Abstract This paper presents a sequence of self-starting deferred correction (DC) schemes built recursively from a modified trapezoidal rule for the numerical solution of general first order ordinary differential equations. It is proven that each scheme is A-stable and that the correction on a scheme DC2j (of order 2j of accuracy) leads to a scheme DC(2j+2) (of order 2j+2). The proof is based on a deferred correction condition (DCC) which guarantees the order of accuracy. Any other scheme (e.g. BDF or RK families) satisfying the DCC can be corrected to increase by two the order of accuracy while preserving the stability property of the corrected scheme. Numerical experiments with standard stiff ODEs are performed with the DC2, ..., DC10 schemes and show that the expected orders of accuracy are achieved together with excellent stability of the method. A-stable backward Euler schemes of arbitrary order resulting from the DC method are also presented.

Keywords ordinary differential equations, high order time-stepping methods, deferred correction, A-stability.

1 Introduction

There exists a vast literature on numerical methods for solving ordinary differential equations (ODEs). Large classes of methods both for stiff problems (problems extremely hard to solve by standard explicit step-by-step methods [18]) and non-stiff problems are proposed, using a diversity of approaches and leading to varying orders of accuracy (see for instance [20] and references therein). Since the second half of the 20th century, the study of stiff ODEs
attracts much attention. Dahlquist introduced the notion of A-stable methods to characterize methods able to solve stiff ODEs and stated the second Dahlquist barrier: «The order of an A-stable linear multi-step method can not exceed 2. The smallest error constant, $c^* = 1/12$, is obtain for the trapezoidal rule,...» (see [21]). To overcome the severe restriction due to the second Dahlquist barrier, a number of numerical methods which are not linear multi-step are proposed (e.g. [1, 13, 14, 22]). In this paper we investigate methods based on Deferred Correction (DC).

The deferred correction method is used to improve the order of accuracy of a numerical method of low order. This method is explored by many authors, e.g. [6, 7, 9, 10, 13]. The approach adopted in [6] is an extension of iterative deferred correction which consists of transforming a continuous nonlinear problem into a discretized one by the mean of asymptotic expansions [15]. The approaches in [7, 10, 13] are quite similar and consist in writing the global error for an existing discrete solution of an ODE into a new ODE or a Picard integral equation. The numerical approximation of the global error equation is then added to the existing approximate solution to improve its order of accuracy. The method in [13], reaching order up to 14, requires sufficiently small time steps for moderately stiff problems while convergence is reduced to order 2 for “very stiff” problems. The method in [9] addresses linear ODEs for which a monotonicity condition is enforced. It consists in a sequence of schemes built recursively from the trapezoidal rule (Crank-Nicholson) via asymptotic expansions of the linear ODEs by central finite difference approximations [11]. Formule (4.6.8) and (5.3.12)]. Numerical experiments on a one-dimensional linear parabolic equation and a one-dimensional linear homogeneous hyperbolic equation are performed and show that the method is effective (orders 2, 4 and 6 are achieved).

The purpose of this paper is to investigate high-order A-stable methods based on deferred correction strategy for the general first order ODE

$$\begin{cases}
\frac{du}{dt} = F(t, u), & t \in [0, T] \\
u(0) = u_0,
\end{cases}$$

where the unknown $u$ is from $[0, T]$ into an arbitrary Banach space and $F$ is any regular function. The choice of the functional space is motivated by the applicability of the results to time-evolution partial differential equations (PDEs) when the space is for example discretized using finite elements [19]. We adopt the approach developed in [9] which only deals with linear ODEs in $\mathbb{R}^d$, under a monotonicity condition. Our result is more general, since the functional space of the solution and the right hand side $F$ are more general. All our results about the order of convergence and A-stability require original arguments for their proof. Since we deal with nonlinear ODEs, we start the correction from a modified trapezoidal rule which shares the same properties (in term of stability and order of convergence) as the trapezoidal rule. The modified trapezoidal rule takes the following form, for a discretization on a uniform grid $t_0 = 0 < t_1 < \cdots < t_n = T$, $t_n = nk$, $k > 0$,
\begin{equation}
\begin{aligned}
\left\{ \begin{array}{l}
u^{2,n+1} - u^{2,n} \\ u^{2,0} = u_0
\end{array} \right. = \frac{k}{2} F \left( t_{n+1/2}, \frac{u^{2,n+1} + u^{2,n}}{2} \right)
\end{aligned}
\end{equation}

In this formula, $u^{2,n}$ represents the approximation of order 2 of the exact solution at time $t = t_n$. Each corrected scheme appears as an advantageous perturbation of the modified trapezoidal rule and inherits the A-stability property of the trapezoidal rule while the order of accuracy increases by two per correction stage. The order of accuracy of the deferred correction schemes is guaranteed by a deferred correction condition (DCC) that holds for the modified trapezoidal rule and each corrected scheme. We prove that, provided the DCC is satisfied, the correction can be made on any time-stepping scheme (such as BDF, Runge-Kutta, ...), increasing the order of accuracy successively by two and preserving the stability properties of the starting scheme. We present also deferred correction for Euler rule which provide a recursive sequence of A-stable schemes of arbitrary order, assuming the backward Euler rule is used.

The paper is organized as follows: in section 2 we present basic formulae from finite difference approximations; in section 3 we introduce the generalized deferred correction schemes for trapezoidal rule; section 4 deals with the analysis of convergence and order of accuracy; in section 5 we introduce DC method for Euler rule; the analysis of absolute stability of DC method is done in section 6. Finally, in section 7 we present numerical results to show the performance of the method.

### 2 Basic formulae for finite difference approximations

We consider $X$ to be a Banach space equipped with the norm $\| \cdot \|$. Given a sufficiently differentiable function $g : X \to X$ we denote the differential of $g$ at $x \in X$ along the direction $h$ by

$$df(x) \cdot h = df(x)(h).$$

Furthermore, from the isomorphism $\mathcal{L}(X; \mathcal{L}(X; X)) \approx \mathcal{L}(X \times X; X)$ between the space of linear maps from $X$ to $\mathcal{L}(X; X)$ and the space of bilinear maps from $X \times X$ into $X$ \cite{2}, the differential of order $m$ of $g$ at $x$ along $(h_1, \cdots, h_m)$ is written as

$$((...d^m g(x) \cdot h_m) \cdots h_2) \cdot h_1 = d^m g(x)(h_1, \cdots, h_m).$$

As in \cite{9} we define the centered, forward and backward difference operators $D$, $D_+$ and $D_-$, respectively, such that, given a function $f$ from $\mathbb{R}$ into $X$ we have

$$D f(x + \Delta x/2) = \frac{f(x + \Delta x) - f(x)}{\Delta x}, \quad D_+ f(x) = \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

As in \cite{9} we define the centered, forward and backward difference operators $D$, $D_+$ and $D_-$, respectively, such that, given a function $f$ from $\mathbb{R}$ into $X$ we have

$$D f(x + \Delta x/2) = \frac{f(x + \Delta x) - f(x)}{\Delta x}, \quad D_+ f(x) = \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

As in \cite{9} we define the centered, forward and backward difference operators $D$, $D_+$ and $D_-$, respectively, such that, given a function $f$ from $\mathbb{R}$ into $X$ we have

$$D f(x + \Delta x/2) = \frac{f(x + \Delta x) - f(x)}{\Delta x}, \quad D_+ f(x) = \frac{f(x + \Delta x) - f(x)}{\Delta x}$$
and
\[ D_+ f(x) = \frac{f(x) - f(x - \Delta x)}{\Delta x}. \]

