbb{P}_m \nabla H$ and $\mathbb{J}^{-1}$ leaves $\mathbb{P}_m\mathcal{Y}$ invariant by lemma 2.5, the corresponding restricted evolution equation reads

$$\dot{u}_m = \mathbb{J}\mathbb{P}_m \nabla H(u_m).$$

Thus, setting $f_m = \mathbb{P}_m F$ and $B_m = \mathbb{P}_m B$, we can write

$$\dot{u}_m \equiv f_m(u_m) = A u_m + B_m(u_m). \tag{4.3}$$

We denote the flow of the projected system on $\mathbb{P}_m\mathcal{Y}$ by ϕ_m^t. For convenience, we set $\Phi_m^t = \phi_m^t \circ \mathbb{P}_m$. Similarly, let w_m denote the stage vector, with w_m^j its components for $j = 1, \dots, s$, let ψ_m^h denote the numerical time-h map obtained by applying an s-stage Runge–Kutta method to the projected semilinear evolution equation (4.3) and abbreviate $W_m^j = w_m^j \circ \mathbb{P}_m$ and $\Psi_m^h = \psi_m^h \circ \mathbb{P}_m$.

In [25], we proved that all of the maps above (the truncated flow Φ_m^t, the components of the stage vector W_m^j and the numerical time-h map Ψ_m^h of the truncated system) are of the same class (i.e. class (2.11)) as the exact flow Φ^t with m-independent bounds. The precise statement is as follows.

THEOREM 4.5 (regularity of flow and numerical method of projected system). *Assume (A), (B1) and (RK1)–(RK3). Then there are positive T_*, h_* and m_* such that for every $m \geqslant m_*$ the flow Φ_m^t is of class (2.11) and the components of the numerical stage vector W_m^j and the numerical time-h map Ψ_m^h are of the same class, but with T_* replaced by h_*, with bounds that are independent of $m \geqslant m_*$. Moreover, Φ_m^t, W_m^j and Ψ_m^h map $\mathcal{D}_{K+1}$ into $\mathcal{D}_K^\delta$.*

Note that Φ_m^t (respectively, Ψ_m^h) are analytic in t (respectively, in h) so long as B is analytic on $\mathcal{Y}$. However, the radius of analyticity is generally non-uniform in m.

Next, we present an exponential error bound for the projection error of the numerical scheme; this is necessary for obtaining an exponentially small embedding error in § 4.3 below.

LEMMA 4.6 (exponential projection error estimate for the numerical scheme). *Assume that the semilinear evolution equation (1.1) satisfies conditions (A) and (B0)–(B2). As before, let Ψ^h and Ψ_m^h denote a single step of a Runge–Kutta method subject to (RK1)–(RK3) applied to the full and projected semilinear evolution equations, (1.1) and (4.3), respectively. Then there are positive constants h_*, m_* and c_Ψ such that, for all $m \geqslant m_*$, $h \in [0, h_*]$ and $U \in \mathcal{D}_{\tau,L+1,q}$,*

$$\|\Psi^h(U) - \Psi_m^h(U)\|_{\mathcal{Y}_1} \leqslant c_\Psi m^{-L} \exp(-\tau m^{1/q}). \tag{4.4}$$

Proof. By theorems 3.1 and 4.5 applied with $\mathcal{Y}_{\tau,L,q}$ in place of $\mathcal{Y}$ and $K = 0$, there exist $h_* > 0$ and $m_* > 0$ such that

$$\Psi^h, \Psi_m^h, W^j, W_m^j \in \mathcal{C}_{\mathrm{b}}(\mathcal{D}_{\tau,L+1,q} \times [0, h_*]; \mathcal{D}_{\tau,L,q}^\delta)$$

for $j = 1, \dots, s$ with bounds that are uniform in $h \in (0, h_*]$ and $m \geqslant m_*$.

We first estimate the difference of the stage vectors $W(U) - W_m(U)$, noting that

$$\begin{aligned} W(U) &= (\mathrm{id} - h\mathsf{a}A)^{-1}(\mathbb{1}U + h\mathsf{a}B(W(U))), \\ W_m(U) &= (\mathrm{id} - h\mathsf{a}A)^{-1}(\mathbb{P}_m\mathbb{1}U + h\mathbb{P}_m\mathsf{a}B(W_m(U))). \end{aligned}$$

Taking the difference of the expressions, we obtain

$$W(U) - W_m(U) = (\mathrm{id} - h\mathsf{a}A)^{-1}(\mathbb{Q}_m\mathbb{1}U + h\mathsf{a}[B(W(U)) - \mathbb{P}_m B(W_m(U))]). \tag{4.5}$$

Setting

$$\|\mathsf{b}\| = \sum_{i=1}^s |\mathsf{b}_i| \quad \text{and} \quad \|\mathsf{a}\| = \max_{i=1,\dots,s} \sum_{j=1}^s |\mathsf{a}_{ij}|,$$

and using estimate (3.5 a), we obtain

$$\|h\mathsf{a}(\mathrm{id}-h\mathsf{a}A)^{-1}(B(W(U))-\mathbb{P}_mB(W_m(U)))\|_{\mathcal{Y}^s} \leqslant h\Lambda\|\mathsf{a}\|\|B(W(U))-\mathbb{P}_mB(W_m(U))\|_{\mathcal{Y}^s}.$$

Using the triangle inequality and the mean-value theorem, we further estimate

$$\begin{aligned}\|B(W(U))-\mathbb{P}_mB(W_m(U))\|_{\mathcal{Y}^s} &\leqslant \|\mathbb{Q}_mB(W(U))\|_{\mathcal{Y}^s}+\|\mathbb{P}_mB(W(U))-\mathbb{P}_mB(W_m(U))\|_{\mathcal{Y}^s}\\ &\leqslant \|\mathbb{Q}_mB(W(U))\|_{\mathcal{Y}^s}+M_0'\|W(U)-W_m(U)\|_{\mathcal{Y}^s}. \end{aligned}\tag{4.6}$$

Taking the $\mathcal{Y}^s$ norm of (4.5), using (3.5 a), and inserting (4.6), we obtain

$$\|W(U)-W_m(U)\|_{\mathcal{Y}^s} \leqslant s\Lambda\|\mathbb{Q}_mU\|_{\mathcal{Y}}+h\|\mathsf{a}\|\Lambda\|\mathbb{Q}_mB(W(U))\|_{\mathcal{Y}^s} + hM_0'\|\mathsf{a}\|\Lambda\|W(U)-W_m(U)\|_{\mathcal{Y}^s}.$$

This proves that, for $h_* < 1/(M_0'\|\mathsf{a}\|\Lambda)$,

$$\|W(U)-W_m(U)\|_{\mathcal{Y}^s} \leqslant s\Lambda\frac{\|\mathbb{Q}_mU\|_{\mathcal{Y}}}{1-h_*M_0'\Lambda\|\mathsf{a}\|}+h_*\|\mathsf{a}\|\Lambda\frac{\|\mathbb{Q}_mB(W(U))\|_{\mathcal{Y}^s}}{1-h_*M_0'\Lambda\|\mathsf{a}\|}.$$

We apply (2.5) to the first term on the right and note that, again by (2.5),

$$\|\mathbb{Q}_mB(W(U))\|_{\mathcal{Y}^s} \leqslant sm^{-L}\exp(-\tau m^{1/q})M_{\tau,L},\tag{4.7}$$

where $M_{\tau,L,q}$ is the bound for the norm of $B\colon \mathcal{D}^{\delta}_{\tau,L,q}\to\mathcal{Y}_{\tau,L,q}$ afforded by assumption (B2). This establishes that there exists a constant c_W such that

$$\|W(U)-W_m(U)\|_{\mathcal{Y}^s} \leqslant c_Wm^{-L}\exp(-\tau m^{1/q})\tag{4.8}$$

for all $m \geqslant m_*$ and $h\in[0,h_*]$. (As in [25], where we considered the case $\tau=0$, we could obtain a stage vector error bound in the $\mathcal{Y}^s_1$-norm by applying A to (4.5) and using (3.5 b), but this is not necessary for what follows.)

To estimate the difference between the Runge–Kutta updates, we write them in the form (3.2 b) such that the respective right-hand sides are (uniformly) bounded operators:

$$\begin{aligned}\Psi^h(U) &= \mathsf{S}(hA)U+h\mathsf{b}^{\mathrm{T}}(\mathrm{id}-h\mathsf{a}A)^{-1}B(W(U)),\\ \Psi^h_m(U) &= \mathsf{S}(hA)\mathbb{P}_mU+h\mathsf{b}^{\mathrm{T}}(\mathrm{id}-h\mathsf{a}A)^{-1}\mathbb{P}_mB(W_m(U)).\end{aligned}$$

Then,

$$\Psi^h(U)-\Psi^h_m(U) = \mathsf{S}(hA)\mathbb{Q}_mU+h\mathsf{b}^{\mathrm{T}}(\mathrm{id}-h\mathsf{a}A)^{-1}[B(W(U))-\mathbb{P}_mB(W_m(U))].$$

Inserting (4.7) and (4.8) back into (4.6), we also find that

$$\|B(W(U))-\mathbb{P}_mB(W_m(U))\|_{\mathcal{Y}^s} \leqslant c_Bm^{-L}\exp(-\tau m^{1/q})\tag{4.9}$$

for some constant $c_B>0$. We note that (2.4) and (3.5 b) imply

$$\|h\mathsf{a}(\mathrm{id}-h\mathsf{a}A)^{-1}\|_{\mathcal{Y}^s\to\mathcal{Y}^s_1} \leqslant 2+\Lambda.$$

This and the invertibility of a, assumption (RK3), then yield

$$\begin{aligned}\|\Psi^h(U) - \Psi^h_m(U)\|_{\mathcal{Y}_1} &\leqslant (1+\sigma h)\|\mathbb{Q}_m U\|_{\mathcal{Y}_1} \\ &\quad + \|\mathsf{b}\|\|\mathsf{a}^{-1}\|\|h\mathsf{a}(\mathrm{id} - h\mathsf{a}A)^{-1}(B(W(U)) - \mathbb{P}_m B(W_m(U)))\|_{\mathcal{Y}_1^s} \\ &\leqslant (1+\sigma h)\|\mathbb{Q}_m U\|_{\mathcal{Y}_1} + (2+\Lambda)\|\mathsf{b}\|\|\mathsf{a}^{-1}\|\|B(W(U)) - \mathbb{P}_m B(W_m(U))\|_{\mathcal{Y}^s}.\end{aligned}$$

