Elsevier Editorial System(tm) for Applied Numerical Mathematics
Manuscript Draft

Manuscript Number: APNUM-D-10-00272

Title: The use of exponentially fitted methods in a deferred correction framework

Article Type: Research paper

Keywords: Mono-implicit Runge-Kutta methods; Boundary value problems; Exponential fitting; Error term; B-series; Parameter selection

Corresponding Author: Mr. Davy Hollevoet,

Corresponding Author's Institution: Ghent University

First Author: Davy Hollevoet

Order of Authors: Davy Hollevoet; Marnix Van Daele; Guido Vanden Berghe

Abstract: The combination of exponential fitting and deferred correction based on mono-implicit Runge-Kutta (MIRK) methods is investigated. The structure of the B-series coefficients of exponentially fitted MIRK methods is compared to that of classical counterparts, and it is shown how a standalone method can be tuned to gain at least one order of accuracy. After that, the coefficients structure of exponentially fitted deferred correction (EFDC) schemes is laid out in a similar fashion. A B-series composition property is applied to show that if $r=p$, a special strategy should be used to tune the EF methods. Several numerical experiments illustrate the annihilation or minimization of the leading error term.

The use of exponentially fitted methods in a deferred correction framework

D. Hollevoet, M. Van Daele, G. Vanden Berghe

Vakgroep Toegepaste Wiskunde en Informatica, Ghent University,
Krijgslaan 281-S9, B-9000 Gent, Belgium

Abstract

The combination of exponential fitting and deferred correction based on mono-implicit Runge-Kutta (MIRK) methods is investigated. The structure of the B-series coefficients of exponentially fitted MIRK methods is compared to that of classical counterparts, and it is shown how a standalone method can be tuned to gain at least one order of accuracy. After that, the coefficients structure of exponentially fitted deferred correction (EFDC) schemes is laid out in a similar fashion. A B-series composition property is applied to show that if $r = p$, a special strategy should be used to tune the EF methods. Several numerical experiments illustrate the annihilation or minimization of the leading error term.

Keywords: Mono-implicit Runge-Kutta methods, Boundary value problems, Exponential fitting, Error term, B-series, Parameter selection

1. Introduction

The technique of (iterated) deferred correction is well-known and often used on general two-point boundary value problems

$$\frac{dy}{dx} = f(x, y(x)), \quad g(y(a), y(b)) = 0, \quad a \leq x \leq b.$$

The idea behind deferred correction is as follows: over all knot points, an initial approximation y to the solution of the problem is computed with a low-order method

$$\phi(y) = 0$$

and is used to construct an estimate of the residual of ϕ, by means of an operator $\hat{\phi}$. This information is used in the second step

$$\phi(\overline{y}) = \hat{\phi}(y)$$

to produce a more accurate solution $\overline{y}$.

In what follows, we will consider deferred correction schemes based on exponentially fitted (EF) variants of mono-implicit Runge-Kutta (MIRK) methods [1], extending the approach suggested in [2] and implemented in the famous TWPBVP code [3]. In this setup, both the base method ϕ and the residual estimator $\hat{\phi} = -\psi$ are (EF)MIRK methods and are combined as

$$\begin{aligned}\phi(y) &= 0\\ \phi(\overline{y}) &= -\psi(y).\end{aligned}$$

The important theorem by Skeel [4] guarantees that if

1. ϕ is a method of order p,
2. ψ is a method of order $p + r$ and
3. $\|\phi(\Delta w) - \psi(\Delta w)\| = \mathcal{O}(h^r)$

with Δw the restriction of any suitable function w to the grid, then the improved solution $\overline{y}$ will be of order $r + p$.

Exponential fitting is a procedure which produces variants of *classical* methods, aimed to solve problems with oscillating solutions more efficiently. Traditional linear multistep and Runge-Kutta methods are based on fitting spaces that contain only polynomials up to a certain order. To construct exponentially fitted methods, one or more pairs of higher-order polynomials are exchanged for pairs of exponentials containing a parameter. In general, the fitting space of a (K, P)-EF method is of the form

$$\{1, x, x^2, \ldots, x^K, e^{\pm\mu x}, xe^{\pm\mu x}, \ldots, x^P e^{\pm\mu x}\}.$$

Problems with solutions that fall within this fitting space, can be solved up to machine accuracy.

The result of the six-step procedure from [5] to produce exponentially fitted methods, is a set of coefficient functions of $h\mu$ to fill the Butcher tableau. The free parameter can for example be used to tune the fitting space to a certain frequency thought to be present in the solution. An other possibility is to choose the value differently in every knot point, in such a way that e.g. the leading error term is annihilated, effectively increasing the order of the method [6].

In this paper, we will examine what can be gained by using deferred correction schemes with exponentially fitted methods and the latter tuning strategy. First, we will look at some structural properties of exponentially fitted MIRK-methods. After stating two auxiliary theorems in section 3, we'll turn our attention to exponentially fitted deferred correction (EFDC) schemes and the effect of error propagation in the initial solution on the error terms of the entire scheme. The final section shows how what was described can be put to work.

2. Exponentially fitted MIRK methods

2.1. MIRK methods

MIRK methods fall within the more general class of parameterized IRK methods [7]:

$$\phi(y_k, y_{k+1}) := y_k - y_{k+1} + h\sum_{i=1}^{s} b_i f(x_k + c_i h, Y_i) \tag{1}$$

$$Y_i := (1 - v_i)y_k + v_i y_{k+1} + h\sum_{j=1}^{s} x_{ij} f(x_k + c_j h, Y_j), \quad i = 1 \ldots s$$

This type of method is characterized by a special tableau (b, X, c, v), related to the well-known Butcher tableau (c, A, b) through $A = X + vb^T$. Restricting X to lower triangular matrices only reveals the class of mono-implicit RK methods. This type of methods is only implicit in the next knot point, which makes them very suitable for solving boundary value problems and using deferred correction, while maintaining a good stability.

It is known from [8] that for any function a, $\phi(y_k, a(y_k))$ can be written as a B-series [9] $B(\boldsymbol{\phi^a}, y_k)$ with

$$\begin{aligned}
\boldsymbol{\phi^a}(t) &= \boldsymbol{a}(t) - \sum_{j=1}^{s} b_j \boldsymbol{k_j^a}(t) \\
\boldsymbol{k_i^a}(t) &= \rho(t)\boldsymbol{g_i^a}(t_1)\boldsymbol{g_i^a}(t_2)\ldots\boldsymbol{g_i^a}(t_m) \\
\boldsymbol{g_i^a}(t) &= v_i\boldsymbol{a}(t) + \sum_{j=1}^{s} x_{ij}\boldsymbol{k_j^a}(t)
\end{aligned} \tag{2}$$

for any $t = [t_1 \ldots t_m]$ and with $\boldsymbol{\phi^a}(\emptyset) = 0$, $\boldsymbol{k_i^a}(\bullet) = 1$. Each method ϕ has an associated operator η which transforms any y_k into $\eta(y_k)$, such that