We denote the average operator by \( E \):
\[ Ef(x + \Delta x/2) = \frac{f(x + \Delta x) + f(x)}{2}. \]

The composites of \( D_+ \) and \( D_- \) are defined recursively. They commute, that is
\[ (D_+ D_-) f(x) = (D_- D_+) f(x) = D_- D_+ f(x), \]
and satisfy the identities
\[ (D_+ D_-)^m f(x) = \frac{1}{\Delta x^{2m}} \sum_{i=0}^{2m} (-1)^i \binom{2m}{i} f(x + (m - i)\Delta x), \] (1)
and
\[ D_-(D_+ D_-)^m f(x) = \frac{1}{\Delta x^{2m+1}} \sum_{i=0}^{2m+1} (-1)^i \binom{2m+1}{i} f(x + (m - i)\Delta x). \] (2)

Formulae (1) and (2) can be proven by a straightforward induction argument.

We introduce the double index \( \alpha^m = (\alpha^m_1, \alpha^m_2) \in \{0, 1, \ldots, m\} \times \{0, 1, \ldots, m\} \) such that
\[ D^{\alpha^m} f(x) = D^{\alpha^m_1} D^{\alpha^m_2} f(x). \] (3)

**Remark 1** If \( |\alpha^m| = \alpha^m_1 + \alpha^m_2 \) is even, then there exists \( x' \) such that
\[ D^{\alpha^m} f(x) = (D_+ D_-)^{|\alpha^m|/2} f(x'). \] (4)

**Example 1** \( D_+ D^3 f(x) = (D_+ D_-)^2 f(x - \Delta x) \) and \( D^4 f(t_n) = (D_+ D_-)^2 f(x - 2\Delta x) \).

### 2.1 Some properties of finite difference approximations

**Proposition 1** Assuming sufficient differentiability of the function \( f \) from \( \mathbb{R} \) to the Banach space \( X \), we have
\[ \| D^{\alpha^m} f(x) \| \leq \max_{a \leq y \leq b} \left\| \frac{d^{\alpha^m}|f(y)}{dx^{\alpha^m}} \right\|, \] (5)
where \( a = x - \alpha^m_1 \Delta x \) and \( b = x + \alpha^m_1 \Delta x \).

**Proof** This inequality can be deduced from [12, Theorem 1, p.249] owing to Remark 1 and the identity \( D^{\alpha^m} f(x) = m! f[x, x + \Delta x, \ldots, x + m\Delta x] \), where \( f[x, x + \Delta x, \ldots, x + m\Delta x] \) is the usual Newton’s divided difference.
Proposition 2 (Finite difference approximation of a composite) Consider two functions $f$ and $u$ with values into Banach spaces such that the composite $f \circ u$ is defined on $\mathbb{R}$ and the differential $df$ is integrable. Then

$$D_f(u(x)) = \int_0^1 df[u(x - \Delta x) + \Delta x D_f u(x) \tau](D_f u(x)) d\tau$$

and

$$D_f(u(x)) = \int_0^1 df[u(x) + \Delta x D_f u(x) \tau](D_f u(x)) d\tau$$

Proof: The function $\varphi(t) = f(u(x - \Delta x) + t \Delta x D_f u(x))$ is differentiable on $[0, 1]$ and we have

$$\varphi(1) - \varphi(0) = \int_0^1 \varphi'(\tau) d\tau = \Delta x \int_0^1 df[u(x - \Delta x) + \Delta x D_f u(x) \tau](D_f u(x)) d\tau.$$ 

Therefore (6) follows from the identity $D_f(u(x)) = (\varphi(1) - \varphi(0))/\Delta x$ and the identity (7) is obtained similarly.

2.2 Central difference approximation

There are various formulae for the approximation of the derivative of a function by finite difference [3–5, 11]. In this subsection we present approximations which comply with the analysis of consistency in section 4. We need the following lemma which proof is an easy induction.

Lemma 1 For each integer $m = 1, 2, \ldots$ and for any real $r$, we have

$$\sum_{j=0}^{m} (-1)^j \binom{m}{j} (m + r - j)^p = \begin{cases} 0, & \text{if } 1 \leq p < m, \\ m!, & \text{if } p = m. \end{cases}$$

(8)

In particular, for any nonnegative integer $p$, we have

$$\sum_{j=0}^{2m} (-1)^j \binom{2m}{j} (m - j)^{2p+1} = 0,$$

and

$$\sum_{j=0}^{2m+1} (-1)^j \binom{2m + 1}{j} (m - j + 1/2)^{2p} = 0,$$

and

$$\sum_{j=0}^{2m} (-1)^j \binom{2m}{j} [(m - j + 1/2)^{2p+1} + (m - j - 1/2)^{2p+1}] = 0.$$ 

(11)

We have the following theorems.
Theorem 1 Let \( p \) be a positive integer and \( f \in C^{2p+2}([0,T], X) \). Let \( 0 = t_0 < t_1 < \ldots < t_N = T \), \( t_n = nk \), be a partition of \([0,T]\). Then, we have

\[
f'(t_{n+1/2}) = \frac{f(t_{n+1}) - f(t_n)}{k} - \sum_{i=1}^{p} c_{2i+1} k^{2i} D(D_k D_-)^i f(t_{n+1/2}) + O(k^{2p+2})
\] (12)

and

\[
f(t_{n+1/2}) = \frac{f(t_{n+1}) + f(t_n)}{2} - \sum_{i=1}^{p} c_{2i} k^{2i} (D_k D_-)^i E f(t_{n+1/2}) + O(k^{2p+2}),
\] (13)

where \( c_2, c_3, \ldots, c_{2p+1}, \ldots \) are scalars independent from \( p \). Table 1 gives the first ten coefficients \( c_i \).

**Table 1** Ten first coefficients of central difference approximations (12) and (13)

| \( c_2 \) | \( c_3 \) | \( c_4 \) | \( c_5 \) | \( c_6 \) | \( c_7 \) | \( c_8 \) | \( c_9 \) | \( c_{10} \) | \( c_{11} \) |
|---|---|---|---|---|---|---|---|---|---|
| \( \frac{1}{8} \) | \( \frac{1}{24} \) | \( -\frac{18}{472} \) | \( \frac{18}{972} \) | \( 450 \) | \( \frac{450}{772} \) | \( 22050 \) | \( \frac{22050}{972} \) | \( -22050 \) | \( \frac{1786050}{10921} \) | \( \frac{1786050}{11921} \) |

Proof The Taylor expansion

\[
f(t_{n+m-j}) = f(t_{n+1/2}) + \sum_{i=1}^{2p+1} \frac{k^i}{i!} (m - j - 1/2)^i f^{(i)}(t_{n+1/2}) + O(k^{2p+2})
\] (14)

together with the formula (2), the identities (8) and (10) imply that

\[
f^{(2q+1)}(t_{n+1/2}) = -\frac{1}{k^{2q+1}} \sum_{i=q+1}^{p} k^{2i+1} (2i+1)! f^{(2i+1)}(t_{n+1/2}) \sum_{j=0}^{2q+1} (-1)^j \binom{2q+1}{j} (q-j+1/2)^{2i+1}
\]

\[+ D(D_k D_-)^q f(t_{n+1/2}) + O(k^{2p-2q+1}).
\] (15)

