Symmetrized local error estimators for time-reversible one-step methods in nonlinear evolution equations

Winfried Auzinger [a], Harald Hofstätter [b], Othmar Koch [b],*

[a] *Technische Universität Wien, Institut für Analysis und Scientific Computing, Wiedner Hauptstrasse 8–10/E101, A-1040 Wien, Austria*
[b] *Universität Wien, Institut für Mathematik, Oskar-Morgenstern-Platz 1, A-1090 Wien, Austria*

A R T I C L E   I N F O

A B S T R A C T

Prior work on computable defect-based local error estimators for (linear) time-reversible integrators is extended to nonlinear and nonautonomous evolution equations. We prove that the asymptotic results from the linear case (Auzinger and Koch, 2018) remain valid, i.e., the modified estimators yield an improved asymptotic order as the step size goes to zero. Typically, the computational effort is only slightly higher than for conventional defect-based estimators, and it may even be lower in some cases. We illustrate this by some examples and present numerical results for evolution equations of Schrödinger type, solved by either time-splitting or Magnus-type integrators. Finally, we demonstrate that adaptive time-stepping schemes can be successfully based on our local error estimators.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

We consider the extension of a defect-based estimator for the local error of self-adjoint time-stepping schemes of even order p, which was introduced in [1] for the linear time-independent case, to nonlinear evolution equations (we set $t_0 = 0$),

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t) = F(u(t)), \quad u(0) = u_0. \tag{1.1}$$

We define a symmetrized version of the defect to serve as the basis for the construction of a local error estimator in the nonlinear case, thus representing an extension of [1]. The error estimator is derived from a representation of the local error in terms of the symmetrized defect, based on a modified nonlinear variation-of-constant formula. Its deviation from the exact local error is one order in the step-size more precise than an analogous error estimator based on the classical defect, for the latter see for instance [2–5]. Our theoretical analysis is based on the assumption that the problem is smooth (the right-hand side is bounded and differentiable with bounded derivatives as required in the analysis) with a unique, smooth solution. In this sense, our treatment is formal and in practical applications with unbounded right-hand side, different techniques are required to deduce the required regularity assumptions in order to establish high-order convergence, see for instance [4].

---

* Corresponding author.
  E-mail addresses: w.auzinger@tuwien.ac.at (W. Auzinger), hofi@harald-hofstaetter.at (H. Hofstätter), othmar@othmar-koch.org (O. Koch).
  URLs: http://www.asc.tuwien.ac.at/~winfried (W. Auzinger), http://www.harald-hofstaetter.at (H. Hofstätter), http://www.othmar-koch.org (O. Koch).

We also point out that in addition to the practical merit of providing a more precise estimator enabling a better choice of adaptive time-steps and a higher-order corrected solution if desired, the approach has potential advantages for theoretical purposes.

In the analysis of local errors and error estimators for self-adjoint schemes, the representation of the local error in terms of the symmetrized defect can be rewritten in a way such that its analysis can be based on an asymptotic expansion in even powers of the stepsize. Applications of this type of analysis will be reported elsewhere.

Outline. In Section 2 we introduce the notions 'classical defect' and the new 'symmetrized defect' associated with one-step integrators for nonlinear evolution equations in the autonomous[1] form (1.1). A well-known integral representation of the local error in terms of the classical defect is obtained from the nonlinear variation-of-constant formula (V.O.C., also referred to as Gröbner–Alexeev-Lemma [6]), this is recapitulated in Theorem 2. Then, in Theorem 3 we present a modified nonlinear V.O.C. formula leading to an integral representation of the local error in terms of the symmetrized defect.

A Hermite-type quadrature approximation to the ensuing integral representation provides a computable defect-based local error estimator, see Section 3. In particular, Theorem 4 shows that the symmetrized error estimator is asymptotically correct, and for the case of a self-adjoint scheme it is of an improved asymptotic quality compared with the analogous classical estimator. Here the required regularity of the problem data and of the exact solution is tacitly assumed.

In Sections 4 and 5 we study the application of these ideas to particular examples of self-adjoint schemes. In Section 4.1, the results are particularized to the implicit midpoint rule to show a concrete example of an implicit one-step method. In Section 4.2, Strang splitting is discussed, and the algorithmic realization for general splitting methods is given in Section 4.3.

In Section 5, the nonautonomous case is considered. In order to illustrate the extension of our ideas to this case, we give details for linear problems with a t-dependent right-hand side. Section 5.1 shows the realization for the exponential midpoint rule, and Section 5.2 contains the algorithmic implementation for general commutator-free Magnus-type and classical Magnus methods.

In Section 6, numerical examples for a splitting approximation to a cubic nonlinear Schrödinger equation and Magnus-type exponential integrators applied to a time-dependent Rosen–Zener model support the theoretical results, and adaptive time-stepping based on the new error estimator is illustrated.

Notation and preliminaries. The flow associated with (1.1) is denoted by $\mathcal{E}(t, u)$, such that the solution of (1.1) is $u(t) = \mathcal{E}(t, u_0)$. By $\partial_1 \mathcal{E}(t, u_0)$ and $\partial_2 \mathcal{E}(t, u_0)$ we denote the derivatives of $\mathcal{E}$ with respect to its first and second arguments, respectively. By definition, $\mathcal{E}(t, u_0)$ satisfies

$$\partial_1 \mathcal{E}(t, u_0) = F(\mathcal{E}(t, u_0)), \quad \mathcal{E}(0, u_0) = u_0.$$

We will repeatedly make use of the following fundamental identity.[2]

Lemma 1.

$$[\,\partial_1 \mathcal{E}(t, u_0) =\,]\; F(\mathcal{E}(t, u_0)) = \partial_2 \mathcal{E}(t, u_0) \cdot F(u_0). \tag{1.2}$$

Proof. (1.2) is a consequence of the first-order variational equation for $\mathcal{E}(t, u)$, see [6, Theorem I.14.3], [4, Appendix A]. The simple direct proof given in [7, (3.7)] proceeds from the identity

$$\mathcal{E}(t + s, u_0) = \mathcal{E}(t, \mathcal{E}(s, u_0)).$$

Differentiation with respect to s gives

$$\frac{\partial}{\partial s} \mathcal{E}(t + s, u_0) = \partial_1 \mathcal{E}(t + s, u_0),$$

$$\frac{\partial}{\partial s} \mathcal{E}(t + s, u_0)\big|_{s=0} = \partial_1 \mathcal{E}(t, u_0) = F(\mathcal{E}(t, u_0)),$$

and on the other hand,

$$\frac{\partial}{\partial s} \mathcal{E}(t, \mathcal{E}(s, u_0)) = \partial_2 \mathcal{E}(t, \mathcal{E}(s, u_0)) \cdot \partial_1 \mathcal{E}(s, u_0),$$

$$\frac{\partial}{\partial s} \mathcal{E}(t, \mathcal{E}(s, u_0))\big|_{s=0} = \partial_2 \mathcal{E}(t, u_0) \cdot \partial_1 \mathcal{E}(0, u_0) = \partial_2 \mathcal{E}(t, u_0) \cdot F(u_0),$$

which completes the proof. □

2. Classical and symmetrized defects for one-step integrators

Consider an approximation to the given problem (1.1) defined by the flow

$$\mathcal{S}(t, u_0) \approx \mathcal{E}(t, u_0), \quad \mathcal{S}(0, u_0) = u_0, \tag{2.1}$$

---

[1] The extension to nonautonomous problems is deferred to Section 5.

[2] For the nonautonomous case see Lemma 5 in Section 5.

of a consistent one-step scheme with stepsize t, starting at $(0, u_0)$. We assume that the scheme has order p, i.e., the local error

$$\mathcal{L}(t, u_0) = \mathcal{S}(t, u_0) - \mathcal{E}(t, u_0) \tag{2.2}$$

satisfies $\mathcal{L}(t, u_0) = \mathscr{O}(t^{p+1})$.

We call

$$\mathcal{D}_{\boldsymbol{c}}(t, u) = \partial_1 \mathcal{S}(t, u) - F(\mathcal{S}(t, u)) = \mathscr{O}(t^p) \tag{2.3}$$

the *classical defect* associated with $\mathcal{S}(t, u)$. The local error can be represented in terms of the classical defect via the well-known nonlinear variation-of-constant formula, the so-called Gröbner–Alekseev Lemma. For convenience we restate this in a form required in our context and also include the proof following[3] [6, Theorem I.14.5] (see also [7, Theorem 3.3]). We formulate it in a concise way making direct use of (1.2).

Theorem 2. *In terms of the classical defect* (2.3), *the local error satisfies the integral representation*

$$\mathcal{L}(t, u_0) = \int_0^t \partial_2 \mathcal{E}(t - s, \mathcal{S}(s, u_0)) \cdot \mathcal{D}_{\boldsymbol{c}}(s, u_0) \, ds. \tag{2.4}$$

Proof. For fixed t, let

$$y(s) = \mathcal{S}(s, u_0),$$
$$z(s) = \mathcal{E}(t - s, y(s)).$$

In this notation, we have

$$z(s) = \mathcal{E}(t - s, \mathcal{S}(s, u_0)),$$
$$\text{satisfying} \quad z(0) = \mathcal{E}(t, u_0), \quad z(t) = \mathcal{S}(t, u_0).$$

Thus,

$$\mathcal{L}(t, u_0) = \mathcal{S}(t, u_0) - \mathcal{E}(t, u_0) = \int_0^t \tfrac{d}{ds} z(s) \, ds, \tag{2.5}$$

with

$$\tfrac{d}{ds} z(s) = -F(z(s)) + \partial_2 \mathcal{E}(t - s, y(s)) \cdot \tfrac{d}{ds} y(s).$$

Now, using (1.2)[4] this can be rewritten in the form

$$\tfrac{d}{ds} z(s) = \underbrace{-F(\mathcal{E}(t - s, y(s))) + \partial_2 \mathcal{E}(t - s, y(s)) \cdot F(y(s))}_{=0}$$
$$+ \partial_2 \mathcal{E}(t - s, y(s)) \cdot \left(\tfrac{d}{ds} y(s) - F(y(s)) \right)$$
$$= \partial_2 \mathcal{E}(t - s, y(s)) \cdot \mathcal{D}_{\boldsymbol{c}}(s, u_0),$$

and together with (2.5), identity (2.4) immediately follows. $\square$

Remark 1. Due to (1.2), an alternative, plausible way to define the defect is

$$\mathcal{D}(t, u) = \partial_1 \mathcal{S}(t, u) - \partial_2 \mathcal{S}(t, u) \cdot F(u). \tag{2.6}$$

Then,

$$\mathcal{L}(t, u_0) = \int_0^t \tfrac{d}{ds} \mathcal{S}(s, \mathcal{E}(t - s, u_0)) \, ds = \int_0^t \mathcal{D}(s, \mathcal{E}(t - s, u_0)) \, ds.$$

Remark 2. We can express the modified defect (2.6) in terms of $\mathcal{D}_{\boldsymbol{c}}(t, u)$ plus a higher-order perturbation,

$$\partial_1 \mathcal{S}(t, u) - \partial_2 \mathcal{S}(t, u) \cdot F(u)$$
$$= \big(\partial_1 \mathcal{S}(t, u) - F(\mathcal{S}(t, u)) \big) + \big(F(\mathcal{S}(t, u)) - \partial_2 \mathcal{S}(t, u) \cdot F(u) \big)$$
$$= \mathcal{D}_{\boldsymbol{c}}(t, u) + \underbrace{\big(F(\mathcal{E}(t, u)) - \partial_2 \mathcal{E}(t, u) \cdot F(u) \big)}_{=0}$$

---

[3] See [6, Figure I.14.1] *Lady Windermere's Fan, Act 2.*
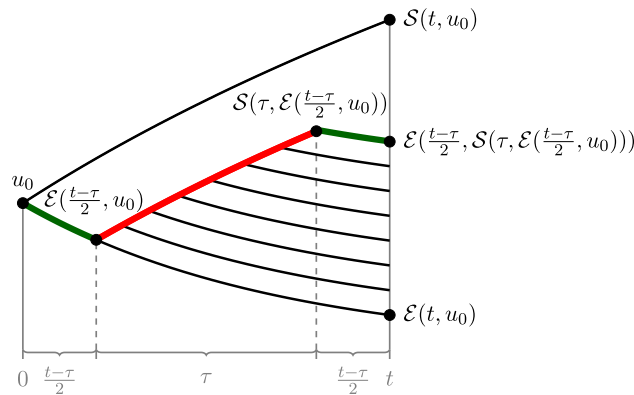[4] *Mutatis mutandis:* $s, t - s$ and $y(s)$ play the role of $0, t$ and u_0 from (1.2).