$$\phi(y_k, \eta(y_k)) = 0. \tag{3}$$

For explicit methods, this means that $\eta(y_k) = \phi(y_k, 0)$, but for (mono-)implicit methods, η cannot be stated in a closed form. Since (3) should hold for any problem and step size, $\boldsymbol{\phi^\eta}(t) = 0$ is required for all t. This in turn allows to derive the structure of $\boldsymbol{\eta}(t)$ in the B-series $B(\boldsymbol{\eta}, y_k)$ of $\eta(y)$ as

$$\begin{aligned}
\boldsymbol{\eta}(t) &= \sum_{j=1}^{s} b_j \boldsymbol{k_j}(t) \\
\boldsymbol{k_i}(t) &= \rho(t)\boldsymbol{g_i}(t_1)\boldsymbol{g_i}(t_2)\ldots\boldsymbol{g_i}(t_m) \\
\boldsymbol{g_i}(t) &= \sum_{j=1}^{s}(v_i b_j + x_{ij})\boldsymbol{k_j}(t)
\end{aligned} \tag{4}$$

for any $t = [t_1 \ldots t_m]$ and with $\boldsymbol{\eta}(\emptyset) = 1$, $\boldsymbol{k_i}(\bullet) = 1$.

Operators $\boldsymbol{\phi}$ and $\boldsymbol{\eta}$ are related in the sense that both are constructed using the same Butcher tableau. The combination of both in the notation of $\boldsymbol{\phi^\eta}$ is redundant when looking at MIRK methods in a standalone setup. However, for the deferred correction schemes that will considered in what follows, other less natural combination will be made. To make it clear that an operator η is related to a method ϕ, we will from now on denote the latter as ϕ^η, even out of the context of B-series.

The first part of table 1 shows in abstracto the properties of the coefficient function $\boldsymbol{\phi^\eta}(t)$, found in the B-series of a MIRK method ϕ^η of order p. Each line shows the evaluation of this function of trees of increasing order. Starting from $\rho(t) = 3$, each cell implicitly contains a vector of values because there are several trees of any order ≥ 3. The assumption that ϕ^η is a method of order p, is represented by the zeros for $\boldsymbol{\phi^\eta}(t)$ up until $\rho(t) > p$. The non-zero evaluations in the shaded cell set the order of the method to p. One could call this gray obstruction the *order boundary* of the method.

$\rho(t)$	$\boldsymbol{\phi}^{\boldsymbol{\eta}}$	$\boldsymbol{\phi}^{\boldsymbol{\eta}}[\boldsymbol{\mu}]$	$+h\mu$	$\dots$	$+(h\mu)^p$
1	0	0	0	0	$\phi^{\eta}_{1,p}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\iddots$	
p	0	0	$\phi^{\eta}_{p,1}$		
$p+1$	$\neq 0$	$\phi^{\eta}_{p+1,0}$			

Table 1: Structure of the coefficient function $\boldsymbol{\phi}^{\boldsymbol{\eta}}(t)$ of a classical method ϕ^{η} (left) and an EF method $\phi^{\eta}[\mu]$ (right).

2.2. Exponentially fitted MIRK methods

As already mentioned in the introduction, the tableau of an exponentially fitted RK method $\phi^{\eta}[\mu]$ contains functions of the product of the free parameter μ and the step size h. For example, the exponentially fitted trapezoid rule in MIRK form can be represented with the following tableau:

$$\begin{array}{c|c|cc} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \hline & & \dfrac{e^{h\mu}-1}{(e^{h\mu}+1)h\mu} & \dfrac{e^{h\mu}-1}{(e^{h\mu}+1)h\mu} \end{array}. \tag{5}$$

As a consequence, through (2) and (4), the operators $\boldsymbol{\eta}[\boldsymbol{\mu}]$ and $\boldsymbol{\phi}^{\boldsymbol{\eta}}[\boldsymbol{\mu}]$ in general no longer produce constants, but again functions of $h\mu$. The right part of table 1 shows the coefficients occurring in a truncated Taylor series development of these $\boldsymbol{\phi}^{\boldsymbol{\eta}}[\boldsymbol{\mu}](t)$, such that

$$\boldsymbol{\phi}^{\boldsymbol{\eta}}(t_j) = \sum_{i=0}^{\infty} \phi^{\eta}_{j,i}(h\mu)^i \tag{6}$$

in which $\rho(t_j) = j$.

The order boundary of an exponentially fitted method is now represented with shaded, upward diagonals: since every series coefficient $\phi^{\eta}_{j,i}$ is multiplied by $(h\mu)^i$ in (6), the contribution of h^{p+1} to $B(\boldsymbol{\phi}^{\boldsymbol{\eta}}[\boldsymbol{\mu}], y_k)$ no longer depends on trees of order $p+1$ only, but also on smaller trees. This dependency colors an entire diagonal in table 1.

In general, the coefficient connected to h^q in $B(\boldsymbol{\phi}^{\boldsymbol{\eta}}[\boldsymbol{\mu}], y_k)$ is spread across the qth diagonal in table 1 and it follows from (6) and the definition of B-series that it has the following form:

$$b_q(\boldsymbol{\phi}^{\boldsymbol{\eta}}[\boldsymbol{\mu}], y_k) := \sum_{i=0}^{q} \sum_{\rho(t)=q-i} \mu^i \phi^{\eta}_{q-i,i}(t) \frac{\alpha(t)}{\rho(t)!} F(t)(y_k). \tag{7}$$

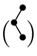
The expression $h^{p+1} b_{p+1}(\boldsymbol{\phi}^{\boldsymbol{\eta}}[\boldsymbol{\mu}], y(x_k))$ represents the leading term of the local truncation error of method ϕ^{η}. If in each knot point an appropriate value can be attributed to μ such that b_{p+1} is annihilated, then the order of accuracy of the method increases by at least 1. Every term in (7) contains an elementary differential $F(t)$, so the question of which values should be used for μ, depends on the problem at hand. Due to symmetry in all constructing parts of (2), there are no odd powers of μ in any $\boldsymbol{\phi}^{\boldsymbol{\eta}}[\boldsymbol{\mu}](t)$ and by extension neither in (7). For (K_0, P_0)-EF methods, the largest occurring power of μ in the order boundary is $2(P_0+1)$.

Example

$\rho(t)$	$\boldsymbol{\phi}^{\boldsymbol{\eta}}$	$\boldsymbol{\phi}^{\boldsymbol{\eta}}[\boldsymbol{\mu}]$	$+h\mu$	$+h^2\mu^2$
1	0	0	0	$\frac{1}{12}$
2	0	0	0	
3	$\frac{-1}{2}, \frac{-1}{2}$	$\frac{-1}{2}, \frac{-1}{2}$		

Table 2: $\boldsymbol{\phi}^{\boldsymbol{\eta}}$ for the (EF) trapezoid rule.