Similarly, (14) together with (11), (8) and (11) yields

\[
f^{(2q)}(t_{n+1/2}) = -\frac{1}{2k^{2q}} \sum_{i=q+1}^{p} k^{2i} (2i)! f^{(2q)}(t_{n+1/2}) \sum_{j=0}^{2q} (-1)^j \binom{2q}{j} \left[ (q-j+1/2)^{2i} + (q-j-1/2)^{2i} \right]
\]

\[+ E(D_k D_-)^q f(t_{n+1/2}) + O(k^{2p-2q+2}).
\] (16)

In particular, formula (14) for \( m-j = 0, 1 \) yields

\[
f(t_{n+1}) = f(t_n) + k f'(t_{n+1/2}) + \sum_{i=1}^{p} d_{1,2i+1} k^{2i+1} (2i+1)! f^{(2i+1)}(t_{n+1/2}) + O(k^{2p+3})
\] (17)
and
\[ f(t_{n+1}) = -f(t_n) + 2f(t_{n+1/2}) + \sum_{i=1}^{p} d_{i,2i} \frac{2^{2i}}{(2i)!} f^{(2i)}(t_{n+1/2}) + O(k^{2p+2}), \quad (18) \]

with \( d_{i,i} = 2^{i-1} \), for \( i = 2, 3, ..., 2p + 1 \). Therefore, substituting successively the derivatives \( f^{(2i)}(t_{n+1/2}) \), \( f^{(2i+1)}(t_{n+1/2}) \), ... and \( f^{(2)}(t_{n+1/2}) \), \( f^{(4)}(t_{n+1/2}) \), ...

by their expansion given by the formulæ [11, Formula 4.6.8 and 10], respectively, into (17) and (18), we derive the identities

\[ f(t_{n+1}) = f(t_n) + k f'(t_{n+1/2}) + d_{1,2} f^{(2)}(t_{n+1/2}) + ... \]

\[ + d_{q,2q+1} 2^{q+1} D(D+D-) f(t_{n+1/2}) + \sum_{i=1}^{p} d_{q+1,2i+1} 2^{2i+1} f^{(2i+1)}(t_{n+1/2}) + O(k^{2p+3}) \]

and

\[ f(t_{n+1}) = f(t_n) + 2f(t_{n+1/2}) + d_{1,2} k^2 D f(t_{n+1/2}) + ... \]

\[ + d_{q,2q+1} 2^{q+1} D(D+D-) f(t_{n+1/2}) + \sum_{i=1}^{p} d_{q+1,2i+1} k^{2i} f^{(2i)}(t_{n+1/2}) + O(k^{2p+2}) \]

where for \( q = 2, ..., p \) and \( i = q, q+1, ..., p \), we have

\[ d_{q,2i+1} = d_{q-1,2i+1} - \frac{d_{q-1,2q-1}}{(2q-1)!} \sum_{j=0}^{2q-1} (-1)^j \begin{pmatrix} 2q-1 \\ j \end{pmatrix} (q-j-1/2)^{2i+1}, \]

and

\[ d_{q,2i} = d_{q-1,2i} - \frac{d_{q-1,2q-2}}{(2q-2)!} \sum_{j=0}^{2q-2} (-1)^j \begin{pmatrix} 2q-2 \\ j \end{pmatrix} (q-j-1/2)^{2i} + (q-j-3/2)^{2i}. \]

Finally, the identities (12) and (13) follow by setting \( c_{2i} = d_{i,2i}/(2i!) \) and \( c_{2i+1} = d_{i,2i+1}/(2i+1)! \), for \( i = 1, 2, ..., p \). The independence of the coefficients \( c_i \) with respect to \( p \) follows from the construction.

**Remark 2** The approximations (12) and (13), from the coefficients \( c_i \) computed in Table 7, are equivalent to the central-difference approximation of the first derivative and the centered Bessel’s formulæ [11, Formulae 4.6.8 and 5.3.12].

**Theorem 2** Let \( f \in C^{p+1}(0, T], X \) and \( (m, r) \) be the couple of integers such that \( p + 1 = 2m + r \), \( r = 0 \) or 1. Let \( 0 = t_0 < t_1 < ... < t_N = T \), \( t_n = nk \), be a partition of the interval \([0, T]\). Then for each \( n = 1, 2, ..., N-1 \) such that \( D^-(D+D-)^{-m} f(t_n) \) is defined, we have

\[ f'(t_n) = \sum_{i=1}^{m+r-1} a_{2i} k^{2i-1} (D+D-) f(t_n) + \sum_{i=0}^{m-1} a_{2i+1} k^{2i} D^{-1} D (D+D-)^{i} f(t_n) + O(k^{p}), \quad (19) \]

where \( a_1, a_2, ..., a_p, ... \) are scalars independent from the order \( p \). Table 2 gives the first ten coefficients \( c_i \).
### Table 2

| $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ | $a_6$ | $a_7$ | $a_8$ | $a_9$ |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1     | 1/2   | 1/3   | 2/3   | 4/3   | 6/3   | 14/3  | 144/3 | 576/3 |

### 3 Deferred correction for trapezoidal rule

Let us consider the Cauchy problem:

\[
\begin{aligned}
\frac{du}{dt} &= F(t, u), \quad t \in [0, T] \\
u(0) &= u_0,
\end{aligned}
\]

where the unknown $u = u(t)$ is a function from $[0, T]$ to the Banach space $X$, $u_0$ and $F = F(t, u)$ are given. We suppose that the function $F$ is $C^{2p+2}$ such that the problem (20) has a unique solution $u \in C^{2p+3}([0, T], X)$, for some positive integer $p$. We investigate approximate solutions $u^n \simeq u(t_n)$ of this problem at the points $0 = t_0 < t_1 < \ldots < t_N = T$, with $t_n = nk$ for a time step $k > 0$. From the formulae (12)-(13) we have the approximation

\[
\frac{du}{dt}(t_n + 1/2) \simeq \frac{u^{n+1} - u^n}{k} - \sum_{i=1}^{p} c_{2i+1} k^{2i} D^i(D_+ D_-)^{i-1} u^{n+1/2}
\]

and

\[
u(t_{n+1/2}) = \frac{u^{n+1} + u^n}{2} - \sum_{i=1}^{p} c_{2i} k^{2i} (D_+ D_-)^{i-1} E u^{n+1/2},
\]

and derive the time-stepping scheme

\[
\frac{u^{n+1} - u^n}{k} = - \sum_{i=1}^{j} c_{2i+1} k^{2i} D^i(D_+ D_-)^{i-1} u^{n+1/2}
\]

\[
= F(t_{n+1/2}, \frac{u^{n+1} + u^n}{2} - \sum_{i=1}^{j} c_{2i} k^{2i} (D_+ D_-)^{i-1} E u^{n+1/2}),
\]

\[j = 1, 2, \ldots, p.\] This is a class of multi-step schemes that required $2j + 1$ initial values, while the Cauchy condition provides only one. We resort to the deferred correction (DC) method to transform (21) to a sequence of one step schemes as follows:

