Noname manuscript No.
(will be inserted by the editor)

High-Order Accurate Local Schemes for Fractional Differential Equations

Daniel Baffet · Jan S. Hesthaven

June 16, 2015

Abstract High-order methods inspired by the multi-step Adams methods are proposed for systems of fractional differential equations. The schemes are based on an expansion in a weighted L^2 space. To obtain the schemes this expansion is terminated after $P+1$ terms. We study the local truncation error and its behavior with respect to the step-size h and P. Building on this analysis, we develop an error indicator based on the Milne device. Methods with fixed and variable step-size are tested numerically on a number of problems, including problems with known solutions, and a fractional version on the Van der Pol equation.

1 Introduction

Interest in fractional calculus has grown over the last decade as it has been demonstrated to provide the right tools for the modeling of anomalous transport and diffusion [1–3]. Such models may describe porous and granular flows, biological processes, and transport in fusion plasmas.

With the discovery of applications employing fractional differential equations (FDEs), comes a need for efficient and reliable numerical methods to approximate their solutions. A number of methods have been proposed for the discretization of spatial fractional operators. A few examples can be found in [4–6] and [7,8]. Some methods have also been proposed for the approximation of time derivatives. The most commonly used are low-order schemes, such as the L1 scheme (e.g., [9] and the references therein) which has been widely used for the approximation of the time fractional diffusion equation. Some high-order methods also exist. For example, the high-order multi-step convolution quadratures (e.g., [10]), proposed in the 80's. More high-order methods for Volterra equations can be found in the monograph [11]. More recently [12,13] a *Discontinuous Galerkin* time-stepping method has been proposed for Volterra equations.

As this paper pertains to time-fractional problems, from this point on we restrict the discussion to this case. This paper pertains to ordinary FDEs on the Caputo form. More precisely, we consider systems of the form

$$D^\alpha u = f(t,u) \ , \tag{1.1}$$

where D^α is the Caputo derivative and $0 < \alpha < 1$. Such systems may describe a process of interest, or may be obtained from a fractional partial differential equation by the discretization of the spatial domain in some way.

The numerical treatment of FDEs must deal with a number of difficulties. Since the fractional derivatives are nonlocal, numerical methods involve global information, and thus require a great deal of computational and memory resources. In time-dependent FDEs, this implies that as the scheme progresses, the required memory and computational effort increase.

D. Baffet, E-mail: daniel.baffet@epfl.ch · J. S. Hesthaven, E-mail: jan.hesthaven@epfl.ch
MATHICSE, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

Another difficulty is due to singular behavior exhibited at the initial time. While a singularity at the initial time can be treated locally, if ignored, it may ruin the accuracy of approximations. One way to treat such a singularity is to grade the step-size near the expected singularity in a particular manner [13]. Another way to overcome this problem, is to pick the time-step adaptively, based on some error indicator.

To deal with the challenges sketched above, a numerical method should exploit any possible advantages. To increase computational efficiency, one may consider high-order methods. Motivated by existing high-order methods for standard ordinary differential equations (ODEs), the methods proposed in this paper are inspired by the high-order multi-step *Adams methods*: the explicit *Adams-Bashforth*, and the implicit *Adams-Moulton* methods (see, e.g., [14]). The difficulty in deriving such methods for FDEs is to approximate the history of the solution. The approximation of past information must be efficient and as accurate as the local approximation. Ideally, we would like to be able to adapt the approximation during the stepping procedure, while maintaining the accuracy.

The methods proposed in this paper are based on an expansion

$$t^{\alpha} D^{\alpha} u(ts) = \sum_{k=0}^{\infty} F_k(t)\, \psi_k(s) \qquad\qquad s \in (0,1) \ , \tag{1.2}$$

in some weighted L^2 space, where $\{\psi_k\}$ is an orthonormal basis in that space. In practice, the infinite sum is truncated after a finite number of terms, say at $k = P$. The coefficients F_k of this expansion are considered as auxiliary variables of the scheme and are advanced in time during the time-stepping procedure. Thus the information used by the scheme is always at the current time, and hence high-order multi-step methods can be employed for the approximation. The orthogonality of $\{\psi_k\}$ and the fact that past information is represented by coefficients at the current time, allow, in principle, to add more terms to the expansion to account for new information. In this work, however, we do not pursue this goal, and focus on developing the basic ideas.

We study the local truncation error of the resulting schemes, and show that it is composed of a local term and a history term. The local term is controlled by the step-size h, and is $O\big(h^{\mu+\alpha}\big)$, and $O\big(h^{\mu+1+\alpha}\big)$ for the explicit and implicit μ-step schemes, respectively. The history term is controlled by P, and can be shown to tend to zero as $P \to \infty$, provided

$$\int_0^t |D^{\alpha} u(\tau)\,|^2\, (t-\tau)^{-1+\alpha}\, \mathrm{d}\tau < \infty \ . \tag{1.3}$$

For applications, however, it is of interest to estimate the convergence rate of this limit. To do this, one must first specify the basis functions ψ_k. We consider the case where ψ_k are the Jacobi polynomials. The derivation of precise estimates on the history term at t requires knowledge of the regularity of $D^{\alpha} u$ in $[0,t]$. At this time, we are not aware of rigorous results connecting properties of f to the necessary regularity of u, and thus unable to complete the analysis in general. Nevertheless, we show that provided (1.3), the history term is $o\big(P^{-\sigma}\big)$, for $0 < \sigma < 2-\alpha$. For specific examples more accurate estimates can be obtained. As an example we show that for $f(t,u) = -u$, the history term is $O\big(P^{-\sigma}\big)$, for $0 < \sigma < 3+\alpha$.

To address the typical singularity at $t = 0$, we employ adaptive time-stepping. This requires an approximation of the local error. Building on the study of the local truncation error, we construct an error indicator, based on the Milne device [14].

Some schemes are tested on several problems. We test two basic schemes, and two adaptive schemes, on problems with known solutions. Finally, we test a high-order adaptive method on a more demanding problem – a fractional version of the Van der Pol equation – and present some numerical results.

The rest of the paper is structured as follows. In §2 we state some basic definitions and results of fractional calculus. In §3 is the derivation of the semi-discrete form, on which the proposed methods are based. The basic methods are presented in §4. We study the local truncation error in §5, and derive an error indicator in §5.3. In §6, we test some methods numerically, and in §7 we present some numerical results obtained for a fractional Van-der-Pol equation. We conclude with some remarks and conclusions in §8. The paper also includes two appendices. Appendix A provides some technical details on polynomial approximation. In Appendix B are details on our implementation of parts of the scheme.

2 Preliminaries

In this section we recall some basic definitions and results which are used in the derivation of the schemes. The reader may wish to consult [15].

For $\alpha > 0$, the fractional integral $\mathcal{I}^\alpha u$ of a function u is defined by

$$\mathcal{I}^\alpha u(t) = \frac{1}{\Gamma(\alpha)} \int_0^t u(\tau)\,(t-\tau)^{-1+\alpha}\,\mathrm{d}\tau \qquad \alpha > 0\ . \tag{2.1}$$

Usually, the fractional integral is defined with a reference to the lower integration limit. For simplicity, in this paper we omit this reference, as the lower integration limit is always zero. The generalization of the methods and results presented in this paper to non-zero lower integration limits is straightforward.

For $0 < \alpha < 1$, the Caputo α-derivative of u is defined by

$$D^\alpha u = \frac{\mathrm{d}}{\mathrm{d}t}\,\mathcal{I}^{1-\alpha}(u - u(0))\ . \tag{2.2}$$

If u is continuous in $[0,T]$ and $u' \in C(0,T] \cap L^1(0,T)$, then

$$D^\alpha u(t) = \mathcal{I}^{1-\alpha} u'(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t u'(\tau)\ (t-\tau)^{-\alpha}\,\mathrm{d}\tau\ , \tag{2.3}$$

which is a more familiar form of the Caputo derivative.

The fractional integral $\mathcal{I}^\alpha$ and Caputo α-derivative D^α, with $0 < \alpha < 1$, satisfy the following relations. If u is continuous in $[0,T]$ then

$$D^\alpha \mathcal{I}^\alpha u = u\ . \tag{2.4}$$

If, in addition, $D^\alpha u$ is continuous in $[0,T]$, then

$$\mathcal{I}^\alpha D^\alpha u = u - u(0)\ . \tag{2.5}$$

Owing to the above, we have the following result: Suppose $0 < \alpha < 1$, $\mathit{\Pi}$ is an open subset of $\mathbb{R}^d$, and $f : [0,T] \times \mathit{\Pi} \to \mathbb{R}^d$ is continuous. If $u : [0,T] \to \mathit{\Pi}$ is continuous and satisfies

$$D^\alpha u = f(t,u) \tag{2.6a}$$

in $(0,T)$, and the initial condition

$$u(0) = u_0\ , \tag{2.6b}$$

then it is also a solution to

$$u = u_0 + \mathcal{I}^\alpha(f(\cdot, u(\cdot))) \tag{2.7}$$

in $(0,T)$. Conversely, if $u : [0,T] \to \mathit{\Pi}$ is $C[0,T]$ and satisfies (2.7), then it is also a solution to the initial value problem (2.6).

Equation (2.7) is a Volterra equation; note that it is also a fractional version of Picard's formula. Picard's formula is used as the starting point for the derivation of Adams methods for standard ODEs. Similarly, the starting point for the derivation of the proposed schemes, which borough from the standard Adams methods, is (2.7).

3 Derivation

In this section we derive a semi-discrete formulation of the proposed schemes. Suppose $0 < \alpha < 1$, and f is continuous. By the results stated in the previous section, the solution to (2.6) is given by (2.7). Hence, the schemes are based on the approximation of (2.7).

In the rest of the paper, I is the interval $I = (0, 1)$, $L^2_\beta(I, \mathbb{R}^m)$ is the space of measurable functions $a : I \to \mathbb{R}^m$, with $\|a\|_\beta < \infty$, where

$$\|a\|_\beta^2 = \int_0^1 |a|^2 \, w_\beta \, \mathrm{d}s \qquad w_\beta(s) = (1-s)^\beta \qquad \beta = -1 + \alpha \ , \tag{3.1}$$

and for $a : I \to \mathbb{R}$, $b : I \to \mathbb{R}^m$,

$$\langle a, b \rangle_\beta = \int_0^1 ab \, w_\beta \, \mathrm{d}s \ . \tag{3.2}$$

(Notice that while $\langle \cdot, \cdot \rangle_\beta$ is an inner product only in the case $m = 1$, (3.2) is well defined for all $m \in \mathbb{N}$.)

Let $u : [0, T] \to \mathbb{R}^d$ be the solution to (2.7), and

$$f^*(t) = f(t, u(t)) \quad \Big(= D^\alpha u(t) \Big) \ . \tag{3.3}$$

Substituting the integration variable $\tau = ts$ into the definition of the fractional integral yields

$$u(t) = u_0 + \frac{t^\alpha}{\Gamma(\alpha)} \int_0^1 f^*(ts) \, (1-s)^{-1+\alpha} \, \mathrm{d}s \ . \tag{3.4}$$

Now, fix $t \in (0, T]$, denote

$$F(s, t) := t^\alpha f^*(ts) \qquad s \in (0, 1) \ , \tag{3.5}$$

and assume

$$F(\cdot, t) = \sum_{k=0}^{\infty} F_k(t) \, \psi_k \tag{3.6}$$

in the $L^2_\beta(I, \mathbb{R}^d)$ norm. Here $\{\psi_k\}$ is an orthonormal basis of $L^2_\beta(I, \mathbb{R})$ such that ψ_0 is a constant. The coefficients F_k of the expansion (3.6) are given by

$$F_k(t) = \langle \psi_k, F(\cdot, t) \rangle_\beta = t^\alpha \int_0^1 f^*(ts) \, \psi_k(s) \, w_\beta(s) \, \mathrm{d}s \ . \tag{3.7}$$

In the schemes studied and tested in this paper, $\{\psi_k\}$ is a sequence of classic orthogonal polynomials (i.e., the Jacobi polynomials translated to the interval $I = (0, 1)$). The reason we assume ψ_0 constant is that the orthogonality of the basis $\{\psi_k\}$ reduces (2.7) to

$$u(t) = u(0) + \frac{1}{\psi_0 \Gamma(\alpha)} F_0(t) \ . \tag{3.8}$$

In the proposed approach, a finite number of coefficients F_k are computed and stored as auxiliary variables, and thus must be updated during the time-stepping procedure. To advance the F_k's, we rely on the following derivation. Fix some positive step-size h, and let

$$\theta = \frac{t}{t+h} \qquad \varphi = 1 - \theta = \frac{h}{t+h} \ . \tag{3.9}$$

By (3.7),

$$\begin{aligned} F_k(t+h) &= (t+h)^\alpha \int_0^1 f^*((t+h)\, s) \, \psi_k(s) \, w_\beta(s) \, \mathrm{d}s \\ &= (t+h)^\alpha \left(\int_0^{\frac{t}{t+h}} + \int_{\frac{t}{t+h}}^1 \right) = I_k^0 + I_k^1 \ . \end{aligned} \tag{3.10}$$

We substitute the integration variable $t\widetilde{s} = (t+h)\,s$ in I_k^0 to obtain

$$I_k^0 = \theta^{1-\alpha} \int_0^1 F(s,t)\,\psi_k(\theta s)\,w_\beta(\theta s)\;\mathrm{d}s\;, \tag{3.11}$$

where the tilde above the integration variable is omitted for the sake of simplicity. Notice that this implies

$$I_k^0 = \Big\langle R_k(\cdot,\theta)\,,F(\cdot,t)\Big\rangle_\beta \tag{3.12}$$

where $R_k(\cdot,\theta) \in L^2_\beta(I,\mathbb{R})$ is given by

$$R_k(s,\theta) = \theta^{1-\alpha}\left(\frac{1-s}{1-\theta s}\right)^{1-\alpha}\psi_k(\theta s) \qquad\qquad s\in(0,1)\;. \tag{3.13}$$