Fig. 1. Lady Windermere's Fan, Act 2 ½.

$$+ \underbrace{\big(F(\mathcal{S}(t, u)) - F(\mathcal{E}(t, u))\big)}_{= \mathcal{O}(t^{p+1})} + \underbrace{\big(\partial_2 \mathcal{S}(t, u) - \partial_2 \mathcal{E}(t, u)\big)}_{= \mathcal{O}(t^{p+1})} \cdot F(u)$$

$$= \mathcal{D}_\mathbf{c}(t, u) + \mathcal{O}(t^{p+1}).$$

Also, e.g., a convex combination of (2.3) and (2.6) represents a plausible defect. In particular, we will consider the arithmetic mean of (2.3) and (2.6) (see (2.9)), and we will introduce a symmetrized variant of Theorem 2, see Theorem 3.

2.1. Symmetrization

The following considerations are relevant for the case where the approximate flow $\mathcal{S}$ is self-adjoint (symmetric, time-reversible),[5] i.e.,

$$\mathcal{S}(-t, \mathcal{S}(t, u)) = u. \tag{2.7}$$

Self-adjoint schemes have an even order p, see [8, Theorem II.3.2].

The identity[6]

$$\partial_1 \mathcal{E}(t, u) = \tfrac{1}{2}\big(F(\mathcal{E}(t, u)) + \partial_2 \mathcal{E}(t, u) \cdot F(u)\big), \tag{2.8}$$

which is valid due to (1.2), motivates the definition of the *symmetrized defect*

$$\mathcal{D}_\mathbf{s}(t, u) = \partial_1 \mathcal{S}(t, u) - \tfrac{1}{2}\big(F(\mathcal{S}(t, u)) + \partial_2 \mathcal{S}(t, u) \cdot F(u)\big), \tag{2.9}$$

satisfying $\mathcal{D}_\mathbf{s}(t, u) = \mathcal{D}_\mathbf{c}(t, u) + \mathcal{O}(t^{p+1})$ (see Remark 2).

Theorem 3. *In terms of the symmetrized defect* (2.9), *the local error has the integral representation*

$$\mathcal{L}(t, u_0) = \int_0^t \partial_2 \mathcal{E}(\tfrac{t-s}{2}, \mathcal{S}(s, \mathcal{E}(\tfrac{t-s}{2}, u_0))) \cdot \mathcal{D}_\mathbf{s}(s, \mathcal{E}(\tfrac{t-s}{2}, u_0)) \, ds. \tag{2.10}$$

---

[5] Definition (2.9) and the assertion of Theorem 3 are independent of this assumption. However, our results derived later on essentially depend on it, in particular Theorem 4.

[6] In the terminology of Lie calculus (cf. for instance [8]), with

$$(D_F G)(u) := G'(u) \cdot F(u) = \tfrac{d}{dt} G(\mathcal{E}(t, u))|_{t=0},$$

and

$$e^{tD_F} G(u) := G(\mathcal{E}(t, u)),$$

we have (set $G = \mathrm{Id}$ and $G = F$, respectively)

$$F(\mathcal{E}(t, u)) = F(e^{tD_F} u) = e^{tD_F} F(u).$$

In this formalism, (2.8) assumes a more 'symmetric flavour', as in the linear case (see [1]),

$$\partial_1 \mathcal{E}(t, u) = \tfrac{1}{2}\big(F(e^{tD_F} u) + e^{tD_F} F(u)\big).$$

However, in the present context this formalism is of little practical use, and we stick to explicit, classical notation.

Proof. We reason in a similar way as in the proof of Theorem 2, but now in the spirit of Fig. 1. For fixed t, let

$$x(s) = \mathcal{E}(\tfrac{t-s}{2}, u_0),$$
$$y(s) = \mathcal{S}(s, x(s)),$$
$$z(s) = \mathcal{E}(\tfrac{t-s}{2}, y(s)).$$

In this notation, we have

$$z(s) = \mathcal{E}(\tfrac{t-s}{2}, \mathcal{S}(s, \mathcal{E}(\tfrac{t-s}{2}, u_0))),$$
$$\text{satisfying} \quad z(0) = \mathcal{E}(t, u_0), \quad z(t) = \mathcal{S}(t, u_0).$$

Thus,

$$\mathcal{L}(t, u_0) = \mathcal{S}(t, u_0) - \mathcal{E}(t, u_0) = \int_0^t \tfrac{\mathrm{d}}{\mathrm{d}s} z(s)\,\mathrm{d}s, \tag{2.11}$$

with

$$\tfrac{\mathrm{d}}{\mathrm{d}s} z(s) = -\tfrac{1}{2}F(z(s)) + \partial_2 \mathcal{E}(\tfrac{t-s}{2}, y(s)) \cdot \tfrac{\mathrm{d}}{\mathrm{d}s} y(s).$$

Now, using (1.2)[7] this can be rewritten in the form

$$\tfrac{\mathrm{d}}{\mathrm{d}s} z(s) = \tfrac{1}{2} \underbrace{\left(-F(\mathcal{E}(\tfrac{t-s}{2}, y(s))) + \partial_2 \mathcal{E}(\tfrac{t-s}{2}, y(s)) \cdot F(y(s))\right)}_{=0}$$
$$+ \partial_2 \mathcal{E}(\tfrac{t-s}{2}, y(s)) \cdot \underbrace{\left(\tfrac{\mathrm{d}}{\mathrm{d}s} y(s) - \tfrac{1}{2}F(y(s))\right)}. \tag{2.12a}$$

Furthermore, from the definition (2.9) of $\mathcal{D}_s(s, u)$, with $u = x(s)$ we obtain

$$\underbrace{\tfrac{\mathrm{d}}{\mathrm{d}s} y(s) - \tfrac{1}{2}F(y(s))}$$
$$= \partial_1 \mathcal{S}(s, x(s)) + \partial_2 \mathcal{S}(s, x(s)) \cdot \left(-\tfrac{1}{2}F(x(s))\right) - \tfrac{1}{2}F(y(s))$$
$$= \partial_1 \mathcal{S}(s, x(s)) - \tfrac{1}{2}\left(F(y(s)) + \partial_2 \mathcal{S}(s, x(s)) \cdot F(x(s))\right) \tag{2.12b}$$
$$= \mathcal{D}_s(s, x(s)).$$

After inserting (2.12b) into (2.12a), together with (2.11) we obtain (2.10). □

3. Classical and symmetrized defect-based local error estimation

Defect-based local error estimate. The idea is due to [1,4]. Let $\mathcal{D}(t, u) = \mathcal{D}_c(t, u)$ or $\mathcal{D}_s(t, u)$, respectively, and denote the integrands in (2.4) respectively (2.10), generically by $\Theta(s)$. Due to order p we have $\mathcal{D}(s, u) = \mathcal{O}(s^p)$ and $\Theta(s) = \mathcal{O}(s^p)$, whence

$$\mathcal{L}(t, u_0) = \int_0^t \Theta(s)\,\mathrm{d}s \approx \int_0^t \tfrac{s^p}{p!}\Theta^{(p)}(0)\,\mathrm{d}s = \tfrac{t^{p+1}}{(p+1)!}\Theta^{(p)}(0)$$
$$\approx \tfrac{t}{p+1}\Theta(t) = \tfrac{t}{p+1}\mathcal{D}(t, u_0). \tag{3.1}$$

Here, '$\approx$' means asymptotic approximation at the level $\mathcal{O}(t^{p+2})$. This approximation can be interpreted as a Hermite-type quadrature of order $p + 1$ for the local error integral, where the quadrature error depends on $\tfrac{\partial^{p+1}}{\partial s^{p+1}}\mathcal{D}(s, u_0) = \mathcal{O}(1)$ due to $\mathcal{D}(s, u_0) = \mathcal{O}(s^p)$, whence

$$\mathcal{L}(t, u_0) = \tfrac{t}{p+1}\mathcal{D}(t, u_0) + \mathcal{O}(t^{p+2}) \quad \text{for } \mathcal{D} = \mathcal{D}_c \text{ or } \mathcal{D} = \mathcal{D}_s.$$

For a precise analysis of the resulting quadrature error based on its Peano representation for the classical case in concrete applications, see for instance [1,3,4].

Next we show that for the self-adjoint case and using the symmetrized defect (2.9) we even have[8]

$$\mathcal{L}(t, u_0) = \tfrac{t}{p+1}\mathcal{D}_s(t, u_0) + \mathcal{O}(t^{p+3}).$$

To this end we consider the corrected scheme

$$\widehat{\mathcal{S}}_s(t, u) = \mathcal{S}(t, u) - \tfrac{t}{p+1}\mathcal{D}_s(t, u), \tag{3.2}$$

and we show that it is of (global) order $p + 2$.

---

[7] *Mutatis mutandis: s, $\tfrac{t-s}{2}$ and $y(s)$ play the role of 0, t and u_0 from (1.2).*
[8] For the linear constant coefficient case see [1, Theorem 1].

Theorem 4. *Consider a self-adjoint one-step scheme of (even) order $p \geq 2$, represented by its flow $\mathcal{S}(t, u)$ satisfying* (2.7), *applied to an evolution equation* (1.1). *Then the corrected scheme* (3.2) *is almost self-adjoint, i.e.,*

$$\widehat{\mathcal{S}}_{\mathbf{s}}(-t, \widehat{\mathcal{S}}_{\mathbf{s}}(t, u_0)) = u_0 + \mathcal{O}(t^{2p+2}). \tag{3.3a}$$

Moreover, the local error $\widehat{\mathcal{L}}_{\mathbf{s}}(t, u) = \widehat{\mathcal{S}}_{\mathbf{s}}(t, u) - \mathcal{E}(t, u)$ of the corrected scheme satisfies

$$\widehat{\mathcal{L}}_{\mathbf{s}}(t, u_0) = \mathcal{O}(t^{p+3}), \tag{3.3b}$$

i.e., $\widehat{\mathcal{S}}_{\mathbf{s}}$ has even order $p + 2$.