For $j = 0$ we have the modified Trapezoidal rule

\[
\frac{u^{2,n+1} - u^{2,n}}{k} = F(t_{n+1/2}, \frac{u^{2,n+1} + u^{2,n}}{2}), \quad u^{2,0} = u_0
\]
For $j \geq 1$,

\[
\frac{u^{2j+2,n+1} - u^{2j+2,n}}{k} = \sum_{i=1}^{j} c_{2i+1} k^{2i} (D_i D_-)^i (u^{2j,n+1/2})^i - \sum_{i=1}^{j} c_{2i} k^{2i} (D_i D_-)^i (u^{2j,n})^i,
\]

\[
= F \left( t_{n+1/2}, \frac{u^{2j+2,n+1} + u^{2j+2,n}}{2} - \sum_{i=1}^{j} c_{2i} k^{2i} (D_i D_-)^i (u^{2j,n+1/2})^i \right). \tag{23}
\]

\[
u^{2j+2,0} = u_0. \tag{24}
\]

The scheme (23)-(24) has unknowns $u^{2j+2,n}$, $n = 1, 2, ..., N$, and is deduced from (21) by substituting the unknown $u^n$ under the summation symbol by $u^{2j,n}$. The index $2j$ indicates that $\{u^{2j,n}\}$ approximates the exact solution with order $2j$ of accuracy. We call the schemes (23)-(24) Deferred Correction of order $2j + 2$ for the trapezoidal rule, denoted DC$(2j+2)$.

**Remark 3** The scheme (23)-(24) involves unknowns $u^{2j-1,-1}, ..., u^{2j,-j}$ which represent an approximate solution of (21) for $t = -k, ..., -jk$. We propose to compute such values to get self-starting methods providing approximate solutions which comply with the expected order of accuracy (this makes sense since the exact solution of such problem exists in the neighborhood of the origin 0). A way to avoid those values corresponding to $t < 0$ is to provide $u^{2j+2,1}, ..., u^{2j+2,j}$ that match the order of accuracy. Usually one takes $u^{2j+2,n} = u^{2j,n}$, for $n = 0, 1, ..., j$, but the latter approach may lead to a loss of accuracy.

For the analysis below we suppose that $u^{2j+2,1}, ..., u^{2j+2,j}$ are given and satisfy

\[
\|u^{2j+2,n} - u(t_n)\| \leq C k^{2j+2}, \quad \text{for } n = 1, 2, ..., j, \tag{25}
\]

where $C$ is a constant independent from $k$.

We give the following definition which provides a sufficient condition for the scheme (23)-(24) to achieve order $2j + 2$ of accuracy:

**Definition 1** Let $u$ be the exact solution of the Cauchy problem (20). Given a positive integer $j$, the sequence $\{u^{2j,n}\}$ is said to satisfy the Deferred Correction Condition (DCC) for the trapezoidal rule if $\{u^{2j,n}\}$ approximates the exact solution $u(t_n)$ with order $2j$ of accuracy and we have

\[
\|D (D_- u^{2j,n+1/2} - u(t_{n+1/2}))\| + \|D_- (u^{2j,n+1} - u(t_{n+1}))\| \leq C k^{2j}, \tag{26}
\]

for $n = 1, 2, ..., N - 2$, where $C$ is a constant depending only on $j$, $T$ and the exact solution $u$. 
Remark 4 The condition (26) is equivalent to
\[ \|D(D_+D_-)^m(u^{2j,n+1/2} - u(t_{n+1/2}))\| + \|(D_+D_-)^m(u^{2j,n+1} - u(t_{n+1}))\| \leq C k^{2j-2m+2}, \]
for each \( m = 1,2,...,j \) and \( n = m,m+1,...,N-m-1 \). This results from (1) together with the identity
\[ D(D_+D_-)^m(u^{2j,n+1/2} - u(t_{n+1/2})) = (D_+D_-)^{m-1}[D(D_+D_-)^m(u^{2j,n+1/2} - u(t_{n+1/2}))]. \]

4 Convergence of the deferred correction schemes

In this section we prove the convergence with order \( 2j+2 \) of the scheme (23)-(24) with (DCC) as sufficient condition. We also present conditions for an approximate solution of (20) to satisfy (DCC).

Theorem 3 Let \( j \) be a positive integer, \( 1 \leq j \leq p \), and \( \{u^{2j,n}\}_n \) a sequence satisfying (DCC) for the trapezoidal rule. Let \( \{u^{2j+2,n}\}_n \) be the solution of (23)-(24), built from \( \{u^{2j,n}\}_n \). Suppose that one of the following four conditions holds:

(i) \( F \) is \( \mu \)-Lipschitz with respect to the second variable \( x \).
(ii) \( X \) is a Hilbert space with inner product \( (\cdot,\cdot) \) and \( F \) satisfies the monotonicity condition
\[ (F(t,x) - F(t,y), x - y) \leq 0, \quad \forall (t,x,y) \in [0,T] \times X \times X. \]
(iii) \( X \) is finite dimensional and \( \{u^{2j+2,n}\}_n \) remains close to the exact solution \( u \) of the problem (20) in the sense that
\[ \|u^{2j+2,n} - u(t_n)\| \leq M, \quad \text{for} \quad n = 0,1,...,N, \]
where \( M \) is a constant independent from \( n \) and \( k \).
(iv) \( \{u^{2j+2,n}\}_n \) converges to the exact solution \( u \) of the problem (20).

Then \( \{u^{2j+2,n}\}_n \) approximates \( u \) with order \( 2j+2 \) of accuracy.

Proof (i) First we consider the case where the function \( F = F(t,x) \) is \( \mu \)-Lipschitz with respect to the second variable \( x \). Combining (20) and (23) we derive the identity
\[ D\Theta^{2j+2,n+1/2} = \sigma^{2j+2,n+1/2} + D(A - \Gamma)(u^{2j,n+1/2} - u(t_{n+1/2})) \]
\[ + F(t_{n+1/2}, \tilde{u}^{2j+2,n+1} - \Gamma \tilde{u}(t_{n+1})) - F(t_{n+1/2}, \tilde{u}(t_{n+1}) - \Gamma \tilde{u}(t_{n+1})), \]
where
\[ \Theta^{2j+2,n} = [u^{2j+2,n} - \Gamma u^{2j,n}] - [u(t_n) - \Gamma u(t_n)], \]
A-stable methods for ODEs via deferred Correction

\[ \Lambda u^{2j,n} = \sum_{i=1}^{j} c_{2i+1} k^{2i}(D_+ D_-)^i u^{2j,n}, \]

\[ \Gamma u^{2j,n} = \sum_{i=1}^{j} c_{2i} k^{2i}(D_+ D_-)^i u^{2j,n}, \]

and

\[ \sigma^{2j+2,n+1/2} = \left[ u'(t_{n+1/2}) - Du(t_{n+1/2}) + D\Lambda u(t_{n+1/2}) \right] \]
\[ - \left[ F(t_{n+1/2}, u(t_{n+1/2})) - F(t_{n+1/2}, \hat{u}(t_{n+1}) - \Gamma \hat{u}(t_{n+1})) \right]. \]