Thus we can substitute (3.6) into (3.11) and change the order of summation and integration to get

$$I_k^0 = \sum_{m=0}^{\infty} R_{km}(\theta)\,F_m(t) \tag{3.14}$$

where

$$R_{km}(\theta) = \Big\langle \psi_m, R_k(\cdot,\theta)\Big\rangle_\beta = \theta^{1-\alpha}\int_0^1 \psi_m(s)\,\psi_k(\theta s)\,w_\beta(\theta s)\;\mathrm{d}s\;. \tag{3.15}$$

Substituting the integration variable $h\widetilde{s} = (t+h)\,s - t$ into I_k^1 yields

$$I_k^1 = h^\alpha \int_0^1 f^*(t+hs)\,\psi_k(\theta+\varphi s)\,w_\beta(s)\;\mathrm{d}s \tag{3.16}$$

(where, again, the tilde above the integration variable is omitted). Combining (3.14) and (3.16) we recover

$$F_k(t+h) = h^\alpha \mathcal{J}_k\big(f^*;t,h\big) + \sum_{m=0}^{\infty} R_{km}(\theta)\,F_m(t) \qquad\qquad k=0,1,\ldots \tag{3.17}$$

where

$$\mathcal{J}_k\big(f^*;t,h\big) = \int_0^1 f^*(t+hs)\,\psi_k(\theta+\varphi s)\;w_\beta(s)\;\mathrm{d}s\;. \tag{3.18}$$

By taking the difference of (3.8) at $t+h$ and at t we recover

$$u(t+h) = u(t) + \frac{1}{\psi_0\Gamma(\alpha)}\left[F_0(t+h) - F_0(t)\right]\;, \tag{3.19}$$

and by substituting (3.17) into (3.19) with $k=0$, we have

$$\begin{aligned} u(t+h) = u(t) &+ \frac{h^\alpha}{\Gamma(\alpha)}\int_0^1 f^*(t+hs)\;w_\beta(s)\;\mathrm{d}s \\ &- \frac{1}{\psi_0\Gamma(\alpha)}\left[F_0(t) - \sum_{m=0}^{\infty} R_{0m}(\theta)\,F_m(t)\right]\;. \end{aligned} \tag{3.20}$$

Explicit schemes are based on the discretization of (3.19) and (3.17), while implicit schemes are based on the discretization of (3.20) and (3.17). The approximation requires a finite number of coefficients, say, $F_0,\ldots,F_P$, and an approximation of (3.18). The schemes also require the computation of the integrals (3.15), but these integrals do not depend on u or $F_0,\ldots,F_P$, and are therefore considered as coefficients of the schemes, in contrast to the approximated dependent variables. The approximation of (3.18), and the coefficients (3.15) is discussed in Appendix B.

Notice that by (3.12),

$$F_k(t+h) = h^\alpha \mathcal{J}_k\big(f^*;t,h\big) + \Big\langle R_k(\cdot,\theta)\,,F(\cdot,t)\Big\rangle_\beta\;, \tag{3.21}$$

and similarly (3.19) and (3.20) can be written as

$$u(t+h) = u(t) + \frac{h^\alpha}{\Gamma(\alpha)} \int_0^1 f^*(t+hs)\; w_\beta(s)\; \mathrm{d}s \;\; - \frac{1}{\Gamma(\alpha)} \Big\langle H(\cdot,\theta)\,, F(\cdot,t) \Big\rangle_\beta \tag{3.22}$$

where

$$H(s,\theta) = 1 - \theta^{1-\alpha} \left(\frac{1-s}{1-\theta s} \right)^{1-\alpha} \quad . \tag{3.23}$$

This is used in the analysis of the local truncation error in §5.

Lastly, a comment should be made on the value of F_k at $t = 0$. While the initial value u_0 of u at the initial time is specified in the statement of the problem, the initial values of F_k must be obtained by other means. Equation (3.7) implies that provided f is bounded near $(0, u_0)$, for each k, there holds $F_k(0) = 0$.

4 Basic schemes

The schemes presented in this section explore the formulations presented in the previous section as a starting point. The explicit and implicit schemes are based on (3.17), (3.19), and (3.17), (3.20) respectively.

To derive numerical schemes from the formulations of the previous section, one has to approximate the integrals (3.18). The orthonormal functions ψ_k must also be specified, although, the precise choice of ψ_k does not affect the formal writing of the schemes. Therefore, in this section we assume $\psi_0, \ldots, \psi_P$ are given.

Since the functionals (3.18) depend only on information originating in the interval $(t, t+h)$, one may consider to adapt to this purpose ideas used in numerical schemes for standard ODEs. The current schemes borough their inspiration from the multi-step Adams methods. We remark that adapting Runga-Kutta methods to this purpose may also be viable, however we have not pursued this.

In the following we use the notation

$$\theta_n = \frac{t_n}{t_n + h} \qquad \varphi_n = 1 - \theta_n = \frac{h}{t_n + h} \qquad c = \frac{1}{\psi_0 \Gamma(\alpha)} \; . \tag{4.1}$$

The approximation of u is denoted by v, and for each $k = 0, \ldots, P$, the approximation of F_k is denoted by G_k. We also write

$$f^n = f(t_n, v^n) \;\; . \tag{4.2}$$

Similar to Adams methods, to approximate (3.18), we replace f^* by an interpolating polynomial f_μ, and thus get the approximation

$$\mathcal{J}_k(f_\mu; t, h) = \int_0^1 f_\mu(t+hs)\, \psi_k(\theta + \varphi s)\; w_\beta(s)\; \mathrm{d}s \;\; . \tag{4.3}$$

The different schemes correspond to different interpolating polynomials.

Here and in the rest of the paper, when referring to the schemes developed here, we avoid the term "order of the scheme". The reason for this is that the definition of a scheme's order traditionally corresponds to the asymptotic behavior of the global error of (convergent) schemes. In the case of standard ODEs, the order of the global error is usually smaller by one than the order of the local truncation error. Since we still have not developed estimates on the global error, we do not know if the same is true of the current schemes and problems.

Below are given the explicit and implicit μ-step schemes:

Explicit μ-step scheme Inspired by the μ-step, μ-order Adams-Bashforth scheme:

$$G_k^{n+1} = h^\alpha \sum_{m=-\mu+1}^{0} f^{n+m}\, J_k^m(\theta_n) + \sum_{m=0}^{P} R_{km}(\theta_n)\, G_m^n \tag{4.4a}$$

$$v^{n+1} = v^n + c\left[G_0^{n+1} - G_0^n\right] \tag{4.4b}$$

Here,

$$J_k^{-m}(\theta) = \int_0^1 \ell_m(s)\,\psi_k(\theta+\varphi s)\,w_\beta(s)\,\mathrm{d}s \qquad m = 0,\ldots,\mu-1 \tag{4.4c}$$

and

$$\ell_m(s) = \prod_{\substack{j=0\\ j\neq m}}^{\mu-1}\left(\frac{s+j}{j-m}\right) = \frac{(-1)^m}{(\mu-1)!}\binom{\mu-1}{m}\prod_{\substack{j=0\\ j\neq m}}^{\mu-1}(s+j) \qquad m = 0,\ldots,\mu-1\ . \tag{4.4d}$$

This scheme is obtained by substituting

$$f_\mu(\tau) = \sum_{m=-\mu+1}^{0} f^{n+m}\ell_{-m}\Big(\frac{\tau-t}{h}\Big)\ . \tag{4.5}$$

into (4.3).

Implicit μ-step scheme Inspired by the μ-step, $\mu+1$-order Adams-Moulton scheme:

$$v^{n+1} = v^n + h^\alpha \sum_{m=-\mu+1}^{1} a_m f^{n+m} - c\left[G_0^n - \sum_{m=0}^{P} R_{0m}(\theta_n)\,G_m^n\right] \tag{4.6a}$$

$$G_k^{n+1} = h^\alpha \sum_{m=-\mu+1}^{1} f^{n+m} J_k^m(\theta_n) + \sum_{m=0}^{P} R_{km}(\theta_n)\,G_m^n \tag{4.6b}$$

$$a_{-m} = \frac{1}{\Gamma(\alpha)}\int_0^1 \ell_m w_\beta\,\mathrm{d}s \tag{4.6c}$$

$$J_k^{-m}(\theta) = \int_0^1 \ell_m(s)\,\psi_k(\theta+\varphi s)\,w_\beta(s)\,\mathrm{d}s \qquad m = -1,\ldots,\mu-1 \tag{4.6d}$$

$$\ell_m(s) = \prod_{\substack{j=-1\\ j\neq m}}^{\mu-1}\left(\frac{s+j}{j-m}\right) = \frac{(-1)^{m+1}}{\mu!}\binom{\mu}{m+1}\prod_{\substack{j=-1\\ j\neq m}}^{\mu-1}(s+j) \qquad m = -1,\ldots,\mu-1 \tag{4.6e}$$

This scheme is obtained by substituting

$$f_\mu(\tau) = \sum_{m=-\mu+1}^{1} f^{n+m}\ell_{-m}\Big(\frac{\tau-t}{h}\Big) \tag{4.7}$$

into (4.3). In this scheme, at each step, first (4.6a) is solved for v^{n+1}, which is then used to calculate G_k^{n+1}, $(k = 0,\ldots,P)$, by (4.6b).

Remark 1 When substituting $\alpha = 1$, the schemes coincide with the Adams-Bashforth and Adams-Moulton schemes for the ODE $u' = f(t,u)$. In that case, $R_{0m}(\theta) = 0$ for $m \neq 0$ and $R_{00}(\theta) = 1$. Therefore P does not matter, and in fact can be set to zero. Of course, since R_{km} are computed numerically, very large values of P may add numerical errors which accumulate, thus the scheme is slowed and the accuracy may only be harmed. So it is advised to set $P = 0$, when $\alpha = 1$.

Remark 2 The size of the algebraic system produced by the implicit schemes is the size of u. In particular, it does not grow with P. In addition, similarly to the properties of the explicit Adams-Bashforth and the implicit Adams-Moulton schemes, the implicit schemes are of higher order and our experience is that implicit schemes are more robust.

Remark 3 In practice, we approximate (4.3) in (3.18) by a Gauss quadrature directly, that is, we do not compute the corresponding sums in (4.4a) and (4.6b). These terms are written as sums with variable coefficients because the coefficients do not depend on v or $G_0,\ldots,G_P$, and therefore can be computed in advance or interpolated from existing data.

Below are some schemes written explicitly:

Explicit 1-step scheme

$$G_k^{n+1} = h^\alpha f^n J_k(\theta_n) + \sum_{m=0}^{P} R_{km}(\theta_n)\, G_m^n \tag{4.8a}$$

$$v^{n+1} = v^n + c\left[G_0^{n+1} - G_0^n\right] \tag{4.8b}$$

$$J_k(\theta) = \int_0^1 \psi_k(\theta + \varphi s)\ (1-s)^{-1+\alpha}\ \mathrm{d}s \tag{4.8c}$$

Explicit 2-step scheme

$$G_k^{n+1} = h^\alpha \sum_{m=-1}^{0} f^{n+m} J_k^m(\theta_n) + \sum_{m=0}^{P} R_{km}(\theta_n)\, G_m^n \tag{4.9a}$$

$$v^{n+1} = v^n + c\left[G_0^{n+1} - G_0^n\right] \tag{4.9b}$$

$$J_k^0(\theta) = \int_0^1 (s+1)\,\psi_k(\theta + \varphi s)\ (1-s)^{-1+\alpha}\ \mathrm{d}s \tag{4.9c}$$

$$J_k^{-1}(\theta) = -\int_0^1 s\,\psi_k(\theta + \varphi s)\ (1-s)^{-1+\alpha}\ \mathrm{d}s \tag{4.9d}$$

Implicit 1-step scheme

$$v^{n+1} = v^n + h^\alpha \sum_{m=0}^{1} a_m f^{n+m} - c\left[G_0^n - \sum_{m=0}^{P} R_{0m}(\theta_n)\, G_m^n\right] \tag{4.10a}$$

$$G_k^{n+1} = h^\alpha \sum_{m=0}^{1} f^{n+m} J_k^m(\theta_n) + \sum_{m=0}^{P} R_{km}(\theta_n)\, G_m^n \tag{4.10b}$$

$$a_0 = \frac{\alpha}{\Gamma(2+\alpha)} \qquad a_1 = \frac{1}{\Gamma(2+\alpha)} \tag{4.10c}$$

$$J_k^0(\theta) = \int_0^1 (1-s)\,\psi_k(\theta + \varphi s)\ (1-s)^{-1+\alpha}\ \mathrm{d}s \tag{4.10d}$$

$$J_k^1(\theta) = \int_0^1 s\,\psi_k(\theta + \varphi s)\ (1-s)^{-1+\alpha}\ \mathrm{d}s \tag{4.10e}$$

5 The local truncation error

5.1 General results

In this section we study the local truncation error of schemes (4.4) and (4.6). We start with (4.4). Let $\mathcal{H}_{k,h}$ be the operator defined by the right hand side of (4.4a), that is

$$G_k^{n+1} = \mathcal{H}_{k,h}(t_n, v, G_0, \ldots, G_P)\ , \tag{5.1}$$

and let

$$T_{k,h}^P(t) = F_k(t+h) - \mathcal{H}_{k,h}(t, u, F_0, \ldots, F_P) \tag{5.2}$$

be the associated local truncation error. Also, let $\mathcal{H}_{v,h}$ be the operator defined by the right hand side of (4.4b),

$$v^{n+1} = \mathcal{H}_{v,h}(t_n, v, G_0, \ldots, G_P) \tag{5.3}$$

and let

$$T^P_{v,h}(t) = u(t+h) - \mathcal{H}_{v,h}(t, u, F_0, \dots, F_P) \tag{5.4}$$

be the associated local truncation error.