Table 2 shows the incarnation of table 1 for the trapezoid rule and the $(-1, 0)$-EF version of the trapezoid rule shown in tableau (5). If in all knot points x_k an appropriate value for μ_k can be found such that

$b_3(\boldsymbol{\phi}^{\boldsymbol{\eta}}[\boldsymbol{\mu_k}], y_k) = \dfrac{-1}{2}\dfrac{1}{3!}F($ $)(y_k) + \dfrac{-1}{2}\dfrac{1}{3!}F($$)(y_k) + \mu_k^2\dfrac{1}{12}F(\bullet)(y_k) = 0,$ (8)

then those values can be used to increase the accuracy of the exponentially fitted trapezoid rule to order 3 for the problem at hand. This is due to the fact that (8) is the coefficient of h^3 in the local truncation error of the method, perhaps more recognizable in terms of total derivatives:

$$b_3(\boldsymbol{\phi}^{\boldsymbol{\eta}}[\boldsymbol{\mu_k}], y_k) = \frac{-1}{12}y^{(3)}(y_k) + \mu_k^2\frac{1}{12}y'(y_k).$$

3. A composition property

In section 4.2, we will use a property of the composition of B-series. In this section, we formulate and prove this property, for both classical and exponentially fitted coefficients.

Theorem 1. *If for given* $\boldsymbol{a} : T \cup \emptyset \mapsto \mathbb{Q}$ *and* $\boldsymbol{b} : T \cup \emptyset \mapsto \mathbb{Q}$ *holds that*

$$\begin{cases} \boldsymbol{a}(\emptyset) = 1 \\ \boldsymbol{a}(t) = 0 \text{ if } 0 < \rho(t) < p \\ \exists t : \rho(t) = p \wedge \boldsymbol{a}(t) \neq 0 \end{cases} \quad \text{and} \quad \begin{cases} \boldsymbol{b}(t) = 0 \text{ if } \rho(t) < q \\ \exists t : \rho(t) = q \wedge \boldsymbol{b}(t) \neq 0 \end{cases},$$

then the composition $\boldsymbol{a} \cdot \boldsymbol{b}$ *has following properties:*

$$\begin{cases} \boldsymbol{a} \cdot \boldsymbol{b}(t) = \boldsymbol{b}(t) \text{ if } \rho(t) < p + q \\ \exists t : \rho(t) = p + q \wedge \boldsymbol{a} \cdot \boldsymbol{b}(t) \neq \boldsymbol{b}(t) \end{cases}.$$

Proof. Assume t_m is the smallest tree for which $\boldsymbol{a} \cdot \boldsymbol{b}(t_m) \neq \boldsymbol{b}(t_m)$. From the recipe for the composition of $\boldsymbol{a}$ and $\boldsymbol{b}$, one learns that the equality can only be restored by adding one or more extra terms to the right hand side. Such term is necessarily the product of some $\boldsymbol{a}(t_1)$ and $\boldsymbol{b}(t_2)$, with $t_1 \neq \emptyset$. The smallest trees t_1 and t_2 such that this product is nonzero, have p and q nodes respectively. Since t_m must be decomposable into t_1 and t_2, it holds that $\rho(t_m) = p + q$. □

When considering exponentially fitted methods, the occurring coefficient functions transform trees into (possibly complex) functions of e.g. $h\mu$. The theorem as above still applies, albeit only to the constant term in the series development of the functions in question. A similar theorem can be stated for the entire functions.

Theorem 2. *If for given* $\boldsymbol{a} : T \cup \emptyset \mapsto (\mathbb{C} \mapsto \mathbb{C})$ *and* $\boldsymbol{b} : T \cup \emptyset \mapsto (\mathbb{C} \mapsto \mathbb{C})$ *holds that*

$$\begin{cases} \boldsymbol{a}(\emptyset)(h) = 1 \\ \boldsymbol{a}(t)(h) = \mathcal{O}(h^{p-\rho(t)}) \text{ if } 0 < \rho(t) \leq p \end{cases} \qquad \boldsymbol{b}(t)(h) = \mathcal{O}(h^{q-\rho(t)}) \text{ if } \rho(t) \leq q$$

then the composition $\boldsymbol{a} \cdot \boldsymbol{b}$ *has following property:*

$$\boldsymbol{a} \cdot \boldsymbol{b}(t)(h) = \boldsymbol{b}(t)(h) + \mathcal{O}(h^{p+q-\rho(t)})$$

Proof. Apart from $b(t)$, each term in $\boldsymbol{a} \cdot \boldsymbol{b}(t)$ is a scaled multiplication of at least one evaluation of $\boldsymbol{a}$ and $\boldsymbol{b}$ each. For this theorem, the most important terms are those of the form $\boldsymbol{a}(t_1)\boldsymbol{b}(t_2)$, since terms with more factors contribute only higher powers of h, given the structure of $\boldsymbol{a}$ and $\boldsymbol{b}$.

The leading term in the series development of such a $\boldsymbol{a}(t_1)\boldsymbol{b}(t_2)$ is $\mathcal{O}(h^{p-\rho(t_1)+q-\rho(t_2)})$, and since t must be decomposable into t_1 and t_2, the smallest possible power of h is $p + q - \rho(t)$.

□

Definition 1. *For* $\boldsymbol{a} : T \cup \emptyset \mapsto (\mathbb{C} \mapsto \mathbb{C})$*:*

$$\boldsymbol{a} \in Tr(c, e) \Leftrightarrow$$

$$\forall t \in T | \rho(t) \leq e : \quad \begin{cases} \boldsymbol{a}(\emptyset)(h) = \boldsymbol{c}(\emptyset)(h) \\ \boldsymbol{a}(t)(h) = \boldsymbol{c}(t)(h) + \mathcal{O}(h^{e-\rho(t)}) \end{cases}$$

This definition allows to restate theorem 2 as

Theorem 3. *If for given* $\boldsymbol{a} : T \cup \emptyset \mapsto (\mathbb{C} \mapsto \mathbb{C})$ *and* $\boldsymbol{b} : T \cup \emptyset \mapsto (\mathbb{C} \mapsto \mathbb{C})$ *holds that*

$$\boldsymbol{a} \in Tr(\boldsymbol{\delta_\emptyset}, p) \quad \text{and} \quad \boldsymbol{b} \in Tr(\boldsymbol{0}, q),$$

then

$$\boldsymbol{a} \cdot \boldsymbol{b} \in Tr(\boldsymbol{b}, p + q)$$

4. The leading error term of an EFDC scheme

4.1. Matching coefficients

The second stage of a DC scheme matches the residual of the base method ϕ^η to an estimate based on a method ψ and an initial solution y_k, previously computed with η:

$$\phi(y(x_k), y(x_{k+1})) = -\psi(y_k, \eta(y_k)) + O(h^q).$$

Translating this to the superscript notation from section 2.1, with $\boldsymbol{1}(t) = 1, \forall t$, one obtains in terms of B-series:

$$B(\boldsymbol{\phi}^{\boldsymbol{1}}, y(x_k)) = -B(\boldsymbol{\psi}^{\boldsymbol{\eta}}, y_k) + O(h^q) \tag{9}$$

The value of q depends on the order of the DC scheme. For a classical MIRK-DC scheme, the residual and the residual estimate agree on coefficients up until $q := p + r + 1$.