We can write

\[ F(t_{n+1/2}, u(t_{n+1/2})) - F(t_{n+1/2}, \hat{u}(t_{n+1}) - \Gamma \hat{u}(t_{n+1})) = \int_0^1 \frac{d_x F(t_{n+1/2}, u(t_{n+1/2}) - \tau U(t_{n+1}))}{(U(t_{n+1}))} d\tau, \]

where

\[ U(t_{n+1}) = u(t_{n+1/2}) - \hat{u}(t_{n+1}) + \sum_{i=1}^{j} c_{2i} k^{2i}(D_+ D_-)^i \hat{u}(t_{n+1}). \]  

(31)

From the identity (13) we have \( U(t_{n+1}) = O(k^{2j+2}) \) and it follows from the continuity of \( d_x F \) that there exists \( k_0 > 0 \) such that \( 0 \leq k \leq k_0 \) implies

\[ \|d_x F(t_{n+1/2}, u(t_{n+1/2}) - \tau U(t_{n+1}))\| \leq 1 + \|d_x F(t_{n+1/2}, u(t_{n+1/2}))\| \]
\[ \leq 1 + \max_{0 \leq t \leq T} \|d_x F(t, u(t))\|. \]

(32)

We then deduce that

\[ \|F(t_{n+1/2}, u(t_{n+1/2})) - F(t_{n+1/2}, \hat{u}(t_{n+1}) - \Gamma \hat{u}(t_{n+1}))\| \leq C k^{2j+2}, \]

where \( C \) is a constant depending only on \( k_0, j, T, F \) and the derivatives of \( u \) up to order \( 2j + 2 \). From the identity (13) we immediately have

\[ \|u'(t_{n+1/2}) - Du(t_{n+1/2}) + D\Lambda u(t_{n+1/2})\| \leq C k^{2j+2}, \]

where \( C \) is a constant depending only on \( j, T \) and the \((2j+3)\)-th derivatives of \( u \). The two last inequalities imply that

\[ \|\sigma^{2j+2,n+1/2}\| \leq C k^{2j+2}, \]  

(33)
where $C$ is a constant depending only on $k_0$, $j$, $T$ and the derivatives of $u$ up to the order $2j + 3$. Since the sequence $\{u^{2j,n}\}_n$ satisfies (DCC), we immediately have

$$
\left\| D(A - \Gamma)(u^{2j,n+1/2} - u(t_{n+1/2})) \right\| \leq Ck^{2j+2}.
$$

From the Lipschitz condition on $F$ we have

$$
\left\| F\left(t_{n+1/2}, \tilde{u}^{2j+2,n+1} - \Gamma \tilde{u}^{2j,n+1}\right) - F\left(t_{n+1/2}, \hat{u}(t_{n+1}) - \Gamma \hat{u}(t_{n+1})\right) \right\| 
\leq \mu \left\| \Theta^{2j+2,n+1} \right\|.
$$

Substituting the inequalities (33)-(35) into (29) we deduce that

$$
\left\| \Theta^{2j+2,n+1} \right\| \leq C\left( k^{2j+3} + \frac{2 + \mu k}{2 - \mu k} \left\| \Theta^{2j+2,n} \right\| \right).
$$

We then deduce by induction on $n$ that

$$
\left\| \Theta^{2j+2,n} \right\| \leq C\left( \frac{1}{2 - \mu k} \left( \frac{2 + \mu k}{2 - \mu k} \right)^{n-j} k^{2j+2} + \left( \frac{2 + \mu k}{2 - \mu k} \right)^{n-j} \left\| \Theta^{2j+2,n} \right\| \right).
$$

From the hypothesis (25) and (DCC) we have

$$
\left\| \Theta^{2j+2,j} \right\| \leq \left\| u^{2j+2,j} - u(t_j) \right\| + \left\| \sum_{i=0}^{j} c_i k^{2i} (D_i^+ D_i^-)^i (u^{2j,i} - u(t_j)) \right\| \leq Ck^{2j+2},
$$

where $C$ is a constant independent from $k$. Moreover, the sequence $\left\{ \left( \frac{2 + \mu k}{2 - \mu k} \right)^n \right\}_n$ is bounded above by $e^{\ln(\frac{1}{1-\varepsilon})(\mu T/2)}$, for $0 \leq \mu k \leq \varepsilon < 1$. Whence

$$
\left\| \Theta^{2j+2,n} \right\| \leq Ck^{2j+2}.
$$

Finally, by the triangle inequality, the identity (30) and (DCC) we get

$$
\left\| u^{2j+2,n} - u(t_n) \right\| \leq Ck^{2j+2} + \left\| \sum_{i=1}^{j} c_i k^{2i} (D_i^+ D_i^-)^i (u^{2j,i} - u(t_n)) \right\| \leq Ck^{2j+2},
$$

where $C$ is a constant depending only on $j$, $T$, the Lipschitz constant $\mu$ and the derivatives of $u$ up to order $2j + 3$. 

Here we consider the case where $X$ is a Hilbert space and $F$ satisfies the monotonicity condition (27). Then, taking the inner product of the identity (29) with $\hat{\Theta}^{2j+2,n+1}$ we deduce the estimate

$$
\left( D\Theta^{2j+2,n+1/2}, \hat{\Theta}^{2j+2,n+1} \right) \leq \left( \sigma^{2j+2,n+1/2}, \hat{\Theta}^{2j+2,n+1} \right) + \left( D(A - \Gamma)(u^{2j,n+1/2} - u(t_{n+1/2})), \hat{\Theta}^{2j+2,n+1} \right),
$$

since, according to (27), we have

$$(F(t_{n+1/2} - \Gamma \hat{u}^{2j,n+1}) - F(t_{n+1/2} - \Gamma \hat{u}(t_{n+1})), \hat{\Theta}^{2j+2,n+1}) \leq 0.$$ 

Inequalities (33)-(34) together with the Cauchy-Schwarz inequality yield

$$\| (\sigma^{2j+2,n+1/2}, \hat{\Theta}^{2j+2,n+1} ) \| \leq C k^{2j+2} \| \hat{\Theta}^{2j+2,n+1} \|$$

and

$$\| (D(A - \Gamma)(u^{2j,n+1/2} - u(t_{n+1/2})), \hat{\Theta}^{2j+2,n+1} ) \| \leq C k^{2j+2} \| \hat{\Theta}^{2j+2,n+1} \|,$$

where $C$ is a constant depending only on $j$, $T$, the function $F$ and the derivatives of $u$ up to order $2j + 3$. Substituting the last inequality into (37), we obtain

$$\left( D\Theta^{2j+2,n+1/2}, \hat{\Theta}^{2j+2,n+1} \right) \leq C k^{2j+2} \| \hat{\Theta}^{2j+2,n+1} \|$$

and deduce from the identity

$$\left( D\Theta^{2j+2,n+1/2}, \hat{\Theta}^{2j+2,n+1} \right) = \frac{1}{2k} \left( \| \theta^{2j+2,n+1} \|^2 - \| \theta^{2j+2,n} \|^2 \right)$$

and the inequality

$$\| \hat{\Theta}^{2j+2,n+1} \| \leq \frac{1}{2} \left( \| \theta^{2j+2,n+1} \| + \| \theta^{2j+2,n} \| \right)$$

that

$$\| \theta^{2j+2,n+1} \| - \| \theta^{2j+2,n} \| \leq C k^{2j+3}.$$ 

It follows by induction that

$$\| \theta^{2j+2,n} \| \leq C \left( \frac{n-j}{N} \right)^T k^{2j+2}$$

and we deduce from (36) that

$$\| \theta^{2j+2,n} \| \leq C \left( 1 + \frac{(n-j)T}{N} \right) k^{2j+2}.$$
Finally, we have
\[ \|u^{2j+2,n} - u(t_n)\| \leq C \left( 1 + \frac{(n-j)T}{N} \right) k^{2j+2}, \] (38)
where \( C \) is a constant depending only on \( j, T \) and the derivative of \( u \) up to order \( 2j+3 \).