Observing that

$$\mathcal{H}_{v,h}(t, v, G_0, \dots, G_P) = v - cG_0(t) + c\mathcal{H}_{0,h}(t, v, G_0, \dots, G_P) \ , \tag{5.5}$$

it follows that

$$T^P_{v,h}(t) = u(t+h) - u(t) - c\,[F_0(t+h) - F_0(t)] + cT^P_{0,h}(t) \tag{5.6}$$

which, by (3.19), implies

$$T^P_{v,h}(t) = cT^P_{0,h}(t) \ . \tag{5.7}$$

Especially, $T^P_{v,h}$ is proportional to $T^P_{0,h}$ and therefore we can concentrate on developing estimates to $T^P_{k,h}$, knowing that the estimate of $T^P_{v,h}$ does not provide new information. Note that this argument also works for scheme (4.6).

In the following, π_P is the operator defined by

$$\pi_P\, a = \sum_{k=0}^{P} \langle \psi_k, a \rangle_\beta \, \psi_k \qquad\qquad a \in L^2_\beta(I, \mathbb{R}^m) \ , \quad (m = 1, d) \ . \tag{5.8}$$

The most general result is as follows.

Proposition 1 *Let $t > 0$, and suppose $D^\alpha u$ has μ continuous derivatives in a neighborhood of t. Then, the local truncation error $T^P_{k,h}$ of (4.4), defined by (5.2), satisfies*

$$T^P_{k,h}(t) \sim d_\mu \psi_k(1)\ h^{\mu+\alpha}\,(D^\alpha u)^{(\mu)}(t) + \Sigma^P_{k,h}(t) \qquad\qquad h \to 0^+ \ , \tag{5.9}$$

where

$$d_\mu = \frac{1}{\mu!} \int_0^1 \left[\prod_{j=0}^{\mu-1} (s+j) \right] w_\beta(s)\ \mathrm{d}s \ , \tag{5.10}$$

and

$$\Sigma^P_{k,h}(t) = \Big\langle\, (1-\pi_P)\, R_k(\cdot, \theta)\,, (1-\pi_P)\ F(\cdot, t) \,\Big\rangle_\beta \ . \tag{5.11}$$

If $D^\alpha u$ satisfies

$$\int_0^t |D^\alpha u(\tau)\,|^2 (t-\tau)^{-1+\alpha}\ \mathrm{d}\tau < \infty \ , \tag{5.12}$$

then

$$\lim_{P\to\infty} \Sigma^P_{k,h}(t) = 0 \ . \tag{5.13}$$

Notice that (5.12) is satisfied, if, for example, $D^\alpha u$ is continuous in $[0, t]$. We remark that a rough estimate of the convergence rate of (5.13) can be obtained from Proposition 3.

Proof By (3.21), we have

$$F_k(t+h) = h^\alpha \mathcal{J}_k\big(f^*; t, h\big) + \Big\langle R_k(\cdot, \theta)\,, F(\cdot, t) \Big\rangle_\beta \ . \tag{5.14}$$

Similarly we get

$$\mathcal{H}_{k,h}(t, u, F_0, \dots, F_P) = h^\alpha \mathcal{J}_k(f_\mu; t, h) + \Big\langle R_k(\cdot, \theta)\,, \pi_P\, F(\cdot, t) \Big\rangle_\beta \ , \tag{5.15}$$

where f_μ is the polynomial

$$f_\mu(\tau) = \sum_{m=-\mu+1}^{0} f^*(t+hm)\, \ell_{-m}\Big(\frac{\tau - t}{h}\Big) \ . \tag{5.16}$$

Here ℓ_m $(m = 0, \ldots, \mu - 1)$ are defined by (4.4d). Since $\mathcal{J}(\cdot; t, h)$ is linear, and π_P satisfies

$$\langle \pi_P a, b \rangle_\beta = \langle a, \pi_P b \rangle_\beta = \langle \pi_P a, \pi_P b \rangle_\beta \ ,$$

$T^P_{k,h}$ is given by

$$T^P_{k,h}(t) = h^\alpha \mathcal{J}_k\big(f^* - f_\mu; t, h\big) + \Sigma^P_{k,h}(t) \ , \tag{5.17}$$

where $\Sigma^P_{k,h}$ is given by (5.11). Next we show (5.9) for scalar equations. Suppose $t > 0$, and $D^\alpha u = f^*$ has μ continuous derivatives in a neighborhood of t. By a standard result on polynomial interpolation [16], we have

$$f^*(t + hs) - f_\mu(t + hs) = \frac{h^\mu}{\mu!} \, (f^*)^{(\mu)}(t + h\zeta_s) \prod_{j=0}^{\mu-1} (s + j) \qquad\qquad s \in (0, 1) \tag{5.18}$$

where $\zeta_s \in (-\mu + 1, 1)$. It follows that

$$\mathcal{J}_k\big(f^* - f_\mu; t, h\big) = \frac{h^\mu}{\mu!} \int_0^1 (f^*)^{(\mu)}(t + h\zeta_s) \, \psi_k(\theta + \varphi s) \left[\prod_{j=0}^{\mu-1} (s + j) \right] w_\beta(s) \, \mathrm{d}s \ . \tag{5.19}$$

We use the integral mean value theorem, and substitute $f^* = D^\alpha u$ to get

$$\mathcal{J}_k\big(f^* - f_\mu; t, h\big) = d_\mu \psi_k(\theta + \varphi\xi) \ h^\mu \, (D^\alpha u)^{(\mu)}(t + h\zeta) \ , \tag{5.20}$$

where $\zeta \in (-\mu + 1, 1)$, $\xi \in (0, 1)$, and d_μ is given by (5.10). Hence, we have (5.9) for the scalar case. To obtain (5.9) for a system, apply the argument above to each entry of $T_{k,h}$ separately.

To show (5.13), we use the Cauchy-Schwartz inequality to get

$$\left| \Sigma^P_{k,h}(t) \right| \le \| \, (1 - \pi_P) \, R_k(\cdot, \theta) \, \|_\beta \, \| \, (1 - \pi_P) \, F(\cdot, t) \, \|_\beta \ . \tag{5.21}$$

We remark that as $\langle \cdot, \cdot \rangle_\beta$ is not an inner product, (5.21) is not the Cauchy-Schwartz inequality but a direct corollary of it. Since $R_k(s, \theta)$ is in $L^2_\beta(I, \mathbb{R})$ and $F(s, t) = t^\alpha D^\alpha u(ts)$ is in $L^2_\beta(I, \mathbb{R}^d)$ (as functions of s),

$$\| \, (1 - \pi_P) \, F(\cdot, t) \, \|_\beta \ , \ \| \, (1 - \pi_P) \, R_k(\cdot, \theta) \, \|_\beta \ \xrightarrow[P \to \infty]{} \ 0 \ . \tag{5.22}$$

□

Estimating the local truncation error for the implicit scheme (4.6) is done in a similar way. The difference is due to the interpolating polynomial f_μ being

$$f_\mu(\tau) = \sum_{m=-\mu+1}^{1} f^*(t + hm) \, \ell_{-m}\left(\frac{\tau - t}{h} \right) \ , \tag{5.23}$$

where ℓ_m $(m = 0, \ldots, \mu - 1)$ are defined by (4.6e). Thus, we have the following result.

Proposition 2 *Let $t > 0$, and suppose $D^\alpha u$ has $\mu + 1$ continuous derivatives in a neighborhood of t. Then, the local truncation error $T^P_{k,h}$ of (4.6) satisfies*

$$T^P_{k,h}(t) \sim d_\mu \psi_k(1) \ h^{\mu+1+\alpha} \, (D^\alpha u)^{(\mu+1)}(t) + \Sigma^P_{k,h}(t) \qquad\qquad h \to 0^+ \ , \tag{5.24}$$

where

$$d_\mu = \frac{1}{(\mu + 1)!} \int_0^1 \left[\prod_{j=-1}^{\mu-1} (s + j) \right] w_\beta(s) \, \mathrm{d}s \ . \tag{5.25}$$

If $D^\alpha u$ satisfies (5.12), then (5.13) holds.

5.2 Estimates on the history term

While the results of §5.1 imply that if $D^\alpha u$ is continuous in $[0,t]$, then

$$\lim_{P\to\infty} \Sigma_{k,h}^P(t) = 0 \ , \tag{5.26}$$

for applications it is of interest to estimate the convergence rate of this limit.

At this time we do not have a general estimate on the history term. The main difficulty in deriving general estimates is that such estimates require information on the regularity of $D^\alpha u$. As u (and therefore $D^\alpha u$) is determined by f, ideally, we would like to have regularity conditions on f ensuring the estimate. Unfortunately, at this time, we are not aware of such results.

Nevertheless, some information regarding the convergence rate of (5.26) can be obtained. By (5.21), to estimate the history term, it suffices to estimate the projection errors of $F(\cdot,t)$, and $R_k(\cdot,\theta)$. Thus the discussion is divided into two parts. In the first we discuss the projection error of $R_k(\cdot,\theta)$, and prove an estimate on this error and present numerical evidence supporting the optimality of the estimate for $k=0$.

The second part pertains to the projection error of $F(\cdot,t)$. Estimating this error requires information on the regularity of $D^\alpha u$. While we do not have the tools to derive such estimates in general, we present an estimate for a specific example of interest, to obtain some intuition of the behavior.

The projection errors depend on the specific basis functions. Here we suppose $\psi_0,\ldots,\psi_P$ are the orthonormal polynomials given by

$$\psi_j(s) = 2^{\alpha/2}\, P_j^{(\beta,0)}(2s-1) \ , \tag{5.27}$$

where $P_j^{(\beta,0)}$ are the classic Jacobi polynomials associated with the weight $w_\beta(\xi) = (1-\xi)^\beta$, normalized such that

$$\int_{-1}^{1} \left(P_j^{(\beta,0)}\right)^2 w_\beta = 1 \ . \tag{5.28}$$

There is no restriction to using other basis functions (as long as ψ_0 is constant), although, other basis functions may yield different estimates.

In the following, A is the classic Sturm-Liouville operator defined by

$$A\psi = -w_\beta^{-1}\Big(s\,(1-s)\,w_\beta\psi'\Big)' = -s\,(1-s)\,\psi'' - (1-(2+\beta)\,s)\,\psi' \ . \tag{5.29}$$

The eigenvalues of A are given by

$$\nu_n = n\,(n+\alpha) \qquad\qquad n=0,1,\ldots \tag{5.30}$$

For $\sigma>0$, $D\Big(A^{\sigma/2}\Big)$ is the domain of $A^{\sigma/2}$.

One way to estimate the projection error of $R_0(\cdot,\theta)$ and $F(\cdot,t)$, is to determine $\overline{\sigma}_1$ and $\overline{\sigma}_2$ such that

$$R_k(\cdot,\theta) \in D\Big(A^{\sigma_1/2}\Big) \qquad\qquad \forall\ 0<\sigma_1<\overline{\sigma}_1 \ , \tag{5.31}$$

and

$$F(\cdot,t) \in D\Big(A^{\sigma_2/2}\Big) \qquad\qquad \forall\ 0<\sigma_1<\overline{\sigma}_2 \ . \tag{5.32}$$

Then, by Lemma A.1,

$$\|\,(1-\pi_P)\,R_k(\cdot,\theta)\,\|_\beta \le P^{-\sigma_1}\ \|A^{\sigma_1/2}R_k(\cdot,\theta)\,\|_\beta \qquad\qquad \forall\ 0<\sigma_1<\overline{\sigma}_1 \ , \tag{5.33}$$

and

$$\|\,(1-\pi_P)\,F(\cdot,t)\,\|_\beta \le P^{-\sigma_2}\ \|A^{\sigma_2/2}F(\cdot,t)\,\|_\beta \qquad\qquad \forall\ 0<\sigma_1<\overline{\sigma}_1 \ , \tag{5.34}$$

which together give

$$\Big|\Sigma_{k,h}^P(t)\Big| \le P^{-\sigma_1-\sigma_2}\ \|A^{\sigma_1/2}R_k(\cdot,\theta)\,\|_\beta\ \|A^{\sigma_2/2}F(\cdot,t)\,\|_\beta \ , \tag{5.35}$$

for all $0<\sigma_1<\overline{\sigma}_1$, and $0<\sigma_2<\overline{\sigma}_2$. The reason we use this approach and not look for a Sobolev space H_β^σ where $R_0(\cdot,\theta)$, or $F(\cdot,t)$ may reside is that H_β^σ is strictly embedded in $D\Big(A^{\sigma/2}\Big)$. Thus, in particular cases, this analysis provides improved estimates [19]. Since R_0 has a singularity at $s=1$, and F typically has a singularity at $s=0$, we expect this approach to yield improved results.

5.2.1 The projection error of $R_k(\cdot,\theta)$

Below we develop estimates on the projection error of $R_k(\cdot,\theta)$, and present some numerical evidence for their optimality.

Proposition 3 *Estimate (5.33) holds with $\overline{\sigma}_1 = 2-\alpha$.*

Proof To prove the proposition, we show (5.31) for $\overline{\sigma}_1 = 2-\alpha$. Let $\overline{\sigma}_1 = 2-\alpha$, and

$$R_k(s,\theta) = g_\theta(s)\, w_{-\beta} \ , \tag{5.36}$$

where

$$g_\theta(s) = \theta^{1-\alpha}\,(1-\theta s)^\beta\, \psi_k(\theta s) \ . \tag{5.37}$$

Notice that for each $\theta \in [0,1)$, g_θ is analytic in a neighborhood of the interval $[0,1]$.