Proof. We consider

$$\begin{aligned}
\widehat{\mathcal{S}}_{\mathbf{s}}(-t, \widehat{\mathcal{S}}_{\mathbf{s}}(t, u_0)) &= \mathcal{S}(-t, \widehat{\mathcal{S}}_{\mathbf{s}}(t, u_0)) + \tfrac{t}{p+1} \mathcal{D}_{\mathbf{s}}(-t, \widehat{\mathcal{S}}_{\mathbf{s}}(t, u_0)) \\
&= \mathcal{S}\left(-t, \mathcal{S}(t, u_0) - \tfrac{t}{p+1} \mathcal{D}_{\mathbf{s}}(t, u_0)\right) \\
&\quad + \tfrac{t}{p+1} \mathcal{D}_{\mathbf{s}}\left(-t, \mathcal{S}(t, u_0) - \tfrac{t}{p+1} \mathcal{D}_{\mathbf{s}}(t, u_0)\right),
\end{aligned}$$

apply Taylor expansion, and make use of the assumption that $\mathcal{S}$ is self-adjoint, and the fact that $t\mathcal{D}_{\mathbf{s}}(t, u_0) = \mathcal{O}(t^{p+1})$:

$$\begin{aligned}
\widehat{\mathcal{S}}_{\mathbf{s}}(-t, \widehat{\mathcal{S}}_{\mathbf{s}}(t, u_0)) &= \underbrace{\mathcal{S}(-t, \mathcal{S}(t, u_0))}_{= u_0} \\
&\quad + \partial_2 \mathcal{S}(-t, \mathcal{S}(t, u_0)) \cdot \left(-\tfrac{t}{p+1} \mathcal{D}_{\mathbf{s}}(t, u_0)\right) + \mathcal{O}(t^{2p+2}) \\
&\quad + \tfrac{t}{p+1} \mathcal{D}_{\mathbf{s}}(-t, \mathcal{S}(t, u_0)) + \mathcal{O}(t^{2p+2}) \\
&= u_0 - \tfrac{t}{p+1} \underbrace{\left(\partial_2 \mathcal{S}(-t, \mathcal{S}(t, u_0)) \cdot \mathcal{D}_{\mathbf{s}}(t, u_0) - \mathcal{D}_{\mathbf{s}}(-t, \mathcal{S}(t, u_0))\right)}_{\text{critical term}} + \mathcal{O}(t^{2p+2}).
\end{aligned} \tag{3.4}$$

Now we collect the contributions to the **critical term**. First, from (2.7) we have[9]

$$0 = \tfrac{\partial}{\partial t} \underbrace{\mathcal{S}(-t, \mathcal{S}(t, u_0))}_{= u_0} = -\partial_1 \mathcal{S}(-t, \mathcal{S}(t, u_0)) + \partial_2 \mathcal{S}(-t, \mathcal{S}(t, u_0)) \cdot \partial_1 \mathcal{S}(t, u_0).$$

This implies

$$\begin{aligned}
\mathcal{D}_{\mathbf{s}}(-t, \mathcal{S}(t, u_0)) &= \\
&= \partial_1 \mathcal{S}(-t, \mathcal{S}(t, u_0)) - \tfrac{1}{2}\left(F(\underbrace{\mathcal{S}(-t, \mathcal{S}(t, u_0))}_{= u_0}) + \partial_2 \mathcal{S}(-t, \mathcal{S}(t, u_0)) \cdot F(\mathcal{S}(t, u_0))\right) \\
&= \partial_2 \mathcal{S}(-t, \mathcal{S}(t, u_0)) \cdot \partial_1 \mathcal{S}(t, u_0) - \tfrac{1}{2} F(u_0) - \tfrac{1}{2} \partial_2 \mathcal{S}(-t, \mathcal{S}(t, u_0)) \cdot F(\mathcal{S}(t, u_0)) \\
&= \partial_2 \mathcal{S}(-t, \mathcal{S}(t, u_0)) \cdot \left(\partial_1 \mathcal{S}(t, u_0) - \tfrac{1}{2} F(\mathcal{S}(t, u_0))\right) - \tfrac{1}{2} F(u_0).
\end{aligned}$$

Summarizing and collecting terms gives

$$\begin{aligned}
\mathbf{critical\ term} &= \\
&= \partial_2 \mathcal{S}(-t, \mathcal{S}(t, u_0)) \cdot \mathcal{D}_{\mathbf{s}}(t, u_0) - \mathcal{D}_{\mathbf{s}}(-t, \mathcal{S}(t, u_0)) \\
&= \partial_2 \mathcal{S}(-t, \mathcal{S}(t, u_0)) \cdot \left(\boxed{\partial_1 \mathcal{S}(t, u_0) - \tfrac{1}{2} F(\mathcal{S}(t, u_0))} - \tfrac{1}{2} \partial_2 \mathcal{S}(t, u_0) \cdot F(u_0)\right) \\
&\quad - \partial_2 \mathcal{S}(-t, \mathcal{S}(t, u_0)) \cdot \left(\boxed{\partial_1 \mathcal{S}(t, u_0) - \tfrac{1}{2} F(\mathcal{S}(t, u_0))}\right) - \tfrac{1}{2} F(u_0) \\
&= -\tfrac{1}{2}\left(\partial_2 \mathcal{S}(-t, \mathcal{S}(t, u_0)) \cdot \partial_2 \mathcal{S}(t, u_0) - \mathrm{Id}\right) \cdot F(u_0) \\
&= -\tfrac{1}{2}\left(\underbrace{\tfrac{\partial}{\partial u_0} \mathcal{S}(-t, \mathcal{S}(t, u_0))}_{= \mathrm{Id}} - \mathrm{Id}\right) \cdot F(u_0) = 0.
\end{aligned}$$

Thus, (3.4) indeed simplifies to (3.3a),

$$\widehat{\mathcal{S}}_{\mathbf{s}}(-t, \widehat{\mathcal{S}}_{\mathbf{s}}(t, u_0)) = u_0 + \mathcal{O}(t^{2p+2}).$$

The proof of (3.3b) now works in the same way as for the linear case [1, proof of Theorem 1], following the argument from [8, Theorem II.3.2]. □

---

[9] Here, $\tfrac{\partial}{\partial t} \mathcal{S}(-t, \mathcal{S}(t, u_0))$ means $\tfrac{\partial}{\partial t} \tilde{\mathcal{S}}(t, u_0)$ with $\tilde{\mathcal{S}}(t, u_0) = \mathcal{S}(-t, \mathcal{S}(t, u_0))$.

Assertion (3.3b) is equivalent to the fact that the symmetrized defect-based local error estimator according to (3.1),

$$\widetilde{\mathcal{L}}_{\mathbf{s}}(t, u_0) := \tfrac{t}{p+1} \mathcal{D}_{\mathbf{s}}(t, u_0) \tag{3.5}$$

is indeed of a better asymptotic quality than the classical defect, with a deviation

$$\widetilde{\mathcal{L}}_{\mathbf{s}}(t, u_0) - \mathcal{L}(t, u_0) = \mathcal{O}(t^{p+3}), \tag{3.6}$$

and not only $\mathcal{O}(t^{p+2})$.

In the following sections we present some examples of self-adjoint methods and show how to evaluate the symmetrized defect $\mathcal{D}_{\mathbf{s}}(t, u_0)$ as the basis for evaluating the local error estimator (3.5).

4. Examples for the autonomous case

4.1. Example: implicit midpoint rule

We illustrate the defect computation for the simplest example of a self-adjoint implicit one-step integrator. The flow of the second order implicit midpoint rule is defined by the relation

$$\mathcal{S}(t, u) = u + t\, F(\tfrac{1}{2}(u + \mathcal{S}(t, u))).$$

With

$$w = \mathcal{S}(t, u) \tag{4.1}$$

we obtain

$$\partial_1 \mathcal{S}(t, u) = \underbrace{F(\tfrac{1}{2}(u + w))}_{= (w - u)/t} + t\, F'\big(\tfrac{1}{2}(u + w)\big) \cdot \tfrac{1}{2} \partial_1 \mathcal{S}(t, u).$$

Thus, $x = \partial_1 \mathcal{S}(t, u)$ is obtained by solving the linear system

$$\big(\mathrm{Id} - \tfrac{t}{2} F'(\tfrac{1}{2}(u + w))\big) \cdot x = F(\tfrac{1}{2}(u + w)). \tag{4.2a}$$

Furthermore,

$$\begin{aligned}
\partial_2 \mathcal{S}(t, u) &= \mathrm{Id} + t F'\big(\tfrac{1}{2}(u + \mathcal{S}(t, u))\big) \cdot \big(\tfrac{1}{2}(\mathrm{Id} + \partial_2 \mathcal{S}(t, u))\big) \\
&= \mathrm{Id} + \tfrac{t}{2} F'(\tfrac{1}{2}(u + w)) + \tfrac{t}{2} F'(\tfrac{1}{2}(u + w)) \cdot \partial_2 \mathcal{S}(t, u),
\end{aligned}$$

whence

$$\big(\mathrm{Id} - \tfrac{t}{2} F'(\tfrac{1}{2}(u + w))\big) \cdot \partial_2 \mathcal{S}(t, u) = \big(\mathrm{Id} + \tfrac{1}{2} F'(\tfrac{1}{2}(u + w))\big).$$

Thus, $y = \partial_2 \mathcal{S}(t, u) \cdot F(u)$ is obtained by solving the linear system

$$\big(\mathrm{Id} - \tfrac{t}{2} F'(\tfrac{1}{2}(u + w))\big) \cdot y = \big(\mathrm{Id} + \tfrac{t}{2} F'(\tfrac{1}{2}(u + w))\big) \cdot F(u), \tag{4.2b}$$

with the same matrix as in (4.2a).

This gives the following defect representations.

- Classical defect:

  $$\mathcal{D}_{\mathbf{c}}(t, u) = x - F(w),$$

  where $x = \partial_1 \mathcal{S}(t, u)$ is the solution of (4.2a) and with w from (4.1).
- Symmetrized defect:

  $$\mathcal{D}_{\mathbf{s}}(t, u) = x - \tfrac{1}{2}(F(w) + y),$$

  where $x = \partial_1 \mathcal{S}(t, u)$ is the solution of (4.2a), and $y = \partial_2 \mathcal{S}(t, u) \cdot F(u)$ is the solution of (4.2b). This can also be written in the form

  $$\mathcal{D}_{\mathbf{s}}(t, u) = z - \tfrac{1}{2} F(w),$$

  where $z = x - \tfrac{1}{2} y$ is the solution of

  $$\big(\mathrm{Id} - \tfrac{t}{2} F'(\tfrac{1}{2}(u + w))\big) \cdot z = F(\tfrac{1}{2}(u + w)) - \tfrac{1}{2} F(u) - \tfrac{t}{4} F'(\tfrac{1}{2}(u + w)) \cdot F(u).$$

  Thus, the computation of the symmetrized defect requires only one additional evaluation of F as compared to the classical version.