(iii) The theorem in the case where \( X \) is finite dimensional and \( \{u^{2j+2,n}\}_n \) satisfies the hypothesis (28) can be deduced from the first case (i) with the Lipschitz constant
\[ \mu = \sup_{0 \leq t \leq T; \|x\| \leq M+R} \|dF_x(t, x)\|, \]
where
\[ R = \left( 1 + \sum_{i=1}^j 2^{2i} |c_{2i}| \right) \max_{0 \leq t \leq T} \|u(t)\|. \]

(iv) Now we consider the case where \( X \) is infinite dimensional and \( \{u^{2j+2,n}\}_n \) converges to the exact solution \( u \). We can write
\[ F(t_{n+1/2}, \tilde{u}^{2j+2,n+1} - \Gamma \tilde{u}(t_{n+1})) = \int_0^1 d_x F\left(t_{n+1/2}, \tilde{u}(t_{n+1}) - \Gamma \tilde{u}(t_{n+1}) + \tau \tilde{\Theta}^{2j+2,n+1}\right) ds. \]
Taking the (DDC) about \( \{u^{2j,n}\}_n \) into account, we have at least
\[ \|d_x F(t_{n+1/2}, \tilde{u}(t_{n+1}) - \Gamma \tilde{u}(t_{n+1}) + \tau \tilde{\Theta}^{2j+2,n+1})\| \leq 1 + \max_{0 \leq t \leq T} \|d_x F(t, u(t))\|, \]
where \( U \) is defined in (31). The theorem is then deduced from the case (i) choosing the Lipschitz constant \( \mu = 1 + \max_{0 \leq t \leq T} \|d_x F(t, u(t))\| \).
The theorem is proven.

**Remark 5** Theorem 3 shows that the correction may be applied from any other scheme satisfying (DCC).

**Remark 6** Theorem 3 under the assumption (ii) together with the estimate (38) shows an unconditional convergence of the DC schemes while we need \( k = O(\frac{1}{T}) \) for the convergence in the case of hypothesis (i), (iii) and (iv).
Remark 7 Under the assumptions (i)-(iv) from Theorem 3 the solution \(\{u^{2,n}\}_n\) of the scheme (22) approximate \(u\) with order 2 of accuracy, that is
\[
\|u(t_n) - u^{2,n}\| \leq Ck^2, \text{ for each } n = 0, 1, 2, \cdots N,
\]
where \(C\) is a constant depending only on \(T, F\) and the derivatives of \(u\) up to order 3.

Before giving conditions for a solution of the scheme (23)-(24) to satisfy (DCC) for the trapezoidal rule we give the following lemma.

Lemma 2 Under the hypothesis (i)-(iv) of Theorem 3, the solution \(\{u^{2,n}\}_n\) of the scheme (22) satisfies the inequality
\[
\|D(D + D_-)^m (u^{2,n+1/2} - u(t_{n+1/2}))\| + \|D(D + D_-)^m (u^{2,n+1} - u(t_{n+1}))\| \leq Ck^2,
\]
for \(m = 0, 1, \cdots, p\) and \(n = m, m+1, \cdots\), where \(C\) is a constant depending only on \(p, F\) and the derivatives of \(u\).

Proof Conditions (i)-(iv) of Theorem 3 are needed only to guarantee the estimate (39). To prove (40), we proceed by induction on the integer \(m = 0, 1, \cdots, p\). As in Theorem 3 we combine (20) and (22) and deduce the identity
\[
D\Theta^{2,n+1/2} = [F(t_{n+1/2}, \hat{u}^{2,n+1}) - F(t_{n+1/2}, \hat{u}(t_{n+1}))] + \sigma^{2,n+1/2},
\]
where
\[
\Theta^{2,n} = u^{2,n} - u(t_n)
\]
and
\[
\sigma^{2,n+1/2} = \left[u'(t_{n+1/2}) - Du(t_{n+1/2})\right] - \left[F(t_{n+1/2}, u(t_{n+1/2})) - F(t_{n+1/2}, \hat{u}(t_{n+1}))\right].
\]
From Taylor’s formula with integral remainder we can write
\[
\sigma^{2,n+1/2} = k^2 g(t_{n+1}),
\]
where \(g\), depending only on \(F\) and the first three derivative of \(u\), is \(C^{2p}([k/2, T], X)\). Proceeding as in Proposition 2 we can also write
\[
F(t_{n+1/2}, \hat{u}^{2,n+1}) - F(t_{n+1/2}, \hat{u}(t_{n+1})) = \int_0^1 dx F(t_{n+1/2}, K_1^{n+1}) (\hat{\Theta}^{2,n+1}) d\tau_1,
\]
where
\[
K_1^{n+1} = \hat{u}(t_{n+1}) + \tau_1 \hat{\Theta}^{2,n+1}.
\]
The last identities substituted into (41) yield
\[
D\Theta^{2,n+1/2} = \int_0^1 dx F(t_{n+1/2}, K_1^{n+1}) (\hat{\Theta}^{2,n+1}) d\tau_1 + k^2 g(t_{n+1}).
\]
(42)
Proceeding as in Theorem 3, we deduce from (39) that
\[
\left\| \int_0^1 dx \left( t_{n+1/2}, K_1^{n+1} \right) \left( \hat{\Theta}^{2,n+1} \right) \right\| \leq C \left\| \hat{\Theta}^{2,n+1} \right\|.
\]

