Version of 27 March 2024

Regularized dynamical parametric approximation

Michael Feischl[1], **Caroline Lasser**[2],
Christian Lubich[3], **Jörg Nick**[4]

Abstract This paper studies the numerical approximation of evolution equations by nonlinear parametrizations $u(t) = \Phi(q(t))$ with time-dependent parameters $q(t)$, which are to be determined in the computation. The motivation comes from approximations in quantum dynamics by multiple Gaussians and approximations of various dynamical problems by tensor networks and neural networks. In all these cases, the parametrization is typically irregular: the derivative $\Phi'(q)$ can have arbitrarily small singular values and may have varying rank. We derive approximation results for a regularized approach in the time-continuous case as well as in time-discretized cases. With a suitable choice of the regularization parameter and the time stepsize, the approach can be successfully applied in irregular situations, even though it runs counter to the basic principle in numerical analysis to avoid solving ill-posed subproblems when aiming for a stable algorithm. Numerical experiments with sums of Gaussians for approximating quantum dynamics and with neural networks for approximating the flow map of a system of ordinary differential equations illustrate and complement the theoretical results.

Keywords. Time-dependent nonlinear parametric approximation, regularization, high-dimensional differential equation, Schrödinger equation, time integration, deep neural network, multi-Gaussian approximation.

[1] Institute for Analysis and Scientific Computing, Wiedner Hauptstraße 8–10, 1040 Wien, TU Wien, Austria. E-mail: michael.feischl@tuwien.ac.at
[2] Department of Mathematics, Boltzmannstraße 3, TU München, D-85748 Garching bei München, Germany. E-mail: classer@ma.tum.de
[3] Mathematisches Institut, Auf der Morgenstelle 10, Univ. Tübingen, D-72076 Tübingen, Germany. E-mail: Lubich@na.uni-tuebingen.de
[4] Seminar für Angewandte Mathematik, Rämistrasse 101, ETH Zürich, CH-8092 Zürich, Switzerland. E-mail: joerg.nick@math.ethz.ch

1 Introduction

1.1 Nonlinear parametric approximation of evolution problems

We consider the numerical approximation of a possibly high-dimensional initial-value problem of ordinary or partial differential equations

$$\dot{y} = f(y), \qquad y(0) = y_0 \tag{1.1}$$

via a nonlinear parametrization

$$y(t) \approx u(t) = \Phi(q(t)), \qquad 0 \le t \le \bar{t}, \tag{1.2}$$

with time-dependent parameters $q(t)$. Here, Φ is a smooth map from a parameter space into the solution space. We are interested in the situation of irregular parametrizations: the derivative matrix $\Phi'(q)$ may have arbitrarily small singular values and possibly also varying rank. This is a typical situation of over-approximation that frequently arises in applications and causes severe numerical difficulties. Our motivation for studying such problems originated from the following areas, for which references are given and discussed in Subsection 1.3.

- Multi-Gaussian approximations in quantum dynamics: Here the parameters are the evolving complex width matrices, positions, momenta, and phases in a sum of complex Gaussians.
- Tensor network approximations in quantum dynamics: Here the parameters are the evolving connection tensors and bases.
- Deep neural network approximations of dynamical problems: Here the parameters are the evolving weight matrices and biases of the various layers of the neural network.

In all these cases, the derivative matrix $\Phi'(q)$ typically has numerous very small singular values.

1.2 Regularized dynamical parametric approximation

In the numerical approach considered in this paper, we determine the time derivatives $\dot{q}(t)$ and $\dot{u}(t) = \Phi'(q(t))\dot{q}(t)$ by solving the regularized linear least squares problem (we omit the argument t)

$$\|\dot{u} - f(u)\|^2 + \varepsilon^2 \|\dot{q}\|^2 = \min! \tag{1.3}$$

with a possibly time-dependent regularization parameter $\varepsilon(t) > 0$. This yields a differential equation for the parameters q, and then $u = \Phi(q)$. We refer to this approximation as a regularized dynamical parametric approximation. The resulting differential equation for q is solved numerically by a standard time-stepping method, where each function evaluation solves a linear least squares problem (1.3).

This algorithmic approach is studied here even though the differential equation for the parameters q is severely ill-conditioned for small ε. Errors in the initial value $q(0)$ can increase by a factor $e^{ct/\varepsilon}$ (with a constant $c > 0$ independent of ε) to errors in $q(t)$ at times $t > 0$. The approach thus appears to run counter to a basic principle of numerical analysis: to avoid solving ill-posed subproblems when aiming for a stable algorithm.

As a consequence of the ill-posedness in the parameters, also the error propagation in the solution approximation $u(t) = \Phi(q(t))$ is ill-behaved, but nevertheless the problem turns out to be what might be called *well-posed up to the order of the defect size* $O(\delta)$, where $\delta = \max_{0\le t\le \bar{t}} \delta(t)$ and $\delta(t)^2$ is the minimum value attained in (1.3) at time t. This beneficial behaviour is found both in the differential equation that results from (1.3) and in its numerical time discretization, and this makes the regularization (1.3) a viable computational approach. Its analysis adds new aspects to that of the underlying differential equation (1.1) and the time discretization.

One possibly obvious finding from our analysis is that good pointwise approximability of the solution $y(t)$ by parametrized functions is not sufficient. It is important that the time derivative $\dot{y}(t)$ can be well approximated in the tangent spaces at parametrized functions $u = \Phi(q)$ near $y(t)$. This suffices when f is Lipschitz-continuous. In the case of a dominant term Ay in $f(y) = Ay + g(y)$ with an unbounded operator A that maps parametrized functions into their tangent space (a situation encountered for the Schrödinger equation), we need that $\dot{y}(t) - Ay(t)$ can be well approximated in the tangent space, and the regularized least-squares problem (1.3) should be slightly modified.

Remark 1.1 (Truncation vs. regularization) As an alternative to regularizing the least-squares problem as in (1.3), small singular values of the matrix $A = \Phi'(q)$ below a threshold $\varepsilon > 0$ are set to zero, yielding a truncated matrix A_ε with smallest nonzero singular value at least ε. Instead of (1.3), the time derivative $\dot{q}$ is then determined to be of minimal norm such that

$$\|A_\varepsilon(q)\dot{q} - f(\Phi(q))\|^2 = \min! \tag{1.4}$$

Because of discontinuities in $A_\varepsilon(\cdot)$ when singular values cross the threshold, the resulting differential equation $\dot{q} = A_\varepsilon(q)^+ f(\Phi(q))$ with the Moore–Penrose pseudo-inverse A_ε^+ has a discontinuous right-hand side and is problematic to analyse. After time discretization, however, this approach can be analysed by arguments that are entirely analogous to the regularized approach (1.3) and it is found to behave in a very similar way, since A_ε^+ behaves similarly to the matrix $(A^\top A + \varepsilon^2 I)^{-1} A^\top$ that appears in the normal equations for (1.3). Both matrices have the same singular vectors and the singular values are σ_i^{-1} if $\sigma_i \ge \varepsilon$ and zero else, and $(\sigma_i^2 + \varepsilon^2)^{-1}\sigma_i$, respectively, where σ_i are the singular values of A.

1.3 Related work

In quantum dynamics, the non-regularized approach (1.3) with $\varepsilon = 0$ is known as the Dirac–Frenkel time-dependent variational principle and has been widely used in computational chemistry and quantum physics over the past decades; see the books [15, 17] for dynamic, geometric and approximation aspects and among countless papers see e.g. [12, 28, 23] with Gaussian wavepackets, [19, 27] and [25, 8] with tensor networks and [4, 11, 24] with artificial neural networks. This approach usually leads to ill-conditioned least squares problems which cause numerical difficulties[1]. *Ad hoc* regularization is often used in computations in such cases, but a systematic foundation and analysis appear to be missing in the literature. The regularization chosen in [23] corresponds precisely to (1.3) but is not further investigated there. The work [13] analyzes and proposes several regularization schemes in the context of Krylov-subspace projections of the original problem. This is of interest if one additionally wants to analyze the solution process of the arising linear systems, which we do not study here.

In recent years, an extensive amount of research has been invested in nonlinear approximations of differential equations, in particular in the context of neural networks. The majority of the literature relies on minimizing the residual in space-time, which yields a nonlinear optimization problem for the parameters globally in time. Prominent examples of such a strategy include the "deep Ritz method" [29] as well as "weak generative adversarial networks" [30] and both were successfully applied to a wide range of partial differential equations (see e.g. [31, 1, 2, 14]). The nonlinear optimization process is, however, difficult to analyse and the existing theory consequently focuses on the expressivity properties of neural networks, i.e., what approximation is theoretically possible (see e.g. [22, 20, 16]). This of course excludes the much harder task of training the network and explaining the success of these methods remains evasive.

An alternative approach in the literature is to use Galerkin based methods to obtain differential equations for the parameters, see e.g. [30]. The Dirac–Frenkel time-dependent variational principle has been used outside the realm of quantum dynamics in the context of deep neural networks, for example in [7] (referred to as "evolutional deep neural network") and in [3] (referred to as "neural Galerkin schemes").

1.4 Outline

After establishing notation and framework in Section 2, we derive *a posteriori* and *a priori* error bounds for the time-continuous regularized approach (1.3)

[1] The case of tree tensor networks, which includes low-rank matrices, Tucker tensors and tensor trains as special cases, is exceptional since its multilinear geometry allows for time integration algorithms that are robust to arbitrarily small singular values [5, 18].

in Section 3, where we also study the sensitivity of approximations to initial values.

In Section 4 we study the error behaviour of the time discretization by the explicit and implicit Euler methods, where we observe in particular the interplay between the time stepsize and the regularization parameter. In Section 5 the result is extended to Runge-Kutta discretizations, for which we prove an optimal-order error bound up to the defect size δ. In Section 6 we propose an adaptive selection of the regularization parameter and the stepsize that is based on the previous error analysis.

In Section 7 we study the effect of enforcing conserved quantities, which are typically not preserved by the regularized parametric approach.

In Section 8 we study the regularization approach for the time-dependent Schrödinger equation as an important exemplary case of an evolutionary partial differential equation.

Finally, in Section 9 we present two numerical examples that illustrate the behaviour of the regularized parametric approximation: first by a neural network to approximate the flow map of a system of ordinary differential equations, and second by a linear combination of complex Gaussians to approximate quantum dynamics.

In an appendix we collect some useful results on regularized least-square problems.

2 Regularized dynamical parametric approximation

We start with the formulation of the framework. Let the state space $\mathcal{H}$ be a Hilbert space (finite- or infinite-dimensional) with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and corresponding norm $\| \cdot \|_{\mathcal{H}}$ and let the parameter space $\mathcal{Q}$ be a finite-dimensional vector space equipped with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{Q}}$ and corresponding norm $\| \cdot \|_{\mathcal{Q}}$. The parametrization map $\Phi : \mathcal{Q} \to \mathcal{H}$ is assumed to be twice continuously differentiable. It need not be one-to-one, not even locally. The derivative $\Phi'(q)$ may have arbitrarily small singular values and a nontrivial nullspace of varying dimension.

The initial value in (1.1) is assumed to be in parametrized form: $y(0) = \Phi(q(0))$ for some $q(0) \in \mathcal{Q}$. We take $u(0) = y(0) = \Phi(q(0))$.

For $t \geq 0$, let $\varepsilon(t) > 0$ be the regularization parameter. For $u(t) \in \mathcal{H}$ and $q(t) \in \mathcal{Q}$ with $u(t) = \Phi(q(t))$, the time derivatives $\dot{u}(t) = \Phi'(q(t))\dot{q}(t) \in \mathcal{H}$ and $\dot{q}(t) \in \mathcal{Q}$ are determined by requiring that

$$\delta(t)^2 := \|\dot{u}(t) - f(u(t))\|_{\mathcal{H}}^2 + \varepsilon(t)^2 \|\dot{q}(t)\|_{\mathcal{Q}}^2 \quad \text{is minimal.} \tag{2.1}$$

In this regularized linear least squares problem, $\dot{q}(t)$ is uniquely determined, which it would not necessarily be without the regularization.

The graph

$$\mathcal{M} = \{(u, q) \,:\, u = \Phi(q)\} \subset \mathcal{H} \times \mathcal{Q}$$

need not be a manifold, but at $(u,q) \in \mathcal{M}$ it has the tangent space

$$T_{(u,q)}\mathcal{M} = \{(\dot u, \dot q) \,:\, \dot u = \Phi'(q)\dot q\} \subset \mathcal{H} \times \mathcal{Q}.$$

Remark 2.1 With respect to the ε-weighted inner product on $\mathcal{H} \times \mathcal{Q}$ that is defined by $\langle \cdot,\cdot\rangle_{\mathcal{H}} + \varepsilon^2 \langle \cdot,\cdot\rangle_{\mathcal{Q}}$, we consider the orthogonal projection onto $T_{(u,q)}\mathcal{M}$, denoted

$$P^{\varepsilon}_{(u,q)} : \mathcal{H} \times \mathcal{Q} \to T_{(u,q)}\mathcal{M}.$$

Then, (2.1) is equivalent to (omitting the argument t)

$$(\dot u, \dot q) = P^{\varepsilon}_{(u,q)}\big(f(u), 0\big). \tag{2.2}$$

While this reformulation is reminiscent of the useful interpretation of the Dirac–Frenkel time-dependent variational principle as a projection, see [17, Section II.1], the difficulty here is that $P^{\varepsilon}_{(u,q)}$ has no moderate Lipschitz constant with respect to (u,q) under our assumptions, as is quantified in Proposition 3.4 below. The Lipschitz-continuity of the projection onto the tangent space was an essential condition in the proof of quasi-optimality of the approximation obtained from the Dirac–Frenkel time-dependent variational principle in [17, Section II.6]. This is not available here, and so the question arises as to what alternative kind of error analysis can still be done.