Inequality (4.4) is now a consequence of (2.5) and (4.9). □

4.3. Approximate embedding of semi-discretizations into a flow

We first review an approximate embedding result for Runge–Kutta discretizations of ODEs and then show how to extend it to PDEs. Consider the autonomous ODE

$$\dot{y} = f(y) \tag{4.10}$$

defined on the closed ball $\mathcal{B}_r^{\mathbb{C}^m}(y^0)$. We write ϕ^t to denote the flow of (4.10) and ψ^h denote the time-h map of an s-stage Runge–Kutta method of the form (3.1) applied to (4.10). When f is analytic, it is known that ψ^h can be expanded in a converging power series in h on a smaller ball, so that we can write

$$\psi^h(y) = y + \sum_{j=1}^{\infty} h^j g^j(y). \tag{4.11}$$

Specifically, as shown in [13, theorem IX.7.2] (with $2R$ there replaced by r here) via Cauchy estimates, (4.11) holds true for $(y,h) \in \mathcal{B}_{r/2}^{\mathbb{C}^m}(y^0) \times [0, r/(4\|\mathsf{a}\|M))$. Moreover, the numerical time-h map ψ^h can be embedded into the flow of a modified vector field up to an exponentially small error. A general form of this result was proved in [1] with specific proofs for the class of Runge–Kutta schemes we consider in [13, 29] (see also [17]). We state the result as follows.

THEOREM 4.7. *In the setting introduced above, there are positive constants η, c_1 and c_2 that depend only on the method such that, for every $r > 0$ and $M > 0$ such that*

$$\|f(y)\| \leqslant M \quad \text{for } y \in \mathcal{B}_r^{\mathbb{C}^m}(y^0)$$

and every $h \in [0, \eta r/M]$, there exists a modified differential equation $\dot{y} = \tilde{f}(y)$, defined on $\mathcal{B}_{r/4}^{\mathbb{C}^m}(y^0)$, whose flow $\tilde{\phi}^t$ satisfies $\tilde{\phi}^t(y^0) \in \mathcal{B}_{r/4}^{\mathbb{C}^m}(y^0)$ for at least $0 \leqslant t \leqslant h$ and which satisfies the estimate

$$\|\psi^h(y^0) - \tilde{\phi}^h(y^0)\| \leqslant h c_1 M \exp\left(-\frac{c_2 r}{hM}\right).$$

The proof will not be repeated in detail here. But we note for later reference that the modified vector field is constructed as a power series in h:

$$\tilde{f}^n(y;h) = f(y) + \sum_{j=p}^{n-1} h^j f^{j+1}(y), \tag{4.12}$$

where p is the order of the numerical method. Its exponential map is then expanded in powers of h and matched term by term with the expansion of the numerical time-h map (4.11). This yields a recursive expression for the coefficient vector fields,

$$f^j(y) = g^j(y) - \sum_{i=2}^{j} \frac{1}{i!} \sum_{k_1+\cdots+k_i=j} (\mathrm{D}_{k_1} \cdots \mathrm{D}_{k_{i-1}} f^{k_i})(y), \tag{4.13}$$

for $j \geqslant 2$, where $k_i \geqslant 1$ for all i (see [1, § 3.1]). We write $\mathrm{D}_i g(y) = \mathrm{D}g(y) f^i(y)$ as short-hand notation for the Lie derivative with respect to the ith coefficient vector field (as in [13, lemma IX.7.3]). The proof of theorem 4.7 proceeds by carefully estimating the growth of the f^j, noting that the optimal truncation is achieved when $n = n(h) = \lfloor c_2 r/(hM) \rfloor$. When referring to the optimally truncated vector field, we write $\tilde{f}_h$ or just $\tilde{f}$.

In addition, we need the following estimate, which guarantees consistency of the truncation. It is a slight generalization of results proved in [13, 29].

Lemma 4.8. *In the notation of theorem 4.7, for every $a \in \mathbb{N}$ there exists a constant $c_3 = c_3(a)$ such that, for every $h \in [0, \eta r/M]$,*

$$\|\tilde{f}^a(y) - \tilde{f}(y)\| \leqslant c_3 r^{-a} M^{a+1} h^a \quad \textit{for } y \in \mathcal{B}^{\mathbb{C}^m}_{r/4}(y^0).$$

Proof. It is known (see [1] for general numerical one-step methods and [13, 29] for the Runge–Kutta methods considered here) that there exist positive constants c_4 and $c_5 \leqslant 1/(c_2 \mathrm{e})$ that depend only on the method such that

$$\|f^j(y)\| \leqslant c_4 M \left(\frac{c_5 M j}{r} \right)^{j-1} \quad \text{for } y \in \mathcal{B}^{\mathbb{C}^m}_{r/4}(y^0). \tag{4.14}$$

Applying this estimate to (4.12) and using that $n \leqslant c_2 r/(hM)$ and therefore $h \leqslant c_2 r/(nM)$, we find that

$$\begin{aligned}
\|\tilde{f}^a(y) - \tilde{f}(y)\| &\leqslant h^a \sum_{j=a}^{n-1} h^{j-a} \|f^{j+1}(y)\| \\
&\leqslant h^a \sum_{j=a}^{n-1} h^{j-a} c_4 M \left(\frac{c_5 M (j+1)}{r} \right)^j \\
&\leqslant c_4 M h^a \sum_{j=a}^{n-1} \left(\frac{c_2 r}{nM} \right)^{j-a} \left(\frac{c_5 M (j+1)}{r} \right)^j \\
&\leqslant c_4 \mathrm{e}^{a+1} M \left(\frac{c_2 h M}{r} \right)^a \sum_{j=a}^{n-1} \left(\frac{j+1}{n} \right)^{j-a} \frac{(j+1)^a}{\mathrm{e}^{j+1}} \\
&\leqslant c_4 \mathrm{e}^{a+1} M \left(\frac{c_2 h M}{r} \right)^a a!,
\end{aligned} \tag{4.15}$$

where in the last inequality we have bounded the first factor inside the sum by 1 and noted that $j^a \mathrm{e}^{-j}$ is decreasing for $j \geqslant a$, so that

$$\sum_{j=a}^{n-1} \frac{(j+1)^a}{\mathrm{e}^{j+1}} \leqslant \int_a^n x^a \mathrm{e}^{-x}\,\mathrm{d}x \leqslant \int_0^\infty x^a \mathrm{e}^{-x}\,\mathrm{d}x = a!.$$

This completes the proof. □

Since $\tilde{f}^p = f$, we note, setting $a = p$, that lemma 4.8 provides a bound on $\|f(y) - \tilde{f}(y)\|$. Hence, by the triangle inequality, for every $h \leqslant \eta r/M$ there is a method-dependent constant $c_{\tilde{f}}$ such that

$$\|\tilde{f}(y)\| \leqslant M c_{\tilde{f}} \quad \text{for } y \in \mathcal{B}^{\mathbb{C}^m}_{r/4}(y^0). \tag{4.16}$$

We now apply theorem 4.7 to the sequence of truncated problems (4.3), where, for each m, we work on the space $\mathbb{P}_m\mathcal{Y}$ endowed with the $\mathcal{Y}_1$-norm. (See remark 4.11 for an explanation of why we work with the $\mathcal{Y}_1$- rather than the $\mathcal{Y}$-norm.)

Let $\mathcal{Y}_1^{\mathbb{C}} = \mathcal{Y}_1 + \mathrm{i}\mathcal{Y}_1$ be the complexification of $\mathcal{Y}_1$. We now choose $R_1 \geqslant 0$ such that $\mathcal{D}_1^\delta \subset \mathcal{B}^{\mathcal{Y}_1}_{R_1}(0)$. Then, by construction, f_m is analytic on $\mathcal{D}^{\mathbb{C}} \cap \mathbb{P}_m\mathcal{B}^{\mathcal{Y}_1^{\mathbb{C}}}_{R_1}(0)$ and satisfies the estimate

$$\begin{aligned}\|f_m(u_m)\|_{\mathcal{Y}_1^{\mathbb{C}}} &\leqslant \|Au_m\|_{\mathcal{Y}_1^{\mathbb{C}}} + \|B_m(u_m)\|_{\mathcal{Y}_1^{\mathbb{C}}} \\ &\leqslant mR_1 + m \sup_{U\in\mathcal{D}^{\mathbb{C}}} \|\mathbb{P}_m B(U)\|_{\mathcal{Y}^{\mathbb{C}}} \leqslant c_F m\end{aligned} \tag{4.17}$$

with $c_F = R_1 + M_0$.

Setting $M = c_F m$, theorem 4.7 asserts that the numerical time-h map can be embedded into a modified flow up to an error that is exponentially small in the step size h, albeit not uniformly in m. If, however, we make the stronger assumption that the initial datum lies in some Gevrey space $\mathcal{Y}_{\tau,L+1,q}$ with $\tau > 0$, lemma 4.6 asserts that the numerical solution of the full semilinear evolution equation (1.1) remains exponentially close in the spectral cut-off m to the numerical solution of the projected system. Thus, we can carefully choose $m = m(h)$ to balance the projection error and the embedding error to obtain an embedding result on the Gevrey space that is still exponential in h but at a lesser rate. This is done in the next lemma, where we also show that the result can be formulated not only on balls as in theorem 4.7, but also on more general m-independent subdomains of $\mathcal{Y}_1$ as needed in the proof of theorem 4.1.