The function \( g \) is also bounded independently from \( k \). Therefore, taking the norm on both side of (39), we deduce by the triangle inequality and (39) that
\[
\left\| D \hat{\Theta}^{2,n+1/2} \right\| \leq C \left\| \hat{\Theta}^{2,n+1} \right\| + k^2 \| g(t_{n+1}) \| \leq C k^2,
\]
where \( C \) is a constant depending only on \( u, F, \) and \( T \). The last inequality combined with (39) implies that (40) is true for \( m = 0 \). Assume that (40) is true for arbitrary integer \( m, \) \( 0 \leq m \leq p - 1 \). We are going to prove that this inequality remains true for \( m + 1 \). For this, we apply \((D_+D_-)^m D_+ \) to (42). This yields
\[
(D_+D_-)^{m+1} \hat{\Theta}^{2,n+1} = (D_+D_-)^m D_+ h(t_{n+1}) + k^2 (D_+D_-)^m D_+ g(t_{n+1}),
\]
where we set
\[
h(t_{n+1}) = \int_0^1 dx \left( t_{n+1/2}, K_1^{n+1} \right) \left( \hat{\Theta}^{2,n+1} \right) d\tau_1.
\]
Since \( g \) is \( C^{2p}([k/2,T],X) \), we deduce from Proposition 4 that
\[
\left\| (D_+D_-)^m D_+ g(t_{n+1}) \right\| \leq C, \quad \text{for } n = m - 1, m, ..., \quad (45)
\]
where \( C \) is a constant depending only on \( T, F \), and the derivatives of \( u \) up to order \( 2m + 4 \). To find a bound for \((D_+D_-)^m D_+ h(t_{n+1})\) we suppose that \( F(t,x) = F(x) \), that is
\[
h(t_{n+1}) = \int_0^1 dF(K_1^{n+1}) \left( \hat{\Theta}^{2,n+1} \right) d\tau_1.
\]
The general case can be deduced from an elementary (but tedious) calculations. From (7) we can successively write
\[
D_+ h(t_n) = \int_0^1 dF(K_1^n) \left( D_+ \hat{\Theta}^{2,n} \right) d\tau_1 + \int_0^1 \int_0^1 d^2 F \left( K_2^n \right) \left( D_+ K_1^n, \hat{\Theta}^{2,n+1} \right) d\tau_1 d\tau_2,
\]
\[
D_+^2 h(t_n) = \int_0^1 dF(K_1^n) \left( D_+^2 \hat{\Theta}^{2,n} \right) d\tau_1 + \int_0^1 \int_0^1 d^2 F \left( K_2^n \right) \left( D_+^2 K_1^n, \hat{\Theta}^{2,n+2} \right) d\tau_1 d\tau_2 + 2 \int_0^1 \int_0^1 d^2 F \left( K_2^n \right) \left( D_+ K_1^n, D_+ \hat{\Theta}^{2,n+1} \right) d\tau_1 d\tau_2 + \int_0^1 \int_0^1 \int_0^1 d^3 F \left( K_3^n \right) \left( D_+ K_1^n, D_+ K_1^{n+1}, \hat{\Theta}^{2,n+2} \right) d\tau_1 d\tau_2 d\tau_3,
\]
where
\[ K_{n+1}^q = K_n^q + \tau_{n+1}(K_{n+1}^{q+1} - K_n^q), \]
and deduce by induction on \( q = 1, 2, \ldots, 2p + 1 \) that
\[
D_k^q h(t_n) = \sum_{i=1}^{q+1} \sum_{|\alpha_i| = q+1} a_{\alpha_i} L_{i, \alpha_i}^n,
\]
where \( \alpha_i = (\alpha_i^1, \cdots, \alpha_i^{i-1}, \alpha_i^i) \in [1, q]^i \) and
\[
L_{i, \alpha_i}^n = \int_{[0,1]} d^i F(K_n^1) \left( D_{i}^{\alpha_i^1} K_{i-1}^{n+r_{i-1}(\alpha_i)}(1), \cdots, D_{i}^{\alpha_i^i} K_1^{n+r_i(\alpha_i)}, \theta^{2n+q-\alpha_i} \right) d\tau,
\]
where \( a_{\alpha_i} \) is a constant and \( r_{i-1}(\alpha_i) \) is a non-negative integer such that \( r_{i-1}(\alpha_i) + \alpha_i^{i-1} \leq q \). For each \( i = 1, 2, \cdots, 2p + 2 \), we can write \( K_i^{n+1} = u(t_{n+1/2}) + O(k) \) and deduce as in (32) that there exists \( k_3 > 0 \) such that \( k \leq k_3 \) implies
\[
\| d^{i} F(K_i^n) \| \leq C, \quad \text{for} \quad i = 1, 2, \ldots, 2p + 2, \quad (47)
\]
where \( C \) is a constant depending only on \( k_3, j, T, F \) and the derivative of \( u \). From the inductions hypothesis \( 10 \) and Proposition \( 1 \) we have
\[
\| D_k^r K_i^n \| \leq C, \quad \text{for} \quad 1 \leq r < i \leq 2m + 1, \quad 1 \leq n \leq N - i \quad (48)
\]
and
\[
\| D_k^r \theta^{2n} \| \leq Ck^2, \quad \text{for} \quad 1 \leq r \leq 2m + 1, \quad 1 \leq n \leq N - r \quad (49)
\]
where \( C \) is a constant depending only on \( p, T, F \) and the derivatives of \( u \). Each \( L^n_{i, \alpha_i} \) being multilinear continuous, we deduce from \( 10 \) \( - \) \( 10 \) that
\[
\| L^n_{i, \alpha_i} \| \leq Ck^2, \quad i = 1, 2, \cdots, 2m + 1
\]
and then, according to Remark \( 4 \) and the triangle inequality, \( 4 \) with \( q = 2m + 1 \) yields
\[
\| (D_+ D_-)^m D_+ h(t_{n+1}) \| = \| D_+^{2m+1} h(t_{n+1-m}) \| \leq Ck^2
\]
where \( C \) is a constant depending only on \( p, T, F \) and the derivatives of \( u \). Passing to the norm into the identity \( 4 \), we deduce from \( 4 \) \( - \) \( 4 \) \( 4 \) \ and the last inequality that
\[
\| (D_+ D_-)^{m+1} \theta^{2n+1} \| \leq Ck^2. \quad (50)
\]
Otherwise, applying \( D_- \) to \( 4 \), the same reasoning taking the induction hypothesis and the inequality \( 4 \) \( - \) \( 4 \) \ into account yields
\[
| D_- (D_+ D_-)^{m+1} \theta^{2n+1} | \leq Ck^2, \quad (51)
\]
where \( C \) is a constant depending only on \( p, T, F \) and the derivatives of \( u \). Inequalities (50) and (51) imply that the induction hypothesis is still true for \( m + 1 \). Finally, we deduce by induction that (40) is true for each integer \( m = 0, 1, \ldots, p \).

We have the following theorem

**Theorem 4** Each solution \( \{u^{2j,n}\}_n \), \( j = 1, 2, \ldots, p \), from (22) and (23)-(27) satisfies (DCC) for the trapezoidal rule. Furthermore, we have the estimate

\[
\|D(D_n)_m(u^{2j,n+1/2} - u(t_{n+1/2}))\| + \|((D_n)_m)^n(u^{2j,n+1} - u(t_{n+1}))\| \leq Ck^2
\]

for \( m = 0, 1, \ldots, p \) and \( n = m + j - 1, \ldots, m + j + \ldots, \) where \( C \) is a constant depending only on \( p, T, F \) and the derivatives of \( u \).

**Proof** We proceed by induction on \( j = 1, 2, \ldots, p \). The case \( j = 1 \) result from Remark 7 and Lemma 2. Suppose that \( \{u^{2j,n}\}_n \) satisfies (DCC) up to an arbitrary order \( j \leq p - 1 \). Let us prove that the theorem is still true for \( j + 1 \). Since \( \{u^{2j,n}\}_n \) satisfies (DCC), from Theorem 3 \( \{u^{2j+2,n}\}_n \) approximate the exact solution \( u \) with order \( 2j + 2 \) of accuracy. Therefore, it is enough to establish (52) for \( j + 1 \). We can rewrite the identity (29) as follows

\[
D\theta^{2j+2,n+\frac{1}{2}} = \tilde{H}(t_{n+1}) + \sigma^{2j+2,n+\frac{1}{2}} + D(A - \Gamma)(u^{2j,n+\frac{1}{2}} - u(t_{n+\frac{1}{2}})),
\]

with

\[
H(t_{n+1}) = \int_0^1 dF \left( t_{n+1/2}, \tilde{u}(t_{n+1}) - \Gamma \tilde{u}(t_{n+1}) + \tau \tilde{\theta}^{2j+2,n+1} \right) d\tau,
\]