By Proposition A.1,

$$w_\gamma \in D\Big(A^{\sigma/2}\Big) \qquad\qquad 0 < \sigma < 1+\beta+2\gamma = \alpha+2\gamma \ . \tag{5.38}$$

In particular, this implies $w_{-\beta} \in D\Big(A^{\sigma/2}\Big)$ for all $0 < \sigma < \overline{\sigma}_1$.

Next we show (5.31). Fix some $0 < \sigma < \overline{\sigma}_1$. Since g_θ is analytic, we have

$$g_\theta(s) = \sum_{j=0}^{2M-1} g_j(\theta)\,(1-s)^j + r_{\theta,2M-1}(s)\,(1-s)^{2M} \tag{5.39}$$

where M is an arbitrarily large integer and $r_{\theta,2M-1}$ is analytic. Then

$$R_k(s,\theta) = \sum_{j=0}^{2M-1} g_j(\theta)\,(1-s)^{j+1-\alpha} + r_{\theta,2M-1}(s)\,(1-s)^{2M+1-\alpha} \tag{5.40}$$

By Proposition A.1, for all $j = 0,\dots,2M-1$, $(1-s)^{j-1+\alpha}$ are in $D\Big(A^{\sigma/2}\Big)$. Thus to complete the proof, it suffices to show that

$$b_{\theta,2M-1}(s) = r_{\theta,2M-1}(s)\,(1-s)^{2M+1-\alpha} \tag{5.41}$$

is also in $D\Big(A^{\sigma/2}\Big)$, for sufficiently large M. However, this is simple, as $b_{\theta,2M-1}$ is $C^{2M}[0,1]$, and the inclusion

$$C^{2M}[0,1] \subset D\Big(A^M\Big) \subset D\Big(A^{\sigma/2}\Big) \ , \tag{5.42}$$

is valid for $M \geq \sigma/2$. Thus, we have (5.31), which by Lemma A.1 implies (5.33), and the proof is complete. □

We remark that it suffices to take $M = 1$ in the proof.

To validate the optimality of this result, consider

$$A^{\sigma/2} R_k(\cdot,\theta) = \sum_{m=0}^{\infty} \nu_m^{\sigma/2} R_{km}(\theta)\,\psi_m \ , \tag{5.43}$$

where $R_{km}(\theta)$ are defined by (3.15). It follows that (5.31) holds, if

$$\sum_{m=1}^{\infty} \nu_m^{\sigma} |R_{km}(\theta)|^2 < \infty \ . \tag{5.44}$$

Recalling that $\nu_m \sim m^2$, estimate (5.33) with $\overline{\sigma}_1 = 2-\alpha$ is optimal, if

$$R_{km}(\theta) = O\Big(m^{-5/2+\alpha}\Big) \ . \tag{5.45}$$

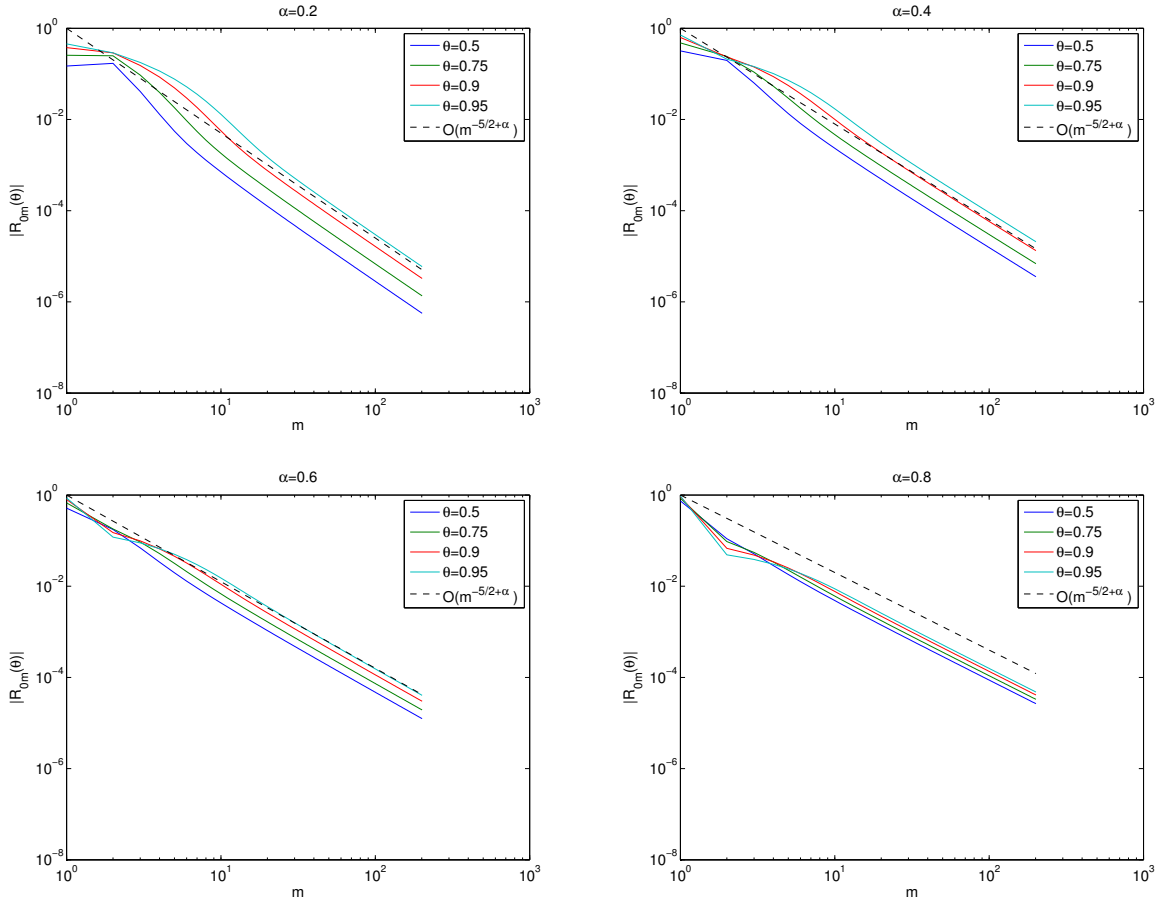

Fig. 1 The absolute value of $R_{0m}(\theta)$ as a function of m, for $\alpha = 0.2, 0.4, 0.6, 0.8$, and several value of θ.

Figure 1 shows the absolute values of the coefficients $R_{0m}(\theta)$ for several values of α and θ. The decay of the coefficients corresponds the predicted rate.

Since, in practice, only a finite number of coefficients is retained, the error of the history term of the highest modes is not controlled by increasing P. That is, there is always a last mode which is not controlled. However, in practice we are only interested in the error of the 0th mode, which can be controlled by the higher modes. To reduce this error we rely on the decay of the coefficients $F_m(t)$ and R_{0m} in m.

5.2.2 An example: the Mittag-Leffler function

Consider the initial value problem

$$D^\alpha u = -u \qquad u(0) = 1 \ . \tag{5.46}$$

The solution u is given by

$$u(t) = E_\alpha(-t^\alpha) \qquad E_\alpha(t) = \sum_{k=0}^{\infty} \frac{t^k}{\Gamma(\alpha k + 1)} \ , \tag{5.47}$$

where E_α is the Mittag-Leffler function. Here,

$$F(s,t) = t^\alpha D^\alpha u(ts) = -t^\alpha E(-t^\alpha s^\alpha) \ . \tag{5.48}$$

Thus, we have

$$F(s,t) = -\sum_{k=0}^{\infty} \frac{t^{\alpha(k+1)}}{\Gamma(\alpha k+1)}\ s^{\alpha k} = -\sum_{k=0}^{M} \frac{t^{\alpha(k+1)}}{\Gamma(\alpha k+1)}\ s^{\alpha k} + t^{\alpha(M+2)} r_M(ts)\ \ s^{\alpha(M+1)}\ . \tag{5.49}$$

By Proposition A.1, the functions $s^{\alpha k}$, with $k = 1, \ldots, M$ are in $D\left(A^{\sigma/2}\right)$ for $0 < \sigma < 1+2\alpha$. In addition, for sufficiently large M, the function

$$-t^{\alpha} + t^{\alpha(M+2)} r_M(ts)\ \ s^{\alpha(M+1)} \tag{5.50}$$

is smooth (as a function of s) in $[0,1]$. Thus, we have (5.32), with $\overline{\sigma}_2 = 1+2\alpha$, which by Lemma A.1 implies (5.34). Together with the decay in the projection error of $R_0(\cdot,\theta)$, we have that $|\Sigma^P_{0,h}|$ decays like $P^{-\sigma}$ for all $0 < \sigma < 3+\alpha$.

5.3 An adaptation of the Milne device

The Milne device is an error indicator used in the numerical approximation of ODEs for adaptive the step-size control. We use the idea of the Milne device to construct an error indicator to our schemes. Suppose G_k^{n+1} is the approximation of $F_k(t_{n+1})$ obtained by substituting the exact solution into the right hand side of (4.6b). Then, by the derivation of the previous section,

$$G_k^{n+1} - F_k(t_{n+1}) \sim c_{\mu k}\, h^{\mu+1+\alpha}\, (D^{\alpha} u)^{(\mu+1)}(t_n) - \Sigma^P_{k,h}(t)\ , \tag{5.51}$$

where

$$c_{\mu k} = \frac{\psi_k(1)}{(\mu+1)!} \int_0^1 (1-s) \left[\prod_{j=0}^{\mu-1} (s+j)\right] w_\beta(s)\ \mathrm{d}s\ . \tag{5.52}$$

Similarly, suppose $\widetilde{G}_k^{n+1}$ is the approximation of $F_k(t_{n+1})$ obtained by substituting the exact solution into the right hand side of (4.4a) with $\mu+1$. Then,

$$\widetilde{G}_k^{n+1} - F_k(t_{n+1}) \sim \widetilde{c}_{\mu+1,k}\, h^{\mu+1+\alpha}\, (D^{\alpha} u)^{(\mu+1)}(t_n) - \Sigma^P_{k,h}(t)\ , \tag{5.53}$$

where

$$\widetilde{c}_{\mu+1,k} = -\frac{\psi_k(1)}{(\mu+1)!} \int_0^1 (s+\mu) \left[\prod_{j=0}^{\mu-1} (s+j)\right] w_\beta(s)\ \mathrm{d}s\ . \tag{5.54}$$

Combining (5.51) and (5.53) we get

$$G_k^{n+1} - F_k(t_{n+1}) \sim \kappa_\mu \left(G_k^{n+1} - \widetilde{G}_k^{n+1}\right) - \Sigma^P_{k,h}(t)\ , \tag{5.55}$$

where

$$\kappa_\mu = \frac{c_{\mu k}}{c_{\mu k} - \widetilde{c}_{\mu+1,k}} = \frac{1}{\mu+1} \frac{\int_0^1 \left[\prod_{j=0}^{\mu-1} (s+j)\right] w_\alpha(s)\ \mathrm{d}s}{\int_0^1 \left[\prod_{j=0}^{\mu-1} (s+j)\right] w_\beta(s)\ \mathrm{d}s}\ . \tag{5.56}$$

Thus, (5.55) with κ_μ defined by (5.56) provides an approximation to the local error. This approximation may be used to control the error by adapting the step-size, and thus make numerical schemes more efficient. The particular way in which this is done in this paper is explained in §6.3.

6 Numerical results

6.1 Test problems

The numerical results are obtained by application of our schemes to the following initial value problems.

Problem 1.

Consider (2.6), with

$$f(t,u) = -u \qquad\qquad u_0 = 1 \tag{6.1}$$

The solution of this problem is given as

$$u(t) = E_\alpha(-t^\alpha) \qquad\qquad E_\alpha(t) = \sum_{k=0}^{\infty} \frac{t^k}{\Gamma(\alpha k + 1)} \ . \tag{6.2}$$

It is smooth in $(0,\infty)$, strictly decreasing, and tends to zero as $t \to \infty$. This makes problem (6.1) relatively simple. However, u and its α-derivative, while continuous, have a singularity at $t = 0$.

Problem 2.

Consider (2.6) with

$$f(t,u) = u^2 - \sin^2(t) + \frac{t^{1-\alpha}}{\Gamma(2-\alpha)}\, {}_1F_2\left(1\,;\ 1+\frac{1-\alpha}{2}\,,\ 1-\frac{\alpha}{2}\,;\ -\frac{t^2}{4}\right) \qquad u_0 = 0\ . \tag{6.3}$$

Here ${}_1F_2$ is a generalized hypergeometric function. This is a nonlinear problem. Its solution is given as $u(t) = \sin(t)$. Here, u is smooth, but its α-derivative has a singularity at $t = 0$.

Problem 3.

Consider (2.6) with

$$f(u) = Au \qquad\qquad A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \ . \tag{6.4a}$$

In this case, we write $u = (x,y)^T$, and similarly, $u_0 = (x_0, y_0)^T$. In our tests we take

$$x_0 = 2 \qquad\qquad y_0 = 0\ . \tag{6.4b}$$

The general solution of (2.6) with (6.4a) is given by

$$u(t) = x_0 \begin{bmatrix} \Re E_\alpha(it^\alpha) \\ -\Im E_\alpha(it^\alpha) \end{bmatrix} + y_0 \begin{bmatrix} \Im E_\alpha(it^\alpha) \\ \Re E_\alpha(it^\alpha) \end{bmatrix} \ . \tag{6.5}$$

6.2 Basic schemes

The results in this section are obtained with the 1-step explicit and implicit schemes. The basis functions ψ_k are orthogonal polynomials given by (5.27). The implicit scheme requires the solution of an algebraic equation at each step. For this, we use a Newton solver.