3 Error bounds

In this section we assume that the vector field $f : \mathcal{H} \to \mathcal{H}$ in the differential equation (1.1) is a Lipschitz-continuous map, with the one-sided Lipschitz constant ℓ and the Lipschitz constant L: for all $v, w \in \mathcal{H}$,

$$\begin{aligned} \operatorname{Re}\langle v-w, f(v)-f(w)\rangle_{\mathcal{H}} &\le \ell\, \|v-w\|^2_{\mathcal{H}}, \\ \|f(v)-f(w)\|_{\mathcal{H}} &\le L\, \|v-w\|_{\mathcal{H}}. \end{aligned} \tag{3.1}$$

The real part is taken in the case of a complex-linear space $\mathcal{H}$.

3.1 A posteriori error bound

For the defect $d(t)$ of the approximation $u = \Phi(q)$ we have

$$\dot u(t) = f(u(t)) + d(t) \qquad \text{with} \quad \|d(t)\|_{\mathcal{H}} \le \delta(t) \tag{3.2}$$

for $\delta(t)$ of (2.1). There is the following error bound in terms of δ.

Proposition 3.1 *In the situation of Section 2 and with the one-sided Lipschitz condition* (3.1)*, the error is bounded by*

$$\|u(t) - y(t)\|_{\mathcal{H}} \le \int_0^t e^{\ell(t-s)}\, \delta(s)\, ds.$$

Proof The proof is standard and is included for the convenience of the reader. We subtract (1.1) from (3.2) and take the inner product with $u - y$ and its real part (if $\mathcal{H}$ is a complex space). On the left-hand side this yields, omitting the omnipresent argument t,

$$\operatorname{Re}\langle u - y, \dot{u} - \dot{y}\rangle = \frac{1}{2}\frac{d}{dt}\|u - y\|^2 = \|u - y\| \cdot \frac{d}{dt}\|u - y\|$$

and on the right-hand side

$$\operatorname{Re}\langle u - y, f(u) - f(y) + d\rangle \le \|u - y\|\big(\ell\,\|u - y\| + \delta\big).$$

Dividing both sides by $\|u - y\|$ and using Gronwall's inequality yields the result. □

While $\delta(t)$ can be monitored during the computation and is thus available *a posteriori*, we have not bounded it *a priori* from known or assumed approximability properties of the exact solution. This is done next.

3.2 A priori error bound

We will bound the defect size $\delta(t)$ by a quantity that measures the approximability of the solution derivative $\dot{y}(t)$ in the tangent spaces $T_{(u,q)}\mathcal{M}$ for all $u = \Phi(q)$ in a neighbourhood of $y(t)$. We fix a radius $\rho > 0$ such that there exist parametrized functions $u = \Phi(q)$ with $\|u - y(t)\|_{\mathcal{H}} \le \rho$, and define

$$\bar{\delta}_\rho(t)^2 := \sup_{q\in\mathcal{Q}:\|\Phi(q)-y(t)\|_{\mathcal{H}}\le\rho} \min_{\dot{q}\in\mathcal{Q}}\Big(\|\Phi'(q)\dot{q} - \dot{y}(t)\|^2 + \varepsilon^2\|\dot{q}\|^2\Big). \tag{3.3}$$

We can bound $\delta(t)$ of (2.1) in terms of $\bar{\delta}_\rho(t)$.

Lemma 3.1 *Provided that $\|u(t) - y(t)\| \le \rho$, we have*

$$\delta(t) \le \bar{\delta}_\rho(t) + \|f(u(t)) - f(y(t))\|_{\mathcal{H}}.$$

Proof We omit the argument t in the following. We have $u = \Phi(q)$ and $\dot{u} = \Phi'(q)\dot{q}$. Let $\dot{q}_+ \in \mathcal{Q}$ be such that

$$\|\Phi'(q)\dot{q}_+ - \dot{y}\|_{\mathcal{H}}^2 + \varepsilon^2\|\dot{q}_+\|_{\mathcal{Q}}^2 \quad \text{is minimal.}$$

We obtain from (2.1), (3.3) and $\dot{y} = f(y)$

$$\begin{aligned}
\delta^2 &= \|\Phi'(q)\dot{q} - f(u)\|_{\mathcal{H}}^2 + \varepsilon^2\|\dot{q}\|_{\mathcal{Q}}^2 \\
&\le \|\Phi'(q)\dot{q}_+ - f(u)\|_{\mathcal{H}}^2 + \varepsilon^2\|\dot{q}_+\|_{\mathcal{Q}}^2 \\
&\le \Big(\|\Phi'(q)\dot{q}_+ - \dot{y}\|_{\mathcal{H}} + \|f(y) - f(u)\|_{\mathcal{H}}\Big)^2 + \varepsilon^2\|\dot{q}_+\|_{\mathcal{Q}}^2 \\
&\le \bar{\delta}_\rho^2 + 2\bar{\delta}_\rho\,\|f(y) - f(u)\|_{\mathcal{H}} + \|f(y) - f(u)\|_{\mathcal{H}}^2 \\
&= \Big(\bar{\delta}_\rho + \|f(y) - f(u)\|_{\mathcal{H}}\Big)^2.
\end{aligned}$$

Taking square roots yields the result. □

We construct a reference approximation $u_*(t) = \Phi(q_*(t))$ with $q_*(t) \in \mathcal{Q}$ from the exact solution such that its derivative is a regularized best approximation in the tangent space at $(u_*(t), q_*(t))$ to the solution derivative $\dot{y}(t)$. The derivatives $\dot{u}_*(t) = \Phi'(q_*(t))\dot{q}_*(t)$ and $\dot{q}_*(t)$ are determined such that

$$\delta_*(t)^2 := \|\dot{u}_*(t) - \dot{y}(t)\|_{\mathcal{H}}^2 + \varepsilon^2 \|\dot{q}_*(t)\|_{\mathcal{Q}}^2 \quad \text{is minimal.} \tag{3.4}$$

Here we have the immediate error bound

$$\|u_*(t) - y(t)\|_{\mathcal{H}} \le \int_0^t \|\dot{u}_*(s) - \dot{y}(s)\|_{\mathcal{H}}\, ds \le \int_0^t \delta_*(s)\, ds \le \int_0^t \bar{\delta}_\rho(s)\, ds \tag{3.5}$$

as long as this bound does not exceed ρ. The following result bounds the error of the numerical approximation $u(t)$ by a multiple of the bound in (3.5).

Proposition 3.2 *In the situation of Section 2 and under the Lipschitz condition* (3.1)*, the error is bounded by*

$$\|u(t) - y(t)\|_{\mathcal{H}} \le e^{(\ell+L)t} \int_0^t \bar{\delta}_\rho(s)\, ds$$

as long as this does not exceed ρ.

Proof The bound follows from Lemma 3.1 inserted into the proof of Proposition 3.1 and using the Lipschitz condition on f and the Gronwall lemma. □

3.3 Sensitivity to initial values

Given two initial values $y_0, \widetilde{y}_0 \in \mathcal{H}$, the difference of the correponding solutions $y(t)$ and $\widetilde{y}(t)$ of the differential equation (1.1) are bounded, under the one-sided Lipschitz condition (3.1), by

$$\|y(t) - \widetilde{y}(t)\|_{\mathcal{H}} \le e^{\ell t}\, \|y_0 - \widetilde{y}_0\|_{\mathcal{H}}, \qquad t \ge 0;$$

see e.g. [10, IV.12]. There is no analogous result for the regularized approximations $u(t) = \Phi(q(t))$ and $\widetilde{u}(t) = \Phi(\widetilde{q}(t))$ with initial values $u_0 = \Phi(q_0)$ and $\widetilde{u}_0 = \Phi(\widetilde{q}_0)$, not even with the Lipschitz constant L instead of ℓ. We only obtain the following bound.

Proposition 3.3 *Under the one-sided Lipschitz condition* (3.1) *we have*

$$\|u(t) - \widetilde{u}(t)\|_{\mathcal{H}} \le e^{\ell t}\, \|u_0 - \widetilde{u}_0\|_{\mathcal{H}} + \int_0^t e^{\ell(t-s)} \big(\delta(s) + \widetilde{\delta}(s)\big) ds, \qquad t \ge 0, \tag{3.6}$$

with the defect size $\delta(t)$ of (2.1) *and analogously $\widetilde{\delta}(t)$ corresponding to $\widetilde{u}(t)$.*

Proof This bound is obtained by the argument of Proposition 3.1 and estimating the difference of the defects $d(t) = \dot{u}(t) - f(u(t))$ of (3.2) and analogously $\widetilde{d}(t)$ in a rough way by $\|d(t) - \widetilde{d}(t)\|_{\mathcal{H}} \le \delta(t) + \widetilde{\delta}(t)$. □

Note that the right-hand side of (3.6) does not tend to zero as $u_0 - \widetilde{u}_0 \to 0$. The difference of the defects is bounded in this rough way because there does not seem to be a better way that does not introduce negative powers of ε into the bound. The difficulty becomes apparent in the following bound for $\dot{u}(t) - \dot{\widetilde{u}}(t)$, where the second term on the right-hand side is a multiple of $\|q(t) - \widetilde{q}(t)\|_{\mathcal{Q}}$, which cannot be avoided and cannot be estimated by an ε-independent multiple of $\|u(t) - \widetilde{u}(t)\|_{\mathcal{H}}$, as we would like.

Proposition 3.4 *As a function of q, we denote by $\dot{q}(q)$ the solution of the regularized least squares problem*

$$\delta(q)^2 = \|\Phi'(q)\dot{q}(q) - f(\Phi(q))\|_{\mathcal{H}}^2 + \varepsilon^2\|\dot{q}(q)\|_{\mathcal{Q}}^2 \quad \textit{is minimal.} \tag{3.7}$$

Then, for all $q, \widetilde{q} \in \mathcal{Q}$ and associated $u = \Phi(q)$, $\widetilde{u} = \Phi(\widetilde{q})$ and $\dot{u} = \Phi'(q)\dot{q}(q)$, $\dot{\widetilde{u}} = \Phi'(\widetilde{q})\dot{q}(\widetilde{q})$,

$$\|\dot{u} - \dot{\widetilde{u}}\|_{\mathcal{H}} \le L\|u - \widetilde{u}\|_{\mathcal{H}} + \tfrac{5}{2}\beta\,\frac{\bar{\delta}}{\varepsilon}\,\|q - \widetilde{q}\|_{\mathcal{Q}},$$

$$\|\varepsilon\dot{q}(q) - \varepsilon\dot{q}(\widetilde{q})\|_{\mathcal{Q}} \le L\|u - \widetilde{u}\|_{\mathcal{H}} + \tfrac{5}{2}\beta\,\frac{\bar{\delta}}{\varepsilon}\,\|q - \widetilde{q}\|_{\mathcal{Q}},$$

$$\textit{and} \qquad |\delta(q) - \delta(\widetilde{q})| \le 3L\|u - \widetilde{u}\|_{\mathcal{H}} + 5\beta\frac{\bar{\delta}}{\varepsilon}\,\|q - \widetilde{q}\|_{\mathcal{Q}},$$

where β is an upper bound on second order derivatives of Φ and $\bar{\delta}$ an upper bound on δ on the line segment between q and $\widetilde{q}$, and L is the Lipschitz constant of f in (3.1).

Proof We use the adjoint $\Phi'(q)^* \in L(\mathcal{H}, \mathcal{Q})$ of the linear map $\Phi'(q) \in L(\mathcal{Q}, \mathcal{H})$ and the normal equations $M_\varepsilon(q)\dot{q}(q) = \Phi'(q)^* f(\Phi(q))$ with Gramian matrix $M_\varepsilon(q) = \Phi'(q)^*\Phi'(q) + \varepsilon^2 I_{\mathcal{Q}}$. We denote

$$\begin{aligned}\Phi'(q)\dot{q}(q) &= \Phi'(q)M_\varepsilon(q)^{-1}\Phi'(q)^* f(\Phi(q))\\ &=: P_\varepsilon(q)f(\Phi(q)).\end{aligned}$$

By Lemma A.2, we have $\|P_\varepsilon(q)\|_{L(\mathcal{H})} \le 1$ and therefore

$$\|P_\varepsilon(\widetilde{q})(f(\Phi(q)) - f(\Phi(\widetilde{q})))\|_{\mathcal{H}} \le L\|\Phi(q) - \Phi(\widetilde{q})\|_{\mathcal{H}}. \tag{3.8}$$

As for the sensitivity of $P_\varepsilon(q)$ with respect to q, we consider the regularized least squares problem with fixed right hand side $b \in \mathcal{H}$ and its normal equation

$$M_\varepsilon(q)\dot{q}_b(q) = \Phi'(q)^* b.$$

We differentiate

$$\partial_q M_\varepsilon(q)\dot{q}_b(q) + M_\varepsilon(q)\partial_q\dot{q}_b(q) = \partial_q\Phi'(q)^* b$$

and obtain

$$\begin{aligned}&\Phi'(q)\partial_q\dot{q}_b(q) = \Phi'(q)M_\varepsilon(q)^{-1}\left(\partial_q\Phi'(q)^* b - \partial_q M_\varepsilon(q)\dot{q}_b(q)\right)\\ &= \Phi'(q)M_\varepsilon(q)^{-1}\left(\partial_q\Phi'(q)^*\left(b - \Phi'(q)\dot{q}_b(q)\right) - \Phi'(q)^*\partial_q\Phi'(q)\dot{q}_b(q)\right).\end{aligned}$$

By Lemma A.2, we have $\|\Phi'(q)M_\varepsilon(q)^{-1}\|_{L(\mathcal{Q},\mathcal{H})} \le \frac{1}{2\varepsilon}$ and consequently

$$\|\Phi'(q)\partial_q \dot q_b(q)\|_{\mathcal{H}} \le \frac{\beta}{2\varepsilon}\|b - \Phi'(q)\dot q_\beta(q)\|_{\mathcal{H}} + \beta\|\dot q_b(q)\|_{\mathcal{Q}} \le \frac{3\beta}{2\varepsilon}\,\delta_b(q)$$

with

$$\delta_b(q)^2 = \|\Phi'(q)\dot q_b(q) - b\|_{\mathcal{H}}^2 + \varepsilon^2\|\dot q_b(q)\|_{\mathcal{Q}}^2.$$