We denote the coefficients of the power series expansion of ψ_m^h by g_m^j, the expansion coefficients of the modified vector field (defined via (4.13)) by f_m^j, and seek an optimally truncated modified vector field of the form

$$\tilde{f}_m^n(u_m; h) = f_m(u_m) + \sum_{j=p}^{n-1} h^j f_m^{j+1}(u_m). \tag{4.18}$$

Lemma 4.9 (embedding lemma for Gevrey class data). *Assume that the semilinear evolution equation (1.1) satisfies conditions (A) and (B0)–(B2). As before, let Ψ^h denote a single step of a Runge–Kutta method subject to (RK1)–(RK3) applied*

to the semilinear evolution equation (1.1). Then there exists $h_* > 0$ such that the choices

$$m(h) = \left(\frac{\chi}{\tau h}\right)^{q/(1+q)} \quad \textit{and} \quad n(h) = \left\lfloor \tau^{q/(1+q)} \left(\frac{\chi}{h}\right)^{1/(1+q)} \right\rfloor \tag{4.19 a}$$

with $\chi = c_2\delta/(4c_F)$ ensure that the modified vector field

$$\tilde{F}(U;h) \equiv \tilde{f}_{m(h)}^{n(h)}(\mathbb{P}_m U; h), \tag{4.19 b}$$

where $\tilde{f}_m^n$ is given by (4.18), has the following properties.

(a) *$\tilde{F}\colon \mathcal{D}_1^{\delta/2} \to \mathcal{Y}_1$ is analytic for every fixed $h \in [0, h_*]$ with bound*

$$\|\tilde{F}\|_{\mathcal{C}_{\mathrm{b}}(\mathcal{D}_1^{\delta/2}, \mathcal{Y}_1)} \leqslant c_{\tilde{F}} m(h) \tag{4.19 c}$$

for some $c_{\tilde{F}} > 0$ independent of h.

(b) *$\tilde{F}$ generates a modified flow $\tilde{\Phi}\colon \mathcal{D}_1 \times [0,h] \to \mathcal{D}_1^{\delta/4}$.*

(c) *There exists a constant $c_{\tilde{\Phi}}$ such that, for every $U \in \mathcal{D}_{\tau,L+1,q}$ and $h \in [0, h_*]$,*

$$\|\Psi^h(U) - \tilde{\Phi}^h(U)\|_{\mathcal{Y}_1} \leqslant c_{\tilde{\Phi}} \exp(-c_* h^{-1/(1+q)}) \tag{4.19 d}$$

with $c_* = \tau^{q/(q+1)}\chi^{1/(q+1)}$.

(d) *For each $a \in \mathbb{N}$ there is a constant $c_a \geqslant 0$ such that, with $\tilde{F}_m^a \equiv \tilde{f}_m^a \circ \mathbb{P}_m$ and $m = m(h)$, we have*

$$\|\tilde{F} - \tilde{F}_m^a\|_{\mathcal{C}_{\mathrm{b}}(\mathcal{D}_1^{\delta/2}, \mathcal{Y}_1)} \leqslant c_a h^a m^{a+1}. \tag{4.19 e}$$

Proof. Set $r = \frac{1}{4}\delta$. As $\mathcal{D}_1^\delta$ is a bounded subset of $\mathcal{Y}_1$, there exists m_* such that, due to (2.5), $\|\mathbb{Q}_m U\|_{\mathcal{Y}} \leqslant m^{-1}\|U\|_{\mathcal{Y}_1} \leqslant \frac{1}{2}\delta - r$ and therefore $\mathbb{P}_m U \in \mathcal{D}^{\delta - r}$ for every $U \in \mathcal{D}_1^{\delta/2}$ and $m \geqslant m_*$. In particular, for any such U and m, the ball

$$\mathcal{B}_r^{\mathbb{P}_m\mathcal{Y}_1^{\mathbb{C}}}(\mathbb{P}_m U) = \{u \in \mathbb{P}_m\mathcal{Y}^{\mathbb{C}}\colon \|u - \mathbb{P}_m U\|_{\mathcal{Y}_1^{\mathbb{C}}} \leqslant r\}$$

is contained in $\mathcal{D}^{\mathbb{C}} \cap \mathbb{P}_m\mathcal{B}_{R_1}^{\mathcal{Y}_1^{\mathbb{C}}}(0)$. Then estimate (4.17) holds true and we can apply theorem 4.7 with $M = c_F m$ and $y^0 = \mathbb{P}_m U$ on this ball. This theorem asserts that, for every $h \in [0, \eta r/(c_F m)]$, the modified vector field

$$\tilde{f}_m = \tilde{f}_m^{n(h)} \quad \text{with } n(h) = \lfloor c_2 r/(h c_F m)\rfloor$$

is defined on $\mathcal{B}_{r/4}^{\mathbb{P}_m\mathcal{Y}_1^{\mathbb{C}}}(\mathbb{P}_m U)$ and analytic as a map from $\mathcal{B}_{r/4}^{\mathbb{P}_m\mathcal{Y}_1^{\mathbb{C}}}(\mathbb{P}_m U)$ to $\mathcal{Y}_1^{\mathbb{C}}$. Its flow $\tilde{\phi}_m^t$ satisfies

$$\tilde{\phi}_m^t(\mathbb{P}_m U) \in \mathcal{B}_{r/4}^{\mathbb{P}_m\mathcal{Y}_1^{\mathbb{C}}}(\mathbb{P}_m U)$$

for at least $0 \leqslant t \leqslant h$, and

$$\|\psi_m^h(\mathbb{P}_m U) - \tilde{\phi}_m^h(\mathbb{P}_m U)\|_{\mathcal{Y}_1^{\mathbb{C}}} \leqslant h c_1 c_F m \exp\left(-\frac{c_2 r}{h c_F m}\right). \tag{4.20}$$

By construction, $\tilde{F}_m = \tilde{f}_m \circ \mathbb{P}_m$ is analytic as a map from $\mathcal{D}_1^{\delta/2}$ to $\mathcal{Y}_1$ with flow map $\tilde{\Phi}_m = \tilde{\phi}_m \circ \mathbb{P}_m + \mathbb{Q}_m$, which is analytic as a map from $\mathcal{D}_1$ to $\mathcal{D}_1^{\delta/4}$ for any choice of $m \geqslant m_*$ and $t \in [0, h]$. Estimates (4.19 *c*) and (4.19 *e*) then follow directly from (4.16) and lemma 4.8, respectively.

Our next step is to estimate the difference between the solution to the modified projected equation and the numerical solution of the full semilinear evolution equation (1.1). We split the error into the projection and the embedding error, the first of which is controlled by lemma 4.6 and the second by (4.20). By Lemma 4.6 there exist $h_* > 0$ and (a possibly increased choice of) $m_* > 0$ such that for $m \geqslant m_*$ and $h \in [0, h_*]$ the projection error estimate (4.4) holds. Increasing m_*, if necessary, to achieve that $m_* \geqslant \eta r/(c_F h_*)$, we ensure that both the embedding error estimate (4.20) and the truncation error estimate (4.4) hold true for every $m \geqslant m_*$ and $h \in [0, \eta r/(c_F m)]$.

Since $\mathcal{D}_{\tau,L+1,q} \subset \mathcal{D}_1$, by splitting the total error into a projection error component and the embedding error on the subspace $\mathbb{P}_m\mathcal{Y}$, we obtain that

$$\begin{aligned}\|\Psi^h(U) - \tilde{\Phi}_m^h(U)\|_{\mathcal{Y}_1} &\leqslant \|\Psi^h(U) - \psi_m^h(\mathbb{P}_m U)\|_{\mathcal{Y}_1} + \|\mathbb{Q}_m U\|_{\mathcal{Y}_1} \\ &\quad + \|\psi_m^h(\mathbb{P}_m U) - \tilde{\phi}_m^h(\mathbb{P}_m U)\|_{\mathcal{Y}_1} \\ &\leqslant (1 + c_\Psi) m^{-\ell} \exp(-\tau m^{1/q}) + h c_1 c_F m \exp\left(-\frac{c_2 r}{h c_F m}\right)\end{aligned}$$

for all $U \in \mathcal{D}_{\tau,L+1,q}$, $h \in [0, \eta r/(c_F m)]$ and $m \geqslant m_*$. The first and second errors decrease with m, whereas the third error increases with m. We now demand that the two exponents on the right coincide. Under the ansatz $m = \zeta h^{-\alpha}$ for some ζ and $\alpha \in (0, 1)$, we obtain

$$\begin{aligned}\|\Psi^h(U) - \tilde{\Phi}_m^h(U)\|_{\mathcal{Y}_1} &\leqslant (1 + c_\Psi) h^{\alpha\ell} \zeta^{-\ell} \exp(-\tau \zeta^{1/q} h^{-\alpha/q}) \\ &\quad + c_1 c_F \zeta h^{1-\alpha} \exp(-\chi \zeta^{-1} h^{\alpha-1})\end{aligned}$$

with $\chi = c_2 r/c_F = c_2 \delta/(4 c_F)$. Then the exponents coincide, provided $\tau \zeta^{1/q} h^{-\alpha/q} = \chi \zeta^{-1} h^{\alpha-1}$, i.e. when

$$\alpha = \frac{q}{1+q} \quad \text{and} \quad \zeta = \left(\frac{\chi}{\tau}\right)^{\alpha}. \tag{4.21}$$

This implies that $m(h)$ is given by (4.19 *a*), and

$$\|\Psi^h(U) - \tilde{\Phi}_{m(h)}^h(U)\|_{\mathcal{Y}_1} \leqslant \tilde{c} h^{\nu} \exp(-c_* h^{-1/(1+q)}) \tag{4.22}$$

with $c_* = \tau \zeta^{1/q} = \tau^{q/(1+q)} \chi^{1/(q+1)}$ and $\nu = \min\{1, q\ell\}/(1+q)$, and where we possibly need to shrink $h_* > 0$ to satisfy $m(h_*) \geqslant m_*$ and $h_* \leqslant \eta r/(c_F m(h_*)) = (\eta/c_2)\chi^{1-\alpha}(\tau h_*)^{\alpha}$. Solving the latter inequality for h_* leads to the restriction that

$$h_* \leqslant \chi \tau^q \left(\frac{\eta}{c_2}\right)^{q+1}.$$

A similar computation yields the form of $n(h)$ stated in (4.19 *a*). The exponential estimate (4.19 *d*) is then obtained by defining the modified vector field by (4.19 *b*) with corresponding modified flow $\tilde{\Phi}^t(U) \equiv \tilde{\Phi}_{m(h)}^t(U)$ and setting $c_\Phi = \tilde{c} h_*^{\nu}$. □

REMARK 4.10. The formal expansions of both the numerical method and the modified vector field contain powers of the unbounded operator A. Therefore, the modified vector field cannot be written as a semilinear Hamiltonian evolution equation of the form (2.12). If we were simply interested in constructing the modified vector field, we could avoid using spatial Galerkin truncation by setting up different spaces for domain and range such that the modified vector field, computed up to a given order, is continuous. In fact, we need such techniques in § 4.4 to obtain the order estimate (4.1 *a*) for the modified Hamiltonian, which is yet to be constructed. In such a setting, however, we do not have a theory of local existence of solutions for the modified differential equation, so we cannot obtain an approximate embedding of the numerical method into a flow.