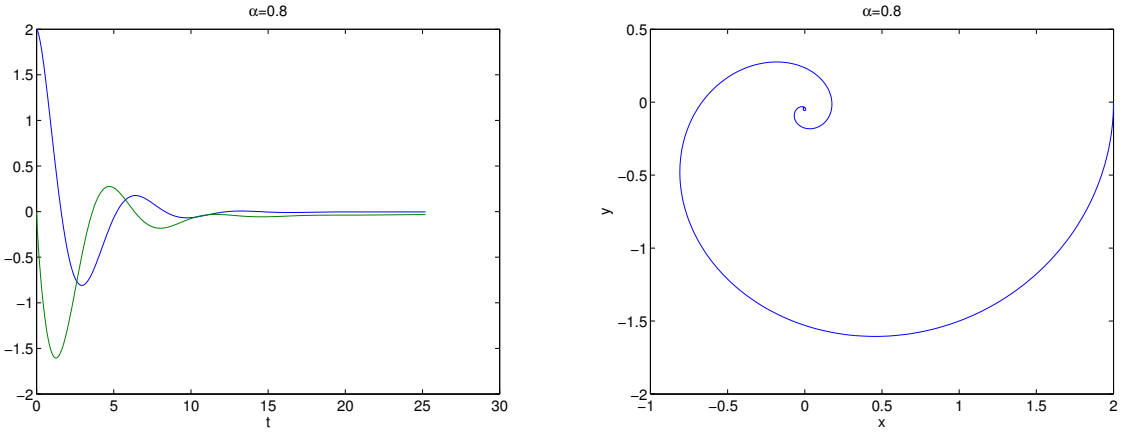

Fig. 2 The solution (6.5) of (6.4) with $\alpha = 0.8$, $x_0 = 2$, and $y_0 = 0$, as computed by Mathematica. On the left are x and y as functions of t, and on the right is the curve $(x(t), y(t))$ in the x-y plane.

Test 1

In this test, the 1-step explicit and 1-step implicit methods are employed to approximate the solution of Problem 1, with $\alpha = 0.2, 0.5$ and 0.8. Figure 3 shows the local error

$$e_h^n = |v^n - u(t_n)| \tag{6.6}$$

as a function of t. Here, v is the numerical solution, and u is the exact solution. The results in the left and right columns are obtained with the explicit and implicit methods, respectively.

Most of the figures show similar behavior of the error, so let us describe this behavior for a particular example. For the discussion, consider the top left figure, corresponding to the approximation of Problem 1, with $\alpha = 0.2$, by the 1-step explicit method.

The figure shows that for $P = 20$, decreasing h from 10^{-1} to 10^{-2} reduces the error as expected. When h is reduced further to 10^{-3}, the error goes down for small t, until at some point it grows. This growth indicates that the value of P limits the accuracy. Indeed, when P is increased to 40, the error is reduced in the entire interval. This behavior can be also observed in the other figures.

The results also show that the error near the initial condition is larger than the error for larger t. This behavior is possibly due to the dissipation in the problem. We conjecture, however, that the large error near the initial condition is caused by the singularity in the α-derivative of u at $t = 0$. Indeed, for larger values of α, where the singularity at $t = 0$ is less significant, the error appears to be more uniform.

A comment should be made regarding the "spikes" showing in the bottom right figure, corresponding to the implicit scheme applied to the problem with $\alpha = 0.8$. The spikes are apparent at $t \approx 1.57$ in all the graphs where the error is of order 10^{-5} or smaller at that t. We note that all these spikes in the error occur at about the same time, and the error does not seem to be affected by these spikes in later times. In particular, the jumps in the graphs of the errors computed with $h = 10^{-3}$ are an order of magnitude larger than the error in a neighborhood of the jump, and otherwise the graphs seem unaffected by the sudden increase of the error. This may indicate that the observed jumps in the errors are caused by a problem with the reference solution computed by Mathematica and not the scheme.

Comparing the left and right columns, we see that the errors of the implicit method are smaller and decay faster than the errors of the explicit method. This is in agreement with the estimates on the local truncation error developed in §5.1. At this point, we do not attempt to measure the global error's decay rate, as it is evident that it is dominated by the error near the initial condition, due to the singularity. We leave these measurements to the next section, where we test schemes employing adaptive time-stepping that can overcome this difficulty.

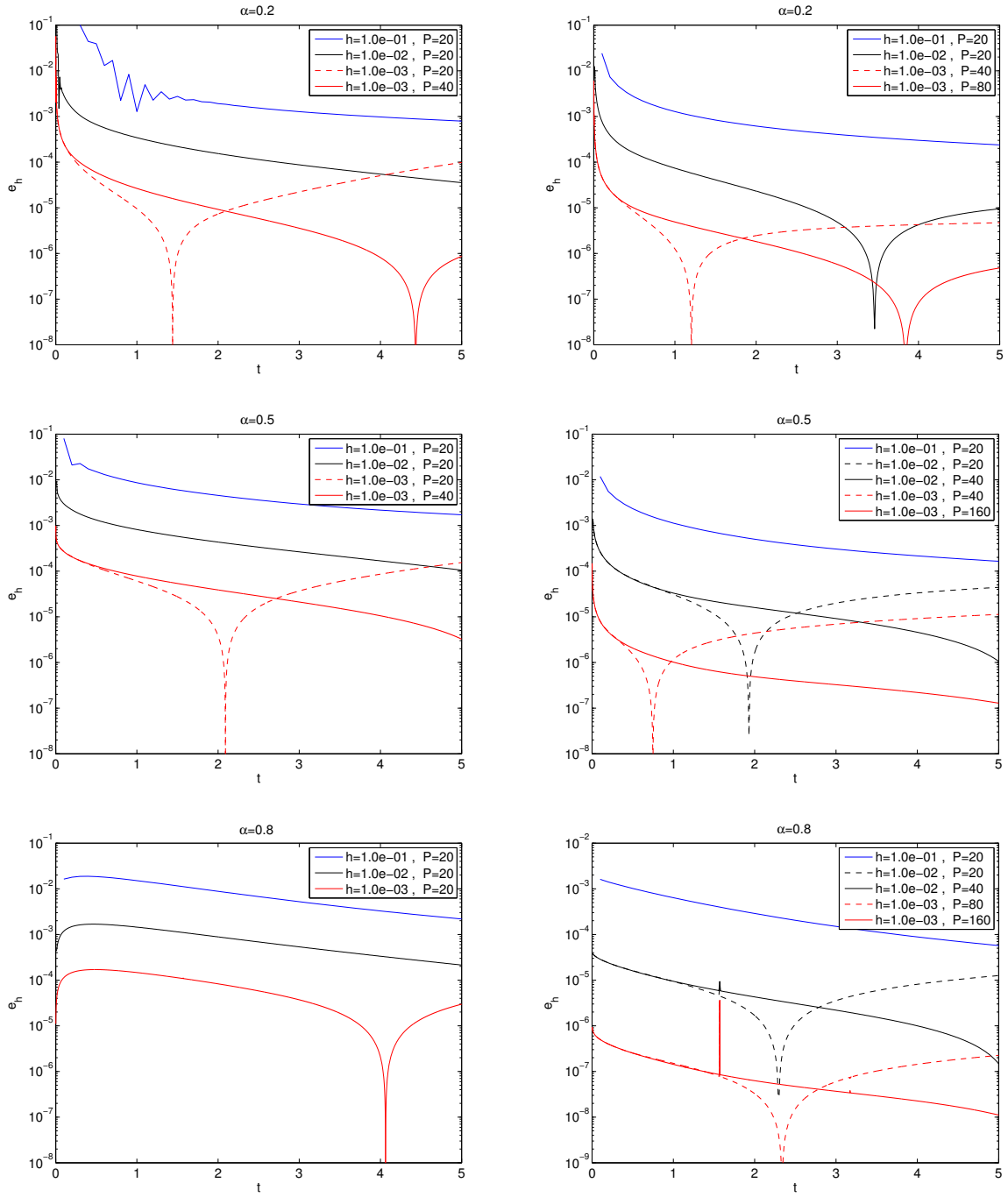

Fig. 3 Test 1. The error e_h of the 1-step explicit method (left column) and 1-step implicit method (right column) applied to (2.6) with (6.1) and different values of α.

6.3 Adaptive time-stepping

The results presented here are obtained with methods employing adaptive time-stepping. The two methods are based on the construction in §5.3, with $\mu = 1$ and $\mu = 3$. The methods are implicit μ-step methods, each utilizes a $\mu + 1$-step explicit method to approximate the local error. We refer to these methods as μ-step adaptive methods, and they should not be confused with the basic schemes, which employ uniform step-size. The error tolerance is enforced by controlling

$$q = \left(\frac{\delta}{\Delta}\right)^{1/(\mu+1+\alpha)} , \tag{6.7}$$

where $\delta > 0$ is the tolerance provided by the user, and

$$\Delta = \kappa \max_{k=0,\dots,P} |\mathcal{H}_{k,h}(t_n, v, G_0, \dots, G_P) - \widetilde{\mathcal{H}}_{k,h}(t_n, v, G_0, \dots, G_P)| . \tag{6.8}$$

Here $\mathcal{H}_{k,h}$ and $\widetilde{\mathcal{H}}_{k,h}$ are the operators defined by the right hand side of (4.6b) and (4.4a), respectively, and κ is given by (5.56). For $\mu = 1$, we have

$$\kappa_1 = \frac{\alpha}{2(2+\alpha)} , \tag{6.9}$$

and for $\mu = 3$,

$$\kappa_3 = \frac{\alpha}{4(4+\alpha)} \frac{27 + 10\alpha + \alpha^2}{18 + 8\alpha + \alpha^2} . \tag{6.10}$$

We remark that when computing the difference $\mathcal{H}_{k,h} - \widetilde{\mathcal{H}}_{k,h}$, the sum $\sum R_{km}(\theta_n)\, G_m^n$ is cancelled out, and we are left with the difference between the approximations to $\mathcal{J}_k$. When $q < 1$ we divide the step-size by two and compute again, and if $q \geq 10$ we multiply the step-size by 2 and advance.

In this section, for a tolerance $\delta > 0$, the local error e_δ is given by

$$e_\delta^n = |v^n - u(t_n)| , \tag{6.11}$$

and the global error $\mathcal{E}_\delta$ is given by

$$\mathcal{E}_\delta = \max_n e_\delta^n . \tag{6.12}$$

The basis functions ψ_k are orthogonal polynomials given by (5.27). The schemes are implicit and thus require the solution of an algebraic equation at each step to advance. For this we use a Newton solver.

Test 2

We apply the 1-step and 3-step adaptive methods to Problem 1, with $\alpha = 0.5$. The results are in Figure 4. The numerical solution v is compared to the exact solution. The reader may wish to compare these results with the results of Test 1 in Figure 3.

On the left and right are results obtained with the 1-step and 3-step methods, respectively. The top row shows the global error $\mathcal{E}_\delta$ as a function of the maximal step-size h_M, the middle row shows the local error e_δ as a function of t, and the bottom row shows the step-size h picked by the program as a function of t. Notice that the step-size never decreases during the simulation. This is because of the regularity and decay of the solution, and occurs, for this problem, whenever P is sufficiently large.

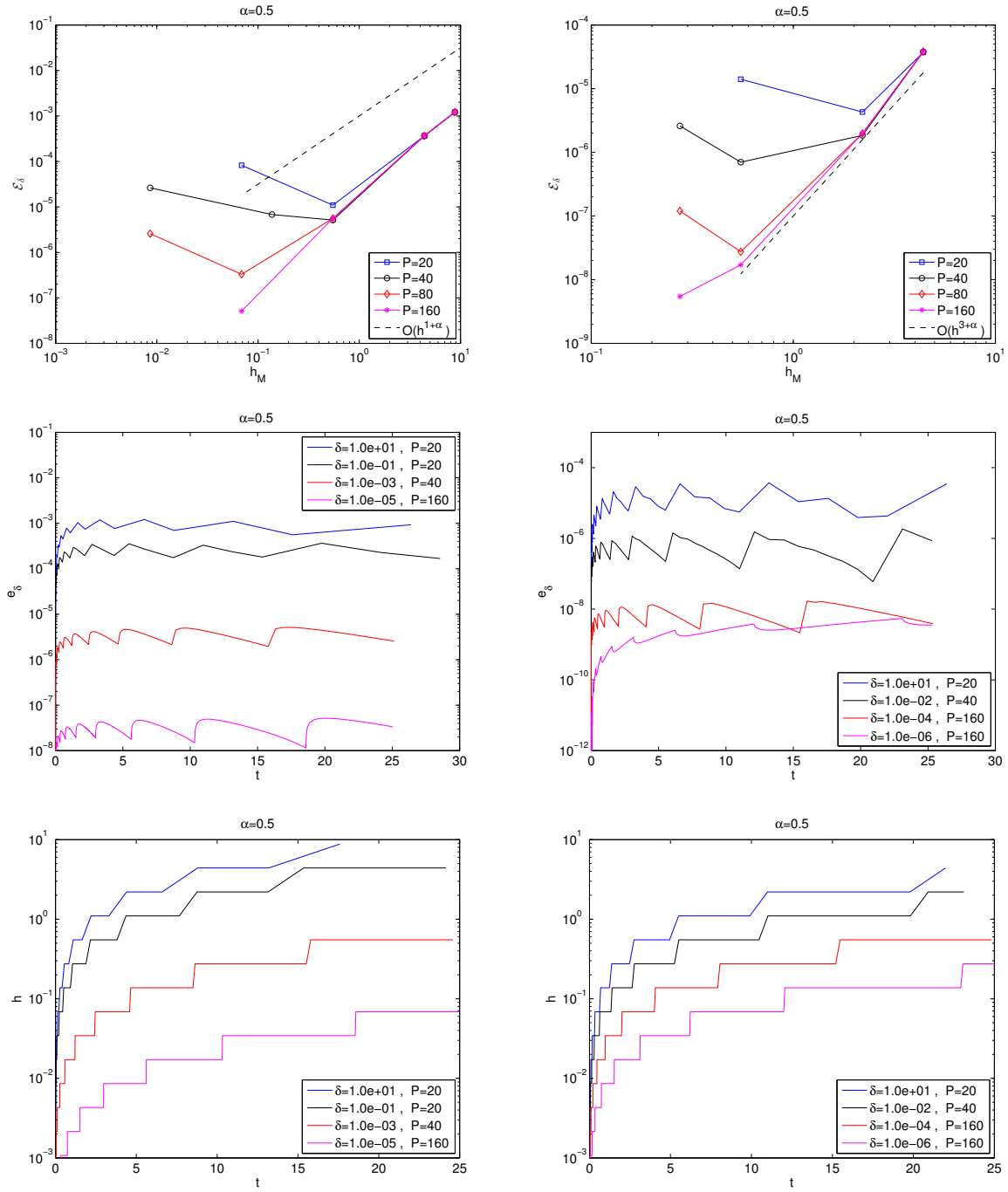

Fig. 4 Test 2. Accuracy tests performed with adaptive methods. Left – 1-step method. Right – 3-step method.