Table 3 shows the coefficient functions occurring in DC scheme (9) evaluated in trees of increasing order. Column $\boldsymbol{\phi^1}$ lists the coefficients of the residual of method ϕ^η. Since it is assumed to be a method of order p, this first column is filled with zeros until $\rho(t) > p$. The second column shows the coefficients of the residual estimate $-\boldsymbol{\psi^\eta}$. The entire deferred correction scheme is of order $p+r$, which is shown by means of matching coefficients up until $\rho(t) > p + r$, marked in light gray. The disagreement at q sets the accuracy of the DC scheme to order q.

$\rho(t)$	$\boldsymbol{\phi^1}$	$-\boldsymbol{\psi^\eta}$	$\boldsymbol{\phi[\mu]^1}$	$+h\mu$	$\dots$	$+h^p\mu^p$	$-\boldsymbol{\psi[\omega]^{\eta[\upsilon]}}$	$+h$	$\dots$	$+h^p$
1	0	0	0	0	$\dots$	$\phi^1_{1,p}$	0	0	$\dots$	$\upsilon^p\phi^1_{1,p}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\iddots$	$\vdots$	$\vdots$	$\vdots$	$\iddots$	$\vdots$
p	0	0	0	$\phi^1_{p,1}$	$\dots$	$\phi^1_{p,p}$	0	$\upsilon\phi^1_{p,1}$	$\dots$	$\upsilon^p\phi^1_{p,p}$
$p+1$	ϕ^1	ϕ^1	$\phi^1_{p+1,0}$	$\phi^1_{p+1,1}$	$\dots$	$\phi_{p+1,p}$	$\phi^1_{p+1,0}$	$\upsilon\phi^1_{p+1,1}$	$\dots$	$\psi^\eta_{p+1,p}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\iddots$		$\vdots$	$\vdots$	$\iddots$	
$p+r$	ϕ^1	ϕ^1	$\phi^1_{p+r,0}$	$\phi^1_{p+r,1}$			$\phi^1_{p+r,0}$	$\psi^\eta_{p+r,1}$		
q	ϕ^1	$\neq \phi^1$	$\phi^1_{q,0}$				$\psi^\eta_{q,0}$			

Table 3: Matching coefficients in a DC (left) and an EFDC scheme (right)

The technique of exponential fitting allows to introduce three parameters into the deferred correction scheme: $\boldsymbol{\phi}$ and $\boldsymbol{\psi^\eta}$ can be replaced with $\boldsymbol{\phi[\mu]}$ and $\boldsymbol{\psi[\omega]^{\eta[\upsilon]}}$. The introduction of parameters here differs from what happened in section 2.2, where only one parameter was added to $\boldsymbol{\phi^\eta}$ as a whole. In this context, $\boldsymbol{\phi}$ and $\boldsymbol{\eta}$ are decoupled and it is less clear whether or not they should contain the same parameter.

The second part of table 3 shows truncated Taylor series of the coefficients occurring in the exponentially fitted version of (9). Switching to EF methods should at least keep the order of the DC scheme at $r + p$, i.e. the light gray areas should agree element-wise. If $r > 0$, this is only possible if $\upsilon = \mu$.

The coefficients in the darker shaded cells together form non-trivial expressions $b_q(\boldsymbol{\phi[\mu]^1}, y_k)$ and $b_q(-\boldsymbol{\psi[\omega]^{\eta[\mu]}}, y_k)$ respectively. A nonzero difference

$$b_q(\boldsymbol{\phi[\mu]^1} + \boldsymbol{\psi[\omega]^{\eta[\mu]}}, y_k) \tag{10}$$

between those expressions sets the order of the DC scheme to $p + r$. If proper values for μ and ω can be found such that (10) is zero at all times, they can be used to increase the order of the DC scheme by one for the problem at hand.

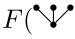
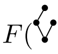
Example

Table 4 lists part of the occurring coefficients in a trapezoid–2-stage Radau I EFDC scheme. The columns for the classical DC scheme are omitted, since they always coincide with first columns of the corresponding exponentially fitted variant. The coefficients marked in light gray already match, which leaves the difference between both dark gray

$\rho(t)$	$\boldsymbol{\phi}[\boldsymbol{\mu}]^{\mathbf{1}}$	$+h^2\mu^2$	$-\boldsymbol{\psi}^{\boldsymbol{\eta}[\boldsymbol{\mu}]}[\boldsymbol{\omega}]$	$+h^2$
1	0	$\frac{1}{12}$	0	$\frac{1}{12}\mu^2$
2	0	$\frac{1}{12}$	0	$\frac{1}{9}\mu^2 - \frac{1}{36}\omega^2$
3	$\frac{-1}{2}, \frac{-1}{2}$		$\frac{-1}{2}, \frac{-1}{2}$	
4	$-1, -1, -1, -1$		$\frac{-10}{9}, \frac{-10}{9}, -1, -1$	

Table 4: Coefficients in a trapezoid–2-stage Radau I EFDC scheme

diagonals as the order boundary of this EFDC scheme. To determine which frequencies both methods should be to tuned to in order to gain an order of accuracy, the expression

$b_4(\boldsymbol{\phi}[\boldsymbol{\mu}]^{\mathbf{1}} + \boldsymbol{\psi}[\boldsymbol{\omega}]^{\boldsymbol{\eta}[\boldsymbol{\mu}]}, y_k) = \frac{1}{72}\Big[(\omega^2 - \mu^2)F($ $)(y_k) + \frac{1}{3}F($  $)(y_k) + F($  $)(y_k)\Big]$

should be annihilated in every knot point.

4.2. The effect of error propagation in the initial solution

It is possible to rewrite part of the right hand side of (9) to take into account the approximate nature of y_k:

$$\begin{aligned} B(\boldsymbol{\phi}^{\mathbf{1}}, y(x_k)) + O(h^q) &= -B(\boldsymbol{\psi}^{\boldsymbol{\eta}}, B(\boldsymbol{e_k}, y(x_k))) \\ &= -B(\boldsymbol{e_k} \cdot \boldsymbol{\psi}^{\boldsymbol{\eta}}, y(x_k)), \end{aligned}$$

with

$$\begin{cases} \boldsymbol{e_k}(\emptyset) = 1 \\ \boldsymbol{e_k}(t) = 0 \text{ if } \rho(t) \leq p \\ \exists t : \rho(t) = p + 1 \wedge \boldsymbol{e_k}(t) \neq 0 \end{cases},$$

which represents the global error due to method ϕ^η in point x_k.

By applying theorem 1, it can be shown that $\boldsymbol{e_k} \cdot \boldsymbol{\psi}^{\boldsymbol{\eta}}(t) = \boldsymbol{\psi}^{\boldsymbol{\eta}}(t)$ if $\rho(t) < 2p + 2$. In other words, the accumulation of local errors in the initial solution provided by η does not affect the DC scheme until dealing with trees of order $\geq 2p + 2$. At that point, the non-constant nature of $\boldsymbol{e_k}$ for trees of order $2p + 2$ affects the contribution of h^{2p+1} to the right hand side of (9). In case $r = p$, this poses a problem when trying to annihilate the coefficient of h^q, since $q = p + r + 1 = 2p + 1$.