REMARK 4.11. The reason for constructing the modified vector field on a subspace of $\mathcal{Y}_1$ rather than $\mathcal{Y}$ is that on general domains we can only maintain a valid domain of definition of the nonlinearity $B(U)$ under Galerkin truncation uniformly in $m \geqslant m_*$, and, in particular, assert estimate (4.17) by dropping down at least one rung on the scale of spaces. Similarly, we require data in $\mathcal{D}_{\tau,L+1,q}$ rather than $\mathcal{D}_{\tau,L,q}$ for the exponential estimates (4.19 *d*) and (4.1 *c*) because we want to define Ψ^h and Ψ^h_m uniformly in h and m on general open sets of Gevrey spaces, not just on open balls. This can only be done when constructing them as maps from $\mathcal{D}_{\tau,L+1,q}$ to $\mathcal{D}^\delta_{\tau,L,q}$ (see [24, 25]).

Next, we show that in the Hamiltonian case the above construction also yields a modified Hamiltonian, which is approximately conserved under the numerical time-h map of the full semilinear evolution equation (2.12).

LEMMA 4.12 (modified Hamiltonian for Gevrey class data). *Under the conditions and in the notation of lemma 4.9, suppose further that (H0)–(H4) and (RK4) hold true. Then, for sufficiently small $h_* > 0$, there exists a modified Hamiltonian $\tilde{H}\colon \mathcal{D}_1^{\delta/2} \times [0, h_*] \to \mathbb{R}$, defined up to a constant of integration, which is analytic in $U \in \mathcal{D}_1^{\delta/2}$ for every $h \in [0, h_*]$ and such that the modified vector field from lemma 4.9 satisfies $\tilde{F} = \mathbb{J}\nabla\tilde{H}$. Moreover, there exist constants $c_* \in (0, \tau^{q/(q+1)}\chi^{1/(q+1)})$, with χ as in lemma 4.9 and $c_{\tilde{H}} > 0$ such that, for every $U \in \mathcal{D}_{\tau,L+1,q}$ and $h \in [0, h_*]$,*

$$|\tilde{H}(\Psi^h(U), h) - \tilde{H}(U, h)| \leqslant c_{\tilde{H}} \exp(-c_* h^{-1/(1+q)}). \tag{4.23}$$

Proof. By assumption (H1), the operator $\mathbb{J}^{-1}A$ is self-adjoint on $\mathcal{Y}$. Since, by lemma 2.5, $\mathbb{J}^{-1}$ and $\mathbb{P}_m$ commute, $\mathbb{J}^{-1}A$ is also self-adjoint on $\mathbb{P}_m\mathcal{Y}$ with respect to the restriction of the $\mathcal{Y}$-inner product to $\mathbb{P}_m\mathcal{Y}$. Hence, the linear part $\dot{u}_m = A_m u_m$ of (4.3) is Hamiltonian on $\mathbb{P}_m\mathcal{Y}$. Moreover, by (H2), the operator $\mathbb{J}^{-1}\mathrm{D}B(U)$ is self-adjoint for each $U \in \mathcal{D}^\delta$. Hence, $\mathbb{J}^{-1}\mathrm{D}B_m(u_m)$ is self-adjoint for each $u_m \in \mathcal{D}^\delta \cap \mathbb{P}_m\mathcal{Y}$ so that, altogether, the vector field f_m from (4.3) is Hamiltonian as a map from $\mathcal{D}^\delta \cap \mathbb{P}_m\mathcal{Y}$ to $\mathbb{P}_m\mathcal{Y}$ with respect to the restriction of the $\mathcal{Y}$-inner product to $\mathbb{P}_m\mathcal{Y}$. Since on each Galerkin subspace $\mathbb{P}_m\mathcal{Y}$ the numerical method ψ^h_m is symplectic, the Taylor coefficients f^j_m of the modified vector field f^j_m are also Hamiltonian (see, for example, [1, 13, 17]). Moreover, the operator $\mathbb{J}^{-1}\mathrm{D}f^j_m(u_m)$ is self-adjoint with respect to the restriction of the $\mathcal{Y}$-inner product to $\mathbb{P}_m\mathcal{Y}$ for each $u_m \in \mathcal{D}^\delta \cap \mathbb{P}_m\mathcal{Y}$ and the same holds true for $\mathbb{J}^{-1}\mathrm{D}\tilde{f}^{n(h)}_{m(h)}(u_m)$.

As we argued in the proof of lemma 4.9, $\mathbb{P}_m U \in \mathcal{D}^\delta$ for $U \in \mathcal{D}_1^{\delta/2}$, so $\mathbb{J}^{-1}\mathrm{D}\tilde{F}(U)$ is self-adjoint with respect to the $\mathcal{Y}$-inner product for each $U \in \mathcal{D}_1^{\delta/2}$. By assumption (H4), the set $\mathcal{D}_1^{\delta/2}$ is simply connected and star shaped.

Therefore, we can proceed as in the proof of lemma 2.6. We fix $U^0 \in \mathcal{D}_1$ such that $\mathcal{D}_1^{\delta/2}$ is star shaped with respect to U^0 and define

$$\tilde{H}(U) = \int_0^1 \langle \mathbb{J}^{-1}\tilde{F}(tU + (1-t)U^0), U - U^0\rangle \,\mathrm{d}t.$$

This modified Hamiltonian $\tilde{H}$ is well defined and analytic on $\mathcal{D}_1^{\delta/2}$. Moreover, the steps taken in (2.16) still apply, so $\tilde{H}$ is invariant under the modified flow with

$$\tilde{F} = \mathbb{J}\nabla\tilde{H}.$$

To prove (4.23), we decrease h_* such that the right-hand side of (4.19 *d*) is smaller than $\delta/4$ for every $U \in \mathcal{D}_{\tau,L+1,q}$. This is to ensure that $\Psi^h(U) \in \mathcal{D}_1^{\delta/2}$, so that $\tilde{H}$ is defined at $\Psi^h(U)$. Then, using the mean-value theorem, the bound on the modified vector field given by (4.19 *b*), (4.19 *d*) and the invertibility of $\mathbb{J}$, we estimate, for $h \in [0, h_*]$ and $U \in \mathcal{D}_{\tau,L+1,q}$, that

$$\begin{aligned}
|\tilde{H}(\Psi^h(U)) - \tilde{H}(\tilde{\Phi}^h(U))| &\leqslant \|\nabla\tilde{H}\|_{\mathcal{C}_\mathrm{b}(\mathcal{D}_1^{\delta/2};\mathcal{Y}_1)} \|\Psi^h(U) - \tilde{\Phi}^h(U)\|_{\mathcal{Y}_1} \\
&\leqslant \|\mathbb{J}^{-1}\tilde{F}\|_{\mathcal{C}_\mathrm{b}(\mathcal{D}_1^{\delta/2};\mathcal{Y}_1)} \|\Psi^h(U) - \tilde{\Phi}^h(U)\|_{\mathcal{Y}_1} \\
&\leqslant \|\mathbb{J}^{-1}\|_{\mathcal{E}(\mathcal{Y}_1)} \|\tilde{F}\|_{\mathcal{C}_\mathrm{b}(\mathcal{D}_1^{\delta/2};\mathcal{Y}_1)} c_{\tilde{\Phi}} \exp(-c_* h^{-1/(1+q)}) \\
&\leqslant \|\mathbb{J}^{-1}\|_{\mathcal{E}(\mathcal{Y}_1)} c_{\tilde{F}} m(h) c_{\tilde{\Phi}} \exp(-c_* h^{-1/(1+q)}).
\end{aligned}$$

Since $\tilde{H}$ is conserved under the modified flow, choosing m as in (4.19 *a*), we obtain

$$|\tilde{H}(\Psi^h(U)) - \tilde{H}(U)| \leqslant \tilde{c} h^{-q/(1+q)} \mathrm{e}^{-c_* h^{-1/(1+q)}}.$$

Dominating the algebraic prefactor by fractional exponential decay, this inequality implies (4.23) with a possibly smaller value for $c_* > 0$ than in lemma 4.9. □

What is still missing is the proof of the $O(h^p)$-closeness of the modified Hamiltonian to the original one. In a first attempt to prove such a result, we write

$$|H(U) - \tilde{H}(U)| \leqslant |H(U) - H(\mathbb{P}_m U)| + |H(\mathbb{P}_m U) - \tilde{H}(U)|, \tag{4.24}$$

where $m = m(h)$ is as in (4.19 *a*). Under the assumptions of lemma 4.12, the first term on the right is exponentially small for $U \in \mathcal{D}_{\tau,L+1,q}$.

To estimate the second term of (4.24), choose some fixed $U^0 \in \mathcal{D}_1$ such that $\mathcal{D}_1^{\delta/2}$ is star shaped with respect to U^0, and set $H(U^0) = \tilde{H}(U^0)$. The naive choice is then to employ (4.19 *e*) with $a = p$ so that $\tilde{F}_m^a = \tilde{f}_m^a \circ \mathbb{P}_m = F \circ \mathbb{P}_m$. Integrating $\tilde{F} - F \circ \mathbb{P}_m$, we obtain the estimate

$$|H(u_m) - \tilde{H}_m(u_m)| \leqslant O(m^{p+1} h^p) = O(h^{(p-q)/(q+1)}) \tag{4.25}$$

for every $u_m = \mathbb{P}_m U$ and $U \in \mathcal{D}_1^{\delta/2}$. This estimate is weaker than the expected $O(h^p)$.