This implies, for

$$\partial_q(P_\varepsilon(q)b) = \partial_q(\Phi'(q)\dot q_b(q)) = \partial_q\Phi'(q)\dot q_b(q) + \Phi'(q)\partial_q\dot q_b(q),$$

the bound

$$\|\partial_q(P_\varepsilon(q)b)\|_{\mathcal{H}} \le \frac{5\beta}{2\varepsilon}\,\delta_b(q).$$

Using this bound for the right-hand side $b = f(\Phi(q))$ and combining it with the Lipschitz estimate (3.8), we obtain

$$\begin{aligned}
&\|\Phi'(q)\dot q(q) - \Phi'(\widetilde q)\dot q(\widetilde q)\|_{\mathcal{H}}\\
&\le \|(P_\varepsilon(q) - P_\varepsilon(\widetilde q))f(\Phi(q))\|_{\mathcal{H}} + \|P_\varepsilon(\widetilde q)(f(\Phi(q)) - f(\Phi(\widetilde q)))\|_{\mathcal{H}}\\
&\le \frac{5\beta\bar\delta}{2\varepsilon}\|q - \widetilde q\|_{\mathcal{Q}} + L\|\Phi(q) - \Phi(\widetilde q)\|_{\mathcal{H}},
\end{aligned}$$

which is the first of the stated bounds. The second one follows in the same way. Finally, with $u = \Phi(q)$ and $\dot u = \Phi'(q)\dot q$ and analogously $\widetilde u$ and $\dot{\widetilde u}$ we have by the Cauchy–Schwarz inequality and (3.7)

$$\begin{aligned}
|\delta(q)^2 - \delta(\widetilde q)^2| &= \big|\big\langle(\dot u - f(u)) + (\dot{\widetilde u} - f(\widetilde u)), (\dot u - \dot{\widetilde u}) - (f(u) - f(\widetilde u))\big\rangle_{\mathcal{H}}\\
&\qquad + \varepsilon^2\big\langle \dot q + \dot{\widetilde q}, \dot q - \dot{\widetilde q}\big\rangle_{\mathcal{Q}}\big|\\
&\le \big(\delta(q) + \delta(\widetilde q)\big)\big(\|\dot u - \dot{\widetilde u}\|_{\mathcal{H}} + \|f(u) - f(\widetilde u)\|_{\mathcal{H}}\big) + \big(\delta(q) + \delta(\widetilde q)\big)\,\varepsilon\|\dot q - \dot{\widetilde q}\|_{\mathcal{Q}},
\end{aligned}$$

which yields

$$|\delta(q) - \delta(\widetilde q)| \le \|\dot u - \dot{\widetilde u}\|_{\mathcal{H}} + L\|u - \widetilde u\|_{\mathcal{H}} + \varepsilon\|\dot q - \dot{\widetilde q}\|_{\mathcal{Q}},$$

and the stated bound follows from the first two bounds. □

Remark 3.1 As far as the inverse powers with respect to ε are concerned, the previous estimates are sharp, as the following one-dimensional example with $\mathcal{H} = \mathcal{Q} = \mathbb{R}$, $\Phi(q) = \frac{1}{2}q^2$, and $f(u) = 1$ illustrates. In this case we have

$$\dot q(q) = \frac{q}{q^2 + \varepsilon^2}, \quad d(q) = \Phi'(q)\dot q(q) - f(u) = -\frac{\varepsilon^2}{q^2 + \varepsilon^2}$$

and

$$\partial_q\dot q(q) = -\frac{q^2}{q^2 + \varepsilon^2} + \frac{\varepsilon^2}{(q^2 + \varepsilon^2)^2} = -\frac{d(q)}{q^2 + \varepsilon^2} - \frac{q^2}{q^2 + \varepsilon^2}.$$

Moreover,

$$\begin{aligned}|\dot{u} - \dot{\widetilde{u}}| &= \left| \frac{q}{q^2+\varepsilon^2} d(\widetilde{q}) + \frac{\widetilde{q}}{\widetilde{q}^2+\varepsilon^2} d(q) \right| |q - \widetilde{q}| \\ &= |\delta(q) - \delta(\widetilde{q})|\end{aligned}$$

which implies

$$\begin{aligned}|\dot{u} - \dot{\widetilde{u}}| = |\delta(q) - \delta(\widetilde{q})| &\le \frac{\bar{\delta}}{\varepsilon} |q - \widetilde{q}| \\ |\varepsilon \dot{q}(q) - \varepsilon \dot{q}(\widetilde{q})| &\le \left(\varepsilon + \frac{\bar{\delta}}{\varepsilon} \right) |q - \widetilde{q}|.\end{aligned}$$

These tight estimates exhibit the factor $\bar{\delta}/\varepsilon$ in the same way as the general bounds of Proposition 3.4.

Remark 3.2 Using Proposition 3.4 and Gronwall's lemma, we obtain the bound

$$\|u(t) - \widetilde{u}(t)\|_{\mathcal{H}} + \varepsilon \|q(t) - \widetilde{q}(t)\|_{\mathcal{Q}} \le e^{\omega t} \left(\|u_0 - \widetilde{u}_0\|_{\mathcal{H}} + \varepsilon \|q_0 - \widetilde{q}_0\|_{\mathcal{Q}} \right)$$

with the exponent $\omega = \max\bigl(\beta\bar{\delta}/\varepsilon^2, L\bigr)$. For nonlinear parametrizations (where $\beta \neq 0$), which is our main interest in this paper, this is not a useful estimate for small ε. For linear parametrizations ($\beta = 0$), however, we even obtain

$$\|u(t) - \widetilde{u}(t)\|_{\mathcal{H}} \le e^{Lt} \|u_0 - \widetilde{u}_0\|_{\mathcal{H}}$$

without any dependence on $\|q_0 - \widetilde{q}_0\|_{\mathcal{Q}}$. Hence, with respect to error propagation, linear and nonlinear near-singular parametrizations behave very differently.

4 Time discretization by the regularized Euler method

We study time-stepping methods in the framework of Sections 2 and 3, first based on the Euler method and in the next section on general Runge–Kutta methods. The unusual feature is that the differential equation for the parameters q, which is the equation that is actually discretized, is ill-behaved, but nevertheless we find better behaviour of $u = \Phi(q)$. We only assume that the vector field f in the original differential equation (1.1) for y is sufficiently differentiable, and that the parametrization map Φ has bounded second derivatives (but a regular parametrization is not assumed). These properties will be used to derive error bounds for the discrete approximations $u_n = \Phi(q_n)$ at $t_n = t_0 + nh$, for stepsizes $h > 0$ that need to be suitably restricted in terms of the defect in the regularized least squares problem and the regularization parameter. No error bounds are derived for the parameters q_n.

4.1 The regularized explicit Euler method

A step of the explicit Euler method applied to the differential equation for the parameters q, starting from q_n at time t_n with the regularization parameter ε_n, reads

$$q_{n+1} = q_n + h\dot{q}_n, \qquad u_{n+1} = \Phi(q_{n+1}), \tag{4.1}$$

where $\dot{q}_n$ is the solution of the regularized linear least squares problem

$$\delta_n^2 := \|\Phi'(q_n)\dot{q}_n - f(u_n)\|_{\mathcal{H}}^2 + \varepsilon_n^2 \|\dot{q}_n\|_{\mathcal{Q}}^2 \quad \text{is minimal.} \tag{4.2}$$

4.2 Local error bound

We first consider the local error, that is the error after one step $u_1 - y(t_1)$, where $y(t)$ is the solution of (1.1) starting from $y(t_0) = u_0 = \Phi(q_0)$.

Lemma 4.1 *Under the stepsize restriction*

$$h\delta_0 \le c\,\varepsilon_0^2, \tag{4.3}$$

the local error of the regularized Euler method starting from $y(t_0) = u_0$ is bounded by

$$\|u_1 - y(t_1)\|_{\mathcal{H}} \le c_1 h\delta_0 + c_2 h^2 \tag{4.4}$$

with $c_1 = 1 + c\beta$, where β is a bound of the second derivative of Φ in a neighbourhood of q_0, and $c_2 = \frac{1}{2}\max_{t_0 \le t \le t_1} \|\ddot{y}(t)\|_{\mathcal{H}}$.

Proof We have by Taylor expansion

$$\begin{aligned} y(t_1) - y(t_0) &= h\dot{y}(t_0) + \int_{t_0}^{t_1} (t_1 - t)\,\ddot{y}(t)\,dt \\ &= hf(u_0) + O(h^2) \end{aligned}$$

and we have

$$\begin{aligned} u_1 - u_0 &= \Phi(q_1) - \Phi(q_0) \\ &= \Phi'(q_0)(q_1 - q_0) + \int_0^1 (1-\theta)\Phi''(q_0 + \theta(q_1 - q_0))[q_1 - q_0, q_1 - q_0]\,d\theta \\ &= \Phi'(q_0)h\dot{q}_0 + O(\|h\dot{q}_0\|_{\mathcal{Q}}^2) \\ &= hf(u_0) + O(h\delta_0) + O\left(\left(\frac{h\delta_0}{\varepsilon_0}\right)^2\right), \end{aligned}$$

where the last line results from the definition of δ_0, which is an upper bound of both $\|\Phi'(q_0)\dot{q}_0 - f(u_0)\|_{\mathcal{H}}$ and $\varepsilon_0\|\dot{q}_0\|_{\mathcal{Q}}$. This implies $\|h\dot{q}_0\|_{\mathcal{Q}} \le h\delta_0/\varepsilon_0$. Under the stepsize restriction $h\delta_0 \le c\varepsilon_0^2$, the last term is also $O(h\delta_0)$. Subtracting the two formulas and tracing the constants in the $O(\cdot)$ terms yields the stated result. □

Remark 4.1 The result of Lemma 4.1 holds without any fixed bound on $\dot{q}_0$. If, however, $\|\dot{q}_0\|_{\mathcal{Q}} \le \gamma$ with a moderate constant γ, which is satisfied if $\delta_0 \le \gamma\varepsilon_0$, then the above proof shows that the error bound (4.4) is satisfied with $c_1 = 1$ and $c_2 = \frac{1}{2}\max_{t_0\le t\le t_1}\|\ddot{y}(t)\|_{\mathcal{H}} + \beta\gamma^2$ with the same β as in the lemma. This clearly indicates that while the stepsize restriction (4.3) is sufficient for the stated error bound, it is not always necessary.

Remark 4.2 A local error estimate of a similar structure, without any restriction on the step size h or the defect δ_0, is readily available by keeping an additional a posteriori error term. Comparing the explicit Euler approximation $y_1 = y_0 + hf(y_0)$ as an intermediate term with $u_1 = \Phi(q_1)$ yields

$$\begin{aligned}
&\|u_1 - y(t_1)\|_{\mathcal{H}} \le \|u_1 - y_1\|_{\mathcal{H}} + \|y_1 - y(t_1)\|_{\mathcal{H}} \\
&\le \|u_1 - y_0 - h\Phi'(q_0)\dot{q}_0\|_{\mathcal{H}} + h\|\Phi'(q_0)\dot{q}_0 - f(u_0)\|_{\mathcal{H}} + \|y_1 - y(t_1)\|_{\mathcal{H}} \\
&\le \|u_1 - y_0 - h\Phi'(q_0)\dot{q}_0\|_{\mathcal{H}} + h\delta_0 + O(h^2). \qquad (4.5)
\end{aligned}$$

The first term in the last line is computable with the quantities derived during the time integration of the scheme and comes at the cost of evaluating an additional $\mathcal{H}$-norm. Under the step size restriction of Lemma 4.1, or in the case of Remark 4.1, the term is of the optimal order $O(h\delta_0 + h^2)$.

4.3 Global error bound (using stable error propagation by the exact flow)

Using the standard argument of Lady Windermere's fan with error propagation by the exact flow, see the book by Hairer, Nørsett & Wanner [9, II.3], we obtain the following global error bound from Lemma 4.1 and condition (3.1).

Proposition 4.1 *Under condition* (3.1) *and the stepsize restriction*

$$h\delta_n \le c\,\varepsilon_n^2, \qquad 0 \le n \le N,$$

the error of the regularized Euler method (4.1)–(4.2) *with initial value* $y_0 = u_0 = \Phi(q_0)$ *is bounded, for* $t_n = nh \le t_N \le \bar{t}$, *by*

$$\|u_n - y(t_n)\|_{\mathcal{H}} = O(\delta + h) \quad \textit{with} \quad \delta = \max_n \delta_n,$$

or more precisely (compare with Proposition 3.1),

$$\|u_n - y(t_n)\|_{\mathcal{H}} \le h\sum_{j=0}^{n-1} e^{\ell(t_{n-1}-t_j)}\,\big(c_1\delta_j + c_2 h\big),$$

where $c_1 = 1 + c\beta$ *with* β *a bound of the second derivative of* Φ *in a neighbourhood of the solution, and* c_2 *is a bound of* $f'f$ *in a neighbourhood of the solution* $y(t)$, *i.e. a bound of second derivatives of solutions* $\widetilde{y}$ *of the differential equation* $\dot{\widetilde{y}} = f(\widetilde{y})$ *with initial values in a neighbourhood of the solution trajectory* $\{y(t) : 0 \le t \le \bar{t}\}$.