One option is to avoid this problem by using a pair of MIRK methods with $r < p$. The polluting effect of $\boldsymbol{e_k}$ still takes place at h^{2p+1}, but $q < 2p + 1$. An alternative is to upgrade the base method to order $p+1$ by annihilating the order boundary with suitable values for μ as described in section 2.2. This delays the effect of $\boldsymbol{e_k}$ to h^{2p+4}, leaving h^q unaffected. Parameter ω from the estimator method can then be used to annihilate the remaining order boundary of the entire EFDC scheme. However, Skeel's theorem provides a shortcut: it says that if both methods can be upgraded to order $p + 1$ and $p + r + 1$ respectively, then the DC scheme also gains one order. Computing the order boundaries is easier for two standalone methods than it is for one standalone method and a DC scheme.

For exponentially fitted methods, $\boldsymbol{e_k} \in Tr(\boldsymbol{\delta_\emptyset}, p+1)$ and $\boldsymbol{\psi}[\boldsymbol{\omega}]^{\boldsymbol{\eta}[\boldsymbol{\mu}]} \in Tr(\boldsymbol{\psi}[\boldsymbol{\omega}]^{\boldsymbol{\eta}[\boldsymbol{\mu}]}, p+1)$, which leads to $\boldsymbol{e_k} \cdot \boldsymbol{\psi}[\boldsymbol{\omega}]^{\boldsymbol{\eta}[\boldsymbol{\mu}]} \in Tr(\boldsymbol{\psi}[\boldsymbol{\omega}]^{\boldsymbol{\eta}[\boldsymbol{\mu}]}, 2p + 2)$. This confirms that the effect of the error build-up starts in the coefficient of h^{2p+2} for EFDC schemes as well.

5. Practical application

In this section, we will apply a few EFDC methods to the test problems below, most of which were taken from [10]. Although problem 11 is stated as an IVP, in our tests it was solved as if it were a BVP.

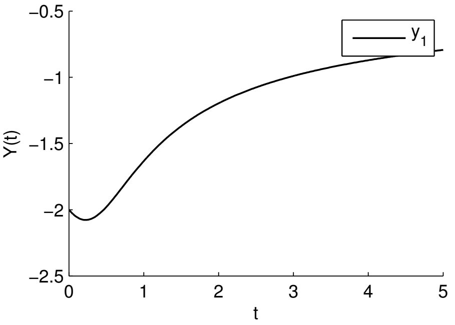
To obtain an estimate for the parameter values that eliminate the order boundary, the problems were each time first solved with classical DC scheme. With this initial guess, the occurring elementary differentials could be approximated, allowing to solve (10) for the frequency parameters.

5.1. Test set

5.1.1. Bernoulli DE

$$\begin{cases} y' = & \dfrac{2y+ty^4}{6} \\ y(0) = & -2 \end{cases} \tag{11}$$

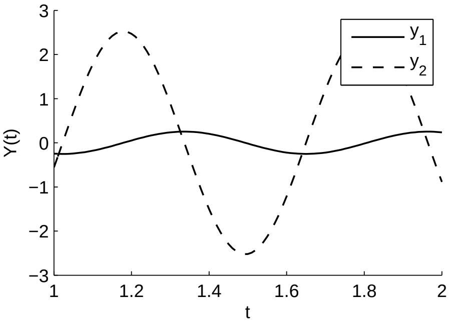
Solution: $y(t) = \dfrac{-2}{(4t-4+5e^{-t})^{\frac{1}{3}}}$



5.1.2. Bessel equation

$$\begin{cases} y'' = & -\left(10+\dfrac{1}{4t^2}\right)y \\ y(1) = & J_0(10) \\ y(2) = & \sqrt{2}J_0(20) \end{cases} \tag{12}$$

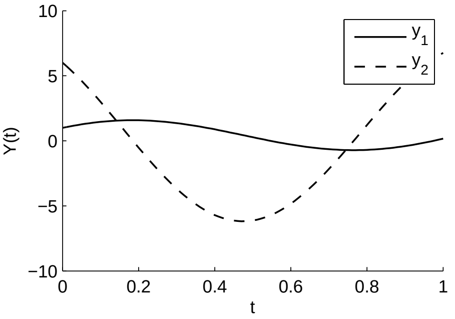
Solution: $y(t) = \sqrt{t}J_0(10t)$



5.1.3. Inhomogeneous equation

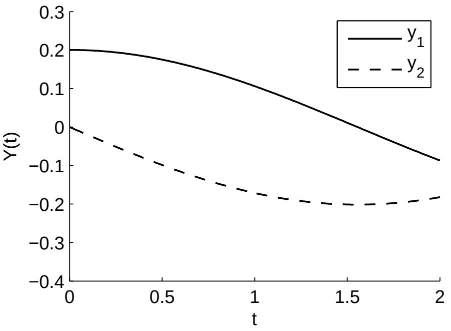
$$\begin{cases} y'' & = -\omega^2 y + (\omega^2-1)\sin(t) \\ y(0) & = 1 \\ y(1) & = \sin(\omega)+\cos(\omega)+\sin(1) \end{cases} \tag{13}$$

Solution: $y(t) = \sin(\omega t) + \cos(\omega t) + \sin(t)$



5.1.4. Duffing equation

$$\begin{cases} y'' &= -y - y^3 + 0.002\cos(1.01t) \\ y(0) &= 0.200426728067 \\ y(2) &= -0.08668702310 \end{cases} \tag{14}$$



Approximative solution from [11]: $y(t) = 0.200179477536\cos(1.01t) + 2.46946143 \times 10^4\cos(3.03t) + 3.04014 \times 10^7\cos(5.05t) + 3.74 \times 10^{10}\cos(7.07t) + \ldots$.

5.2. Interesting cases

In order to actually gain an order of accuracy in a EFDC scheme, appropriate values for μ and ω should be found such that

$$b_q(\boldsymbol{\phi}[\boldsymbol{\mu}]^{\mathbf{1}}, y(x_k)) = -b_q(\psi[\boldsymbol{\omega}]^{\boldsymbol{\eta}[\boldsymbol{\mu}]}, y(x_k)) \tag{15}$$

in every knot point.

Each side of this equation is an expression in one (left) or two (right) unknowns, the coefficients are scaled elementary differentials. If the problem to be solved is a system of n differential equations, these elementary differentials are vectors with n components. It follows that (15) is also a system of n equations.

5.2.1. $r = p$, $n = 1$

As shown in section 4.2, the contribution of h^q to the error term is polluted by the error build-up in the initial solution provided by η. While there are two unknowns in (15), at least one unknown will have to be sacrificed to mitigate or postpone the effect of $\boldsymbol{e_k}$. At most one parameter remains to perform the actual trick: a solution to (15) can only be guaranteed for scalar problems.

Figure 1 shows the results of the application of a trapezoid–3 stage Lobatto IIIA EFDC scheme to problem 5.1.1. The values for parameters μ and ω were chosen such that both methods would gain an order of accuracy for this problem when used in a standalone fashion. Section 4.2 explained how the EFDC scheme then also gains one order of accuracy.

5.2.2. $r < p$, $n \leq 2$

If the order gain of a DC scheme is not maximized, there is no such pollution at h^q. Both free parameters in (15) can be used to balance the equation, which allows a solution for problems with two components.