Table 1

Coefficients of the self-adjoint splitting method from [9, Emb 4/3 AK s].

i	a_i	b_i
1	0.267171359000977615	−0.361837907604416033
2	−0.033827909669505667	0.861837907604416033
3	0.533313101337056104	0.861837907604416033
4	−0.033827909669505667	−0.361837907604416033
5	0.2671171359000977615	0

4.2. Example: Strang splitting applied to a semilinear evolution equation

We consider a semilinear problem of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t) = F(u(t)) = Au(t) + B(u(t)), \quad u(0) = u_0.$$

Denoting the flow of the nonlinear part by $\mathcal{E}_B(t, u)$, the second order self-adjoint Strang splitting scheme is given by

$$\mathcal{S}(t, u) = \mathrm{e}^{\frac{t}{2}A}\mathcal{E}_B\big(t, \mathrm{e}^{\frac{t}{2}A}u\big).$$

Let

$$v_1 = \mathrm{e}^{\frac{t}{2}A}u, \quad v_2 = \mathcal{E}_B(t, v_1), \quad w = \mathrm{e}^{\frac{t}{2}A}v_2 = \mathcal{S}(t, u).$$

Then,

$$\partial_1\mathcal{S}(t, u) = \tfrac{1}{2}A\mathcal{S}(t, u) + \mathrm{e}^{\frac{t}{2}A}\big(\partial_1\mathcal{E}_B(t, v_1) + \partial_2\mathcal{E}_B(t, v_1)(\tfrac{1}{2}Av_1)\big)$$
$$= \tfrac{1}{2}Aw + \mathrm{e}^{\frac{t}{2}A}\big(B(v_2) + \tfrac{1}{2}\partial_2\mathcal{E}_B(t, v_1)(Av_1)\big),$$

and

$$\partial_2\mathcal{S}(t, u)(\xi) = \mathrm{e}^{\frac{t}{2}A}\partial_2\mathcal{E}_B(t, v_1)\big(\mathrm{e}^{\frac{t}{2}A}\xi\big).$$

This gives the following defect representations.

- Classical defect:

$$\mathcal{D}_c(t, u) = \partial_1\mathcal{S}(t, u) - F(\mathcal{S}(t, u))$$
$$= \mathrm{e}^{\frac{t}{2}A}\big(B(v_2) + \tfrac{1}{2}\partial_2\mathcal{E}_B(t, v_1)\cdot(Av_1)\big) - \tfrac{1}{2}Aw - B(w). \tag{4.3}$$

- Symmetrized defect:

$$\mathcal{D}_s(t, u) = \partial_1\mathcal{S}(t, u) - \tfrac{1}{2}\big(F(\mathcal{S}(t, u)) + \partial_2\mathcal{S}(t, u)\cdot F(u)\big)$$
$$= \tfrac{1}{2}\boxed{Aw} + \mathrm{e}^{\frac{t}{2}A}\big(B(v_2) + \tfrac{1}{2}\partial_2\mathcal{E}_B(t, v_1)\cdot(Av_1)\big)$$
$$\quad - \tfrac{1}{2}\Big(\boxed{Aw} + B(w) + \mathrm{e}^{\frac{t}{2}A}\partial_2\mathcal{E}_B(t, v_1)\cdot\big(\mathrm{e}^{\frac{t}{2}A}(Au + B(u))\big)\Big)$$
$$= \mathrm{e}^{\frac{t}{2}A}B(v_2) + \boxed{\tfrac{1}{2}\mathrm{e}^{\frac{t}{2}A}\partial_2\mathcal{E}_B(t, v_1)(Av_1)}$$
$$\quad - \tfrac{1}{2}B(w) - \boxed{\tfrac{1}{2}\mathrm{e}^{\frac{t}{2}A}\partial_2\mathcal{E}_B(t, v_1)(Av_1)} - \tfrac{1}{2}\mathrm{e}^{\frac{t}{2}A}\partial_2\mathcal{E}_B(t, v_1)\big(\mathrm{e}^{\frac{t}{2}A}B(u)\big)$$
$$= \mathrm{e}^{\frac{t}{2}A}\Big(B(v_2) - \tfrac{1}{2}\partial_2\mathcal{E}_B(t, v_1)\big(\mathrm{e}^{\frac{t}{2}A}B(u)\big)\Big) - \tfrac{1}{2}B(w). \tag{4.4}$$

Thus, (4.3) resp. (4.4) require one evaluation of $\partial_2\mathcal{E}_B(t, v_1)\cdot(\,\cdot\,)$, and either one or two evaluations of $\mathrm{e}^{\frac{t}{2}A}(\,\cdot\,)$, respectively.

4.3. Algorithmic realization for higher order splitting methods

In Fig. 2, we give pseudocodes for the economical algorithmic realization of the symmetrized defect when it is employed in the context of splitting methods involving an arbitrary number of J compositions. If we denote the subflow of the nonlinear operator by $\mathcal{E}_B(t, u_0)$, an n-stage splitting approximation is defined by a composition of the two subflows,

$$\mathcal{S}(t, u_0) = \mathcal{E}_B(b_J t, \ldots \mathrm{e}^{a_2 tA}\mathcal{E}_B(b_1 t, \mathrm{e}^{a_1 tA}u_0)\cdots).$$

An optimized fourth order method we will use in Section 6.1 has the coefficient tableau given in Table 1.

The algorithms in Fig. 2 have the splitting approximation $u = \mathcal{S}(t, u_0)$ and the symmetrized defect $d = \mathcal{D}_s(t, u_0)$ as the output; for efficiency, u and d are evaluated simultaneously. The left algorithm refers to the situation where the operator A is linear, and on the right the general nonlinear case is elaborated.
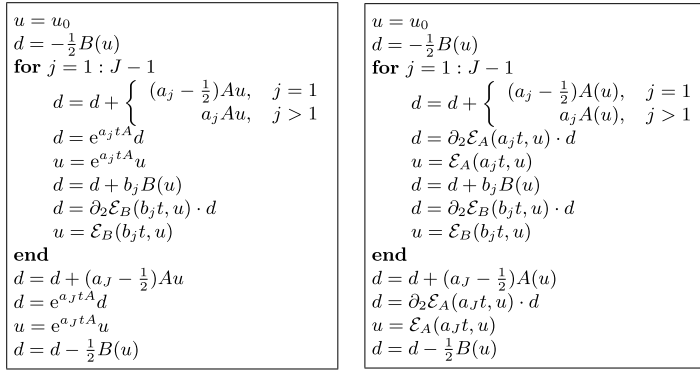
$$
\begin{array}{|l|}
\hline
u = u_0 \\
d = -\frac{1}{2}B(u) \\
\textbf{for } j = 1 : J - 1 \\
\quad d = d + \begin{cases} (a_j - \frac{1}{2})Au, & j = 1 \\ a_j Au, & j > 1 \end{cases} \\
\quad d = \mathrm{e}^{a_j t A} d \\
\quad u = \mathrm{e}^{a_j t A} u \\
\quad d = d + b_j B(u) \\
\quad d = \partial_2 \mathcal{E}_B(b_j t, u) \cdot d \\
\quad u = \mathcal{E}_B(b_j t, u) \\
\textbf{end} \\
d = d + (a_J - \frac{1}{2})Au \\
d = \mathrm{e}^{a_J t A} d \\
u = \mathrm{e}^{a_J t A} u \\
d = d - \frac{1}{2}B(u) \\
\hline
\end{array}
\qquad
\begin{array}{|l|}
\hline
u = u_0 \\
d = -\frac{1}{2}B(u) \\
\textbf{for } j = 1 : J - 1 \\
\quad d = d + \begin{cases} (a_j - \frac{1}{2})A(u), & j = 1 \\ a_j A(u), & j > 1 \end{cases} \\
\quad d = \partial_2 \mathcal{E}_A(a_j t, u) \cdot d \\
\quad u = \mathcal{E}_A(a_j t, u) \\
\quad d = d + b_j B(u) \\
\quad d = \partial_2 \mathcal{E}_B(b_j t, u) \cdot d \\
\quad u = \mathcal{E}_B(b_j t, u) \\
\textbf{end} \\
d = d + (a_J - \frac{1}{2})A(u) \\
d = \partial_2 \mathcal{E}_A(a_J t, u) \cdot d \\
u = \mathcal{E}_A(a_J t, u) \\
d = d - \frac{1}{2}B(u) \\
\hline
\end{array}
$$

Fig. 2. Algorithmic realization of the symmetrized defect for splitting methods. Left: semilinear case. Right: nonlinear case.

5. The nonautonomous case, with examples

The results from Sections 2 and 3 carry over to nonautonomous evolution equations

$$
\tfrac{\mathrm{d}}{\mathrm{d}t} u(t) = F(t, u(t)), \quad u(t_0) = u_0. \tag{5.1a}
$$

For our purpose it is notationally more favorable to introduce the 'local' variable τ, such that $t = t_0 + \tau$, and reformulate (5.1a) in the form

$$
\tfrac{\mathrm{d}}{\mathrm{d}\tau} u(t_0 + \tau) = F(t_0 + \tau, u(t_0 + \tau)), \quad u(t_0) = u_0. \tag{5.1b}
$$

The exact flow associated with (5.1) is denoted by $\mathcal{E}(\tau, t_0, u)$. It satisfies[10]

$$
\partial_1 \mathcal{E}(\tau, t_0, u_0) = F(t_0 + \tau, \mathcal{E}(\tau, t_0, u_0)), \quad \mathcal{E}(0, t_0, u_0) = u_0.
$$

To infer the appropriate definition of the symmetrized defect in this case there are two approaches, which we both discuss for the sake of completeness. The first one relies on a direct extension of the fundamental identity (1.2) (Lemma 1), see Lemma 5. The other approach is based on reformulating (5.1) in autonomous form in the usual way, leading to the same conclusion and showing that the theoretical background based on Theorems 3 and 4 directly carries over to the nonautonomous case.

Lemma 5.

$$
\begin{aligned}
&[\partial_1 \mathcal{E}(\tau, t_0, u_0) =] \\
&\quad F(t_0 + \tau, \mathcal{E}(\tau, t_0, u_0)) = \partial_2 \mathcal{E}(\tau, t_0, u_0) + \partial_3 \mathcal{E}(\tau, t_0, u_0) \cdot F(t_0, u_0).
\end{aligned} \tag{5.2}
$$

Proof. The idea is the same as in the proof of Lemma 1. We proceed from the identity

$$
\mathcal{E}(\tau + \sigma, t_0, u_0) = \mathcal{E}(\tau, t_0 + \sigma, \mathcal{E}(\sigma, t_0, u_0)).
$$

Differentiation with respect to σ gives

$$
\begin{aligned}
&\tfrac{\partial}{\partial \sigma} \mathcal{E}(\tau + \sigma, t_0, u_0) = \partial_1 \mathcal{E}(\tau + \sigma, t_0, u_0), \\
&\tfrac{\partial}{\partial \sigma} \mathcal{E}(\tau + \sigma, t_0, u_0)\big|_{\sigma=0} = \partial_1 \mathcal{E}(\tau, t_0, u_0) = F(t_0 + \tau, \mathcal{E}(\tau, t_0, u_0)),
\end{aligned}
$$

---

[10] Again, $\partial_1 \mathcal{E}(\tau, t_0, u_0)$ denotes $\tfrac{\mathrm{d}}{\mathrm{d}\tau} \mathcal{E}(\tau, t_0, u_0)$, and ∂_2, ∂_3 are defined analogously.