A closer inspection reveals that, in the context of the semilinear evolution equation (2.12), the second inequality of (4.15) in the proof of lemma 4.8 is too weak:

when estimating the Taylor coefficients f_m^j of the modified vector field in some fixed Hilbert space norm, the unboundedness of the operators A contained therein will introduce a factor m^j (see (4.14)). This propagates into the proof of (4.25). Note, however, that these estimates are simply about consistency, not about constructing a flow. Thus, we can afford to lose smoothness rather than order. In other words, we can estimate the Taylor coefficients of the modified vector field as maps from one Hilbert space into another with a weaker norm. This will be detailed in the next section.

4.4. Modified vector fields on Hilbert spaces

In this section, we present a more subtle estimate on the difference between the original Hamiltonian and the modified Hamiltonian. The main difference to the derivation of (4.25) in the previous section is that we consider the expansion coefficients of the numerical method and of the modified vector field as maps between different rungs on our scale of Hilbert spaces such that the loss of smoothness is carefully accounted for.

We begin by establishing the necessary functional setting for the analysis of modified vector fields on Hilbert spaces. We then review a result from [25] on the Galerkin projection error for the numerical time-h maps. This estimate is then propagated into an estimate on the difference between the full and the modified vector field, which finally implies a corresponding estimate on the difference between the exact Hamiltonian and the modified Hamiltonian.

In this section, we work directly with the standard construction of the modified vector field. Namely, for $\ell = 1, \dots, K+1$, we write

$$G^\ell = \frac{\partial_h^\ell \Psi^h}{\ell!}\bigg|_{h=0} \tag{4.26}$$

to denote the ℓth coefficient of the expansion of Ψ^h in powers of h, and define the analogues of the expansion coefficients (4.13) for the modified vector field on Hilbert spaces as follows: we set $F^1 \equiv G^1$ and define

$$F^\ell = G^\ell - \sum_{i=2}^{\ell} \frac{1}{i!} \sum_{\ell_1+\cdots+\ell_i=\ell} \mathrm{D}_{\ell_1} \cdots \mathrm{D}_{\ell_{i-1}} F^{\ell_i} \tag{4.27}$$

for $\ell = 2, \dots, K+1$, where the sum ranges over indices $\ell_i \geqslant 1$ for all i, where $\mathrm{D}_j G = \mathrm{D}GF^j$. We also recall from § 4.3 that g_m^ℓ and f_m^ℓ denote the ℓth coefficients of the expansions of the projected numerical method and the corresponding modified vector field, respectively, and set $G_m^\ell \equiv g_m^\ell \circ \mathbb{P}_m$ and $F_m^\ell \equiv f_m^\ell \circ \mathbb{P}_m$.

In the notation of condition (B1), we set $\mathcal{U}_\kappa = \mathcal{D}_\kappa$ for $\kappa = 1, \dots, K+1$. Then the regularity results given by theorem 3.1 on Ψ^h and theorem 4.5 on Ψ_m^h imply that, in particular, there exists m_* such that, for all $m \geqslant m_*$ and $\ell = 1, \dots, K+1$,

$$G^\ell, G_m^\ell \in \bigcap_{\substack{j+k \leqslant N \\ \ell \leqslant k \leqslant K+1}} \mathcal{C}_{\mathrm{b}}^j(\mathcal{U}_k; \mathcal{Y}_{k-\ell}). \tag{4.28}$$

Moreover, bounds in the norms associated with (4.28) are uniform in $m \geqslant m_*$.

In the following, we shall state such bounds on vector fields in terms of the three-parameter family of norms

$$|g|_{N,K,S} = \max_{\substack{j+k\leqslant N\\ S\leqslant k\leqslant K}} \|\mathrm{D}^j g\|_{\mathcal{C}(\mathcal{U}_k;\mathcal{E}^j(\mathcal{Y}_k,\mathcal{Y}_{k-S}))} \tag{4.29}$$

for $1 \leqslant S \leqslant K \leqslant N$. The parameter S plays the role of a loss-of-smoothness index as it forces the image of the map be estimated at least S rungs down the scale. We can then prove a simple result on the regularity of the expansion coefficients of the modified vector field.

LEMMA 4.13 (modified vector field on a scale of Hilbert spaces). *Assume that G^1, $\ldots, G^{K+1}$ are of class (4.28). Then the vector fields $F^1, \ldots, F^{K+1}$ defined by (4.27) are also of class (4.28).*

Proof. The proof is based on the simple fact that $F \in \mathcal{C}^n_{\mathrm{b}}(\mathcal{Y}_{i+j}, \mathcal{Y}_j)$ and $G \in \mathcal{C}^{n+1}_{\mathrm{b}}(\mathcal{Y}_j, \mathcal{Y})$ imply that $\mathrm{D}GF \in \mathcal{C}^n_{\mathrm{b}}(\mathcal{Y}_{i+j}, \mathcal{Y})$. Thus, it remains to observe that their repeated application to the terms in the inner sum of (4.27) causes all loss indices to always sum to ℓ. We proceed by induction in ℓ. The case $\ell = 1$ does not require proof. Assume therefore that $\ell > 1$ and the lemma is proved up to index $\ell - 1$. For $\ell > \ell_1 \geqslant 1$, we estimate, with $G \equiv \mathrm{D}_{\ell_2} \cdots \mathrm{D}_{\ell_{i-1}} F^{\ell_i}$ that

$$\begin{aligned}
|\mathrm{D}_{\ell_1} G|_{N,K+1,\ell} &= |\mathrm{D}GF^{\ell_1}|_{N,K+1,\ell} \\
&= \max_{\substack{j+k\leqslant N\\ \ell\leqslant k\leqslant K+1}} \|\mathrm{D}^j(\mathrm{D}GF^{\ell_1})\|_{\mathcal{C}(\mathcal{U}_k;\mathcal{E}^j(\mathcal{Y}_k,\mathcal{Y}_{k-\ell}))} \\
&\leqslant 2^N \max_{\substack{j+k\leqslant N\\ \ell\leqslant k\leqslant K+1}} \|\mathrm{D}^{j+1} G\|_{\mathcal{C}(\mathcal{U}_{k-\ell_1};\mathcal{E}^{j+1}(\mathcal{Y}_{k-\ell_1},\mathcal{Y}_{k-\ell}))} \\
&\qquad \times \max_{\substack{j+k\leqslant N\\ \ell_1\leqslant k\leqslant K+1}} \|\mathrm{D}^j F^{\ell_1}\|_{\mathcal{C}(\mathcal{U}_k;\mathcal{E}^j(\mathcal{Y}_k,\mathcal{Y}_{k-\ell_1}))} \\
&= 2^N \max_{\substack{j+k\leqslant N+1-\ell_1\\ \ell-\ell_1\leqslant k\leqslant K+1-\ell_1}} \|\mathrm{D}^j G\|_{\mathcal{C}(\mathcal{U}_k;\mathcal{E}^j(\mathcal{Y}_k,\mathcal{Y}_{k-(\ell-\ell_1)}))} |F^{\ell_1}|_{N,K+1,\ell_1} \\
&\leqslant 2^N |G|_{N,K+1-\ell_1,\ell-\ell_1} |F^{\ell_1}|_{N,K+1,\ell_1},
\end{aligned} \tag{4.30}$$

provided that G is of class (4.28) with K replaced by $K-\ell_1$ and ℓ replaced by $\ell-\ell_1$. Here the first inequality is based on the product rule and selective weakening of the norm on the domain spaces, thereby increasing the respective operator norms. The identity between the fourth and the fifth lines is achieved by redefining $j+1$ as j and $k-\ell_1$ as k. The final inequality holds because $\ell_1 \geqslant 1$, so that we are strictly extending the range of the running indices. We note that the second term in the final line of (4.30) is bounded by the induction hypothesis. The estimation of the first term in the final line of (4.30) can now be made recursively to resolve the entire product $\mathrm{D}_{\ell_1} \cdots \mathrm{D}_{\ell_{i-1}} F^{\ell_i}$ from the inner sum of (4.27) in terms of quantities that are bounded by the induction hypothesis. This is always possible because at the kth step of this process we lose ℓ_k rungs of smoothness, and the sum of the loss indices satisfies $\ell_1 + \cdots + \ell_i = \ell$ by construction. □

We now aim to derive estimates on the difference between F^ℓ and F^ℓ_m with respect to the same type of norm. In [25], we obtained a related result on the difference

between Ψ^h and Ψ^h_m from which we can start. Setting $\mathcal{I} = (0, h_*)$, $\mathcal{U} = \mathcal{D}_{K+1}$ and $\mathcal{X} = \mathcal{Y}_{K+1}$, we define the norm

$$\|\Psi\|_{N,K} = \max_{\substack{j+k\leqslant N\\ \ell\leqslant k\leqslant K}} \|\mathrm{D}_U^j \partial_h^\ell \Psi\|_{\mathcal{L}_\infty(\mathcal{U}\times\mathcal{I};\mathcal{E}^j(\mathcal{X};\mathcal{Y}_{k-\ell}))} \tag{4.31}$$

for $0 \leqslant K \leqslant N$. Then the stability of the numerical method on a scale of Hilbert spaces under spectral truncation can be formulated as follows.

LEMMA 4.14 (Oliver and Wulff [25, theorem 3.7]). *Assume (A), (B1) and (RK1)–(RK3). Then there is $h_* > 0$ such that, for every $0 \leqslant S \leqslant K+1$,*

$$\|\Psi^h - \Psi^h_m\|_{N-1-S,K+1-S} = O(m^{-S}) \tag{4.32}$$

as $m \to \infty$. The order constants depend only on the bounds afforded by (B1), (2.10), (3.5), on the coefficients of the method and on δ.