Figure 2 shows how this performs using a 2 stage Radau I–3 stage Lobatto IIIA EFDC scheme on problems 5.1.3 and 5.1.4. Since there was no explicit effort to increase the accuracy of the base method, the subresults obtained with the initial Radau I method (EFR3) remains of order 3. Only after completing the entire EFDC scheme, the profit

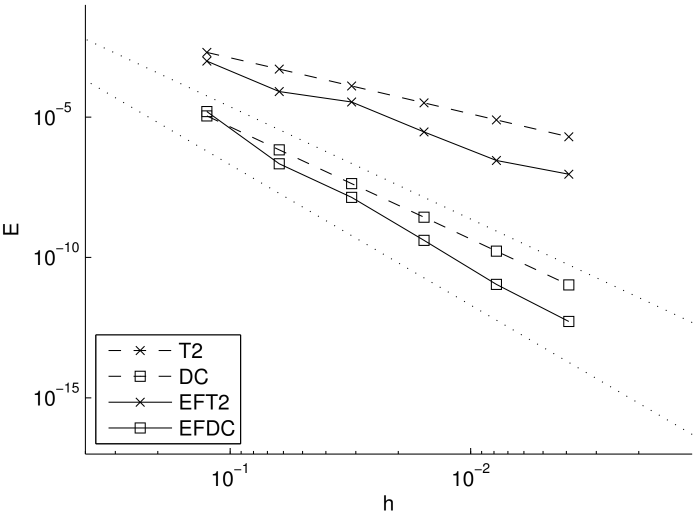


Figure 1: Error for problem 5.1.1, solved with a trapezoid–3 stage Lobatto IIIA EFDC scheme (i.e. $r = p$). Dotted lines mark fourth and fifth order. Using exponentially fitted methods (solid lines) allows to gain an order of accuracy.

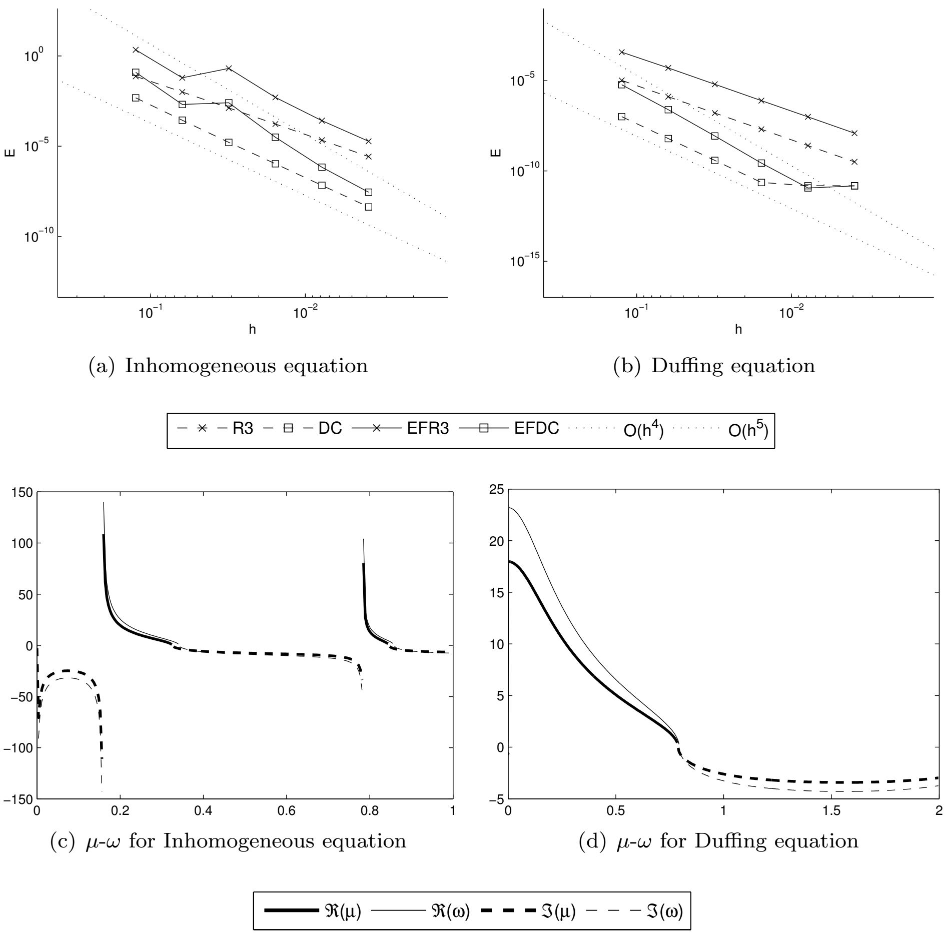
becomes apparent: the end result is of order 5. It appears however that the solution is less accurate: annihilating the coefficient of h^5 in the error required large values for μ and ω, inflating the coefficient of h^6. Plots 2(d) and 2(d) of figure 2 show the different values attributed to μ and ω for $h = 1/256$.

Since both free parameters can be used freely, it is possible to attribute a constant 0 to either μ or ω. This means that it is possible to combine a classical and an exponentially fitted method in an EFDC scheme. Figure 3 shows the error in the solution to problem 5.1.1 computed with a 3 stage Lobatto IIIA–5 stage-order 6 (from [1], with $c_3 = 1/4$, $c_4 = 3/4$) DC scheme. Only the former method was applied in exponentially fitted form. Even though the results obtained with the EF Lobatto IIIA method (EFL4) are less accurate than those of the classical version (L4), the end result of the EFDC scheme is nearly of order 7. Apparently, this combination of methods does not lead to an amplification of the remaining error terms.

$\rho(t)$	$\phi[\mu]^1$	$+h^2\mu^2$	$-\phi[\omega]^{\eta[\upsilon]}$	$+h^2$
0	0	0	0	0
1	0	$\frac{1}{36}$	0	$\frac{1}{36}\upsilon^2 - \frac{1}{36}\omega^2$
2	0		0	
3	$\frac{1}{9}, \frac{1}{9}, \frac{-1}{3}, \frac{-1}{3}$		0	

Table 5: Coefficients table for a 2 stage Radau I^2-EFDC scheme

Perhaps an interesting pathological case is the combination of two identical methods, in which $r = 0$. Although this construction has no meaning for classical MIRK-DC



(a) Inhomogeneous equation

(b) Duffing equation

(c) μ-ω for Inhomogeneous equation

(d) μ-ω for Duffing equation

Figure 2: Error obtained with a EFDC scheme with $r = 1 < p = 3$ (2 stage Radau I–3 stage Lobatto IIIA). The dotted lines marks fourth and fifth order. Using exponentially fitted methods allows to gain an order of accuracy, although the size of the error actually increases. This is due to the large values attributed to μ and ω, as shown in (c)-(d) for $h = 1/256$.