and on the other hand,

$$\tfrac{\partial}{\partial \sigma} \mathcal{E}(\tau, t_0 + \sigma, \mathcal{E}(\sigma, t_0, u_0))$$
$$= \partial_2 \mathcal{E}(\tau, t_0 + \sigma, \mathcal{E}(\sigma, t_0, u_0)) + \partial_3 \mathcal{E}(\tau, t_0 + \sigma, \mathcal{E}(\sigma, t_0, u_0)) \cdot \partial_1 \mathcal{E}(\sigma, t_0, u_0),$$

$$\tfrac{\partial}{\partial \sigma} \mathcal{E}(\tau, t_0 + \sigma, \mathcal{E}(\sigma, t_0, u_0))\big|_{\sigma=0}$$
$$= \partial_2 \mathcal{E}(\tau, t_0, \mathcal{E}(0, t_0, u_0)) + \partial_3 \mathcal{E}(\tau, t_0, \mathcal{E}(0, t_0, u_0)) \cdot \partial_1 \mathcal{E}(0, t_0, u_0)$$
$$= \partial_2 \mathcal{E}(\tau, t_0, u_0) + \partial_3 \mathcal{E}(\tau, t_0, u_0) \cdot F(t_0, u_0),$$

which completes the proof. $\square$

Alternatively, we can reformulate (5.1b) in autonomous form, defining

$$U = \begin{bmatrix} t_0 + \tau \\ u \end{bmatrix}, \quad \boldsymbol{F}(U) = \begin{bmatrix} 1 \\ F(t_0 + \tau, u) \end{bmatrix}$$

whence

$$\tfrac{d}{d\tau} U(\tau) = \boldsymbol{F}(U(\tau)), \quad U(0) = \begin{bmatrix} t_0 \\ u_0 \end{bmatrix},$$

and with the flow

$$\boldsymbol{\mathcal{E}}(\tau, U) = \boldsymbol{\mathcal{E}}(\tau, t_0, u) = \begin{bmatrix} t_0 + \tau \\ \mathcal{E}(\tau, t_0, u) \end{bmatrix}$$

satisfying the fundamental identity according to Lemma 1,

$$[\partial_1 \boldsymbol{\mathcal{E}}(\tau, U) =] \; \boldsymbol{F}(\boldsymbol{\mathcal{E}}(\tau, U)) = \partial_2 \boldsymbol{\mathcal{E}}(\tau, U) \cdot \boldsymbol{F}(U). \tag{5.3}$$

With $U_0 = (t_0, u_0)$ we have

$$\partial_1 \boldsymbol{\mathcal{E}}(\tau, U_0) = \boldsymbol{F}(\boldsymbol{\mathcal{E}}(\tau, U_0)) = \begin{bmatrix} 1 \\ F(t_0 + \tau, \mathcal{E}(\tau, t_0, u_0)) \end{bmatrix}, \quad \boldsymbol{\mathcal{E}}(0, U_0) = U_0,$$

and

$$\partial_2 \boldsymbol{\mathcal{E}}(\tau, U_0) = \begin{bmatrix} 1 & 0 \\ \partial_2 \mathcal{E}(\tau, t_0, u_0) & \partial_3 \mathcal{E}(\tau, t_0, u_0) \end{bmatrix}.$$

Using (5.3) and evaluating the second component again gives (5.2).

For a one-step approximation represented by $\mathcal{S}(\tau, t_0, u_0) \approx \mathcal{E}(\tau, t_0, u_0)$, relation (5.2) again motivates the definition of the symmetrized defect

$$\mathcal{D}_{\boldsymbol{s}}(\tau, t_0, u_0) = \partial_1 \mathcal{S}(\tau, t_0, u_0)$$
$$- \tfrac{1}{2}\big(F(t_0 + \tau, \mathcal{S}(\tau, t_0, u_0)) + \partial_2 \mathcal{S}(\tau, t_0, u_0) + \partial_3 \mathcal{S}(\tau, t_0, u_0) F(t_0, u_0)\big) \tag{5.4}$$
$$= \big(\partial_1 - \tfrac{1}{2}\partial_2\big)\mathcal{S}(\tau, t_0, u_0) - \tfrac{1}{2}\big(F(t_0 + \tau, \mathcal{S}(\tau, t_0, u_0)) + \partial_3 \mathcal{S}(\tau, t_0, u_0)F(t_0, u_0)\big).$$

The linear nonautonomous case. Now we consider the case of a linear time-dependent problem

$$\tfrac{d}{d\tau} u(t_0 + \tau) = A(t_0 + \tau)u(t_0 + \tau), \quad u(t_0) = u_0. \tag{5.5}$$

Since in the present case the flow is linear in u_0, we write it in the simplified form[11]

$$\mathcal{E}(\tau, t_0, u_0) =: \mathcal{E}(\tau, t_0)u_0, \tag{5.6}$$

satisfying

$$\partial_1 \mathcal{E}(\tau, t_0) = A(t_0 + \tau)\, \mathcal{E}(\tau, t_0), \quad \mathcal{E}(0, t_0) = \mathrm{Id}.$$

Note that

$$\mathcal{E}(-\tau, t_0 + \tau)\, \mathcal{E}(\tau, t_0) = \mathrm{Id}. \tag{5.7}$$

A one-step approximation $\mathcal{S}(\tau, t_0, u_0) \approx \mathcal{E}(\tau, t_0, u_0)$, is also typically linear in u_0,

$$\mathcal{S}(\tau, t_0, u_0) =: \mathcal{S}(\tau, t_0)u_0 \approx \mathcal{E}(\tau, t_0)u_0.$$

In particular, we again focus on self-adjoint schemes which are characterized by the identity (cf. (5.7))

$$\mathcal{S}(-\tau, t_0 + \tau)\, \mathcal{S}(\tau, t_0) = \mathrm{Id}. \tag{5.8}$$

For $\mathcal{S}(\tau, t_0)u_0$ we obtain the following defect representations.

---

[11] (5.6) is a minor abuse of notation. Note that $\mathcal{E}(\tau, t_0)$ can be expressed as a matrix exponential via the so-called Magnus expansion, see for instance [1,10].

- Classical defect:
$$\mathcal{D}_{\mathbf{c}}(\tau, t_0, u_0) =: \mathcal{D}_{\mathbf{c}}(\tau, t_0)u_0,$$
  with
$$\mathcal{D}_{\mathbf{c}}(\tau, t_0) = \partial_1 \mathcal{S}(\tau, t_0) - A(t_0 + \tau)\mathcal{S}(\tau, t_0). \tag{5.9}$$

- Symmetrized defect (5.4):
$$\mathcal{D}_{\mathbf{s}}(\tau, t_0, u_0) =: \mathcal{D}_{\mathbf{s}}(\tau, t_0)u_0,$$
  with
$$\begin{aligned}
\mathcal{D}_{\mathbf{s}}(\tau, t_0) &= \partial_1 \mathcal{S}(\tau, t_0) - \tfrac{1}{2}\big(A(t_0 + \tau)\mathcal{S}(\tau, t_0) + \partial_2 \mathcal{S}(\tau, t_0) + \mathcal{S}(\tau, t_0)A(t_0)\big) \\
&= \big(\partial_1 - \tfrac{1}{2}\partial_2\big)\mathcal{S}(\tau, t_0) - \tfrac{1}{2}\big(A(t_0 + \tau)\mathcal{S}(\tau, t_0) + \mathcal{S}(\tau, t_0)A(t_0)\big).
\end{aligned} \tag{5.10}$$

5.1. Example: Exponential midpoint rule

The self-adjoint second order exponential midpoint rule applied to (5.5) is given by
$$S(\tau, t_0) = e^{\tau A(t_0 + \frac{\tau}{2})}.$$
Let
$$\mathcal{R}(\tau, t_0)(\,\cdot\,) = \tfrac{\mathrm{d}}{\mathrm{d}\Omega}e^{\Omega}\big|_{\Omega = \tau A(t_0 + \frac{\tau}{2})}(\,\cdot\,),$$
where $\tfrac{\mathrm{d}}{\mathrm{d}\Omega}e^{\Omega}$ denotes the Fréchet derivative of the matrix exponential, see (5.13). Then,
$$\begin{aligned}
\partial_1 \mathcal{S}(\tau, t_0) &= \mathcal{R}(\tau, t_0)\big(\tfrac{\partial}{\partial\tau}(\tau A(t_0 + \tfrac{\tau}{2}))\big) \\
&= \mathcal{R}(\tau, t_0)\big(A(t_0 + \tfrac{\tau}{2}) + \tfrac{1}{2}\tau A'(t_0 + \tfrac{\tau}{2})\big), \\
\partial_2 \mathcal{S}(\tau, t_0) &= \mathcal{R}(\tau, t_0)\big(\tfrac{\partial}{\partial t_0}(\tau A(t_0 + \tfrac{\tau}{2}))\big) \\
&= \mathcal{R}(\tau, t_0)\big(\tau A'(t_0 + \tfrac{\tau}{2})\big).
\end{aligned}$$
This gives the following defect representations.

- Classical defect (5.9):
$$\mathcal{D}_{\mathbf{c}}(\tau, t_0) = \mathcal{R}(\tau, t_0)\big(A(t_0 + \tfrac{\tau}{2}) + \tfrac{1}{2}\tau A'(t_0 + \tfrac{\tau}{2})\big) - A(t_0 + \tau)\mathcal{S}(\tau, t_0). \tag{5.11}$$

- Symmetrized defect (5.10):
$$\begin{aligned}
\mathcal{D}_{\mathbf{s}}(\tau, t_0) &= \mathcal{R}(\tau, t_0)\big(A(t_0 + \tfrac{\tau}{2}) + \boxed{\tfrac{1}{2}\tau A'(t_0 + \tfrac{\tau}{2})}\,\big) \\
&\quad - \tfrac{1}{2}\big(A(t_0 + \tau)\mathcal{S}(\tau, t_0) + \mathcal{R}(\tau, t_0)\big(\boxed{\tau A'(t_0 + \tfrac{\tau}{2})} - \mathcal{S}(\tau, t_0)A(t_0)\big) \\
&= \mathcal{R}(\tau, t_0)\big(A(t_0 + \tfrac{\tau}{2})\big) - \tfrac{1}{2}\big(A(t_0 + \tau)\mathcal{S}(\tau, t_0) + \mathcal{S}(\tau, t_0)A(t_0)\big).
\end{aligned} \tag{5.12}$$

Here, the explicit representation
$$\begin{aligned}
\mathcal{R}(\tau, t_0)(V) &= \int_0^1 e^{\sigma\tau A(t_0 + \frac{\tau}{2})} V e^{(1-\sigma)\tau A(t_0 + \frac{\tau}{2})}\,\mathrm{d}\sigma \\
&= \int_0^1 e^{\sigma\tau A(t_0 + \frac{\tau}{2})} V e^{-\sigma\tau A(t_0 + \frac{\tau}{2})}\,\mathrm{d}\sigma \cdot \mathcal{S}(\tau, t_0)
\end{aligned} \tag{5.13}$$
follows from [11, (10.15)]. For evaluating (5.11), a sufficiently accurate quadrature approximation for the integral according to (5.13) is required. This involves evaluation of A' and the commutator $[A, A']$, see [1]. In contrast, the relevant term from (5.12) simplifies to
$$\begin{aligned}
\mathcal{R}(\tau, t_0)\big(A(t_0 + \tfrac{\tau}{2})\big) &= \int_0^1 e^{\sigma\tau A(t_0 + \frac{\tau}{2})} A(t_0 + \tfrac{\tau}{2}) e^{-\sigma\tau A(t_0 + \frac{\tau}{2})}\,\mathrm{d}\sigma \cdot \mathcal{S}(\tau, t_0) \\
&= A(t_0 + \tfrac{\tau}{2})\mathcal{S}(\tau, t_0) = \mathcal{S}(\tau, t_0)A(t_0 + \tfrac{\tau}{2}),
\end{aligned}$$
whence the symmetrized defect (5.12) can be evaluated exactly,
$$\begin{aligned}
\mathcal{D}_{\mathbf{s}}(\tau, t_0) &= \big(A(t_0 + \tfrac{\tau}{2}) - \tfrac{1}{2}A(t_0 + \tau)\big)\mathcal{S}(\tau, t_0) - \tfrac{1}{2}\mathcal{S}(\tau, t_0)A(t_0) \\
&= \mathcal{S}(\tau, t_0)\big(A(t_0 + \tfrac{\tau}{2}) - \tfrac{1}{2}A(t_0)\big) - \tfrac{1}{2}A(t_0 + \tau)\mathcal{S}(\tau, t_0).
\end{aligned} \tag{5.14}$$
This involves an additional application of $\mathcal{S}(\tau, t_0)$, but it does not require evaluation of the derivative A' or of a commutator expression. We also note that the applications of $\mathcal{S}$ from left and right can be evaluated in parallel.