4.4 Error propagation by the numerical method

The Euler method applied to (1.1) is stable in the sense that the numerical results y_1 and $\widetilde{y}_1$ obtained from starting values y_0 and $\widetilde{y}_0$, respectively, differ by

$$\|y_1 - \widetilde{y}_1\|_{\mathcal{H}} \leq (1 + hL)\|y_0 - \widetilde{y}_0\|_{\mathcal{H}},$$

where L is again the Lipschitz constant of f. For the regularized Euler method we only obtain, similarly to the proof of Lemma 4.1, a discrete analogue of (3.6):

$$\|u_1 - \widetilde{u}_1\|_{\mathcal{H}} \leq (1 + hL)\|u_0 - \widetilde{u}_0\|_{\mathcal{H}} + h\delta_0 + h\widetilde{\delta}_0. \tag{4.6}$$

This is not good enough to be used in Lady Windermere's fan with error propagation by the numerical method, which is the most common way to prove error bounds for numerical methods for nonstiff differential equations; see [9].

4.5 A priori error bound for the regularized explicit Euler method

The bound in Proposition 4.1 is partly *a priori* (in its dependence on ℓ, c_1 and c_2) and partly *a posteriori* (in its dependence on the defect sizes δ_j). In terms of the method-independent defect sizes $\bar{\delta}_\rho(t)$ ρ-near to the exact solution as defined in (3.3), we have the following discrete analogue of Proposition 3.2.

Proposition 4.2 *In the situation of Proposition 4.1 we have the* a priori *error bound, for* $0 \leq t_n = nh \leq \bar{t}$,

$$\|u_n - y(t_n)\|_{\mathcal{H}} \leq C_1 h \sum_{j=1}^{n} \bar{\delta}_\rho(t_j) + C_2 h,$$

where C_1 *and* C_2 *depend on* $L\bar{t}$ *but are independent of* h *and* ε. *This bound is valid as long as the right-hand side does not exceed* ρ.

Proof We still have the bound of Lemma 3.1,

$$\delta_n \leq \bar{\delta}_\rho(t_n) + \|f(u_n) - f(y(t_n))\|_{\mathcal{H}}.$$

Inserting this bound in the local error bound of Proposition 4.1 and using the Lipschitz condition on f yields

$$\|u_n - y(t_n)\|_{\mathcal{H}} \leq c_1 h \sum_{j=1}^{n} \bar{\delta}_\rho(t_j) + c_1 h \sum_{j=1}^{n} L\,\|u_j - y(t_j)\|_{\mathcal{H}} + c_2 h,$$

and the discrete Gronwall inequality yields the result. □

4.6 Regularized implicit Euler method

A step of the implicit Euler method, starting from q_n at time t_n reads

$$q_{n+1} = q_n + h\dot{q}_{n+1}, \qquad u_{n+1} = \Phi(q_{n+1}),$$

where $\dot{q}_{n+1}$ is the solution of the regularized linear least squares problem

$$\delta_{n+1}^2 = \|\Phi'(q_{n+1})\dot{q}_{n+1} - f(u_{n+1})\|_{\mathcal{H}}^2 + \varepsilon_{n+1}^2\|\dot{q}_{n+1}\|_{\mathcal{Q}}^2 \quad \text{is minimal.}$$

The defining equation for q_{n+1} is a fixed point equation $q_{n+1} = \varphi_n(q_{n+1})$ for the function

$$\varphi_n(q) := q_n + h \operatorname{argmin}_{\dot{q}\in\mathcal{Q}} \left(\|\Phi'(q)\dot{q} - f(\Phi(q))\|_{\mathcal{H}}^2 + \varepsilon_{n+1}^2\|\dot{q}\|_{\mathcal{Q}}^2\right).$$

This is a contraction for h sufficiently small due to the bound of Proposition 3.4, which implies

$$\|\varphi_n(q) - \varphi_n(\widetilde{q})\|_{\mathcal{Q}} \le h\left(\frac{\beta_2\bar{\delta}}{\varepsilon_{n+1}^2} + \frac{\beta_1 L}{\varepsilon_{n+1}}\right)\|q - \widetilde{q}\|_{\mathcal{Q}}$$

for $\bar{\delta} > 0$ an upper bound on the residual size in a neighbourhood of $q, \widetilde{q}$. The constants $\beta_1, \beta_2 > 0$ depend on first order and second order derivatives of Φ, respectively. Hence, with a step-size restriction of the form

$$h\left(\frac{\beta_2\bar{\delta}}{\varepsilon_{n+1}^2} + \frac{\beta_1 L}{\varepsilon_{n+1}}\right) < 1,$$

where $\bar{\delta} > 0$ is an upper bound on the residual size in the vicinity of the numerical solution, the fixed point iteration is contractive and thus converging.

The local error satisfies

$$\|u_1 - y(t_1)\|_{\mathcal{H}} \le \beta h^2\|\dot{q}_1\|_{\mathcal{H}}^2 + h\delta_1 + c_2 h^2,$$

as is seen by repeating the proof of Lemma 4.1 with obvious adaptations. This leads to global error bounds as in Propositions 4.1 and 4.2.

5 Regularized explicit Runge–Kutta methods

A step of an explicit Runge–Kutta method of order p with coefficients a_{ij} and b_j applied to the differential equation for the parameters q reads as follows. Starting from q_n at time t_n with the regularization parameter ε_n, we first compute consecutively the internal stages (for $i = 1, \dots, s$)

$$q_{n,i} = q_n + h\sum_{j=1}^{i-1} a_{ij}\dot{q}_{n,j}, \qquad u_{n,i} = \Phi(q_{n,i}), \tag{5.1}$$

and $\dot{q}_{n,i}$ as the solution of the regularized linear least squares problem

$$\delta_{n,i}^2 := \|\Phi'(q_{n,i})\dot{q}_{n,i} - f(u_{n,i})\|_{\mathcal{H}}^2 + \varepsilon_n^2\|\dot{q}_{n,i}\|_{\mathcal{Q}}^2 \quad \text{is minimal.} \tag{5.2}$$

Finally, the new value is computed as

$$q_{n+1} = q_n + h \sum_{j=1}^{s} b_j \dot{q}_{n,j}, \qquad u_{n+1} = \Phi(q_{n+1}). \tag{5.3}$$

We first bound the local error.

Lemma 5.1 *Let $\delta_0 = \max_i \delta_{0,i}$. Under the stepsize restriction*

$$h\delta_0 \le c\,\varepsilon_0^2, \tag{5.4}$$

the local error of the regularized p-th order Runge–Kutta method starting from $y(t_0) = u_0 = \Phi(q_0)$ is bounded by

$$\|u_1 - y(t_1)\|_{\mathcal{H}} \le c_1 h\delta_0 + c_2 h^{p+1}, \tag{5.5}$$

where c_1 is a constant times $1 + c\beta$ with the bound β of the second derivative of Φ in a neighbourhood of q_0, and $c_2 h^{p+1}$ is the bound for the local error of the p-th order Runge–Kutta method applied to (1.1).

Proof We note that (5.2) implies $\|\varepsilon_0 \dot{q}_{0,j}\|_{\mathcal{Q}} \le \delta_0$ and hence $\|h\dot{q}_{0,j}\|_{\mathcal{Q}} \le h\delta_0/\varepsilon_0$. Moreover, (5.2) also implies $\|\Phi'(q_{0,j})\dot{q}_{0,j} - f(u_{0,j})\|_{\mathcal{H}} \le \delta_0$. This yields

$$\begin{aligned}
u_{0,i} - u_0 &= \Phi(q_{0,i}) - \Phi(q_0) = \Phi'(q_0) h \sum_{j=1}^{i-1} a_{ij} \dot{q}_{0,j} + O\bigg(\Big(\frac{h\delta_0}{\varepsilon_0}\Big)^2\bigg) \\
&= h \sum_{j=1}^{i-1} a_{ij} \Phi'(q_{0,j}) \dot{q}_{0,j} - \sum_{j=1}^{i-1} a_{ij} \big(\Phi'(q_{0,j}) - \Phi'(q_0)\big) h\dot{q}_{0,j} + O\bigg(\Big(\frac{h\delta_0}{\varepsilon_0}\Big)^2\bigg) \\
&= h \sum_{j=1}^{i-1} a_{ij} f(u_{0,j}) + O(h\delta_0) + O\bigg(\Big(\frac{h\delta_0}{\varepsilon_0}\Big)^2\bigg) + O\bigg(\Big(\frac{h\delta_0}{\varepsilon_0}\Big)^2\bigg)
\end{aligned}$$

and in the same way

$$u_1 - u_0 = h \sum_{j=1}^{s} b_j f(u_{0,j}) + O(h\delta_0) + O\bigg(\Big(\frac{h\delta_0}{\varepsilon_0}\Big)^2\bigg).$$

Apart from the $O(\cdot)$ terms, these formulae are those that define the result y_1 of one step of the Runge–Kutta method applied to (1.1). So we obtain, using also the Lipschitz continuity of f,

$$u_1 - y_1 = O(h\delta_0) + O\bigg(\Big(\frac{h\delta_0}{\varepsilon_0}\Big)^2\bigg).$$

Since the method is of order p and f is sufficiently differentiable, we have

$$y_1 - y(t_1) = O(h^{p+1}).$$

Noting that under the stepsize restriction (5.4) we have $(h\delta_0/\varepsilon_0)^2 \le ch\delta_0$, the error bound (5.5) follows. □

As before, using the local error bound in Lady Windermere's fan with error propagation by the exact solutions, we obtain the following global error bound for the p-th order Runge method. We formulate the result for variable stepsizes h_n, so that $t_{n+1} = t_n + h_n$ and $t_N = \bar{t}$.

Proposition 5.1 *Under condition* (3.1) *and the stepsize restriction*

$$h_n \delta_{n,i} \le c\,\varepsilon_n^2, \qquad 0 \le n \le N, \quad i = 1, \dots, s,$$

the error of the regularized p-th order Runge–Kutta method (5.1)–(5.3) *with initial value* $y_0 = u_0 = \Phi(q_0)$ *is bounded, for* $t_n \le \bar{t}$, *by*

$$\|u_n - y(t_n)\|_{\mathcal{H}} = O(\delta + h^p)$$

with $\delta = \max_{n,i} \delta_{n,i}$ *and* $h = \max_n h_n$. *The constants symbolized by the O-notation are independent of* δ *and the regularization parameters* ε_n *(under the given stepsize restriction), and of the stepsize sequence* (h_k) *and* n *with* $t_n \le \bar{t}$.

6 Choice of the regularization parameter and the stepsize

The algorithm below chooses the regularization parameter ε_n in the nth time step as large as possible such that the defect δ_n is within a given factor of the defect attained for tiny ε or within a prescribed tolerance. The stepsize h_n is chosen such that the critical quadratic error terms are of size $h_n \delta_n$.

For arbitrary $\varepsilon > 0$, in the nth time step we let $\dot{q}_n(\varepsilon)$ be the solution of the regularized linear lest squares problem with regularization parameter ε such that

$$\delta_n(\varepsilon)^2 := \|\Phi'(q_n)\dot{q}_n(\varepsilon) - f(u_n)\|_{\mathcal{H}}^2 + \varepsilon^2 \|\dot{q}_n(\varepsilon)\|_{\mathcal{Q}}^2 \quad \text{is minimal.}$$

Let ε_n^0 and h_n^0 be initializations of the regularization parameter and the stepsize, respectively. These might be the values from the previous time step. Let $\varepsilon_\star$ be a tiny reference parameter such that $\delta_n(\varepsilon_\star)$ can still be computed reliably (just so). Let $\delta_{\min} > 0$ be a given threshold.

6.1 Choice of the regularization parameter ε_n

Compute $\delta_n(\varepsilon_\star)$ and set $\delta_n^{\mathrm{tol}} = \max(17\,\delta_n(\varepsilon_\star), \delta_{\min})$ as the target defect size[2]. We aim at having $\delta_n(\varepsilon_n) \approx \delta_n^{\mathrm{tol}}$. We use 1 or 2 Newton iterations for the equation $\delta_n(\varepsilon_n)^2 - (\delta_n^{\mathrm{tol}})^2 = 0$ (except in the very first step where a good starting value is not available and more iterations might be needed). In view of Lemma A.1, which shows that this is a monotonically increasing concave

[2] The factor 17, chosen in honour of Gauß, can be replaced by a different factor *ad libitum*.

smooth function with the derivative $d\delta_n^2/d\varepsilon^2 = \|\dot q_n(\varepsilon)\|_{\mathcal{Q}}^2$, we have the Newton iteration

$$\left(\varepsilon_n^{k+1}\right)^2 = \left(\varepsilon_n^k\right)^2 - \frac{\delta_n(\varepsilon_n^k)^2 - \left(\delta_n^{\mathrm{tol}}\right)^2}{\|\dot q_n(\varepsilon_n^k)\|_{\mathcal{Q}}^2}.$$

Alternatively, we can proceed in a Levenberg–Marquardt style, cf. [21]:
Initialize $\varepsilon_n = \varepsilon_n^0$.
If $\delta_n(\varepsilon_n) \le \delta_n^{\mathrm{tol}}$
  then while $\delta_n(3\varepsilon_n) \le \delta_n^{\mathrm{tol}}$ set $\varepsilon_n := 3\varepsilon_n$
  else while $\delta_n(\varepsilon_n) > \delta_n^{\mathrm{tol}}$ set $\varepsilon_n := \varepsilon_n/3$.

In the following we set $\dot q_n = \dot q_n(\varepsilon_n)$ and $\delta_n = \delta_n(\varepsilon_n)$.

6.2 Choice of the stepsize h_n

The stepsize h_n is chosen such that $\|\Phi''(q_n)[h_n\dot q_n, h_n\dot q_n]\|_{\mathcal{H}} \approx h_n\delta_n$. This is satisfied for the choice

$$h_n = \frac{h_n^0\delta_n}{\|(\Phi'(q_n + h_n^0\dot q_n) - \Phi'(q_n))\dot q_n\|_{\mathcal{H}}}.$$

6.3 Regularized Runge–Kutta step with ε_n and h_n

We compute q_{n+1} and $u_{n+1} = \Phi(q_{n+1})$ by a regularized Runge–Kutta step (see Section 5) with the proposed regularization parameter ε_n and the proposed stepsize h_n. We can use an embedded pair of Runge–Kutta methods that gives a local error estimate that should not substantially exceed $h_n\delta_n$ (else the step is rejected and repeated with a reduced stepsize).