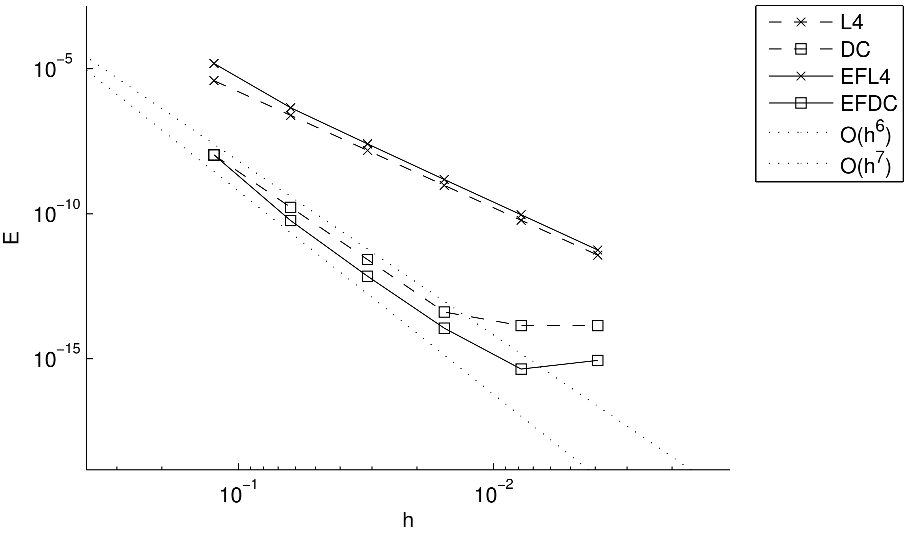


Figure 3: Error for problem 5.1.1 with an $r = 2 < p = 4$ scheme. The first stage of the EFDC-scheme remains of order 4, but the entire EFDC procedure leads a solution of order 7. The application of exponential fitting in this case leads to the gain of one order of accuracy and a smaller error.

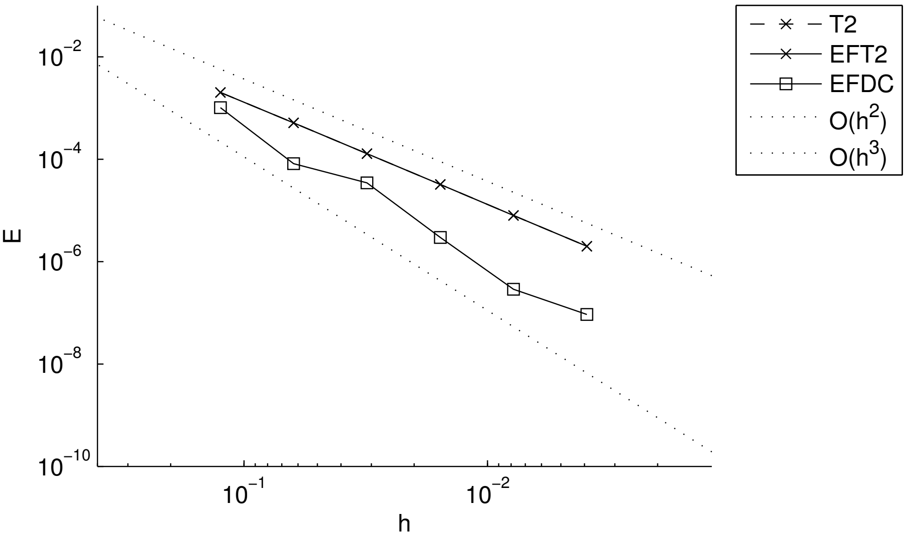
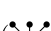
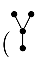
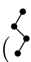


Figure 4: Error for problem 5.1.1 with $r = 0, p = 2$: the combination of two 2 stage Radau I methods with different parameters. The first stage of the EFDC-scheme remains of order 2, but the entire EFDC procedure leads a solution of order 3. The application of exponential fitting in this case leads to the gain of one order of accuracy and a smaller error.

schemes, the order boundary is non-trivial for an exponentially fitted scheme tuned to three distinct frequencies. Table 5 is the equivalent of table 3 for this setup, using a 3 stage Radau I method with parameter μ as base method and with parameter ω in the residual estimator. The requirement that the same parameter should be used for $\boldsymbol{\phi}$ and $\boldsymbol{\eta}$ no longer holds since $r = 0$, so a new parameter υ is used in the exponentially fitted $\boldsymbol{\eta}$. The order boundary constructed from both methods is

$b_4(\boldsymbol{\phi}[\boldsymbol{\mu}]^{\mathbf{1}} + \boldsymbol{\psi}[\boldsymbol{\omega}]^{\boldsymbol{\eta}[\boldsymbol{\upsilon}]}, y_k) = \frac{1}{72}\Big[(\omega^2 - \mu^2 + \upsilon^2)F($ $)(y_k) +$

$\frac{1}{3}F($  $)(y_k) + F($ $)(y_k) - F($  $)(y_k) - F($  $)(y_k)\Big]$

This equation contains three parameters, albeit not very usefully interconnected: there is only one degree of freedom available for leading error term annihilating purposes.

5.2.3. $n > 1 + \delta_{r<p}$

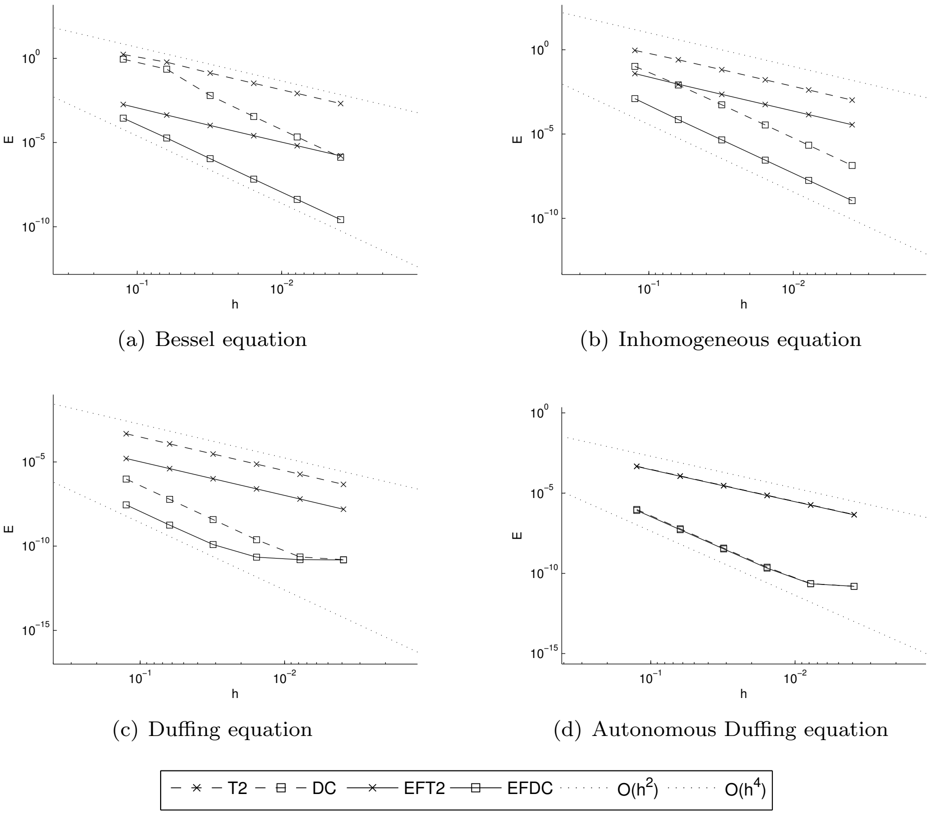
In this general case, there are more components than available free parameters. System (15) is overdetermined and usually has no solution. One thing that can be done with the available parameters is minimizing the difference between both sides with a least squares procedure.