5.2. Algorithmic realization for higher order magnus-type methods

The integrators which we consider for the numerical approximation of (6.3) are commutator-free Magnus-type methods (CFM) and classical Magnus integrators.

In contrast to the special case of the exponential midpoint rule, for practical evaluation the defect needs to be approximated in an asymptotically correct way. To this end we require an approximation scheme which preserves the desired order $p+2$ of the corrected scheme (3.2), or equivalently, the asymptotic quality (3.6) of the local error estimator is not affected by such an approximation.

Various versions of the resulting classical defect-based error estimators for these exponential integrators are presented in [12]. We now follow two of these approaches. To keep the presentation self-contained within reason, we briefly recapitulate the underlying material from [12, Section 3], and we introduce the corresponding symmetrized defect approximations.

5.2.1. Commutator-free Magnus-type integrators

As the basic integrator we consider a *commutator-free Magnus-type (CFM)* method [13],

$$\mathcal{S}(\tau, t_0) = \mathcal{S}_J(\tau, t_0) \cdots \mathcal{S}_1(\tau, t_0), \tag{5.15a}$$

where

$$S_j(\tau, t_0) = e^{\Omega_j(\tau, t_0)} = e^{\tau B_j(\tau, t_0)},$$

$$\text{with } B_j(\tau, t_0) = \sum_{k=1}^{K} a_{jk} A(t_0 + c_k \tau), \tag{5.15b}$$

where the coefficients c_k and a_{jk} are chosen in such a way that a desired order of consistency is obtained. Note that the assumption of symmetry of the scheme also implies symmetry of the coefficients in the following sense,

$$c_k - \tfrac{1}{2} = \tfrac{1}{2} - c_{K+1-k}, \quad k = 1, \ldots, K, \tag{5.16a}$$

and

$$a_{jk} = a_{J+1-j, K+1-k}, \quad j = 1, \ldots, J, \quad k = 1, \ldots, K. \tag{5.16b}$$

Our construction involves evaluation of the derivatives

$$\tfrac{\partial}{\partial \tau} e^{\Omega_j(\tau, t_0)} = \Gamma_{\tau, j}(\tau, t_0) e^{\Omega_j(\tau, t_0)}, \quad \tfrac{\partial}{\partial t_0} e^{\Omega_j(\tau, t_0)} = \Gamma_{t_0, j}(\tau, t_0) e^{\Omega_j(\tau, t_0)},$$

where

$$\Gamma_{\tau, j}(\tau, t_0) = B_j(\tau, t_0) + \sum_{m \geq 0} \tfrac{1}{(m+1)!} \tau^{m+1} \text{ad}^m_{B_j(\tau, t_0)}(\tfrac{\partial}{\partial \tau} B_j(\tau, t_0)),$$

and

$$\Gamma_{t_0, j}(\tau, t_0) = \sum_{m \geq 0} \tfrac{1}{(m+1)!} \tau^{m+1} \text{ad}^m_{B_j(\tau, t_0)}(\partial_2 B_j(\tau, t_0)).$$

Applying the product rule to $\mathcal{S}(\tau, t_0)$ defined in (5.15) we see that the symmetrized defect (5.10) of the numerical approximation is an expression involving the derivatives

$$\left(\partial_1 - \tfrac{1}{2}\partial_2\right)\mathcal{S}_j(\tau, t_0) = \Gamma_j(\tau, t_0)\,\mathcal{S}_j(\tau, t_0), \tag{5.17}$$

with

$$\Gamma_j(\tau, t_0) = \Gamma_{\tau, j}(\tau, t_0) - \tfrac{1}{2}\Gamma_{t_0, j}(\tau, t_0)$$
$$= B_j(\tau, t_0) + \sum_{m \geq 0} \tfrac{1}{(m+1)!} \tau^{m+1} \text{ad}^m_{B_j(\tau, t_0)}(\check{B}_j(\tau, t_0)), \tag{5.18}$$

where we have defined

$$\check{B}_j(\tau, t_0) = \left(\partial_1 - \tfrac{1}{2}\partial_2\right) B_j(\tau, t_0) = \sum_{k=1}^{K} a_{jk}(c_k - \tfrac{1}{2}) A'(t_0 + c_k \tau).$$

One possible computable approximation is obtained by truncating the series (5.18); we will refer to the resulting procedure as *Taylor variant*. The procedure in conjunction with the classical defect is given in detail in [12, Section 3].

We remark at this point that symmetry of the basic CFM integrator implies that truncation of the series (5.18) at $m = p$, i.e., approximating $\Gamma_j(\tau, t_0)$ by[12]

$$\widetilde{\Gamma}_j(\tau, t_0) = B_j(\tau, t_0) + \sum_{m=0}^{p-1} \frac{1}{(m+1)!} \tau^{m+1} \mathrm{ad}^m_{B_j(\tau, t_0)}(\check{B}_j(\tau, t_0)) \tag{5.19}$$

is already sufficient to obtain a defect approximation of accuracy $p + 2$, as is demonstrated in the following.

Proposition 6. *Let $\mathcal{D}_s$ be the symmetrized defect of a self-adjoint CFM integrator of order p, and $\widetilde{\mathcal{D}}_s$ its approximation constructed via the truncated Taylor variant according to (5.19). Then,*

$$\mathcal{D}_s(\tau, u_0) - \widetilde{\mathcal{D}}_s(\tau, u_0) = \mathcal{O}(\tau^{p+2}).$$

Proof. Observe that

$$B_j(\tau, t_0) = X_j A(t_0) + \mathcal{O}(\tau), \quad \check{B}_j(\tau, t_0) = Y_j A'(t_0) + \mathcal{O}(\tau),$$

where

$$X_j = \sum_{k=1}^{K} a_{jk}, \quad Y_j = \sum_{k=1}^{K} a_{jk}(c_k - \tfrac{1}{2}).$$

Thus,

$$\Gamma_j(\tau, t_0) - \widetilde{\Gamma}_j(\tau, t_0) = \frac{1}{(p+1)!} \tau^{p+1} X_j^p Y_j \, \mathrm{ad}^p_{A(t_0)}(A'(t_0)) + \mathcal{O}(\tau^{p+2}).$$

Inserting this in the computational algorithm given in Fig. 3 (left) and taking into account that

$$e^{\tau B_j(\tau, t_0)} = \mathrm{Id} + \mathcal{O}(\tau),$$

the total error resulting from substitution of the exact defect $\mathcal{D}_s$ by the truncated Taylor approximation of Γ_j is

$$\mathcal{D}_s(\tau, t_0) - \widetilde{\mathcal{D}}_s(\tau, t_0) = \frac{1}{(p+1)!} \tau^{p+1} Z \, \mathrm{ad}^p_{A(t_0)}(A'(t_0)) + \mathcal{O}(\tau^{p+2}),$$

with

$$Z = \sum_{j=1}^{J} X_j^p Y_j.$$

To establish the assertion of the proposition we now show $Z = 0$: From (5.16),

$$X_j^p Y_j = -X_{J+1-j}^p Y_{J+1-j}, \quad j = 1, \ldots, J, \quad X_{\lfloor J/2 \rfloor + 1}^p Y_{\lfloor J/2 \rfloor + 1} = 0 \ \text{if } J \text{ is odd},$$

whence

$$Z = \sum_{j}^{\lfloor J/2 \rfloor} (X_j^p Y_j + X_{J+1-j}^p Y_{J+1-j}) \ \big[+ X_{\lfloor J/2 \rfloor + 1}^p Y_{\lfloor J/2 \rfloor + 1} \ \text{if } J \text{ is odd} \big] = 0,$$

which completes the proof. □

As an alternative to the series representation (5.18), we may use the integral representation which follows from [11, (10.15)],

$$\Gamma_j(\tau, t_0) = B_j(\tau, t_0) + \int_0^{\tau} e^{\sigma B_j(\tau, t_0)} \check{B}_j(\tau, t_0) e^{-\sigma B_j(\tau, t_0)} \, d\sigma,$$

and apply a p-th order two-sided Hermite-type quadrature (see [12, Section 3]) to approximate the integral. We will refer to the resulting procedure as *Hermite variant*. The procedure in conjunction with the classical defect was also introduced in [12, Section 3]. Similarly as for the Taylor variant, it can be shown that quadrature of order p is sufficient to obtain a defect approximation of order $p + 2$.

These two sketched strategies result in the procedures given as pseudocode in Fig. 3 where the defect $d = \mathcal{D}_s(\tau, t_0)u_0$ is computed as the output along with the basic approximation $u = \mathcal{S}(\tau, t_0)u_0$. Then, for order $p = 4$, for instance, for the Taylor variant we have

$$\widetilde{\Gamma}_j(\tau, t_0) = B_j(\tau, t_0) + \tau \check{B}_j(\tau, t_0) + \tfrac{1}{2}\tau^2[B_j(\tau, t_0), \check{B}_j(\tau, t_0)]$$
$$+ \tfrac{1}{6}\tau^3[B_j(\tau, t_0), [B_j(\tau, t_0), \check{B}_j(\tau, t_0)]]$$
$$+ \tfrac{1}{24}\tau^4[B_j(\tau, t_0), [B_j(\tau, t_0), [B_j(\tau, t_0), \check{B}_j(\tau, t_0)]]],$$

---

[12] A priori one would expect that it is required to include the term of degree $p + 1$ also.

$$
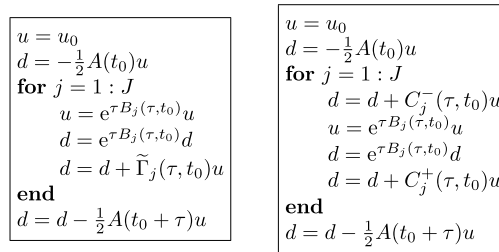\begin{array}{|l|}
\hline
u = u_0 \\
d = -\tfrac{1}{2}A(t_0)u \\
\textbf{for } j = 1 : J \\
\quad u = e^{\tau B_j(\tau, t_0)} u \\
\quad d = e^{\tau B_j(\tau, t_0)} d \\
\quad d = d + \widetilde{\Gamma}_j(\tau, t_0)u \\
\textbf{end} \\
d = d - \tfrac{1}{2}A(t_0 + \tau)u \\
\hline
\end{array}
\qquad
\begin{array}{|l|}
\hline
u = u_0 \\
d = -\tfrac{1}{2}A(t_0)u \\
\textbf{for } j = 1 : J \\
\quad d = d + C_j^-(\tau, t_0)u \\
\quad u = e^{\tau B_j(\tau, t_0)} u \\
\quad d = e^{\tau B_j(\tau, t_0)} d \\
\quad d = d + C_j^+(\tau, t_0)u \\
\textbf{end} \\
d = d - \tfrac{1}{2}A(t_0 + \tau)u \\
\hline
\end{array}
$$

Fig. 3. Algorithmic realization of the symmetrized defect for CFM methods. Left: Taylor variant. Right: Hermite variant.

and for the Hermite variant,

$$
C_j^\pm(\tau, t_0) = \tfrac{1}{2}\big(B_j(\tau, t_0) + \tau \check{B}_j(\tau, t_0)\big) \pm \tfrac{1}{12}\tau^2[B_j(\tau, t_0), \check{B}_j(\tau, t_0)].
$$