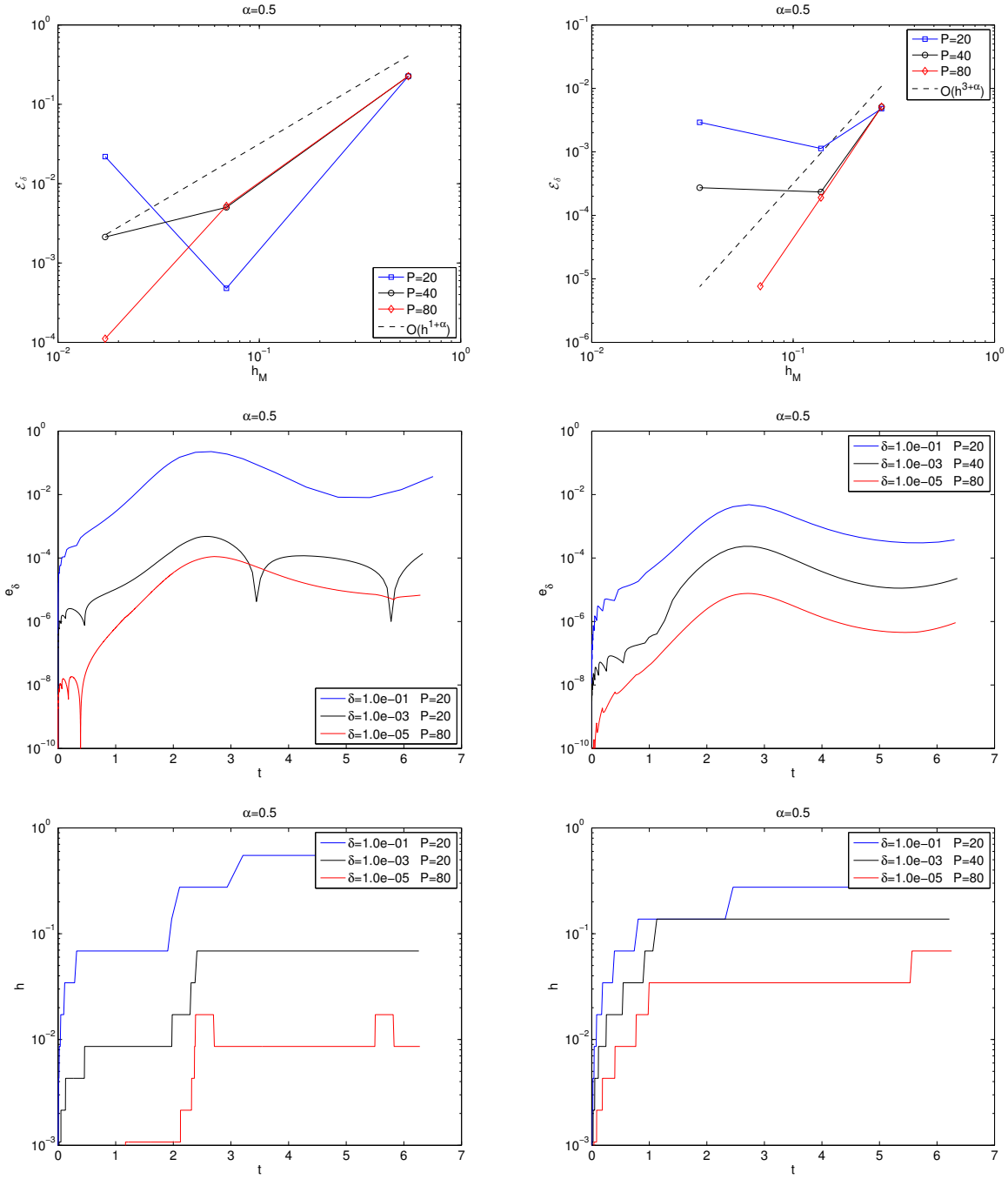

Fig. 5 Test 3. Accuracy tests performed with adaptive methods. Left – 1-step method. Right – 3-step method.

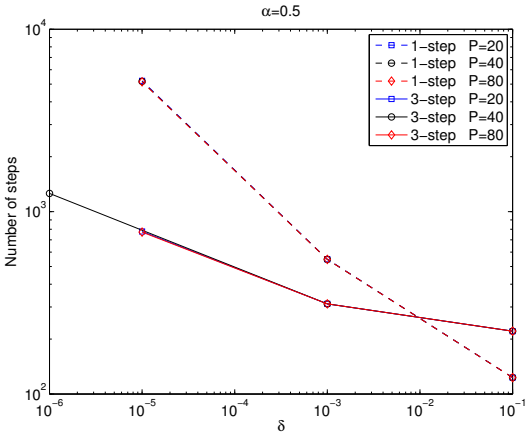

Fig. 6 Test 3. Number of steps as a function of δ.

Test 3

The 1-step and 3-step adaptive methods are applied to Problem 2, with $\alpha = 0.5$, in $t \in (0, 2\pi)$. The results are in figures 5 and 6. Figure 5 shows the global error $\mathcal{E}_\delta$ as a function of h_M (top row), the local error e_δ as a function of t (middle row), and the step-size h as a function of t (bottom row). The left and right columns show results obtained with the 1-step and 3-step schemes, respectively. Except for the test with the 1-step method, $\delta = 10^{-3}$, and $P = 20$, where the error is smaller than expected, the results exhibit the expected behavior described previously.

Figure 6 shows the number of steps as a function of δ. Since the 3-step method does not require much more computational effort than the 1-step method, and in this example, the program rejected very few steps, this plot may be viewed as a measure of the work required by the schemes. The figure shows that the number of steps is not greatly affected by P. It is also evident that the number of steps required by the 1-step method grows faster than the number of steps required for the 3-step method.

Test 4

In this test we apply the 3-step method to Problem 3, with $\alpha = 0.8$. Here, the simulation time is $T = 25$. The results are in figures 7 and 8. The numerical approximation v is compared to the exact solution (6.5). Figure 7 shows the global error $\mathcal{E}_\delta$. On the left $\mathcal{E}_\delta$ is plotted as a function of the maximal step-size. The figure shows that when P is fixed, and the error tolerance δ decreases, the global error goes down until the error reaches a minimal value and then starts increasing. This may be because P is too small. Indeed, when P increases, the error can be reduced further. On the right $\mathcal{E}_\delta$ is plotted as a function of P. This figure shows that for a fixed tolerance δ, when increasing P the error goes down, until the error caused by h dominates. At that point, the error can not be decreased further without enforcing a stricter condition on h. Figure 8 shows the step-size h as a function of t for some values of P and some error tolerances δ.

7 Numerical tests with a fractional Van der Pol equation

Consider the fractional Van der Pol equation

$$\left(D^\alpha\right)^2 x - \varepsilon\left(1 - x^2\right) D^\alpha x + x = 0 \tag{7.1a}$$

in $(0, T)$, and initial conditions

$$x(0) = x_0 \qquad D^\alpha x(0) = y_0 \ . \tag{7.1b}$$

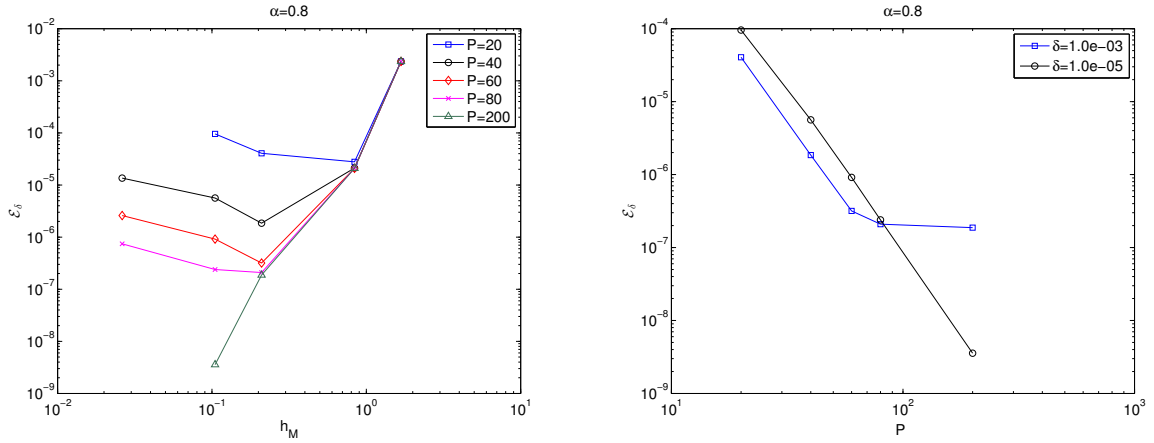

Fig. 7 Test 4. The global error $\mathcal{E}_\delta$ of the adaptive 3-step method applied to Problem 3, with $\alpha = 0.8$, $x_0 = 2$, $y_0 = 0$ and $T = 25$. On the left $\mathcal{E}_\delta$ is plotted as a function of the maximal time-step $h_M = \max_n h_n$, and the different graphs correspond to different P. On the right $\mathcal{E}_\delta$ is plotted as a function of P, and the different graphs correspond to different δ.

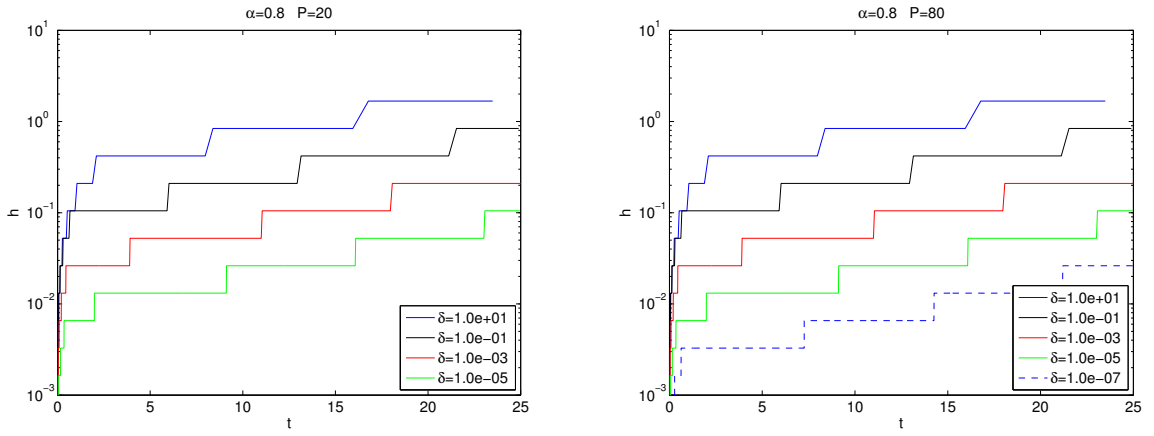

Fig. 8 Test 4. The step-size as a function of t.

Here ε is a nonnegative constant, and $x_0, y_0 \in \mathbb{R}$. For $\alpha = 1$, (7.1a) is the classical Van der Pol equation. In this case it can be shown to have a stable periodic solution. To apply the scheme we write (7.1a) as a system by substituting $y = D^\alpha u$. Thus, we have

$$D^\alpha x = y \tag{7.2a}$$

$$D^\alpha y = \varepsilon \left(1 - x^2\right) y - x \ , \tag{7.2b}$$

in $(0, T)$ subject to the initial condition

$$x(0) = x_0 \qquad y(0) = y_0 \ . \tag{7.2c}$$

For $\varepsilon = 0$, we recover Problem 3.