We note that lemma 4.6 above already provided us with an exponential estimate on $\Psi^h - \Psi^h_m$ for Gevrey-regular data, whereas lemma 4.14 here also asserts bounds on derivatives with respect to h and U; the proof is correspondingly more complicated even in spaces of finite order of smoothness and can be found in [25].

To proceed, we observe that lemma 4.14 holds with $K+1$ in the statement of the lemma replaced by any κ between $S+1$ and $K+1$, with K as defined in condition (B1). Note that in the definition of the norm (4.31) we must correspondingly read $\mathcal{X}$ and $\mathcal{U}$ as $\mathcal{Y}_\kappa$ and $\mathcal{U}_\kappa$, respectively. Then, specializing to the particular value $k = \kappa - S$ in the definition of the norm appearing in (4.32), we obtain

$$\max_{\substack{j+\kappa\leqslant N-1\\ S+\ell\leqslant\kappa}} \|\mathrm{D}_U^j \partial_h^\ell (\Psi^h - \Psi^h_m)\|_{\mathcal{L}_\infty(\mathcal{U}_\kappa\times\mathcal{I};\mathcal{E}^j(\mathcal{Y}_\kappa;\mathcal{Y}_{\kappa-S-\ell}))} = O(m^{-S}).$$

Thus, fixing $\ell \in 1, \ldots, K+1-S$ and taking the maximum over the allowed range $\kappa = S+\ell, \ldots, K+1$, we can write

$$\max_{\substack{j+\kappa\leqslant N-1\\ S+\ell\leqslant\kappa\leqslant K+1}} \|\mathrm{D}_U^j \partial_h^\ell (\Psi^h - \Psi^h_m)\|_{\mathcal{L}_\infty(\mathcal{U}_\kappa\times\mathcal{I};\mathcal{E}^j(\mathcal{Y}_\kappa;\mathcal{Y}_{\kappa-S-\ell}))} = O(m^{-S}). \tag{4.33}$$

Due to the definition of G^ℓ in (4.26), this directly implies that

$$|G^\ell - G^\ell_m|_{N-1,K+1,S+\ell} = O(m^{-S}). \tag{4.34}$$

LEMMA 4.15 (stability of the modified vector field). *Suppose that $G^1, \ldots, G^{K+1}$ and $\bar{G}^1, \ldots, \bar{G}^{K+1}$ are of class (4.28), and let F^ℓ and $\bar{F}^\ell$ denote expansion coefficients of the respective associated modified vector fields defined via (4.27). Then, for $S \in 1, \ldots, K$ and every $\ell \in 1, \ldots, K+1-S$,*

$$|F^\ell - \bar{F}^\ell|_{N-1,K+1,S+\ell} \leqslant c \max_{1\leqslant k\leqslant \ell} |G^k - \bar{G}^k|_{N-1,K+1,S+k},$$

where the constant c depends on the bounds on G^k and $\bar{G}^k$ in the norm $|\cdot|_{N-1,K+1,k}$ for $k = 1, \ldots, \ell$.

Proof. The proof follows the same steps as that of lemma 4.13. We set $\bar{\mathrm{D}}_\ell G \equiv \mathrm{D}G\bar{F}_\ell$. The crucial estimate corresponding to (4.30) then takes the form

$$\begin{aligned}|\mathrm{D}_{\ell_1}G - \bar{\mathrm{D}}_{\ell_1}\bar{G}|_{N-1,K+1,S+\ell} &\leqslant |\mathrm{D}G(F^{\ell_1} - \bar{F}^{\ell_1})|_{N-1,K+1,S+\ell} \\ &\quad + |\mathrm{D}(G-\bar{G})\bar{F}^{\ell_1}|_{N-1,K+1,S+\ell} \\ &\leqslant 2^N |G|_{N-1,K+1-S-\ell_1,\ell-\ell_1}|F^{\ell_1} - \bar{F}^{\ell_1}|_{N-1,K+1,S+\ell_1} \\ &\quad + 2^N |G-\bar{G}|_{N-1,K+1-\ell_1,S+\ell-\ell_1}|\bar{F}^{\ell_1}|_{N-1,K+1,\ell_1},\end{aligned}$$

where the estimates in the last inequality follow from (4.30). This reasoning can again be applied iteratively to resolve the entire difference of products of the form $\mathrm{D}_{\ell_1}\cdots\mathrm{D}_{\ell_{i-1}}F^{\ell_i}$, where we note that the loss indices now add up to exactly $S+\ell$, as required. □

We now turn our attention to the optimally truncated modified vector field $\tilde{F} \equiv \tilde{f}_m \circ \mathbb{P}_m$, where $m = m(h)$ is as in (4.19 *a*). This is the same modified vector field that gives rise to the modified Hamiltonian in lemma 4.12. Then the difference between $\tilde{F}$ and F, the exact vector field of the semilinear evolution equation (1.1), can be estimated as follows.

Lemma 4.16. *Suppose conditions (A), (B0)–(B2) and (RK1)–(RK3) are satisfied with $K+1 \geqslant P$, where*

$$P = \left\lceil \frac{p(q+1)^2}{q} + q \right\rceil + 1, \tag{4.35}$$

q is defined in (B2) and p is the order of the numerical method. Then the difference between the original vector field F and the modified vector field $\tilde{F}$ from lemma 4.9 satisfies

$$\|\tilde{F} - F\|_{\mathcal{C}_\mathrm{b}(\mathcal{D}_P;\mathcal{Y})} = O(h^p). \tag{4.36}$$

Proof. We define two intermediate vector fields. Let $\tilde{F}^a$ denote the modified vector field of the numerical method applied to the original semilinear evolution equation (1.1) computed up to order $a \leqslant K+1$. Formally, as in the finite-dimensional case, it has an expansion of the form (4.12):

$$\tilde{F}^a = F + \sum_{j=p}^{a-1} h^j F^{j+1}. \tag{4.37}$$

Note that $F = \tilde{F}^p$. Due to (4.28) and lemma 4.13, all coefficients F^k in the expansion (4.37), and consequently $\tilde{F}^a$, are also of class (4.28) with $\ell = a$.

We now choose h_* and m_* as in the proof of lemma 4.9 and suppose $m \geqslant m_*$, so that the projected modified vector fields are well defined. Let $\tilde{F}^a_m \equiv \tilde{f}^a_m \circ \mathbb{P}_m$ denote the corresponding modified vector field of the projected system (4.3), again up to order a, with $m = O(h^{-q/(q+1)})$ as in (4.19 *a*). By lemma 4.13 applied to G^ℓ_m, the vector field $\tilde{F}^a_m$ is also of class (4.28).

We now decompose

$$F - \tilde{F} = (F - \tilde{F}^a) + (\tilde{F}^a - \tilde{F}^a_m) + (\tilde{F}^a_m - \tilde{F}). \tag{4.38}$$

We show that when a is chosen as

$$a = \lceil p(q+1) + q \rceil, \tag{4.39}$$

each term on the right is $O(h^p)$ in appropriate norms.

A bound on the first difference on the right-hand side of (4.38) follows directly from the definition of $\tilde{F}^a$ in (4.37). Using the norms defined in (4.29),

$$|F - \tilde{F}^a|_{N-1,K+1,a} \leqslant h^p \sum_{j=p}^{a-1} h^{j-p} |F^{j+1}|_{N-1,K+1,a}.$$

By lemma 4.13, all norms in the right-hand sum are finite, so that, for $a \in 1, \ldots, K+1$,

$$|F - \tilde{F}^a|_{N-1,K+1,a} = O(h^p). \tag{4.40}$$

To estimate the second difference on the right-hand side of (4.38), we apply lemma 4.15 with $\bar{G}^\ell = G^\ell_m$, which we recall are also of class (4.28), so that (4.34) applies, yielding

$$|\tilde{F}^a - \tilde{F}^a_m|_{N-1,K+1,a+S} = O(m^{-S}) \tag{4.41}$$

for $S \in 1, \ldots, K+1-a$.

A bound on the third difference on the right-hand side (4.38) is provided by (4.19 *e*), namely

$$\|\tilde{F}^a_{m(h)} - \tilde{F}\|_{\mathcal{C}(\mathcal{D}_1^{\delta/2};\mathcal{Y}_1)} = O(h^a m(h)^{a+1}). \tag{4.42}$$

We now seek conditions under which the estimates (4.41) and (4.42) are of order h^p. Due to (4.19 *a*), $m = m(h) = O(h^{-q/(1+q)})$, so that the requirement $m^{-S} = O(h^p)$ leads to the choice

$$S = \left\lceil p \frac{1+q}{q} \right\rceil.$$

Similarly, the requirement $O(h^a m^{a+1}) = O(h^p)$ is equivalent to

$$O(h^a m^{a+1}) = O(h^{a-q(a+1)/(q+1)}) = O(h^{a/(q+1)-q/(q+1)}) = O(h^p),$$

which leads to the choice (4.39).

Since P is defined such that $P \geqslant S + a$ (see (4.35)), the above estimates for the three terms of the decomposition (4.38) imply (4.36). □

Remark 4.17. If we can change $\tilde{F}$ so that its leading order linear part is A rather than $A\mathbb{P}_{m(h)}$, then theorem 4.1 still applies. This is true since, by (2.5), for $U^0 \in \mathcal{Y}_{\tau,L+1,q}$ the differences between the two modified vector fields and their flows up to time h are exponentially small in the $\mathcal{Y}_1$-norm.

In the Hamiltonian case, the previous result on $O(h^p)$-closeness of true and modified vector field carries over to a statement on $O(h^p)$-closeness of the corresponding Hamiltonians.