5.2.2. Classical magnus integrators

As an example we consider the classical fourth order Magnus integrator based on quadrature at Gaussian points (see [12]),

$$
\mathcal{S}(\tau, t_0) = e^{\Omega(\tau, t_0)} = e^{\tau B(\tau, t_0)}, \tag{5.20a}
$$

where $\Omega(\tau, t_0) = \tau B(\tau, t_0)$ approximates the Magnus series $\Omega(\tau, t_0)$,

$$
\begin{aligned}
B(\tau, t_0) &= \tfrac{1}{2}\big(A(t_0 + c_1\tau) + A(t_0 + c_2\tau)\big) - \tfrac{\sqrt{3}}{12}\tau[A(t_0 + c_1\tau), A(t_0 + c_2\tau)], \\
c_{1,2} &= \tfrac{1}{2} \pm \tfrac{\sqrt{3}}{6}.
\end{aligned} \tag{5.20b}
$$

Following [12, Section 3] for the classical defect, the symmetrized defect (5.10) is now given by

$$
\mathcal{D}_s(\tau, t_0) = \big(\Gamma(\tau, t_0) - \tfrac{1}{2}A(t_0 + \tau)\big)\mathcal{S}(\tau, t_0) - \tfrac{1}{2}\mathcal{S}(\tau, t_0)A(t_0),
$$

where $\Gamma(\tau, t_0)$ has a series representation analogous to (5.18). To approximate $\mathcal{D}_s(\tau, t_0)$ in an asymptotically correct way, we again truncate the series defining $\Gamma(\tau, t_0)$ and obtain the *Taylor variant*

$$
\mathcal{D}_s(\tau, t_0) \approx \big(\widetilde{\Gamma}(\tau, t_0) - \tfrac{1}{2}A(t_0 + \tau)\big)\mathcal{S}(\tau, t_0) - \tfrac{1}{2}\mathcal{S}(\tau, t_0)A(t_0),
$$

where

$$
\begin{aligned}
\widetilde{\Gamma}(\tau, t_0) &= B(\tau, t_0) + \tau \check{B}(\tau, t_0) + \tfrac{1}{2}\tau^2[B(\tau, t_0), \check{B}_j(\tau, t_0)] \\
&\quad + \tfrac{1}{6}\tau^3[B(\tau, t_0), [B(\tau, t_0), \check{B}(\tau, t_0)]] \\
&\quad + \tfrac{1}{24}\tau^4[B(\tau, t_0), [B(\tau, t_0), [B(\tau, t_0), \check{B}(\tau, t_0)]]],
\end{aligned}
$$

with

$$
\begin{aligned}
\check{B}(\tau, t_0) &= \big(\partial_1 - \tfrac{1}{2}\partial_2\big)B(\tau, t_0) \\
&= \tfrac{1}{2}\big((c_1 - \tfrac{1}{2})A'(t_0 + c_1\tau) + (c_2 - \tfrac{1}{2})A'(t_0 + c_2\tau)\big) \\
&\quad - \tfrac{\sqrt{3}}{12}[A(t_0 + c_1\tau), A(t_0 + c_2\tau)] \\
&\quad - \tfrac{\sqrt{3}}{12}(c_1 - \tfrac{1}{2})\tau[A'(t_0 + c_1\tau), A(t_0 + c_2\tau)] \\
&\quad - \tfrac{\sqrt{3}}{12}(c_2 - \tfrac{1}{2})\tau[A(t_0 + c_1\tau), A'(t_0 + c_2\tau)].
\end{aligned} \tag{5.21}
$$

Due to $c_1 + c_2 = 1$ it follows by expansion in τ that $\check{B}(\tau, t_0) = \mathcal{O}(\tau)$. Thus, truncation after $p = 4$ again yields a sufficiently accurate approximation. Alternatively, application of fourth order two-sided Hermite quadrature for the approximation of $\Gamma(\tau, t_0)$ yields the *Hermite variant*

$$
\mathcal{D}_s(\tau, t_0) \approx \big(C^+(\tau, t_0) - \tfrac{1}{2}A(t_0 + \tau)\big)\mathcal{S}(\tau, t_0) + \mathcal{S}(\tau, t_0)\big(C^-(\tau, t_0) - \tfrac{1}{2}A(t_0)\big),
$$

where

$$
C^\pm(\tau, t_0) = \tfrac{1}{2}\big(B(\tau, t_0) + \tau \check{B}(\tau, t_0)\big) \pm \tfrac{1}{12}\tau^2[B(\tau, t_0), \check{B}(\tau, t_0)],
$$

with $\check{B}(\tau, t_0)$ as in (5.21).

Table 2
Local error and deviation of the symmetrized defect-based error estimator for the second order Strang splitting applied to (6.1).

τ	$\|\mathcal{L}(\tau, u_0)\|_2$	Order	$\|\widetilde{\mathcal{L}}_s(\tau, u_0) - \mathcal{L}(\tau, u_0)\|_2$	Order
1.563e−02	3.791e−05		3.377e−07	
7.813e−03	4.753e−06	3.00	1.161e−08	4.86
3.906e−03	5.946e−07	3.00	3.726e−10	4.96
1.953e−03	7.434e−08	3.00	1.172e−11	4.99
9.766e−04	9.293e−09	3.00	3.669e−13	5.00
4.883e−04	1.162e−09	3.00	1.160e−14	4.98

Table 3
Global error and corrected solution for the second order Strang splitting applied to (6.1).

τ	Global error	Order	Error of corrected solution	Order
1.563e−02	2.539e−04		5.703e−07	
7.813e−03	6.354e−05	2.00	3.634e−08	3.97
3.906e−03	1.589e−05	2.00	2.283e−09	3.99
1.953e−03	3.972e−06	2.00	1.428e−10	4.00
9.766e−04	9.931e−07	2.00	8.928e−12	4.00
4.883e−04	2.483e−07	2.00	5.611e−13	3.99

Table 4
Local error and deviation of the symmetrized defect-based error estimator for the fourth order integrator from [9, Emb 4/3 AK s] applied to (6.1).

τ	$\|\mathcal{L}(\tau, u_0)\|_2$	Order	$\|\widetilde{\mathcal{L}}_s(\tau, u_0) - \mathcal{L}(\tau, u_0)\|_2$	Order
3.125e−02	7.017e−06		3.420e−07	
1.563e−02	2.282e−07	4.94	2.646e−09	7.01
7.813e−03	7.164e−09	4.99	2.123e−11	6.96
3.906e−03	2.240e−10	5.00	1.706e−13	6.96

6. Numerical examples

We illustrate the theoretical analysis of the deviation of the symmetrized error estimator by showing the orders of the error of the basic integrator and of the deviation of the error estimator from the true error. We will consider splitting methods for a cubic nonlinear Schrödinger equation and commutator-free and classical Magnus-type integrators for a Rosen–Zener model.

6.1. Cubic Schrödinger equation

We solve the cubic nonlinear Schrödinger equation on the real line $x \in \mathbb{R}$

$$\begin{aligned} \mathrm{i}\, \partial_t \psi(x, t) &= -\tfrac{1}{2} \partial_x^2\, \psi(x, t) - |\psi(x, t)|^2\, \psi(x, t), \quad t > 0, \\ \psi(x, 0) &= \psi_0(x) \end{aligned}$$

(6.1)

by splitting methods. Here, a soliton solution exists,

$$\psi(x, t) = 2\, \mathrm{e}^{\mathrm{i}(\frac{3}{2}t - x)}\, \mathrm{sech}(2(t + x))$$

Our initial condition is chosen commensurate with this solution, and we truncate the spatial domain to $x \in [-16, 16]$ and impose periodic boundary conditions. Spectral collocation at 512 equidistant mesh points leads to an ODE system of the form

$$\frac{\mathrm{d}}{\mathrm{d}t} \Psi(t) = F(\Psi(t)) = A\Psi(t) + B(\Psi(t)), \quad \Psi(0) = \Psi_0,$$

with $A\Psi \sim \frac{\mathrm{i}}{2} \partial_x^2 \psi$ and $B(\Psi) \sim \mathrm{i}\, |\psi|^2 \psi$. We solve this by the second order Strang splitting and by the self-adjoint fourth-order method represented by the higher-order method in the embedded pair referred to as Emb 4/3 AK s in the collection [9], recapitulated for easy reference in Table 1 in Section 4.3. The A-part is solved via [I]FFT, while the B-part can be integrated directly on the given mesh.

In Table 2, we give the local error of the Strang splitting and the error of our symmetrized error estimator as compared to the exact errors. Table 3 shows the global errors on the interval [0, 1/8] of the basic integrator and of the solution corrected by adding the error estimate. In accordance with our theory, we observe local orders three and five, respectively, and the expected orders two and four for the global errors. Likewise, Table 4 shows orders five and seven for the local errors of the fourth order integrator from [9, Emb 4/3 AK s], and Table 5 shows the matching global errors.

Table 5
Global error and corrected solution for the fourth order integrator from [9, Emb 4/3 AK s] applied to (6.1).

τ	Global error	Order	Error of corrected solution	Order
3.125e−02	7.894e−06		6.859e−07	
1.563e−02	4.035e−07	4.29	2.771e−09	7.95
7.813e−03	2.471e−08	4.03	2.987e−11	6.54
3.906e−03	1.537e−09	4.01	4.622e−13	6.01
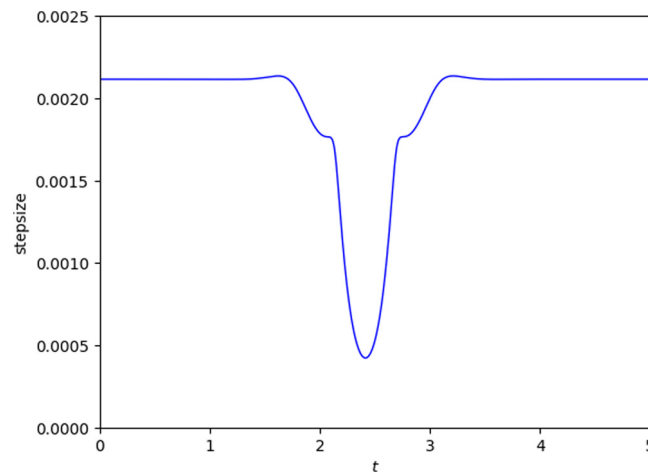
Fig. 4. Step-sizes generated by an adaptive strategy based on the symmetric error estimator for the integrator from [9, Emb 4/3 AK s] for the problem (6.1) with crossing solitons.

Adaptive time-stepping. The error estimators introduced in this paper are intended to be used as the basis for an adaptive time-stepping procedure to enhance the efficiency. To illustrate this aspect, we show step-sizes generated by the standard step-size selection strategy [6]. We solve problem (6.1) with the initial condition

$$\psi(x, 0) = \sum_{j=1}^{2} \frac{a_j e^{-ib_j x}}{\cosh(a_j(x - c_j))}$$

with $a_1 = a_2 = 2$, $b_1 = 1$, $b_2 = -3$, $c_1 = 5$, $c_2 = -5$, and a space discretization at 512 points on the interval $[-16, 16]$. Time integration is effected by the integrator from [9, Emb 4/3 AK s]. This example features two solitons which cross at $t \approx 2.3$, at which point the unsmooth solution demands smaller stepsizes. If we prescribe a tolerance of 10^{-10} on the local error, we obtain the stepsizes shown in Fig. 4. It is found that the stepsizes indeed decrease in the region where the solitons cross, which corresponds with the behavior observed for adaptive time-stepping based on standard error estimators in [14].