7 Conserved quantities

7.1 Conserved quantities and regularized dynamical approximation

Let $g = (g_1, \dots, g_m)^\top : \mathcal{H} \to \mathbb{R}^m$ be an m-vector of real-valued functions that are conserved along the flow of the differential equation (1.1):

$$g(y(t)) = g(y(0)) \qquad \text{for all } t$$

for every choice of initial value $y(0) \in \mathcal{H}$. Differentiating this equation w.r.t. t, this is seen to be equivalent to

$$G(y(t))\dot y(t) = 0,$$

where $G = g' = \partial g/\partial y$, and hence to

$$G(y)f(y) = 0 \qquad \text{for all } y.$$

However, g is no longer a conserved quantity for the regularized dynamical approximation (2.1), not even for linear g. We use the notation

$$A = \Phi', \quad M_\varepsilon = A^\top A + \varepsilon^2 I, \quad P_\varepsilon = A M_\varepsilon^{-1} A^\top$$

where we omitted the argument q for all appearing matrices[3]. For the regularized least squares problem (2.1) we have the normal equations (now omitting the argument t)

$$M_\varepsilon(q)\dot{q} = A(q)^\top f(\Phi(q)) \tag{7.1}$$

and for $u = \Phi(q)$, the time derivative $\dot{u} = \Phi'(q)\dot{q}$ becomes

$$\dot{u} = P_\varepsilon(q) f(u) = f(u) + d \quad \text{with} \quad d = -(I - P_\varepsilon(q)) f(u), \tag{7.2}$$

where the defect d is bounded by $\|d\|_{\mathcal{H}} \le \delta$ in view of (2.1). As $G(u)f(u) = 0$, we obtain

$$\frac{d}{dt}\, g(u) = G(u)\dot{u} = G(u)d,$$

which in general is different from zero, so that $g(u(t))$ is not conserved. We note, however, the bound

$$|g(u(t)) - g(u(0))| \le K \int_0^t \delta(t)\, dt,$$

where K is an upper bound of the norm of $G(u)$ along the trajectory $u(\cdot)$.

7.2 Enforcing conservation

We can enforce conservation of g along the approximation $u(t) = \Phi(q(t))$ by adding the condition $G(u(t))\dot{u}(t) = 0$ as a constraint, i.e. (omitting the argument t)

$$G(\Phi(q))A(q)\dot{q} = 0,$$

and we minimize in (2.1) under this constraint. With the notation

$$C(q) := G(\Phi(q))A(q) = g'(u)\Phi'(q),$$

we obtain instead of (7.1) the constrained system with a Lagrange multiplier $\lambda(t) \in \mathbb{R}^m$,

$$\begin{aligned} M_\varepsilon(q)\dot{q} + C(q)^\top \lambda &= A(q)^\top f(\Phi(q)) \\ C(q)\dot{q} \qquad\qquad &= 0. \end{aligned} \tag{7.3}$$

Inserting $\dot{q}$ from the first equation into the second equation, we find λ from the equation (omitting the argument q or u of the matrices)

$$C M_\varepsilon^{-1} C^\top \lambda = C M_\varepsilon^{-1} A^\top f(u)$$

[3] In the infinite-dimensional case, the linear map $A : \mathbb{R}^k \to \mathcal{H}$ can be viewed as a quasi-matrix $A = (a_1, \ldots, a_k)$ with $a_i \in \mathcal{H}$, see e.g. [26], for which $A^\top$ is to be interpreted as the adjoint of A. For $v, w \in \mathcal{H}$ we interpret $v^\top w$ as the inner product $(v, w)_{\mathcal{H}}$. In the following we will use the familiar matrix notation throughout.

or equivalently, since $C = GA$ and $P_\varepsilon = AM_\varepsilon^{-1}A^\top$ imply $CM_\varepsilon^{-1}C^\top = GP_\varepsilon G^\top$ and $CM_\varepsilon^{-1}A^\top = GP_\varepsilon$, and since $(P_\varepsilon - I)f(u) = d$ and $G(u)f(u) = 0$, we find

$$\big(G(u)P_\varepsilon(q)G(u)^\top\big)\lambda = G(u)d. \tag{7.4}$$

The symmetric positive semi-definite matrix $P_\varepsilon = AM_\varepsilon^{-1}A^\top$ has the eigenvalues $\lambda_i = \sigma_i^2/(\sigma_i^2+\varepsilon^2)$ and 0, where σ_i are the singular values of A. Eigenvalues are very small if they correspond to very small singular values $\sigma_i \ll \varepsilon$ of A, but are larger than $\frac{1}{2}$ for $\sigma_i \geq \varepsilon$. To understand under which condition the symmetric positive semi-definite matrix $GP_\varepsilon G^\top$ has a moderately bounded inverse, let Λ be the diagonal matrix of eigenvalues of P_ε and U the orthogonal matrix of eigenvectors, and for $\theta > 0$ let U_θ be the matrix composed of those eigenvectors of P_ε that correspond to the eigenvalues $\lambda_i \geq \theta$. If the smallest singular value of GU_θ equals $\rho > 0$, then

$$\|(GP_\varepsilon G^\top)^{-1}\|_2 \leq \frac{1}{\theta\rho^2}, \tag{7.5}$$

because $v^\top GU\Lambda(GU)^\top v \geq v^\top GU_\theta\Lambda_\theta(GU_\theta)^\top v \geq \theta\|(GU_\theta)^\top v\|_2^2 \geq \theta\rho^2\|v\|_2^2$ for all $v \in \mathbb{R}^m$. Here, Λ_θ is the diagonal matrix of those eigenvalues of P_ε that are larger than θ. We remark that in the case of just one conserved quantity ($m = 1$), the inverse of $GP_\varepsilon G^\top \in \mathbb{R}$ is moderately bounded if ∇g is not near-orthogonal to all those singular vectors of A that correspond to singular values $\sigma_i \geq \varepsilon$. This appears to be a very mild requirement.

Inserting λ from (7.4) into the first equation of (7.3) and using the definition of the defect d yields $\dot u = A\dot q$ as

$$\dot u + P_\varepsilon G^\top(GP_\varepsilon G^\top)^{-1}Gd = f(u) + d.$$

With the projection onto the null-space of $G(u)$ for $u = \Phi(q)$ that is given by

$$\Pi(q) = I - P_\varepsilon(q)G(u)^\top(G(u)P_\varepsilon(q)G(u)^\top)^{-1}G(u),$$

we thus obtain for $u(t) = \Phi(q(t))$, instead of (7.2), the differential equation with the projected defect,

$$\dot u = f(u) + \Pi(q)d. \tag{7.6}$$

7.3 Constrained regularized Euler method

We now ensure the condition $g(u_{n+1}) = g(u_n)$ by adding it as a constraint to the regularized least squares problem (4.2). A step of the constrained regularized Euler method, starting from q_n at time t_n with the regularization parameter ε_n, reads

$$q_{n+1} = q_n + h\dot q_n, \qquad u_{n+1} = \Phi(q_{n+1}), \tag{7.7}$$

where $\dot q_n$ is the solution of the constrained regularized linear least squares problem

$$\begin{aligned} &\widehat\delta_n^2 := \|\varPhi'(q_n)\dot q_n - f(u_n)\|_{\mathcal H}^2 + \varepsilon_n^2 \|\dot q_n\|_{\mathcal Q}^2 \quad \text{is minimal} \\ &\text{subject to} \quad g(u_{n+1}) = g(u_n). \end{aligned} \tag{7.8}$$

Note that while δ_n of (4.2) depends only on q_n, the defect size $\widehat\delta_n \ge \delta_n$ depends also on the stepsize h. With the notation of the previous subsections, a step of the unconstrained regularized Euler method of Section 4 reads (with $\varepsilon = \varepsilon_n$)

$$M_\varepsilon(q_n)\dot q_n = A(q_n)^\top f(u_n)$$

together with $\widetilde q_{n+1} = q_n + h\dot q_n$ and $\widetilde u_{n+1} = \varPhi(q_{n+1})$. The minimality condition for the constrained problem (7.8) now determines (a different) $\dot q_n$ together with the Lagrange multiplier λ_{n+1} from the nonlinear system of equations

$$\begin{aligned} &M_\varepsilon(q_n)\dot q_n + C(q_n)^\top \lambda_{n+1} = A(q_n)^\top f(u_n) \\ &g\big(\varPhi(q_n + h\dot q_n)\big) = g(u_n). \end{aligned} \tag{7.9}$$

We then set $q_{n+1} = q_n + h\dot q_n$ and $u_{n+1} = \varPhi(q_{n+1})$. Inserting $\dot q_n$ from the first equation into the second equation, we get a nonlinear equation for λ_{n+1}: with $\widetilde q_{n+1} = q_n + hM_\varepsilon(q_n)^{-1}A(q_n)^\top f(u_n)$ (which is the result of the unconstrained Euler method) we have

$$g\big(\varPhi(\widetilde q_{n+1} - hM_\varepsilon(q_n)^{-1}C(q_n)^\top \lambda_{n+1})\big) - g(u_n) = 0.$$

A modified Newton method applied to this equation determines the $(k+1)$st iterate $\lambda_{n+1}^{(k+1)} = \lambda_{n+1}^{(k)} + \varDelta\lambda_{n+1}^{(k)}$ by solving the linear system

$$\begin{aligned} &- hC(q_n)M_\varepsilon(q_n)^{-1}C(q_n)^\top \varDelta\lambda_{n+1}^{(k)} \\ &\qquad = -g\big(\varPhi(\widetilde q_{n+1} - hM_\varepsilon(q_n)^{-1}C(q_n)^\top \lambda_{n+1}^{(k)})\big) + g(u_n). \end{aligned} \tag{7.10}$$

The starting value is chosen as $\lambda_{n+1}^{(0)} = 0$. The matrix $CM_\varepsilon^{-1}C^\top = GP_\varepsilon G^\top$ is the same as in (7.4). It is assumed to be invertible, see the bound (7.5) of the inverse.

Lemma 7.1 *If the matrix $G(u_n)P_\varepsilon(q_n)G(u_n)^\top$ has a moderately bounded inverse, then the modified Newton iteration* (7.10) *with starting value $\lambda_{n+1}^{(0)} = 0$ converges under the stepsize restriction*

$$h(\delta_n + h) \le c\varepsilon$$

with a sufficiently small c that is independent of h, ε and δ. Moreover,

$$\lambda_{n+1} = O(\delta_n + h).$$

Proof The modified Newton iteration is a fixed-point iteration for the map (omitting the argument u_n of G and q_n of $A, M_\varepsilon, P_\varepsilon$ and letting $z = h\lambda$)

$$\varphi(z) = z - (GP_\varepsilon G^\top)^{-1}\Big(g\big(\Phi(\widetilde{q}_{n+1} - M_\varepsilon^{-1}A^\top G^\top z\big) - g(u_n)\Big).$$

Using the $O(h(\delta_n + h))$ local error bound of the regularized Euler method as given in Lemma 4.1 and the conservation of g by the exact flow from t_n to t_{n+1}, we obtain for $\widetilde{u}_{n+1} = \Phi(\widetilde{q}_{n+1})$ that $g(\widetilde{u}_{n+1}) - g(u_n) = O(h(\delta_n + h))$ and hence $z^{(1)} = O(h(\delta_n + h))$ for the starting value $z^{(0)} = 0$. Using that $\|M_\varepsilon^{-1}A^\top\| \le 1/(2\varepsilon)$, we find that in a ball of radius $O(h(\delta_n + h))$ centered at 0 we have

$$\varphi'(z) = O\big(h(\delta_n + h)/\varepsilon\big),$$

which is strictly smaller than 1 under the given stepsize restriction. The stated result then follows with the Banach fixed-point theorem. □

7.4 Error analysis

The local error has a bound similar to Lemma 4.1 with the only difference that the constants now also depend on a bound of the inverse of the matrix in (7.4) and on bounds of derivatives of g. Note that the following local error bound is in terms of the defect size δ_0 of the *unconstrained* regularized Euler method, as in Section 4, not just of the larger $\widehat{\delta}_0$ of the constrained method (7.8).

Lemma 7.2 *Assume that the matrix $G(u_0)P_\varepsilon(q_0)G(u_0)^\top$ (with $\varepsilon = \varepsilon_0$) has an inverse bounded by $\big\|\big(G(u_0)P_\varepsilon(q_0)G(u_0)^\top\big)^{-1}\big\| \cdot \|G(u_0)\|^2 \le \gamma$. Under the stepsize restriction (cf.* (4.3)*)*

$$h\widehat{\delta}_0 \le c\varepsilon_0^2 \tag{7.11}$$

with a sufficiently small c (inversely proportional to γ), we have

$$\widehat{\delta}_0 \le \widehat{c}\,\delta_0,$$

where $\widehat{c}$ is proportional to γ but independent of h and ε_0. The local error of the regularized Euler method starting from $y(t_0) = u_0 = \Phi(q_0)$ is then bounded by

$$\|u_1 - y(t_1)\|_{\mathcal{H}} \le \widehat{c}_1 h\delta_0 + \widehat{c}_2 h^2, \tag{7.12}$$

where $\widehat{c}_2$ equals c_2 of Lemma 4.1 and $\widehat{c}_1$ is proportional to γ, depends on a bound of the second derivative of Φ in a neighbourhood of q_0 and on a bound of the second derivative of g in a neighbourhood of u_0.