Lemma 4.18. *Under the assumptions of lemma 4.16 suppose that, in addition, the semilinear evolution equation is Hamiltonian satisfying (H0)–(H4) and that the*

numerical method is symplectic, i.e. satisfies (RK4), and is of order p. Then the modified Hamiltonian $\tilde{H}$ from lemma 4.12 can be chosen such that

$$\|H - \tilde{H}\|_{\mathcal{C}_\mathrm{b}(\mathcal{D}_P;\mathbb{R})} = O(h^p).$$

Proof. By (H4), $\mathcal{D}_P$ is star shaped with respect to some $U^0 \in \mathcal{D}_P$. Since the modified Hamiltonian from lemma 4.12 is defined only up to a constant, we can choose the constant of integration such that $H(U_0) = \tilde{H}(U_0)$. Then, following the proof of lemma 2.6, we estimate, for any $U \in \mathcal{D}_P$,

$$\begin{aligned}|\tilde{H}(U) - H(U)| &\leqslant \left| \int_0^1 \langle \mathbb{J}^{-1}(F - \tilde{F})(tU + (1-t)U_0), U - U_0 \rangle \,\mathrm{d}t \right| \\ &\leqslant \|\mathbb{J}^{-1}\|_{\mathcal{E}(\mathcal{Y})} \|F - \tilde{F}\|_{\mathcal{C}_\mathrm{b}(\mathcal{D}_P;\mathcal{Y})} \sup_{U \in \mathcal{D}_P} \|U - U^0\|_{\mathcal{Y}}.\end{aligned}$$

Since $\mathcal{D}_P$ is $\mathcal{Y}$-bounded, lemma 4.16 implies that the right-hand side is $O(h^p)$. □

5. Lower estimates in an example: a nonlinear Schrödinger equation

In this section we set up a counter-example, motivated by [21], which shows that analyticity of the initial data is necessary to achieve an embedding of the implicit midpoint rule into a Hamiltonian flow.

5.1. Model evolution equation

We work with functions $u\colon [0,\infty) \to \ell_2(\mathbb{N}_0;\mathbb{C})$, whose components may be interpreted as the Fourier coefficients of a square-integrable function defined on the circle. Further, we write $u = v + \mathrm{i}w$ to identify the real and imaginary parts of the components of u.

We now define the Hamiltonian

$$H = \frac{1}{2}\sum_{k=1}^{\infty} \omega_k |u_k|^2 + w_0 \operatorname{Re} f(u)$$

with $\omega_k \geqslant 0$ and

$$f(u) = f(u_1, u_2, \dots) = \sum_{\alpha} c_\alpha u^\alpha,$$

where $\alpha = (\alpha_1, \alpha_2, \dots) \in \mathbb{N}_0^{\mathbb{N}}$ is a multi-index, each c_α is a real coefficient, $u^\alpha = u_1^{\alpha_1} u_2^{\alpha_2} \cdots$ as usual, and the summation is over all multi-indices with $\alpha_0 = 0$ and a finite number of non-zero coefficients α_k for $k \geqslant 1$.

In terms of v and w, this defines a Hamiltonian system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} v \\ w \end{pmatrix} = \mathbb{J}\nabla H(v,w), \tag{5.1}$$

where $\mathbb{J}$ is the standard symplectic structure matrix

$$\mathbb{J} = \begin{pmatrix} 0 & \mathrm{id} \\ -\mathrm{id} & 0 \end{pmatrix}.$$

For $k = 0$, we obtain by direct computation that

$$\dot{v}_0 = \frac{\partial H}{\partial w_0} = \operatorname{Re} f(u), \tag{5.2 a}$$

$$\dot{w}_0 = -\frac{\partial H}{\partial v_0} = 0. \tag{5.2 b}$$

For $k \geqslant 1$,

$$\dot{v}_k = \frac{\partial H}{\partial w_k} = \omega_k w_k + w_0 \operatorname{Re} \frac{\partial f(u)}{\partial w_k}, \tag{5.2 c}$$

$$\dot{w}_k = \frac{\partial H}{\partial v_k} = -\omega_k v_k - w_0 \operatorname{Re} \frac{\partial f(u)}{\partial v_k}. \tag{5.2 d}$$

In all of the following, we assume the initial condition $w_0(0) = 0$ so that, due to (5.2 *b*), $w_0(t) = 0$ for all $t \geqslant 0$. Then (5.2 *c*) and (5.2 *d*) combine into

$$\dot{u}_k = \mathrm{i}\omega_k u_k$$

for $k \geqslant 1$, which is immediately solved as

$$u_k(t) = u_k(0)\mathrm{e}^{\mathrm{i}\omega_k t}. \tag{5.3}$$

Substituting (5.3) back into (5.2 *a*) and integrating in time, we obtain

$$\begin{aligned} v_0(t) &= v_0(0) + \operatorname{Re} \sum_\alpha c_\alpha u^\alpha(0) \int_0^t \exp\left(\mathrm{i} \sum_k \omega_k \alpha_k t\right) \mathrm{d}t \\ &= v_0(0) + \operatorname{Re} \sum_\alpha c_\alpha u^\alpha(0) \frac{\exp(\mathrm{i}\sum_k \omega_k \alpha_k t) - 1}{\mathrm{i}\sum_k \omega_k \alpha_k}. \end{aligned}$$

5.2. Implicit midpoint time discretization

We write $u^n = v^n + \mathrm{i}w^n$ to denote a time discretization of (5.2). Specifically, one step of the implicit midpoint time discretization with step size h applied to (5.2 *a*) and (5.2 *b*) takes the form

$$v_0^1 = v_0^0 + h \operatorname{Re} f(\tfrac{1}{2}(u^1 + u^0)), \tag{5.4 a}$$

$$w_0^1 = w_0^0. \tag{5.4 b}$$

Thus, assuming that w_0 is zero initially, w_0^n remains zero in every step. Then, for $k \geqslant 1$, we obtain

$$u_k^1 = u_k^0 + h\mathrm{i}\omega_k \frac{u_k^1 + u_k^0}{2} + h\frac{w_0^1 + w_0^0}{2}(\cdots)$$

so that, noting that the last term is zero, we can write

$$u^1 = \mathsf{S}(hA)u^0, \tag{5.5}$$

where $\mathsf{S}(hA) = \operatorname{diag}(\mathsf{s}_1, \mathsf{s}_2, \dots)$ with

$$\mathsf{s}_k = \mathsf{s}_k(h) = \frac{1 + \frac{1}{2}\mathrm{i}\omega_k h}{1 - \frac{1}{2}\mathrm{i}\omega_k h}. \tag{5.6}$$

Plugging (5.5) into (5.4 *a*), we can write

$$v_0^1 = v_0^0 + h\,\mathrm{Re}\,f(\mathsf{G}u^0) = v_0^0 + h\,\mathrm{Re}\sum_\alpha c_\alpha(\mathsf{G}u^0)^\alpha, \tag{5.7}$$

where $\mathsf{G} = \mathsf{G}(h) = \mathrm{diag}(\mathsf{g}_1, \mathsf{g}_2, \dots)$ with

$$\mathsf{g}_k = \mathsf{g}_k(h) = \frac{1+\mathsf{s}_k}{2} = \frac{1}{1-\frac{1}{2}\mathrm{i}\omega_k h}. \tag{5.8}$$

5.3. Modified vector field

To obtain an expression for the modified vector field for initial data $w_0(0) = 0$, we make the ansatz

$$\dot{\tilde{v}}_0 = \mathrm{Re}(\tilde{f}(\tilde{u})) \quad \text{with } \tilde{f}(u) = \sum_\alpha \tilde{c}_\alpha u^\alpha \tag{5.9}$$

and, for $k \geqslant 1$,

$$\dot{\tilde{u}}_k = \mathrm{i}\tilde{\omega}_k\tilde{u}_k. \tag{5.10}$$

Integrating (5.10) from 0 to h, where $\tilde{u}(0) \equiv u^0$ and equating the resulting expression with (5.5), we immediately obtain that

$$\mathsf{s}_k = \mathrm{e}^{\mathrm{i}\tilde{\omega}_k h}. \tag{5.11}$$

Similarly, inserting the solution of (5.10) into (5.9), integrating from 0 to h and equating the resulting expression with (5.7), we find

$$h\,\mathrm{Re}[c_\alpha(\mathsf{G}u^0)^\alpha] = \mathrm{Re}\left[\tilde{c}_\alpha\frac{\exp(\mathrm{i}\sum_k \alpha_k\tilde{\omega}_k h) - 1}{\mathrm{i}\sum_k \alpha_k\tilde{\omega}_k}(u^0)^\alpha\right]. \tag{5.12}$$

Therefore, to satisfy (5.12), we have to find $\tilde{c}_\alpha$ such that

$$hc_\alpha\mathsf{g}^\alpha = \tilde{c}_\alpha\frac{\exp(\mathrm{i}\sum_k \alpha_k\tilde{\omega}_k h) - 1}{\mathrm{i}\sum_k \alpha_k\tilde{\omega}_k}. \tag{5.13}$$

Note that the left-hand side of this equation only vanishes when $h = 0$ or $c_\alpha = 0$.

5.4. Resonances

Equation (5.13) can be solved for $\tilde{c}_\alpha$ unless the numerator on the right-hand side is zero. Let us call this a resonance. Due to (5.11), we can write the condition for resonance as $\mathsf{s}^\alpha = 1$. Let us look for particular resonances between three consecutive 'wavenumbers' such that

$$\mathsf{s}_{k-1}\mathsf{s}_k\mathsf{s}_{k+1} = 1. \tag{5.14}$$

Using the definition of s_k in (5.6), this condition reads

$$4(\omega_k + \omega_{k+1} + \omega_{k-1}) = h^2\omega_{k-1}\omega_k\omega_{k+1}. \tag{5.15}$$

We consider the case $\omega_k = k^2$ so that (5.2) is a nonlinear Schrödinger equation. We let $u \equiv (v, w) \in \hbar_1 \times \hbar_1 = \mathcal{Y}$, where $\hbar_\ell = \hbar_\ell(\mathbb{N}_0; \mathbb{R})$. Thus, $\mathcal{Y}$ is the Sobolev space $\mathcal{H}_\ell(\mathbb{S}^1; \mathbb{C})$ in Fourier coordinates. Further, (5.2) is of the form (1.1) with A

the Laplacian in Fourier coordinates, i.e. $(Au)_k = \mathrm{i}\omega_k u_k$. Consequently, (A), (H0) and (H1) hold as described in § 2.6; recall that we require $u \in \hbar_1(\mathbb{N}_0;\mathbb{C})$ to ensure that $H(u)$ is finite. The nonlinearity B from (1.1) is then defined by

$$B(u)_k = \begin{cases} \begin{pmatrix} w_0 \operatorname{Re} \dfrac{\partial f(u)}{\partial w_k} \\ -w_0 \operatorname{Re} \dfrac{\partial f(u)}{\partial v_k} \end{pmatrix} & \text{for } k \geqslant 1, \\ \begin{pmatrix} \operatorname{Re} f(u) \\ 0 \end{pmatrix} & \text{for } k = 0. \end{cases} \tag{5.16}$$