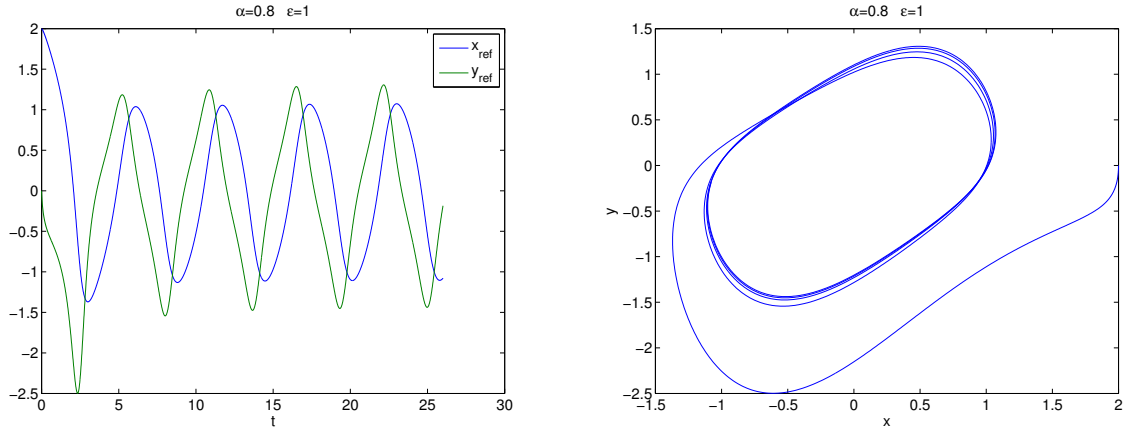

Fig. 9 Test 5. The reference solution v_{ref} computed with $P = 200$, and $\delta = 10^{-5}$, for the problem (7.2) with $\alpha = 0.8$, $\varepsilon = 1$, $x_0 = 2$, $y_0 = 0$. On the left are x_{ref} and y_{ref} as functions of t, and on the right is the curve $\left(x_{ref}, y_{ref}\right)$ in the (x, y) plane.

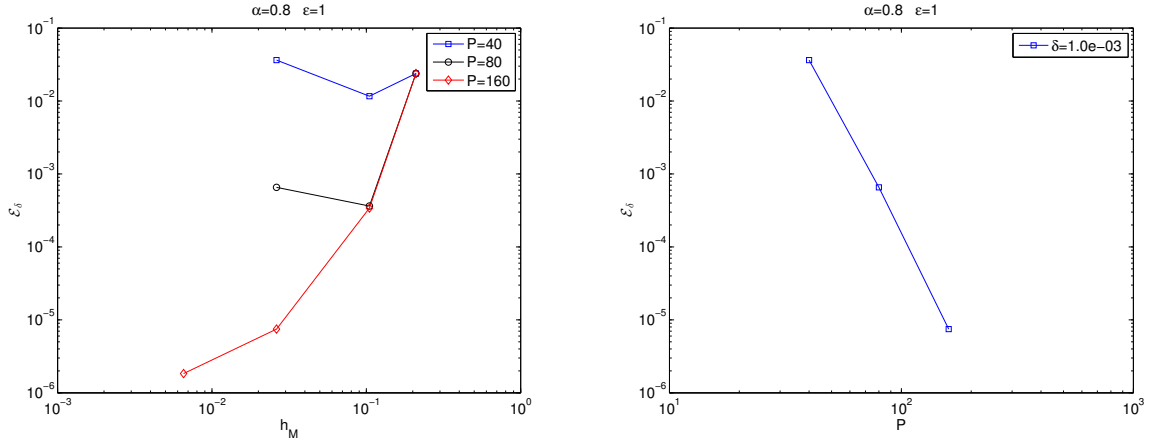

Fig. 10 Test 5. Numerical approximations of (7.2) with $\alpha = 0.8$, $\varepsilon = 1$, $x_0 = 2$, $y_0 = 0$ are compared with a reference solution. The figure shows the global error $\mathcal{E}_\delta$ in $(0, 25)$. On the left $\mathcal{E}_\delta$ is plotted as a function of the maximal time-step h_M, and the different graphs correspond to different P. On the right $\mathcal{E}_\delta$ is plotted as a function of P.

7.1 Results

The approximation is obtained with a 3-step adaptive method (see §5.3). The basis functions ψ_k are orthogonal polynomials given as (5.27). The scheme is implicit and thus requires the solution of an algebraic equation at each step to advance. For this we use a Newton solver.

Test 5

In this test $\alpha = 0.8$, $\varepsilon = 1$, $x_0 = 2$, $y_0 = 0$ and $T = 25$. The approximations v are compared with a reference solution v_{ref} computed with $P = 200$, and $\delta = 10^{-5}$. To measure the error we interpolate the reference solution by a cubic spline approximation. Figure 9 shows the reference solution $v_{ref} = \left(x_{ref}, y_{ref}\right)^T$. On the left are x_{ref} and y_{ref} as functions of t, and on the right is the curve $\left(x_{ref}, y_{ref}\right)$ in the (x, y) plane. The results are in Figure 10. On the left the global error $\mathcal{E}_\delta$ is plotted as a function of the maximal time-step h_M, and the different graphs correspond to different P. On the right $\mathcal{E}_\delta$ is plotted as a function of P.

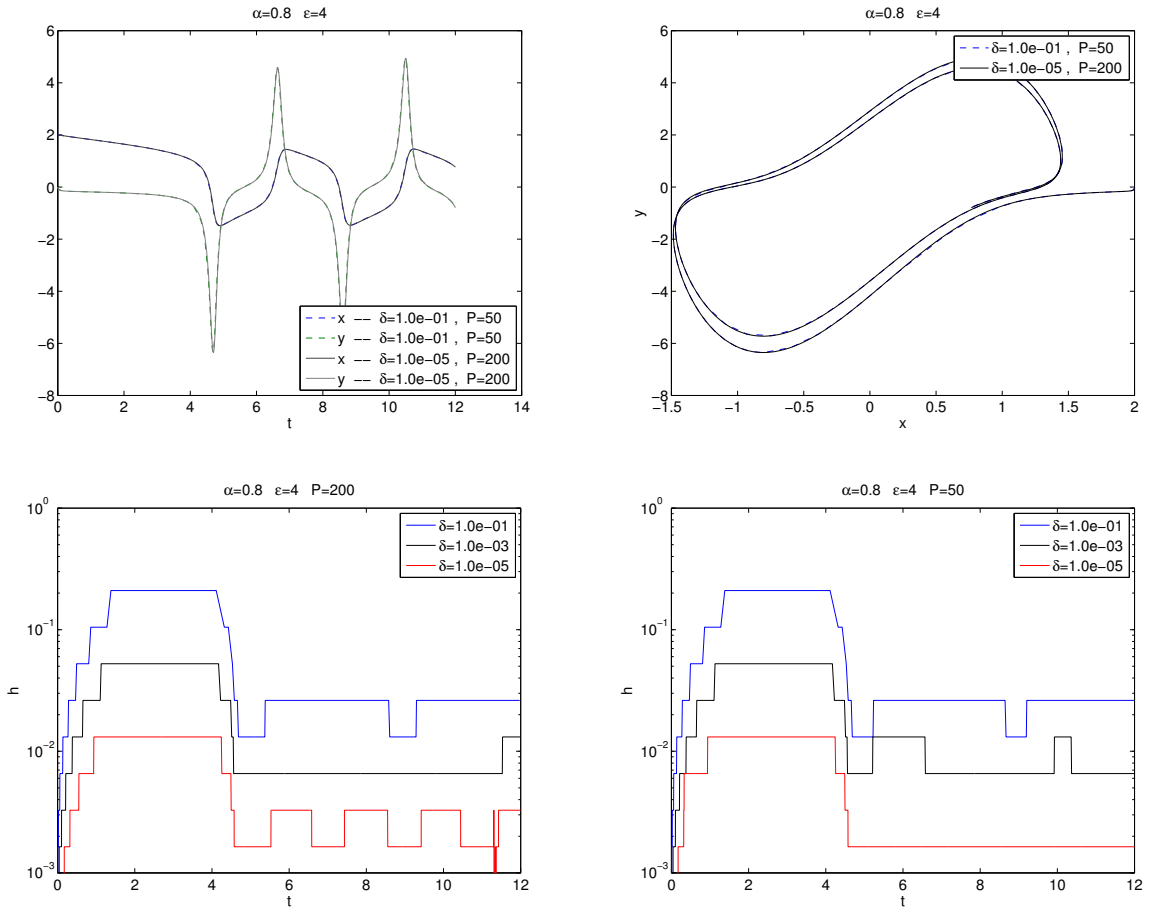

Fig. 11 Test 6. In the top row numerical solutions computed with $\delta = 10^{-1}$, $P = 50$, and $\delta = 10^{-5}$, $P = 200$ are compared. In the bottom row, are the step-sizes as functions of t.

Test 6

In this test $\alpha = 0.8$, $\varepsilon = 4$, $x_0 = 2$, $y_0 = 0$ and $T = 12$. The results are in Figure 11. In the top row numerical solutions computed with $\delta = 10^{-1}$, $P = 50$, and $\delta = 10^{-5}$, $P = 200$ are compared. In the bottom row is the step-size h as a function of t. The two approximations show good agreement throughout. In this example we can see that the program manages, with some success, to capture the rapid changes in the solution and adapts the step-size accordingly.

Test 7

This test illustrates an issue we have encountered with the implementation. Figure 12 shows the step-sizes picked by the program for $P = 100$, and different values of δ. The figure shows that for $\delta = 10^{-1}$, when the solution starts to oscillate, the program reduces the step-size. After that, the program still varies the step-size, but not at every rapid change of the solution. For $\delta = 10^{-3}$, the program captures all the changes in the solution. For $\delta = 10^{-5}$, however, the program is unable to complete the test. Typically, this means that P is too small. As the results of Test 6 in Figure 11 show, for $P = 200$, the program is able to complete the test. What is surprising, is that for $P = 50$, the program also completes the test. At this time, we do not fully understand this issue.

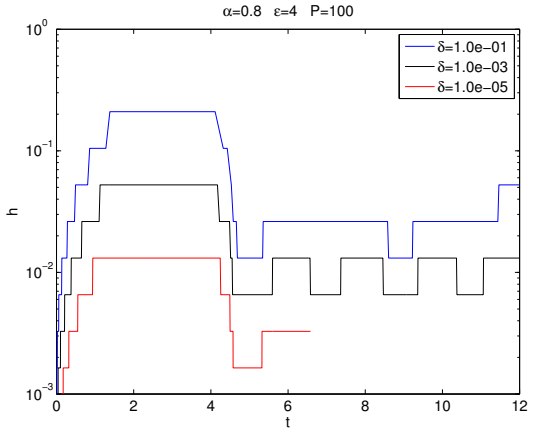

Fig. 12 Test 7. The step-size h picked by the program for $P = 100$, and different values of δ.

8 Concluding remarks

We have presented a new family of high-order accurate schemes for FDEs. The schemes are based on the multi-step Adams methods and rely on an expansion in some weighted L^2 space. We have studied the local truncation error and have shown that it can be written as a sum of two terms: a local term, controlled by h, and a history term, controlled by P. The local term of the μ-step explicit and implicit schemes is $O(h^\mu)$ and $O\big(h^{\mu+1}\big)$, respectively. We have also studied the history term for schemes employing the Jacobi polynomials. Due to lack of rigorous results connecting properties of f and the regularity of u, we have been unable to complete the analysis in general. Nevertheless, we have shown that provided (1.3), the history term is $o\big(P^{-\sigma}\big)$, for any $0 < \sigma < 2 - \alpha$. This estimate is not optimal, and for specific examples more accurate estimates can be obtained. As an example we have shown that for $f(t, u) = -u$, the history term is $O\big(P^{-\sigma}\big)$, for any $0 < \sigma < 3 + \alpha$. Building on the study of the local truncation error, we have derived an error indicator based on the Milne device.

We have also presented some numerical results, showing the performance of the proposed methods. The results indicate that to improve accuracy, while decreasing h, P must increase in some way to compensate for the the error accumulating at the highest modes during the additional steps; otherwise the accuracy may be harmed. This issue should be resolved. Ideally, we would like to have a uniform approximation to the history term; an approximation independent of the step-size. If a way to resolve this issue completely is not found, a different approach would be to accommodate it in some manner. One way to accommodate this problem is to estimate the optimal value P for a given h (or δ). Such an estimate may also allow to pick P adaptively, and increase it only when necessary, thus making computations more efficient.

Other topics and extensions may also be included in future work. Here are some possibilities: The stability of the methods and the behavior of the global error may be studied in order to improve the understanding of the schemes. To improve the convergence rate of the history term, other basis functions and other scalings of the fractional integral could be explored. Developing multi-stage methods, similar to Runge-Kutta methods, may also be of interest, given the desirable properties such methods for standard ODEs have.