Figure 5 shows what can be gained by using an implementation of this approach on some of the test problems. In this setup, the order boundary of the base method and the estimator method were minimized separately. As a result, the size of the error in both phases of the EFDC scheme is reduced. Figure 5(d) is an exception: the exponentially fitted trapezoid rule is not able to reconstruct the artificial spatial component accurately, unless $\mu = 0$. The size of the leading error term depends on the accuracy of this component, so trying to minimize the order boundary pulls the μ towards zero, away from values more suitable for the other components.

An other approach is to minimize the difference between the order boundary of the EFDC scheme as a whole. The results obtained with an trapezoid–3 stage Lobatto IIIA-EFDC scheme using this tuning is shown in figure 6. Sadly, only one plot shows a reduction in error size. This is due to the polluting effect of $\boldsymbol{e_k}$ in the initial solution. Since $r = p = 2$, the effect of this error accumulation affects the coefficient of h^5, which was supposed to be minimized. Although a combination of methods with $r < p$ would be unaffected by this, it can still suffer from (often severe) error inflation, as already shown in figure 2.

6. Conclusion

We have considered using exponentially fitted methods in MIRK deferred correction schemes. The B-series coefficients structure was examined to reveal the order boundary of both standalone EF methods and EFDC schemes. Of the latter, it was shown that special attention should be paid to the base method if the order gain is maximized, i.e. $r = p$. A viable approach is to minimize the leading error term of both methods separately, as suggested by Skeel's theorem. If $r < p$, an order of accuracy can be gained for problems with two components, although the size of the error actually increased.



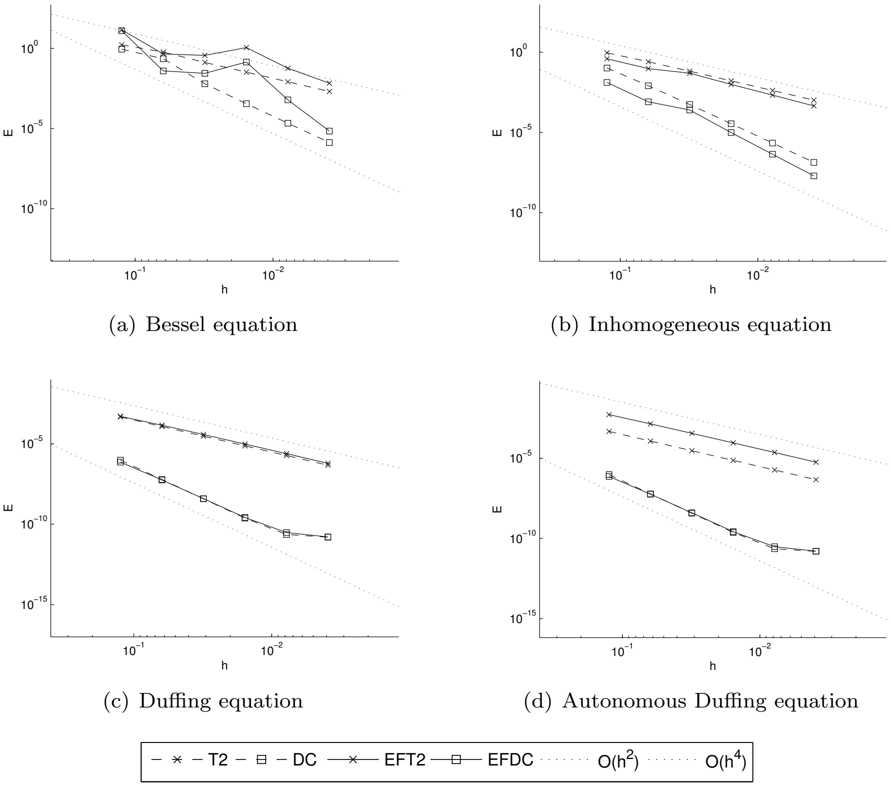
(a) Bessel equation

(b) Inhomogeneous equation

(c) Duffing equation

(d) Autonomous Duffing equation

Figure 5: Error for a trapezoid–3 stage Lobatto IIIA-EFDC scheme applied to several test problems. The leading error terms of both methods were minimized separately. All but one plots show a smaller error in each stage of the EFDC scheme. The error size in figure (d) is not reduced because the values for μ are pulled towards zero by the spatial component.



(a) Bessel equation

(b) Inhomogeneous equation

(c) Duffing equation

(d) Autonomous Duffing equation

Figure 6: Error for a trapezoid–3 stage Lobatto IIIA-EFDC scheme applied to several test problems. The leading error term of the entire EFDC-scheme was minimized by setting μ and ω to an appropriate value, but only in (b) there is an actual improvement in error size.

References

[1] K. Burrage, F. H. Chipman, P. H. Muir, Order results for mono-implicit Runge-Kutta methods, SIAM J. Num. Anal 31 (1993) 876–891.

[2] J. R. Cash, Numerical integration of non-linear two-point boundary-value problems using iterated deferred corrections–I : A survey and comparison of some one-step formulae, Comput. Math. Appl. 12 (10, Part 1) (1986) 1029–1048.

[3] `http://www2.imperial.ac.uk/~jcash/BVP_software/readme.php`.

[4] R. D. Skeel, A theoretical framework for proving accuracy results for deferred corrections, SIAM J. Num. Anal. 19 (1) (1982) 171–196.

[5] G. Ixaru, G. Vanden Berghe, Exponential fitting, Kluwer Academic Publishers, Dordrecht, 2004.

[6] D. Hollevoet, M. Van Daele, G. Vanden Berghe, The optimal exponentially-fitted numerov method for solving two-point boundary value problems, J. Comput. Appl. Math. 230 (1) (2009) 260–269.

[7] W. H. Enright, P. H. Muir, Efficient classes of Runge-Kutta methods for two-point boundary value problems, Computing 37 (4) (1986) 315–334.

[8] M. Van Daele, J. R. Cash, Superconvergent deferred correction methods for first order systems of nonlinear two-point boundary value problems, SIAM J. Sci. Comput. 22 (5) (2000) 1697–1716.

[9] E. Hairer, S. Nørsett, G. Wanner, Solving Ordinary Differential Equations: Nonstiff Problems, Springer, New York, 1993.

[10] C. Tsitouras, T. E. Simos, Optimized Runge-Kutta pairs for problems with oscillating solutions, J. Comput. Appl. Math. 147 (2) (2002) 397–409.

[11] R. Van Dooren, Stabilization of Cowells classical finite difference methods for numerical integration, J. Comput. Phys. 16 (1974) 186–192.