Proof As in the proof of Lemma 4.1, we write $y(t_1) - y(t_0) = hf(u_0) + O(h^2)$, and we have again

$$u_1 - u_0 = A(q_0)h\dot{q}_0 + O(h^2\|\dot{q}_0\|_{\mathcal{Q}}^2),$$

and further

$$0 = g(u_1) - g(u_0) = G(u_0)(u_1 - u_0) + O(h^2\|\dot{q}_0\|_{\mathcal{Q}}^2).$$

From the first equation of (7.9) we have

$$A(q_0)\dot{q}_0 + P_\varepsilon(q_0)G(u_0)^\top\lambda_1 = f(u_0) + d_0,$$

where $d_0 = -(I - P_\varepsilon(q_0))f(u_0)$ is the defect of the *unconstrained* regularized Euler method, which is bounded by δ_0. From the constraint, using $G(u_0)f(u_0) = 0$, we thus find

$$\begin{aligned} 0 = g(u_1) - g(u_0) &= G(u_0)A(q_0)h\dot{q}_0 + O(h^2\|\dot{q}_0\|_{\mathcal{Q}}^2) \\ &= -hG(u_0)P_\varepsilon(q_0)G(u_0)^\top\lambda_1 + hG(u_0)d_0 + O(h^2\|\dot{q}_0\|_{\mathcal{Q}}^2), \end{aligned}$$

which yields $\lambda_1 = O(\delta_0) + O(h\|\dot{q}_0\|_{\mathcal{Q}}^2)$ and, with the projection $\varPi(q)$ appearing in (7.6),

$$u_1 - u_0 = hf(u_0) + h\varPi(q_0)d_0 + O(h^2\|\dot{q}_0\|_{\mathcal{Q}}^2).$$

Hence we obtain the error bound

$$u_1 - y(t_1) = O(h\delta_0) + O(h^2\|\dot{q}_0\|_{\mathcal{Q}}^2) + O(h^2). \tag{7.13}$$

It remains to show that the second term on the right-hand side is also $O(h\delta_0)$ under the stepsize restriction (7.11). So far we only know from (7.8) that $\varepsilon\|\dot{q}_0\|_{\mathcal{Q}} \le \widehat{\delta}_0$ and $\delta_0 \le \widehat{\delta}_0$. From the first equation of (7.9) we obtain

$$\varepsilon\dot{q}_0 + \varepsilon M_\varepsilon(q_0)^{-1}A(q_0)^\top G(u_0)^\top\lambda_1 = \varepsilon M_\varepsilon(q_0)^{-1}A(q_0)^\top f(u_0).$$

We estimate

$$\|\varepsilon M_\varepsilon(q_0)^{-1}A(q_0)^\top G(u_0)^\top\lambda_1\|_{\mathcal{Q}} \le \|G(u_0)^\top\lambda_1\|_{\mathcal{H}} = O(\delta_0) + O(h\|\dot{q}_0\|_{\mathcal{Q}}^2).$$

We write $f(u_0) = A(q_0)\dot{q}_0^{\mathrm{uncon}} - d_0$, where $\dot{q}_0^{\mathrm{uncon}}$ is the derivative approximation in the unconstrained regularized Euler method, which is bounded by $\varepsilon\|\dot{q}_0^{\mathrm{uncon}}\|_{\mathcal{Q}} \le \delta_0$. Using further that $\|M_\varepsilon(q_0)^{-1}A(q_0)^\top A(q_0)\| \le 1$, this yields the bound

$$\varepsilon\|\dot{q}_0\|_{\mathcal{Q}} \le \|G(u_0)^\top\lambda_1\|_{\mathcal{H}} + \varepsilon\|\dot{q}_0^{\mathrm{uncon}}\|_{\mathcal{Q}} + \|d_0\|_{\mathcal{H}} = O(\delta_0) + O(h\|\dot{q}_0\|_{\mathcal{Q}}^2).$$

Under the stepsize restriction (7.11) we have

$$h\|\dot{q}_0\|_{\mathcal{Q}}^2 \le h(\widehat{\delta}_0/\varepsilon)^2 \le c\widehat{\delta}_0 \quad \text{with a small factor } c,$$

so that

$$\varepsilon\|\dot{q}_0\|_{\mathcal{Q}} \le C\delta_0 + \tfrac{1}{4}\widehat{\delta}_0.$$

We further note that the obtained bound on λ_1 implies

$$A(q_0)\dot{q}_0 - f(u_0) = -P_\varepsilon(q_0)G(u_0)^\top\lambda_1 + d_0 = O(\delta_0) + O(h\|\dot{q}_0\|_{\mathcal{Q}}^2),$$

so that also

$$\|A(q_0)\dot{q}_0 - f(u_0)\|_{\mathcal{H}} \le C\delta_0 + \tfrac{1}{4}\widehat{\delta}_0.$$

Together with the estimate for $\varepsilon\dot{q}_0$ this shows that

$$\widehat{\delta}_0^2 = \|A(q_0)\dot{q}_0 - f(u_0)\|_{\mathcal{H}}^2 + \varepsilon^2\|\dot{q}_0\|_{\mathcal{Q}}^2 \le C'\delta_0^2 + \tfrac{1}{2}\widehat{\delta}_0^2,$$

so that

$$\widehat{\delta}_0 = O(\delta_0).$$

Tracing the constants in the O-notation yields the stated result. □

From this local error bound we again obtain a global $O(h+\delta)$ error bound as in Proposition 4.1, using Lady Windermere's fan with error propagation by the exact flow.

8 Case study: the Schrödinger equation

The time-dependent Schrödinger equation is arguably the evolution equation for which nonlinear approximations have been first and most often used, ever since Dirac's paper of 1930 [6]. Gaussians and tensor networks are nowadays the most prominent examples of nonlinear approximations in quantum dynamics. As a partial differential equation with an unbounded operator, the Schrödinger equation does not fall into the Lipschitz framework considered so far. In this section we study what remains and what needs to be changed in the regularized approach.

8.1 Preparation

The Schrödinger equation determines the complex-valued wave function $\psi(x,t)$ that depends on spatial variables $x \in \mathbb{R}^d$ and time t:

$$\mathrm{i}\dot{\psi} = -\Delta\psi + V\psi, \tag{8.1}$$

where i is the imaginary unit, $\dot{\psi} = \partial_t\psi$ is the time derivative, Δ is the Laplacian on $\mathbb{R}^d$ and $V = V(x)$ is a real-valued potential that multiplies the wave function. The Schrödinger equation is considered as an evolution equation on the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^d)$ for the wave function $\psi(t) = \psi(\cdot,t) \in \mathcal{H}$.

Consider a continuously differentiable map Φ from a parameter space $\mathcal{Q}$ into $\mathcal{H}$. We aim to approximate

$$\psi(t) \approx u(t) = \Phi(q(t)) \in \mathcal{H} \quad \text{for some} \quad q(t) \in \mathcal{Q}$$

by the regularized dynamical nonlinear approximation (2.1) with the linear operator $f(u) = \mathrm{i}\Delta u - \mathrm{i}Vu$. Since the Laplacian is an unbounded operator,

the Lipschitz framework of the previous section does not apply here. However, we still have the one-sided Lipschitz condition with $\ell = 0$ and from this we obtain the *a posteriori* error bound of Proposition 3.1, with the same proof.

To obtain an *a priori* error bound, we need to modify the construction of the regularized approximation $u = \Phi(q)$. We will use the property that the Laplacian maps into the tangent space:

$$\text{If } u = \Phi(q), \text{ then } \Delta u = \Phi'(q)q^{\Delta} \text{ for some } q^{\Delta} \in \mathcal{Q}. \tag{8.2}$$

This holds true for Gaussians and tensor networks but not for neural networks. We assume (8.2) throughout this section.

8.2 A modified regularized dynamical approximation

Instead of (2.1), we now choose (omitting the argument t) $\dot{u} = \Phi'(q)\dot{q}$ and $\dot{q} \in \mathcal{Q}$ such that

$$\dot{u} - \mathrm{i}\Delta u = v \quad \text{and} \quad \dot{q} - \mathrm{i}q^{\Delta} = p, \tag{8.3}$$

where $v = \Phi'(q)p$ with $p \in \mathcal{Q}$ is chosen such that

$$\delta^2 := \|v + \mathrm{i}Vu\|_{\mathcal{H}}^2 + \varepsilon^2\|p\|_{\mathcal{Q}}^2 \quad \text{is minimal.} \tag{8.4}$$

Inserting (8.3) into (8.4) and comparing with (2.1), we find that the term $\varepsilon^2\|\dot{q}\|^2$ in (2.1) is now replaced by $\varepsilon^2\|\dot{q} - \mathrm{i}q^{\Delta}\|^2$, everything else being equal. In contrast to (2.1), the free Schrödinger equation (i.e., with $V = 0$) is solved exactly with (8.3)–(8.4) for every $\varepsilon > 0$. Indeed, $p = 0$ provides $\dot{q} = iq^{\Delta}$ and $\dot{u} = i\Delta u$. We have

$$\partial_t(u - \psi) = \mathrm{i}\Delta(u - \psi) - \mathrm{i}V(u - \psi) + d \quad \text{with} \quad \|d\|_{\mathcal{H}} \le \delta,$$

and as in the proof of Proposition 3.1 (with $\ell = 0$), we obtain the *a posteriori* error bound

$$\|u(t) - \psi(t)\|_{\mathcal{H}} \le \int_0^t \delta(s)\, ds. \tag{8.5}$$

Remark 8.1 Multiplying a sum of complex Gaussians $u = \Phi(q) = \sum_j \varphi(z_j)$ by a subquadratic potential V, we have

$$Vu = \sum_j V\varphi(z_j) = \sum_j (U_j + W_j)\varphi(z_j),$$

where U_j denotes the second order Taylor polynomial of V centered around the position center of the jth Gaussian and W_j the cubic remainder. Therefore,

$$Vu = \Phi'(q)q^U + \chi(q) \text{ for some } q^U \in \mathcal{Q}$$

and $\chi(q) = \sum_j W_j\varphi(z_j)$ with $\|\chi(q)\|_{\mathcal{H}} \le \beta_3\|(1 + |x|^2)u\|_{\mathcal{H}}$ for some constant $\beta_3 > 0$ depending on third order derivative bounds of V. Working with $q^{\Delta} + q^U$ instead of q^{Δ} extends the above approximation and makes it exact for harmonic oscillators.

8.3 A priori error bound

We can bound the defect size $\delta(t)$ by a quantity that measures the uniform approximability of $\dot\psi - \mathrm{i}\Delta\psi$ in the tangent spaces $T_{(u,q)}\mathcal{M}$ for all $u = \Phi(q)$ in a neighbourhood of $\psi(t)$. We fix a radius $\rho > 0$ and define

$$\bar\delta_\rho(t)^2 := \sup_{q\in\mathcal{Q}:\|\Phi(q)-\psi(t)\|_{\mathcal{H}}\le\rho} \min_{\dot q\in\mathcal{Q}} \Big(\|\Phi'(q)\dot q - (\dot\psi - \mathrm{i}\Delta\psi)\|_{\mathcal{H}}^2 + \varepsilon^2\|\dot q\|_{\mathcal{Q}}^2 \Big). \quad (8.6)$$

We can bound the defect size $\delta(t)$ of (8.4) in terms of $\bar\delta_\rho(t)$.

Lemma 8.1 *Provided that $\|u(t) - \psi(t)\|_{\mathcal{H}} \le \rho$, we have*

$$\delta(t) \le \bar\delta_\rho(t) + \|Vu(t) - V\psi(t))\|_{\mathcal{H}}.$$

Proof The proof is similar to that of Lemma 3.1. We omit the argument t in the following. We have $u = \Phi(q)$ and $\dot u = \Phi'(q)\dot q$. Let $p_+ \in \mathcal{Q}$ be such that

$$\|\Phi'(q)p_+ - (\dot\psi - \mathrm{i}\Delta\psi)\|_{\mathcal{H}}^2 + \varepsilon^2\|p_+\|_{\mathcal{Q}}^2 \quad \text{is minimal.}$$

By (8.1)–(8.4),

$$\begin{aligned}
\delta^2 &= \|\Phi'(q)p + \mathrm{i}Vu\|_{\mathcal{H}}^2 + \varepsilon^2\|p\|_{\mathcal{Q}}^2 \\
&\le \|\Phi'(q)p_+ + \mathrm{i}Vu\|_{\mathcal{H}}^2 + \varepsilon^2\|p_+\|_{\mathcal{Q}}^2 \\
&\le \Big(\|\Phi'(q)p_+ - (\dot\psi - \mathrm{i}\Delta\psi)\|_{\mathcal{H}} + \|-\mathrm{i}V\psi + \mathrm{i}Vu\|_{\mathcal{H}} \Big)^2 + \varepsilon^2\|p_+\|_{\mathcal{Q}}^2 \\
&\le \Big(\bar\delta_\rho + \|Vu - V\psi\|_{\mathcal{H}} \Big)^2.
\end{aligned}$$

Taking square roots yields the result. □

We construct a reference approximation $u_* = \Phi(q_*)$ from the exact wave function ψ by choosing $\dot u_* = \Phi'(q_*)\dot q_*$ and $\dot q_* \in \mathcal{Q}$ such that

$$\dot u_* - \mathrm{i}\Delta u_* = v_* \quad \text{and} \quad \dot q_* - \mathrm{i}q_*^\Delta = p_*, \quad (8.7)$$

where $v_* = \Phi'(q_*)p_*$ with $p_* \in \mathcal{Q}$ is chosen as a regularized best approximation to $\dot\psi - \mathrm{i}\Delta\psi$ in the tangent space:

$$\delta_*^2 := \|v_* - (\dot\psi - \mathrm{i}\Delta\psi)\|_{\mathcal{H}}^2 + \varepsilon^2\|p_*\|_{\mathcal{Q}}^2 \quad \text{is minimal.} \quad (8.8)$$

With (8.8) we have

$$\partial_t(u_* - \psi) = \mathrm{i}\Delta(u_* - \psi) + d_* \quad \text{with} \quad \|d_*\|_{\mathcal{H}} \le \delta_*,$$

and as before it follows that

$$\|u_*(t) - \psi(t)\|_{\mathcal{H}} \le \int_0^t \delta_*(s)\,ds \le \int_0^t \bar\delta_\rho(s)\,ds \quad (8.9)$$

as long as this is bounded by ρ. The error of the numerical approximation $u(t) = \Phi(q(t))$ is bounded by a multiple of the bound in (8.9).