A Polynomial approximation

A.1 Jacobi polynomials

In this section we suppose $\alpha > -1$. Let $P_n^{(\alpha,0)}$ be the Jacobi polynomial of degree n corresponding the weight $w_\alpha(\xi) = (1-\xi)^\alpha$, normalized such that $\|P_n^{(\alpha,0)}\|_\alpha^2 = 1$, where

$$\|f\|_\alpha^2 = \int_{-1}^{1} |f|^2 \, w_\alpha \ . \tag{A.1}$$

Let $I = (-1, 1)$,

$$\langle f, g \rangle_\alpha = \int_{-1}^{1} f g \, w_\alpha \ , \tag{A.2}$$

for $f : I \to \mathbb{R}$, $g : I \to \mathbb{R}^d$, and $L^2_\alpha(I, \mathbb{R}^d)$ the space of measurable functions $f : I \to \mathbb{R}^d$ such that $\|f\|_\alpha < \infty$. The following can be found in [17], for example. The Jacobi polynomials $P_n^{(\alpha,0)}$ are given by Rodrigues' formula

$$P_n^{(\alpha,0)}(\xi) = \frac{\sqrt{2n+\alpha+1}}{2^{(\alpha+1)/2}} \, \frac{(-1)^n}{2^n n!} \, w_\alpha^{-1} \, \frac{\mathrm{d}^n}{\mathrm{d}\xi^n} \Big(\left(1-\xi^2\right)^n w_\alpha(\xi) \Big) \ , \tag{A.3}$$

and are the eigenfunctions of the Sturm-Liouville problem

$$A\, v = \nu_n v \tag{A.4}$$

where

$$A\, v = -w_\alpha^{-1} \Big(\left(1-\xi^2\right) w_\alpha v' \Big)' \qquad \nu_n = n\,(n+\alpha+1) \ . \tag{A.5}$$

The operator $A : D(A) \to L^2_\alpha(I, \mathbb{R})$ is self adjoint. Let $f \in L^2_\alpha(I, \mathbb{R}^d)$, and

$$f_n = \left\langle P_n^{(\alpha,0)}, f \right\rangle_\alpha \ . \tag{A.6}$$

It can be shown that

$$f = \sum_{n=0}^{\infty} f_n P_n^{(\alpha,0)} \tag{A.7}$$

in the $L^2_\alpha(I, \mathbb{R}^d)$ norm. Equivalently, there holds

$$\lim_{N\to\infty} \|f - \pi_N f\|_\alpha = 0 \ , \tag{A.8}$$

where π_N is given by

$$\pi_N f = \sum_{n=0}^{N} f_n P_n^{(\alpha,0)} \ . \tag{A.9}$$

Parseval's identity holds:

$$\|f\|_\alpha^2 = \sum_{n=0}^{\infty} |f_n|^2 \ . \tag{A.10}$$

A.2 Approximation of $D\left(A^{\sigma/2}\right)$ functions

In this section some results regarding polynomial approximation of functions in $L^2_\alpha(I, \mathbb{R}^d)$ are presented. In particular, the results of this section concern the approximation of functions which have singularities at the interval's boundaries. For such a function f the approach taken here provides improved estimates compared to the estimates obtained by finding σ such that $f \in H^\sigma_\alpha$. This approach can also be found in [18,19].

For $0 < r \in \mathbb{R}$, define

$$A^r f = \sum_{n=1}^{\infty} \nu_n^r f_n P_n^{(\alpha,0)} \ . \tag{A.11}$$

The domain $D(A^r)$ of A^r is the space of functions $f \in L^2_\alpha(I, \mathbb{R}^d)$, such that

$$\|A^r f\|_\alpha^2 = \sum_{n=1}^{\infty} \nu_n^{2r} |f_n|^2 < \infty \ . \tag{A.12}$$

Lemma A.1 *Suppose* $0 < \sigma \in \mathbb{R}$, *and* $f \in D\big(A^{\sigma/2}\big)$. *Then,*

$$\| \left(1 - \pi_N\right) f\|_\alpha \leq (N+1)^{-\sigma} \|A^{\sigma/2} f\|_\alpha \qquad\qquad N \geq 0 \ . \tag{A.13}$$

Proof Suppose $f \in D\big(A^{\sigma/2}\big)$, and $N \geq 0$. Then

$$\| \left(1 - \pi_N\right) f\|_\alpha^2 = \sum_{n=N+1}^{\infty} |f_n|^2 \ . \tag{A.14}$$

Owing to

$$1 = \nu_n^{-\sigma} \nu_n^\sigma \leq (N+1)^{-2\sigma} \nu_n^\sigma \qquad\qquad n \geq N+1 \ , \tag{A.15}$$

we get

$$\begin{aligned} \| \left(1 - \pi_N\right) f\|_\alpha^2 &\leq (N+1)^{-2\sigma} \sum_{n=N+1}^{\infty} \nu_n^\sigma |f_n|^2 \\ &\leq (N+1)^{-2\sigma} \|A^{\sigma/2} f\|_\alpha^2 \end{aligned} \tag{A.16}$$

and thus the conclusion. □

Proposition A.1 *Suppose* $\gamma > 0$, $f(\xi) = (1+\xi)^\gamma$, *and* $g(\xi) = (1-\xi)^\gamma$. *Then,*

$$f \in D\Big(A^{\sigma/2}\Big) \qquad\qquad 0 < \sigma < 1 + 2\gamma \ , \tag{A.17}$$

and

$$g \in D\Big(A^{\sigma/2}\Big) \qquad\qquad 0 < \sigma < 1 + \alpha + 2\gamma \ . \tag{A.18}$$

Proof Here we only prove (A.17). The proof of (A.18) is similar and can be also found in [19]. We have

$$A^{\sigma/2} f = \sum_{n=1}^{\infty} \nu_n^{\sigma/2} f_n P_n^{(\alpha,0)} \tag{A.19}$$

where

$$f_n = \int_{-1}^{1} (1+\xi)^\gamma \, P_n^{(\alpha,0)}(\xi) \, w_\alpha(\xi) \, \mathrm{d}\xi \ . \tag{A.20}$$

It follows that (A.17) is valid if and only if

$$\|A^{\sigma/2} f\|_\alpha^2 = \sum_{n=1}^{\infty} |\nu_n|^\sigma \, |f_n|^2 < \infty \ . \tag{A.21}$$

Thus we require an estimate on f_n. By Rodrigues' formula,

$$f_n = \frac{(-1)^n \sqrt{2n+\alpha+1}}{2^{n+\alpha/2+1/2} n!} \int_{-1}^{1} (1+\xi)^\gamma \, \Big(\big(1-\xi^2\big)^n \, w_\alpha \Big)^{(n)} \, \mathrm{d}\xi \ . \tag{A.22}$$

We integrate by parts to get

$$f_n = \frac{(-1)^n \sqrt{2n+\alpha+1}}{2^{n+\alpha/2+1/2} \, n!} \, \frac{\Gamma(n-\gamma)}{\Gamma(-\gamma)} \int_{-1}^{1} (1+\xi)^\gamma \, (1-\xi)^{n+\alpha} \, \mathrm{d}\xi \tag{A.23}$$

which yields

$$f_n = 2^{\alpha/2+\gamma+1/2} (-1)^n \frac{\Gamma(1+\gamma)}{\Gamma(-\gamma)} \, \frac{\Gamma(n-\gamma) \, \Gamma(n+\alpha+1)}{n! \Gamma(n+\alpha+\gamma+2)} \sqrt{2n+\alpha+1} \ . \tag{A.24}$$

We use Stirling's approximation [17],

$$\Gamma(x) \sim \sqrt{\frac{2\pi}{x}} \left(\frac{x}{e}\right)^x \qquad\qquad x \to \infty \tag{A.25}$$

to get

$$f_n \sim c_{\alpha\gamma} n^{-3/2-2\gamma} \qquad\qquad n \to \infty \ . \tag{A.26}$$

So, $A^{\sigma/2} f \in L_\alpha^2(I, \mathbb{R})$ if and only if $\sigma < 1 + 2\gamma$, and thus the conclusion. □

B Computing R_{km} and $\mathcal{J}_k$

Here, $0 < \alpha < 1$, $\beta = -1 + \alpha$, $w_\beta(s) = (1-s)^\beta$, and $P_j^{(\beta,0)}$ are the Jacobi polynomials associated with the weight w_β, normalized such that their norm is one. We have

$$\psi_j(s) = 2^{\alpha/2}\, P_j^{(\beta,0)}(2s-1) \ , \tag{B.1}$$

and conversely

$$2^{-\alpha/2}\, \psi_j\left(\frac{1+\xi}{2}\right) = P_j^{(\beta,0)}(\xi) \ . \tag{B.2}$$

We transform the expressions for R_{km} and $\mathcal{J}_k$ into integrals over $(-1,1)$: we get

$$\begin{aligned} R_{km}(\theta) &= \frac{\theta^{1-\alpha}}{2} \int_{-1}^{1} \psi_m\left(\frac{1+\xi}{2}\right)\, \psi_k\left(\theta\frac{1+\xi}{2}\right)\, \left(1-\theta\frac{1+\xi}{2}\right)^\beta \,\mathrm{d}\xi \\ &= \theta^{1-\alpha} \int_{-1}^{1} P_m^{(\beta,0)}(\xi)\, P_k^{(\beta,0)}(\theta\xi-\varphi)\, \Big(2-\theta\,(1+\xi)\Big)^\beta \,\mathrm{d}\xi \end{aligned} \tag{B.3}$$

and

$$\begin{aligned} \mathcal{J}_k(f;t,h) &= \int_0^1 f(t+hs)\, \psi_k(\theta+\varphi s)\, w_\beta(s)\, \mathrm{d}s \\ &= \frac{1}{2^{\alpha/2}} \int_{-1}^{1} f\left(t+h\frac{\xi+1}{2}\right) P_k^{(\beta,0)}(\theta+\varphi\xi)\, w_\beta(\xi)\, \mathrm{d}\xi \ . \end{aligned} \tag{B.4}$$

In our implementation, the integrals above are approximated by a Gauss quadrature. Precisely, R_{km} is computed with the Gauss-Jacobi quadrature associated with the weight w_β, and $\mathcal{J}_k$ is computed with the Gauss-Legendre quadrature.

The approximation of the matrix $\mathcal{R}(\theta) = (R_{km}(\theta))$ requires the computation of some values every time θ changes. To make the computation more efficient, the part of $\mathcal{R}$ that does not require adaptation can be stored. The Gauss quadrature provides

$$R_{km}(\theta) \approx \theta^{1-\alpha} \sum_{j=1}^{N_q} P_m^{(\beta,0)}(\xi_j)\, P_k^{(\beta,0)}(\theta\xi_j-\varphi)\, \Big(2-\theta\,(1+\xi_j)\Big)^\beta \omega_j \tag{B.5}$$

where ω_j, and ξ_j are the quadrature weights and nodes, respectively. The last equation can be written as a matrix product

$$\mathcal{R}(\theta) = \mathcal{R}_2^T(\theta)\, \mathcal{R}_1 \ . \tag{B.6}$$

Notice that R_1 does not change during the time-stepping, and can be stored and reused, while $\mathcal{R}_2$ must be computed whenever θ changes. We have

$$(\mathcal{R}_1)_{j,m+1} = P_m^{(\beta,0)}(\xi_j)\, w_j \qquad j = 1,\dots,N_q \quad m = 0,\dots,P \tag{B.7}$$

$$(\mathcal{R}_2)_{j,k+1}(\theta) = P_k^{(\beta,0)}(\theta\xi_j-\varphi)\, \Big(2-\theta\,(1+\xi_j)\Big)^\beta \qquad j = 1,\dots,N_q \quad k = 0,\dots,P \ . \tag{B.8}$$

Conflict of interest

This work was partially supported by the NSF DMS-1115416 and by OSD/AFOSRFA9550-09-1-0613.

References

1. E. Barkai, *Fractional Fokker-Planck Equation, Solution, and Application*, Phys. Rev. E. 63 (2001), 046118.
2. R. Metzler, J. Klafter, *The Random Walk's Guide to Anomalous Diffusion: A Fractional Dynamics Approach*, Phys. Rep. 339 (2000), 1-77.
3. R. Metzler, J. Klafter, *The Restaurant at the End of the Random Walk: Recent Developments in the Description of Anomalous Transport by Fractional Dynamics*, J. Phys. A 37 (2004), 161-208.
4. C. Tadjeran, M. M. Meerschaert, *A second-order accurate numerical method for two dimensional fractional diffusion equation*, J. Comput. Phys. 220 (2007), 813-823.
5. W. H. Deng, *Finite Element Method for the Space and Time Fractional Fokker-Planck Equation*, SIAM J. Numer. Anal. 47 (2009), 204-226.
6. V. J. Ervin, J. P. Roop, *Variational Formulation for the Stationary Fractional Advection Dispersion Equation*, Numer. Methods Partial Differential Equations, 22 (2006), 558-576.
7. Q. Xu, J. S. Hesthaven, *Stable Multi-domain Spectral Penalty Methods for Fractional Partial Differential Equations*, J. Comput. Phys. 257 (2014), 241-258.

8. W. H. Deng, J. S. Hesthaven, *Local Discontinuous Galerkin Methods for Fractional Diffusion Equations*, ESAIM Math. Model. Numer. Anal. 47 (2013), 1845-1864.
9. Y. Zhang, Z. Sun, H. Liao, *Finite Difference Methods for the Time Fractional Diffusion Equation on Non-Uniform Meshes*, J. Comput. Phys. 256 (2014), 195-210.
10. C. Lubich, *Convolution Quadrature and Discretized Operational Calculus I*, Numer. Math. 52 (1988), 129-145.
11. H. Brunner, P.J. van der Houwen, *The Numerical Solution of Volterra Equations*, North Holland, 1986.
12. H. Brunner, D. Schötazau, *hp Discontinuous Galerkin Time-Stepping for Volterra Integrodifferential Equations*, SIAM J. Numer. Anal., Vol. 44, No. 1, pp. 224-245.
13. K. Mustapha, H. Brunner, H. Mustapha, D. Schötazau, *An hp-Version Discontinuous Galerkin Method for Integro-Differential Equations of Parabolic Type*, SIAM J. Numer. Anal., Vol. 49, No. 4, pp. 1369-1396.
14. A. Iserles, *A First Course in the Numerical Analysis of Differential Equations*, Second Edition, Cambridge University Press, 2008.
15. I. Podlubny, *Fractional Differential Equations*, Academic Press, San Diego, 1999.
16. E. Süli, D. F. Mayers, *An introduction to Numerical Analysis*, Cambridge University Press, 2003.
17. M. Abramowitz, I. A. Stegun, *Handbook of Mathematical Functions*, Tenth Printing, 1972, U.S. Department of Commerce, NIST.
18. C. Bernardi, Y. Maday, *Spectral Methods*, Handbook of Numerical Analysis, vol. V, Techniques of Scientific Computing (Part 2), Elsavier Science, 1997.
19. C. Bernardi, Y. Maday, *Polynomial approximation of Some Singular Functions*, Appl. Anal., Vol. 42 (1991), 1-32.