6.2. Rosen–Zener model

As a second example, we solve a Rosen–Zener model from [15] by Magnus-type methods. The associated Schrödinger equation in the interaction picture is given by

$$i\dot{\psi}(t) = H(t)\psi(t) \tag{6.2}$$

with

$$H(t) = f_1(t)\sigma_1 \otimes I_{k \times k} + f_2(t)\sigma_2 \otimes R \in \mathbb{C}^{2k \times 2k}, \quad k = 50,$$

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \tag{6.3}$$

$$R = \text{tridiag}(1, 0, 1) \in \mathbb{R}^{k \times k}, \qquad f_1(t) = V_0 \cos(\omega t)(\cosh(t/T_0))^{-1},$$

$$f_2(t) = V_0 \sin(\omega t)(\cosh(t/T_0))^{-1}, \qquad \omega = \tfrac{1}{2},\ T_0 = 1,\ V_0 = 1,$$

subject to the initial condition $\psi(0) = (1, \ldots, 1)^T$.

In Tables 6 and 13, we give the local errors and deviation of the symmetrized error estimators for the test problem (6.2). Table 6 gives the results for the exponential midpoint rule, where the symmetrized defect can be evaluated exactly.

Table 6
Local error and deviation of the symmetrized defect-based error estimator for the second order exponential midpoint rule applied to (6.2).

τ	$\|\mathcal{L}(\tau, u_0)\|_2$	Order	$\|\widetilde{\mathcal{L}}_s(\tau, u_0) - \mathcal{L}(\tau, u_0)\|_2$	Order
1.250e−01	3.343e−03		7.157e−06	
6.250e−02	4.198e−04	2.99	2.251e−07	4.99
3.125e−02	5.254e−05	3.00	7.047e−09	5.00
1.563e−02	6.569e−06	3.00	2.203e−10	5.00
7.813e−03	8.212e−07	3.00	6.885e−12	5.00
3.906e−03	1.026e−07	3.00	2.157e−13	5.00

Table 7
Global error and corrected solution for the exponential midpoint rule applied to (6.2).

τ	Global error	Order	Error of corrected solution	Order
5.000e−01	2.713e−01		7.652e−03	
2.500e−01	6.618e−02	2.04	4.638e−04	4.04
1.250e−01	1.645e−02	2.01	2.880e−05	4.01
6.250e−02	4.106e−03	2.00	1.797e−06	4.00
3.125e−02	1.026e−03	2.00	1.123e−07	4.00
1.563e−02	2.565e−04	2.00	7.018e−09	4.00

Table 8
Local error and deviation of the symmetrized defect-based error estimator for the fourth order CFM integrator [13, CF4:2 in Table 2] applied to (6.2), defect evaluation by Taylor variant.

τ	$\|\mathcal{L}(\tau, u_0)\|_2$	Order	$\|\widetilde{\mathcal{L}}_s(\tau, u_0) - \mathcal{L}(\tau, u_0)\|_2$	Order
5.000e−01	1.884e−03		5.854e−05	
2.500e−01	6.029e−05	4.97	4.875e−07	6.91
1.250e−01	1.892e−06	4.99	3.868e−09	6.98
6.250e−02	5.918e−08	5.00	3.033e−11	6.99
3.125e−02	1.850e−09	5.00	2.373e−13	7.00

Table 9
Global error and corrected solution for the fourth order CFM integrator [13, CF4:2 in Table 2] applied to (6.2), defect evaluation by Taylor variant.

τ	Global error	Order	Error of corrected solution	Order
5.000e−01	2.098e−03		5.330e−05	
2.500e−01	1.212e−04	4.11	7.419e−07	6.17
1.250e−01	7.443e−06	4.03	1.126e−08	6.04
6.250e−02	4.632e−07	4.01	1.745e−10	6.01
3.125e−02	2.892e−08	4.00	2.768e−12	5.98
1.563e−02	1.807e−09	4.00	1.175e−13	4.56

Tables 8 and 10 give the empirical convergence orders for the commutator-free fourth order Magnus-type integrator [13, CF4:2 in Table 2] in conjunction with the symmetrized defect-based error estimator, evaluated by means of the Taylor variant in Table 8 and the Hermite variant in Table 10, respectively (see Fig. 3). Finally, Table 12 gives the result for the classical fourth order Magnus integrator, where the error estimator is evaluated by means of the Hermite variant. Tables 7, 9, 11 and 13 give the corresponding global errors on the interval [0, 1] of the basic solution and of the solution corrected by the symmetric error estimate. In all cases, the theoretical results are well reflected in the numerical experiments.

7. Conclusion

We have discussed a symmetrized defect-based estimator for self-adjoint time discretizations of nonlinear evolution equations. We have introduced the general construction principle extending the ideas from [1], and have elaborated the algorithms for an implicit Runge–Kutta method, for splitting methods and for exponential Magnus-type integrators for time-dependent linear problems. We have proven that the deviation of the estimated error from the true error is two orders in the step-size smaller than the basic integrator, and illustrated the theoretical result for two examples solved by either splitting methods or exponential Magnus-type integrators of different orders.

It can be expected that in adaptive simulations, where choice of the step-size is delicate, the improved accuracy of the error estimator may add to the reliability and efficiency of the integrator. However, this topic exceeds the scope

Table 10

Local error and deviation of the symmetrized defect-based error estimator for the fourth order CFM integrator [13, CF4:2 in Table 2] applied to (6.2), defect evaluation by Hermite variant.

τ	$\|\mathcal{L}(\tau, u_0)\|_2$	Order	$\|\widetilde{\mathcal{L}_s}(\tau, u_0) - \mathcal{L}(\tau, u_0)\|_2$	Order
5.000e−01	1.884e−03		4.008e−05	
2.500e−01	6.029e−05	4.97	3.277e−07	6.93
1.250e−01	1.892e−06	4.99	2.584e−09	6.99
6.250e−02	5.918e−08	5.00	2.023e−11	7.00
3.125e−02	1.850e−09	5.00	1.583e−13	7.00

Table 11

Global error and corrected solution for the fourth order CFM integrator [13, CF4:2 in Table 2] applied to (6.2), defect evaluation by Hermite variant.

τ	Global error	Order	Error of corrected solution	Order
5.000e−01	2.098e−03		3.203e−05	
2.500e−01	1.212e−04	4.11	4.402e−07	6.19
1.250e−01	7.443e−06	4.03	6.702e−09	6.04
6.250e−02	4.632e−07	4.01	1.041e−10	6.01
3.125e−02	2.892e−08	4.00	1.676e−12	5.96
1.563e−02	1.807e−09	4.00	1.052e−13	3.99

Table 12

Local error and deviation of the symmetrized defect-based error estimator for the fourth order classical Magnus integrator (5.20) applied to (6.2), defect evaluation by Hermite variant.

τ	$\|\mathcal{L}(\tau, u_0)\|_2$	Order	$\|\widetilde{\mathcal{L}_s}(\tau, u_0) - \mathcal{L}(\tau, u_0)\|_2$	Order
5.000e−01	4.788e−03		1.214e−04	
2.500e−01	1.618e−04	4.89	1.126e−06	6.75
1.250e−01	5.154e−06	4.97	9.201e−09	6.94
6.250e−02	1.618e−07	4.99	7.269e−11	6.98
3.125e−02	5.064e−09	5.00	5.693e−13	7.00

Table 13

Global error and corrected solution for the fourth order classical Magnus integrator (5.20) applied to (6.2), defect evaluation by Hermite variant.

τ	Global error	Order	Error of corrected solution	Order
5.000e−01	6.957e−03		1.536e−04	
2.500e−01	4.362e−04	4.00	2.452e−06	5.97
1.250e−01	2.728e−05	4.00	3.853e−08	5.99
6.250e−02	1.705e−06	4.00	6.029e−10	6.00
3.125e−02	1.066e−07	4.00	9.419e−12	6.00
1.563e−02	6.662e−09	4.00	1.688e−13	5.80

of the present work and will be explored elsewhere. Here, we have confined ourselves to a numerical illustration that our error estimators induce adaptive step-sizes commensurate with the solution behavior. Note, moreover, that the numerical approximation based on a scheme of order p and corrected by our error estimator (see (3.2)) is very close to self-adjoint and has improved convergence order $p + 2$ (see Theorem 3), thus providing a nearly self-adjoint higher order approximation at moderate computational cost. Since the additive correction is of high order, no stability problems will arise for the corrected scheme (3.2).

Acknowledgments

References

[1] W. Auzinger, O. Koch, An improved local error estimator for symmetric time-stepping schemes, Appl. Math. Lett. 82 (2018) 106–110, http://dx.doi.org/10.1016/j.aml.2018.03.001.

[2] W. Auzinger, O. Koch, M. Thalhammer, Defect-based local error estimators for splitting methods, with application to Schrödinger equations, Part I: The linear case, J. Comput. Appl. Math. 236 (2012) 2643–2659.

[3] W. Auzinger, O. Koch, M. Thalhammer, Defect-based local error estimators for splitting methods, with application to Schrödinger equations, Part II: Higher-order methods for linear problems, J. Comput. Appl. Math. 255 (2013) 384–403.

[4] W. Auzinger, H. Hofstätter, O. Koch, M. Thalhammer, Defect-based local error estimators for splitting methods, with application to Schrödinger equations, Part III: The nonlinear case, J. Comput. Appl. Math. 273 (2014) 182–204.

[5] W. Auzinger, O. Koch, M. Thalhammer, Defect-based local error estimators for high-order splitting methods involving three linear operators, Numer. Algorithms 70 (2015) 61–91.

[6] E. Hairer, S. Nørsett, G. Wanner, Solving Ordinary Differential Equations I, Springer-Verlag, Berlin, Heidelberg, New York, 1987.

[7] S. Descombes, M. Thalhammer, The Lie–Trotter splitting for nonlinear evolutionary problems with critical parameters: A compact local error representation and application to nonlinear Schrödinger equations in the semiclassical regime, IMA J. Numer. Anal. 33 (2012) 722–745.

[8] E. Hairer, C. Lubich, G. Wanner, Geometric Numerical Integration, Springer-Verlag, Berlin, Heidelberg, New York, 2002.

[9] W. Auzinger, H. Hofstätter, O. Koch, Coefficients of various splitting methods, 2017, http://www.asc.tuwien.ac.at/winfried/splitting/.

[10] S. Blanes, F. Casas, J. Oteo, J. Ros, The Magnus expansion and some of its applications, Phys. Rep. 470 (2008) 151–238.

[11] N. Higham, Functions of Matrices: Theory and Computations, SIAM, Philadelphia, PA, 2008.

[12] W. Auzinger, H. Hofstätter, O. Koch, M. Thalhammer, A posteriori error estimation for Magnus-type integrators, M2AN Math. Model. Numer. Anal. (2019) in press https://doi.org/10.1051/m2an/2018050.

[13] A. Alverman, H. Fehske, High-order commutator-free exponential time-propagation of driven quantum systems, J. Comput. Phys. 230 (2011) 5930–5956.

[14] W. Auzinger, I. Březinová, H. Hofstätter, M. Quell, Practical splitting methods for the adaptive integration of nonlinear evolution equations. Part II: Comparisons of local error estimation and step-selection strategies for nonlinear Schrödinger and wave equations, Comput. Phys. Comm. 234 (2019) 55–71.

[15] S. Blanes, F. Casas, M. Thalhammer, High-order commutator-free quasi–Magnus integrators for non-autonomous linear evolution equations, Comput. Phys. Comm. 220 (2017) 243–262.