Proposition 8.1 *If condition* (8.2) *is satisfied and* $\sup_x |V(x)| \le \nu$*, then the error of* $u(t)$ *defined by* (8.3)–(8.4) *is bounded by*

$$\|u(t) - \psi(t)\|_{\mathcal{H}} \le e^{\nu t} \int_0^t \bar{\delta}_\rho(s)\, ds$$

as long as this is bounded by ρ*.*

Proof The bound follows from Lemma 8.1 inserted into (8.5) and using the Gronwall lemma. □

8.4 Energy conservation

With the Hamiltonian $H = -\Delta + V$, which is a self-adjoint linear operator on $\mathcal{H} = L^2(\mathbb{R}^d)$ with domain $D(H) = H^2(\mathbb{R}^d)$, the total energy is defined as $\langle u, Hu\rangle_{\mathcal{H}}$ for $u \in D(H)$. The energy is conserved along the exact wave function $\psi(t) \in D(H)$ of (8.1), since (omitting in the following the argument t and the subscript $\mathcal{H}$ in the inner product)

$$\frac{d}{dt}\langle \psi, H\psi\rangle = 2\,\mathrm{Re}\,\langle H\psi, \dot{\psi}\rangle = 2\,\mathrm{Re}\,\langle H\psi, -\mathrm{i}H\psi\rangle = 0.$$

We show that the energy is also conserved along the regularized approximation (2.1). We adapt the notation of Section (7) to the complex setting and let $P_\varepsilon(q) = A(q)M_\varepsilon(q)^{-1}A(q)^*$ with $A = \Phi'$ and $M_\varepsilon = A^*A + \varepsilon^2 I$. We note that (2.1) yields $\mathrm{i}\dot{u} = P_\varepsilon(q)Hu$. Since both H and $P_\varepsilon(q)$ are self-adjoint, we have (omitting the arguments $q(t)$ or t)

$$\frac{d}{dt}\langle u, Hu\rangle = 2\,\mathrm{Re}\,\langle Hu, \dot{u}\rangle = 2\,\mathrm{Re}\,\langle Hu, -\mathrm{i}P_\varepsilon Hu\rangle = 0.$$

However, the modified approximation u defined by (8.3)–(8.4) satisfies the differential equation $\mathrm{i}\dot{u} = -\Delta u + P_\varepsilon V u$ and thus

$$\begin{aligned}\frac{d}{dt}\langle u, Hu\rangle &= 2\,\mathrm{Re}\,\langle Hu, \dot{u}\rangle = 2\,\mathrm{Re}\,\langle -\Delta u + Vu, \mathrm{i}\Delta u - \mathrm{i}P_\varepsilon V u\rangle \\ &= 2\,\mathrm{Re}\,\langle \Delta u, \mathrm{i}P_\varepsilon V u\rangle + 2\,\mathrm{Re}\,\langle Vu, \mathrm{i}\Delta u\rangle \\ &= 2\,\mathrm{Re}\,\langle \Delta u, \mathrm{i}(I - P_\varepsilon)Vu\rangle,\end{aligned}$$

which in general is nonzero. Moreover, the norm $\|u(t)\|_{\mathcal{H}}$ is not preserved by either regularized approximation. Norm and energy conservation can be enforced simultaneously as described in Section 7.

9 Numerical experiments

We close the paper with a collection of numerical experiments, both for the approximation of flow maps for ODEs and the Schrödinger equation.

9.1 Approximating the flow map of a Lotka–Volterra model

As a simple nonlinear initial value problem, we consider the classical predator–prey model

$$\begin{aligned} \dot{x} &= \alpha x - \beta xy. \\ \dot{y} &= \delta xy - \gamma y, \end{aligned} \tag{9.1}$$

where $\alpha, \beta, \gamma, \delta$ are positive constants, which are all set to 1 in our experiments. The region of interest in our experiments is the square $D = \left[\frac{1}{2}, \frac{5}{2}\right]^2$.

The flow map $\varphi_t : D \to \mathbb{R}^2$ at time t, which to every initial value $(x_0, y_0) \in D$ associates the corresponding solution value $(x(t), y(t))$ of (9.1), is considered as an element of the Hilbert space $\mathcal{H} = L^2(D)^2$. It satisfies the differential equation on $\mathcal{H}$

$$\frac{d}{dt}\varphi_t = f(\varphi_t), \qquad \varphi_0 = \mathrm{Id},$$

where f is given by the right-hand side of (9.1) and Id is the identity on $\mathcal{H}$.

We use the Tensorflow library to approximate the flow map φ_t by a small feedforward neural network with three fully connected hidden layers, each with a depth of four neurons. Overall, the network architecture requires only 62 parameters. As the activation function on each layer, we use the sigmoid function

$$\sigma(x) = \frac{e^x}{1 + e^x}.$$

The metric installed on the weight space $\mathcal{Q}$ is the standard euclidian norm.

As the initial nonlinear parametrization at $t = 0$, we require a network that approximates the identity on D with a high accuracy. For our experiments, this was realized by pretraining an initial approximation $u_0 \colon D \to \mathbb{R}^2$ with a standard optimizer and then applying the regularized procedure (2.1) to the initial value-problem

$$\dot{u}(\cdot, t) = \mathrm{Id} - u_0$$

with time-independent right-hand side. At $t = 1$, we then obtain an approximation $u(\cdot, 1) \approx \mathrm{Id}$. For the presented experiments, we used the classical Runge–Kutta method of order 4 with the regularization parameter $\varepsilon = 10^{-6}$ and $N = 2000$ time steps.

Remark 9.1 (Connection to the Gauß–Newton method) We note that the approach above can be used to construct neural networks (or any nonlinear parameterization) approximating any given arbitrary function g, instead of the identity Id. The implementation of such methods generalizes the Gauß–Newton method applied to $\|\Phi(q) - g\| = \min$, which would be obtained by applying the Euler method (with $\varepsilon = 0$ and $h = 1$) to the differential equation that results from the (non-regularized) least squares problem $\|\Phi'(q)\dot{q} - (g - \Phi(q))\| = \min$.

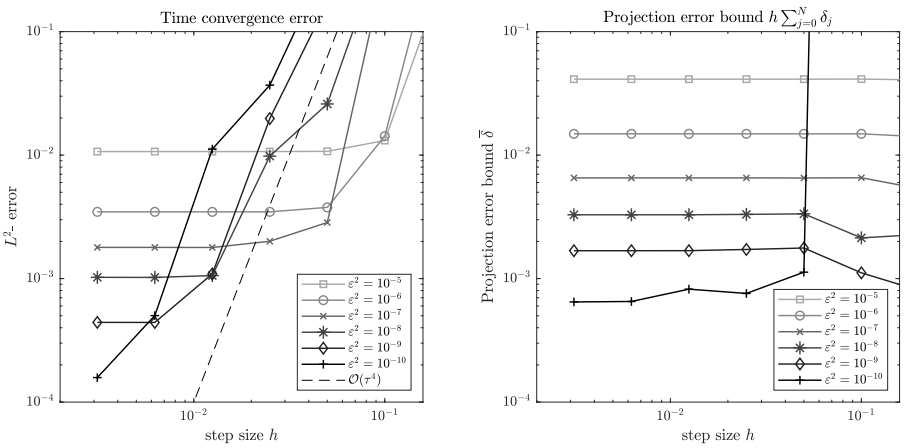

Fig. 9.1 Time convergence plot for the Lotka–Volterra system, computed with a fixed neural network architecture with three hidden layers and four neurons each, which is fully described by $q \in \mathbb{R}^{62}$. We fix the regularization parameter ε and observe the error behaviour of the classical Runge–Kutta approximation to the regularized flow (2.1). On the right-hand side, we plot the projection error term of the error bound described in Proposition 5.1.

The assembly of the Jacobian $\Phi'(q_{n,i})$, for the given parameters at the internal stages $q_{n,i}$ of the Runge–Kutta method, is efficiently realized by the automatic differentiation routines provided by the Tensorflow framework. For the numerical quadrature, we choose a composite Gaussian quadrature with 4 nodes on each subinterval and 10 subintervals in each direction.

We now apply the classical Runge–Kutta method of order 4 and observe the error behaviour for varying step size h and regularization parameter ε.

In Figure 9.1, we fix several values of the regularization parameter ε and vary the time step size to observe the time convergence behaviour. The $\mathcal{H}$-norm (i.e. the L^2-norm) on D is the natural error measure, which is taken at the fixed time $t = 1$. As predicted by the theory, we observe a step size restriction depending on the parameter ε. For smaller values of ε, we require a smaller time step size h in order to achieve convergence. The observed time step restriction is, however, milder than the restriction in Proposition 5.1. On the right-hand side, we visualize the a posteriori bounds for the projections. As expected, these bounds are quite stable with respect to the time step size and estimate the possible accuracy for a fixed parameter ε and the underlying nonlinear approximation.

In Figure 9.2, we conversely fix the number of time steps and observe the error and the projection errors for a varying regularization parameter ε. Overall, we observe a convergence of the order of $\mathcal{O}(\varepsilon)$, when the number of time steps is sufficiently large. When the number of time steps is not sufficiently large, we again observe the effect of the time step restriction. The a posteriori terms of the error bound capture the effects of the regularization parameter ε, but are by construction almost invariant with respect to the time discretization.

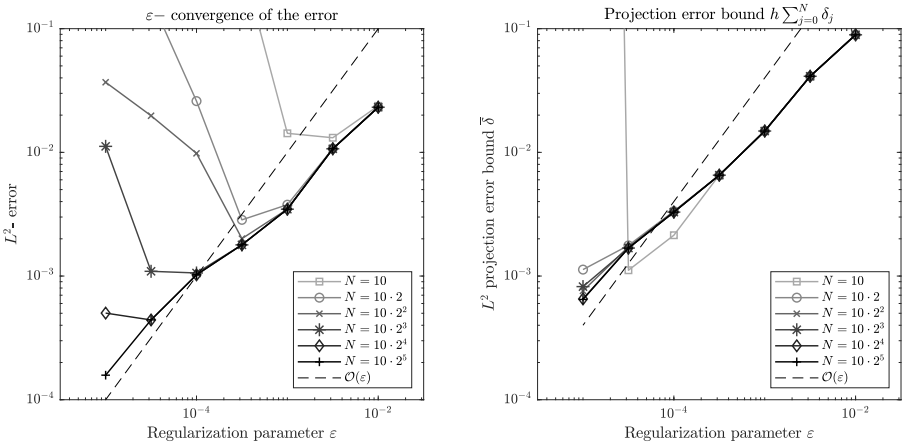

Fig. 9.2 The $\varepsilon-$ convergence of the same network architecture, with the same time discretization. We fix the number of time steps and vary the regularization parameter.

9.2 Approximating double-well quantum dynamics

We consider a one-dimensional Schrödinger equation $\mathrm{i}\dot{\psi} = H\psi$ formulated within the setting of the complex Hilbert space $\mathcal{H} = L^2(\mathbb{R}, \mathbb{C})$. The equation serves as a model for tunneling dynamics. The Schrödinger operator

$$H = -\frac{1}{2}\partial_x^2 + \alpha_2 x^2 + \alpha_4 x^4, \quad \psi_0(x) = \pi^{-1/4}\mathrm{e}^{-(x-q_\ell)^2/2},$$

contains a quartic double-well potential with polynomial parameters $\alpha_2 = -\frac{1}{8}$ and $\alpha_4 = \alpha_2^2$. The initial condition ψ_0 is a single normalized Gaussian, whose width stems from the standard harmonic oscillator $H_0 = -\frac{1}{2}\partial_x^2 + \frac{1}{2}x^2$, placed at the left minimum $q_\ell = -2$ of the double well potential. During the time interval $[0, T] = [0, 12]$ the wave packet travels from the left to the right well, see also [12]. The approximation ansatz is a frozen sum of $M = 36$ Gaussians,

$$u(t, x) = \sum_{m=1}^{M} c_m(t)\mathrm{e}^{-x^2/2-\kappa_m(t)x},$$

with $2M$ complex parameters $(c_m(t), \kappa_m(t)) \in \mathcal{Q} = \mathbb{C}^2$. For the initialization $u_0 \in \mathcal{M}$, we put a non-uniform grid of Gauß–Hermite quadrature nodes (x_i, ξ_j) with origin at $(q_\ell, 0)$ on $\mathbb{R}^2$, reformulate the corresponding Gaussian wave packets $\mathrm{e}^{-(x-x_i)^2/2+\mathrm{i}\xi_j(x-x_i)}$ in the complex algebraic format $\mathrm{e}^{-x^2/2-\kappa_m(0)x}$, and determine the optimal linear expansion coefficients $c_m(0)$ by solving the linear least squares problem

$$\|u(0) - \psi_0\|_{\mathcal{H}} = \min!$$

The matrices for the initial minimization and the ones involving the parameter Jacobian $\Phi'(q)$ are evaluated via analytical formulas for Gaussian integrals of

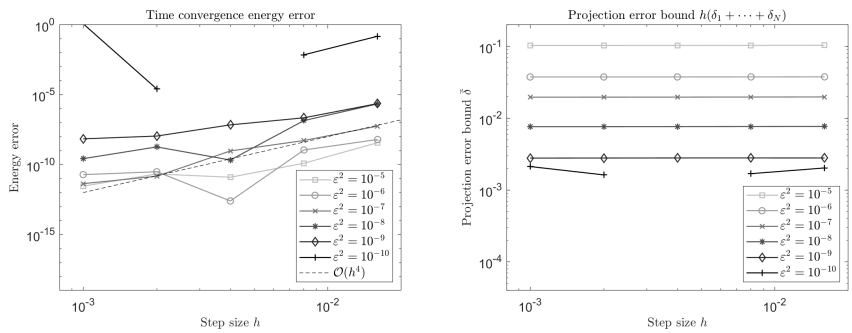

Fig. 9.3 Time convergence plot for the double-well system, computed with a sum of $M = 36$ complex Gaussians, which is fully described by $q \in \mathbb{C}^{72}$. We fix the regularization parameter ε and observe the error behaviour of the classical Runge–Kutta approximation to the regularized flow (2.1). On the left-hand side, we plot the energy error $|\mathcal{E}(u(T)) - \mathcal{E}(u(0))|$, on the right-hand side the projection error term of the error bound described in Proposition 5.1.

the type

$$\int_{\mathbb{R}} x^k \mathrm{e}^{-\beta x^2 - \lambda x}, \quad \beta > 0, \quad \lambda \in \mathbb{C}.$$

The time integrator is the classical Runge–Kutta scheme of order four. As before, we observe the error behaviour for varying step size h and regularization parameter ε.