We wish to satisfy (B0)–(B2) and (H2–4) with $\mathcal{D}_k = \mathcal{B}_R^{\mathcal{Y}_k}(0)$ and $\mathcal{D}_{\tau,L,q} = \mathcal{B}_R^{\mathcal{Y}_{\tau,L,q}}(0)$ for some positive R, τ and L. Let us look at one such f, where all interactions are between triples of consecutive wavenumbers, namely

$$f(u) = \sum_{j=2}^{\infty} u_{j-1}u_j u_{j+1}.$$

Due to the Hölder inequality,

$$|f(u)| \leqslant \sum_{j=2}^{\infty} |u_{j-1}u_j u_{j+1}| \leqslant \|u\|_{\ell_2}^2 \|u\|_{\ell_\infty} \leqslant \|u\|_{\ell_2}^3.$$

so that $f\colon \ell_2(\mathbb{N}_0;\mathbb{C}) \to \mathbb{C}$ is a bounded trilinear form on ℓ_2, and hence analytic as a function on ℓ_2 and therefore also on $\hbar_1$.

Similarly, we can check that $B(u)$ defined as in (5.16) is an analytic map from $\hbar_k$ to itself for all $k \in \mathbb{N}_0$ so that (B0)–(B2) hold for any $L \geqslant 0$, $\tau > 0$ and $q = 2$. As in § 2.6, we consider the case $q = 2$ because this Gevrey space consists of sequences $\{u_k\}$ whose Fourier series $u(x) = \sum_{k\in\mathbb{Z}} u_k \mathrm{e}^{\mathrm{i}kx}$ is analytic in x.

Let us now consider the resonance condition (5.15) for our example, where $\omega_k = k^2$. We obtain

$$4((k-1)^2 + k^2 + (k+1)^2) = h^2(k-1)^2k^2(k+1)^2$$

or, after simplification,

$$12 + \frac{8}{k^2} = h^2(k^2-1)^2.$$

For h sufficiently small, there is exactly one positive root $k(h)$. Clearly, $k(h) = O(h^{-1/2})$ as $h \to 0$, and there is a resonance whenever $k(h) \in \mathbb{Z}$.

At a resonance, the embedding error in the v_0 component is given by the left-hand side of (5.12). For our example, $c_\alpha = 1$, so that the embedding error reads

$$e(h) = h \operatorname{Re}(g_{k-1}g_k g_{k+1} u^0_{k-1}u^0_k u^0_{k+1}).$$

Since $k = O(h^{-1/2})$, (5.8) shows that $g_j = O(1)$ for $j = k-1, k, k+1$. Clearly, the embedding error can only be exponentially small provided u_k^0 decays exponentially as $k \to \infty$. As a specific example, take

$$u_k^0 = \frac{\mathrm{e}^{-\tau k}}{k^{\ell+2}}$$

for $k \geqslant 1$ and $u_0^0 = 0$. It is easily seen that $u^0 \in \mathcal{Y}_{\tau,\ell,2}$ for any $\tau, \ell \geqslant 0$. For this initial condition, the embedding error satisfies

$$e(h) = O\left(\frac{h\mathrm{e}^{-3\tau k}}{k^{3\ell+6}}\right) = O(\mathrm{e}^{-ch^{-1/2}}) \tag{5.17}$$

for some $c > 0$ as in [21].

Remark 5.1. If in the setting above we choose

$$f(u) = \frac{1}{u_0} + \sum_{j=2}^{\infty} u_{j-1}u_j u_{j+1},$$

then the nonlinearity B defined by $f(u)$ via (5.16) can be considered on the open half-balls

$$\mathcal{D}_k = \mathrm{int}(\mathcal{B}_R^{\mathcal{Y}_k}(0)) \cap \{u\colon u_0 > \varepsilon\}$$

for some $\varepsilon > 0$ fixed. Then $\mathcal{D}_k$ is convex, and hence star shaped with respect to any $u \in \mathcal{D}_k$. Defining $\mathcal{D}_{\tau,\ell,q}$ analogously, we obtain a nested domain hierarchy on which B satisfies (B0)–(B2) and (H2)–(H4); (H0), (H1) and (A) hold true as before. This example shows that it makes sense to consider domains $\mathcal{D}_k$ different from balls.

Acknowledgements

The work of C.W. was supported by the Nuffield Foundation, the Leverhulme Foundation and by EPSRC Grant no. EP/D063906/1. M.O. was supported by a Max Kade Fellowship, by the ESF network 'Harmonic and Complex Analysis and Applications (HCAA)' and by German Science Foundation Grant no. OL 155/5-1.

Data management

No additional research data beyond the data presented and cited in this work are needed to validate the research findings in this work.

References

1 G. Benettin and A. Giorgilli. On the Hamiltonian interpolation of near-to-the-identity symplectic mappings with application to symplectic integration algorithms. *J. Statist. Phys.* **74** (1994), 1117–1143.

2 B. Cano. Conserved quantities of some Hamiltonian wave equations after full discretization. *Numer. Math.* **103** (2006), 197–223.

3 K. D. Cherednichenko and V. P. Smyshlyaev. On full two-scale expansion of the solutions of nonlinear periodic rapidly oscillating problems and higher-order homogenized variational problems. *Arch. Ration. Mech. Analysis* **174** (2004), 385–442.

4 D. Cohen, E. Hairer and C. Lubich. Conservation of energy, momentum and actions in numerical discretizations of nonlinear wave equations. *Numer. Math.* **110** (2008), 113–143.

5 A. Debussche and E. Faou. Modified energy for split-step methods applied to the linear Schrödinger equation. *SIAM J. Numer. Analysis* **47** (2009), 3705–3719.

6 G. Dujardin and E. Faou. Normal form and long time analysis of splitting schemes for the linear Schrödinger equation with small potential. *Numer. Math.* **106** (2007), 223–262.

7 E. Faou and B. Grébert. Hamiltonian interpolation of splitting approximations for nonlinear PDEs. *Found. Comput. Math.* **11** (2011), 381–415.

8 E. Faou, B. Grébert and E. Paturel. Birkhoff normal form for splitting methods applied to semilinear Hamiltonian PDEs. I. Finite dimensional discretization. *Numer. Math.* **114** (2010), 429–458.

9 E. Faou, B. Grébert and E. Paturel. Birkhoff normal form for splitting methods applied to semilinear Hamiltonian PDEs. II. Abstract splitting. *Numer. Math.* **114** (2010), 459–490.

10 A. B. Ferrari and E. S. Titi. Gevrey regularity for nonlinear analytic parabolic equations. *Commun. PDEs* **23** (1998), 1–16.

11 B. Fiedler and J. Scheurle. *Discretization of homoclinic orbits, rapid forcing and 'invisible' chaos* (Providence, RI: American Mathematical Society, 1996).

12 L. Gauckler and C. Lubich. Splitting integrators for nonlinear Schrödinger equations over long times. *Found. Comput. Math.* **10** (2010), 275–302.

13 E. Hairer, C. Lubich and G. Wanner. *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations* (Springer, 2002).

14 A. Iserles. *First course in the numerical analysis of differential equations* (Cambridge University Press, 1996).

15 A. L. Islas and C. M. Schober. Backward error analysis for multisymplectic discretizations of Hamiltonian PDEs. *Math. Comput. Simulat.* **69** (2005), 290–303.

16 V. Kamotski, K. Matthies and V. Smyshlyaev. Exponential homogenization of linear second order elliptic problems with periodic coefficients. *SIAM J. Math. Analysis* **38** (2007), 1565–1587.

17 B. Leimkuhler and S. Reich. *Simulating Hamiltonian dynamics* (Cambridge University Press, 2004).

18 J. E. Marsden and T. S. Ratiu. *Introduction to mechanics and symmetry* (Springer, 1994).

19 K. Matthies. Time-averaging under fast periodic forcing of parabolic partial differential equations: exponential estimates. *J. Diff. Eqns* **174** (2001), 88–133.

20 K. Matthies. Backward error analysis of a full discretisation scheme for a class of parabolic partial differential equations. *Nonlin. Analysis TMA* **52** (2003), 805–826.

21 K. Matthies and A. Scheel. Exponential averaging of Hamiltonian evolution equations. *Trans. Am. Math. Soc.* **355** (2003), 747–773.

22 B. Moore and S. Reich. Backward error analysis for Hamiltonian PDEs with applications to nonlinear wave equations. *Numer. Math.* **95** (2003), 625–652.

23 A. I. Neishtadt. On the separation of motions in systems with rapidly rotating phase. *J. Appl. Math. Mech.* **48** (1984), 134–139.

24 M. Oliver and C. Wulff. A-stable Runge–Kutta methods for semilinear evolution equations. *J. Funct. Analysis* **263** (2012), 1981–2023.

25 M. Oliver and C. Wulff. Stability under Galerkin truncation of A-stable Runge–Kutta discretizations in time. *Proc. R. Soc. Edinb.* A **144** (2014), 603–636.

26 M. Oliver, M. West and C. Wulff. Approximate momentum conservation for spatial semi-discretizations of nonlinear wave equations. *Numer. Math.* **97** (2004), 493–535.

27 A. Pazy. *Semigroups of linear operators and applications to partial differential equations* (Springer, 1983).

28 M. Reed and B. Simon. *Methods of modern mathematical physics*, volume I: *Functional analysis* (San Diego, CA: Academic Press, 1972).

29 S. Reich. Backward error analysis for numerical integrators. *SIAM J. Numer. Analysis* **36** (1999), 1549–1570.

30 J. M. Sanz-Serna and M. P. Calvo. *Numerical Hamiltonian problems* (London: Chapman-Hall, 1994).