In Figure 9.3, we fix several values of the regularization parameter ε and vary the time step size h to observe the time convergence behaviour. Since energy $\mathcal{E}(\psi(t)) = \langle \psi(t), H\psi(t)\rangle_{\mathcal{H}}$, $t \in \mathbb{R}$, is a conserved quantity of the Schrödinger evolution, we evaluate the approximate energy $\mathcal{E}(u(t))$ at the final time $t = T$ and compare with the value at initial time $t = 0$, see the left-hand side. We observe, that the simulations with the smallest regularization parameter $\varepsilon = 10^{-5}$ have large errors and even prematurely terminate for the particular step size $h = 4 \cdot 10^{-3}$. For the other choices of the regularization parameter, the errors follow the order of the time integrator without the predicted step size restriction. On the right-hand side, we show the a posteriori error bounds for the projections. As before for the Lotka–Volterra model, these bounds are stable with respect to the time step size and estimate the accuracy of the underlying approximation.

In Figure 9.4, we conversely fix the size of the time step h and present errors for a varying regularization parameter ε. The Schrödinger dynamics are unitary, and thus conserve the norm $\|\psi(t)\|_{\mathcal{H}}$, $t \in \mathbb{R}$, of the solution (mass conservation). On the left hand-side we compare the approximate norm $\|u(T)\|_{\mathcal{H}}$ at the final time T with the exact unit value. We observe decay of the mass error only for relatively large values of the regularization parameter. Depending on the time step size, less regularization results in larger errors. The right-hand side shows the a posteriori terms of the error bound. Again, they capture the effects of the regularization parameter ε, but are by construction insensitive with respect to the time discretization.

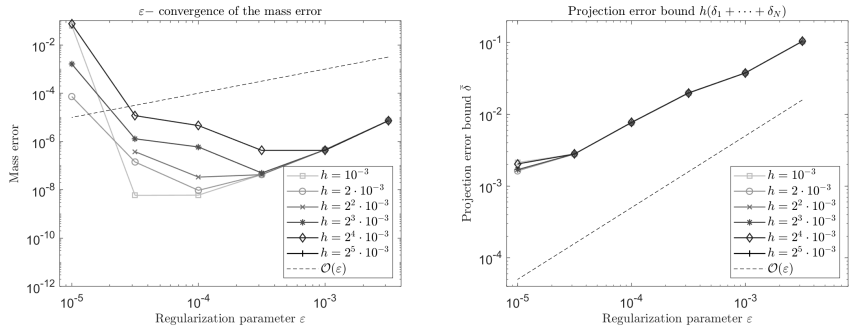

Fig. 9.4 The $\varepsilon-$convergence of the same sum of Gaussians, with the same time discretization. We fix the number of time steps and vary the regularization parameter. On the left-hand side, we plot the norm error $|\|u(T)\|_{\mathcal{H}}^2 - 1|$, on the right-hand side the projection error.

Acknowledgements This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under projects SFB 1173 – Project-ID 258734477 and TRR 352 – Project-ID 470903074 as well as the Austrian Science Fund (FWF) under the special research program Taming complexity in PDE systems (grant SFB F65).

References

1. G. Bai, U. Koley, S. Mishra, and R. Molinaro. Physics informed neural networks (PINNs) for approximating nonlinear dispersive PDEs. *J. Comput. Math.*, 39(6):816–847, 2021.
2. C. Beck, S. Becker, P. Grohs, N. Jaafari, and A. Jentzen. Solving the Kolmogorov PDE by means of deep learning. *J. Sci. Comput.*, 88(3):Paper No. 73, 28, 2021.
3. J. Bruna, B. Peherstorfer, and E. Vanden-Eijnden. Neural Galerkin schemes with active learning for high-dimensional evolution equations. *J. Comput. Phys.*, 496:Paper No. 112588, 22, 2024.
4. G. Carleo and M. Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.
5. G. Ceruti, C. Lubich, and H. Walach. Time integration of tree tensor networks. *SIAM J. Numer. Anal.*, 59(1):289–313, 2021.
6. P. A. Dirac. Note on exchange phenomena in the Thomas atom. *Proc. Cambridge Phil. Soc.*, 26(3):376–385, 1930.
7. Y. Du and T. A. Zaki. Evolutional deep neural network. *Phys. Rev. E*, 104(4):Paper No. 045303, 14, 2021.
8. J. Haegeman, C. Lubich, I. Oseledets, B. Vandereycken, and F. Verstraete. Unifying time evolution and optimization with matrix product states. *Phys. Rev. B*, 94(16):165116, 2016.
9. E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I. Nonstiff Problems.* Springer, Berlin, 2nd edition, 1993.
10. E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems.* Springer, Berlin, 2nd edition, 1996.

11. M. J. Hartmann and G. Carleo. Neural-network approach to dissipative quantum many-body dynamics. *Phys. Rev. Letters*, 122(25):250502, 2019.
12. L. Joubert-Doriol. Variational approach for linearly dependent moving bases in quantum dynamics: Application to gaussian functions. *Journal of Chemical Theory and Computation*, 18(10):5799–5809, 2022.
13. M. E. Kilmer and D. P. O'Leary. Choosing regularization parameters in iterative methods for ill-posed problems. *SIAM J. Matrix Anal. Appl.*, 22(4):1204–1221, 2001.
14. N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *J. Machine Learning Res.*, 24(89):1–97, 2023.
15. P. Kramer and M. Saraceno. *Geometry of the time-dependent variational principle in quantum mechanics*, volume 140 of *Lecture Notes in Physics*. Springer, Berlin, 1981.
16. G. Kutyniok, P. Petersen, M. Raslan, and R. Schneider. A theoretical analysis of deep neural networks and parametric PDEs. *Constr. Approx.*, 55:73–125, 2022.
17. C. Lubich. *From quantum to classical molecular dynamics: reduced models and numerical analysis*. European Mathematical Society, 2008.
18. C. Lubich, I. V. Oseledets, and B. Vandereycken. Time integration of tensor trains. *SIAM J. Numer. Anal.*, 53(2):917–941, 2015.
19. H.-D. Meyer, U. Manthe, and L. S. Cederbaum. The multi-configurational time-dependent Hartree approach. *Chem. Phys. Letters*, 165(1):73–78, 1990.
20. S. Mishra and R. Molinaro. Estimates on the generalization error of physics-informed neural networks for approximating PDEs. *IMA J. Numer. Anal.*, 43(1):1–43, 2023.
21. J. J. Moré. The Levenberg–Marquardt algorithm: implementation and theory. In *Numerical analysis: proceedings of the biennial Conference held at Dundee, June 28–July 1, 1977*, pages 105–116. Springer, 2006.
22. J. A. A. Opschoor, P. C. Petersen, and C. Schwab. Deep ReLU networks and high-order finite element methods. *Anal. Appl. (Singap.)*, 18(5):715–770, 2020.
23. G. W. Richings, I. Polyak, K. E. Spinlove, G. A. Worth, I. Burghardt, and B. Lasorne. Quantum dynamics simulations using Gaussian wavepackets: the vMCG method. *Int. Rev. Phys. Chem.*, 34(2):269–308, 2015.
24. M. Schmitt and M. Heyl. Quantum many-body dynamics in two dimensions with artificial neural networks. *Phys. Rev. Letters*, 125(10):100503, 2020.
25. Y.-Y. Shi, L.-M. Duan, and G. Vidal. Classical simulation of quantum many-body systems with a tree tensor network. *Phys. Rev. A*, 74(2):022320, 2006.
26. A. Townsend and L. N. Trefethen. Continuous analogues of matrix factorizations. *Proc. Royal Soc. A: Math. Phys. Eng. Sci.*, 471(2173):20140585, 2015.

27. H. Wang and M. Thoss. Multilayer formulation of the multiconfiguration time-dependent Hartree theory. *J. Chem. Phys.*, 119(3):1289–1299, 2003.
28. G. Worth, M. Robb, and I. Burghardt. A novel algorithm for non-adiabatic direct dynamics using variational Gaussian wavepackets. *Faraday discussions*, 127:307–323, 2004.
29. B. Yu and W. E. The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems. *Comm. Math. Stat.*, 6(1):1–12, 2018.
30. Y. Zang, G. Bao, X. Ye, and H. Zhou. Weak adversarial networks for high-dimensional partial differential equations. *J. Comput. Phys.*, 411:109409, 14, 2020.
31. Y. Zhu, N. Zabaras, P.-S. Koutsourelakis, and P. Perdikaris. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. *J. Comput. Phys.*, 394:56–81, 2019.

A Appendix: Note on the regularized least squares problem

To study the sensitivity with respect to the regularization parameter, we consider the linear least squares problem to find $x = x(\alpha)$ such that

$$\vartheta(\alpha) := \|Ax - b\|^2 + \alpha\|x\|^2 \quad \text{is minimal,}$$

where we take the Euclidean norms. Note that $\vartheta = \delta^2$ and $\alpha = \varepsilon^2$ in the setting of the paper. The dependence of ϑ on α is remarkably simple.

Lemma A.1 *We have*

$$\vartheta'(\alpha) = \|x(\alpha)\|^2.$$

Moreover,

$$\vartheta''(\alpha) = \frac{d}{d\alpha}\|x(\alpha)\|^2 \le 0 \quad \textit{and} \quad \frac{d}{d\alpha}\frac{\vartheta(\alpha)}{\alpha} \le 0.$$

Proof With $M = M(\alpha) = (A^\top A + \alpha I)$ we have the normal equations $Mx = A^\top b$ and hence

$$x = M^{-1}A^\top b.$$

Since $M' = I$, we have $Mx' + x = 0$ and hence

$$x' = -M^{-1}x.$$

We have

$$\vartheta' = 2\langle Ax - b, Ax'\rangle + 2\alpha\langle x, x'\rangle + \|x\|^2,$$

which becomes

$$\begin{aligned}\langle Ax - b, Ax'\rangle &= \langle Ax - b, -AM^{-1}x\rangle = -\langle x, A^\top AM^{-1}x\rangle + \langle M^{-1}A^\top b, x\rangle \\ &= -\langle x, A^\top AM^{-1}x\rangle + \langle x, x\rangle = \langle x, (I - A^\top AM^{-1})x\rangle = \langle x, \alpha M^{-1}x\rangle,\end{aligned}$$

and

$$\alpha\langle x, x'\rangle = \langle x, -\alpha M^{-1}x\rangle,$$

so that

$$\vartheta' = 2\langle x, \alpha M^{-1}x\rangle + 2\langle x, -\alpha M^{-1}x\rangle + \|x\|^2 = \|x\|^2,$$

which is the stated result for the first derivative. The second derivative is

$$\vartheta'' = \frac{d}{d\alpha}\,\|x(\alpha)\|^2 = 2\langle x, x'\rangle = -2\langle x, M^{-1}x\rangle,$$

which is negative (unless $x = 0$), since M is positive definite. Finally,

$$\left(\frac{\vartheta}{\alpha}\right)' = \frac{\alpha\vartheta' - \vartheta}{\alpha^2} = \frac{\alpha\|x\|^2 - \vartheta}{\alpha^2} = -\frac{\|Ax - b\|^2}{\alpha^2},$$

which is non-positive. □

The following matrix estimates are often used in the paper.

Lemma A.2 *Denote $M_\varepsilon = A^\top A + \varepsilon^2 I$. We have in the matrix 2-norm*

$$\|AM_\varepsilon^{-1}A^\top\| \le 1, \qquad \|AM_\varepsilon^{-1}\| \le \frac{1}{2\varepsilon}, \qquad \|M_\varepsilon^{-1}\| \le \frac{1}{\varepsilon^2}.$$

Proof The bounds follow from the singular value decomposition $A = U\Sigma V^\top$ with U a linear isometry, V unitary and Σ diagonal. Then,

$$AM^{-1}A^\top = U\Sigma\left(\Sigma^2 + \varepsilon^2 I\right)^{-1}\Sigma U^\top.$$

Since U is a linear isometry, we have

$$\|AM_\varepsilon^{-1}A^\top\| \le \|\Sigma\left(\Sigma^2 + \varepsilon^2 I\right)^{-1}\Sigma\| \le \sup_{\sigma\ge 0}\frac{\sigma^2}{\sigma^2 + \varepsilon^2} = 1.$$

Similarly, $AM^{-1} = U\Sigma(\Sigma^2 + \varepsilon^2 I)^{-1}U^\top$, so that

$$\|AM_\varepsilon^{-1}\| = \|\Sigma(\Sigma^2 + \varepsilon^2)^{-1}\| \le \sup_{\sigma\ge 0}\frac{\sigma}{\sigma^2 + \varepsilon^2} = \frac{1}{2\varepsilon},$$

and further

$$\|M_\varepsilon^{-1}\| = \|(\Sigma^2 + \varepsilon^2)^{-1}\| \le \sup_{\sigma\ge 0}\frac{1}{\sigma^2 + \varepsilon^2} = \frac{1}{\varepsilon^2},$$

as stated. □