where \( \theta^{2j+2,n} \) and \( \sigma^{2j+2,n+1/2} \) are as in Theorem 3. Proceeding as in Lemma 2 and using Theorem 1, we can write

\[
\sigma^{2j+2,n+1/2} = k^{2j+2} \varepsilon_1(t_{n+1}),
\]

where \( \varepsilon_1 \) is \( C^{2p-2j+2} \), depending only on \( u \) and \( F \). From Proposition 3 we have

\[
\|D^m \varepsilon_1(t_{n+1})\| \leq \hat{C} \max_{0 \leq t \leq T} \left\| \frac{d^{m-1} \varepsilon_1(t)}{dt^{m-1}} \right\| \leq C, \text{ for } |m| \leq 2p - 2j + 2,
\]

where \( C \) is a constant depending only on \( p, T, u \) and \( F \). According to the inequality (52) from the induction hypothesis we may write

\[
D(A - \Gamma)(u^{2j,n+1/2} - u(t_{n+1/2})) = k^{2j+2} \varepsilon_2(t_{n+1}),
\]

where

\[
\|D^m \varepsilon_2(t_{n+1})\| \leq C, \text{ for } |m| \leq 2p - 2j + 2.
\]

Therefore, writing (53) as follows
\[ D_+ \Theta^{2j, n+1} = H(t_{n+1}) + k^{2j, n+1} G(t_{n+1}), \]

with
\[ G(t_{n+1}) = \varepsilon_1(t_{n+1}) + \varepsilon_2(t_{n+1}), \]

the induction hypothesis and the reasoning from Lemma 2, substituting the functions \( h \) and \( g \) respectively by \( H \) and \( G \), \( \hat{\Theta}^{2j, n+1} \) by \( \hat{\Theta}^{2j, n+1} \) and \( k^2 \) by \( k^{2j, n+1} \), yields
\[ \| D(D_+ D_-)^m \hat{\Theta}^{2j, n+1} \| + \| (D_+ D_-)^m \hat{\Theta}^{2j, n+1} \| \leq C k^{2j, n+1} \]

for \( m = 0, 1, \ldots, p \) and \( n = m + j - 1, m + j, \ldots \), where \( C \) is a constant depending only on \( p \), \( T \), \( F \) and the derivatives of \( u \). Inequality (52) holds for \( \{ u^{2j, n+1} \} \) by the triangle inequality from the last inequality.

5 Deferred correction for Euler rule

Again we consider the Cauchy problem (20) and, owing to the approximate (19), we construct by induction on \( j = 1, \ldots, m \), the sequence \( \{ u_{j,n} \} \) of approximate solution of (20) as follows:

\[ \frac{u_{1,n+1} - u_{1,n}}{k} = F(t_n, u_{1,n}), \quad (54) \]

\[ \frac{u_{2,n+1} - u_{2,n}}{k} - \frac{k}{2} D_+ D_- u_{1,n} = F(t_n, u_{2,n}), \quad (55) \]

and for \( j = 2, 3, \ldots, \)

\[ \frac{u_{j+1,n+1} - u_{j+1,n}}{k} - \sum_{i=1}^{j} a_{i+1} k^i D_+ D_-^{i-1} (D_+ D_-)^{i-1} u^{j,n} = F(t_n, u^{j+1,n}), \quad (56) \]

with
\[ u^{j,0} = u_0, \text{ for } j = 1, 2, \ldots. \quad (57) \]

Here \([x]\) is the integer part of \( x \in \mathbb{R} \). This scheme can also be written for the backward Euler method. The index \( j + 1 \) indicates that \( u^{j+1,n} \) approximates \( u(t_n) \) with order \( j + 1 \) of accuracy. We call the schemes (54)-(56) deferred correction schemes for Euler rule. Remark 3 applies to deferred correction schemes for Euler rule.

As in the previous section we give the following definition

**Definition 2** Let \( u \) be the exact solution of (20). For a positive integer \( j = 1, 2, \ldots \), a sequence \( \{ u^{j,n} \} \) is said to satisfy the Deferred Correction Condition (DCC) for Euler rule if \( \{ u^{j,n} \} \) approximates \( u(t_n) \) with order \( j \) of accuracy and

\[ |D_+ D_- (u^{j,n+1} - u(t_{n+1}))| \leq C k^j, \quad (58) \]

where \( C \) is a constant depending only on \( j \), \( u \) and \( T \).
Theorem 5 of section 4 becomes

**Theorem 5** Let \( j \) be a positive integer, \( 1 \leq j \leq 2p \), and \( \{w^{j,n}\}_n \) a sequence satisfying (DCC) for the Euler rule. Let \( \{w^{j+1,n}\}_n \) be the solution of (57) and (59)-(57), built from \( \{w^{j,n}\}_n \). Suppose that one of the following four conditions holds:

(i) \( F \) is \( \mu \)-Lipschitz with respect to the second variable \( x \),

(ii) \( X \) is a Hilbert space with inner product \((.,.)\) and \( F \) satisfies the monotonicity condition

\[
(F(t,x) - F(t,y), x - y) \leq 0, \quad \forall (t,x,y) \in [0,T] \times X \times X.
\]  

(iii) \( X \) is finite dimensional and \( \{w^{2j+2,n}\}_n \) remains close to the exact solution \( u \) of the problem (20) in the sense that

\[
\|w^{2j+2,n} - u(t_n)\| \leq M, \quad \text{for} \quad n = 0, 1, ..., N,
\]  

where \( M \) is a constant independent from \( n \) and \( k \).

(iv) \( \{w^{j+1,n}\}_n \) converges to the exact solution \( u \) of the problem (20).

Then \( \{w^{j+1,n}\}_n \) approximates \( u \) with order \( j + 1 \) of accuracy.

**Proof** Same proof as in Theorem 3.

As in Theorem 4, one can show that the (DCC) for Euler rule is satisfied for each sequence for Euler rule.

### 6 Absolute stability

In this section we propose to prove absolute stability result for the DCC schemes (22) and (23)-(24).

A numerical method is said to be \( A \)-stable if it has no stability restrictions with the test problem

\[
\begin{cases}
  u' = \lambda u \\
  u(0) = 1,
\end{cases}
\]  

for \( \Re(\lambda) < 0 \), see [20, p.40]. The exact solution of (44) is \( u(t) = e^{\lambda t} \) and satisfies \( \lim_{t \to +\infty} |u(t)| = 0 \), if \( \Re(\lambda) < 0 \). More generally, we have the following definition [16].

**Definition 3** A numerical method is said to be absolutely stable if the corresponding solution for the problem (67) for fixed \( k > 0 \) and some \( \Re(\lambda) < 0 \) is such that

\[
\lim_{n \to +\infty} |u^n| = 0.
\]  

The region of absolute stability of a numerical method is defined as the subset of the complex plane

\[
\mathcal{A} = \{ z = h\lambda \in \mathbb{C} : \text{\text{(62) is satisfied}} \}.
\]  

If \( \mathcal{A} \cap \mathbb{C}_- = \mathbb{C}_- \), \( \mathbb{C}_- = \{ \lambda \in \mathbb{C} : \Re(\lambda) < 0 \} \), the numerical method is said to be \( A \)-stable.
Before establishing absolute stability result for the deferred correction methods (22) and (23)-(24), we recall the following result.

**Lemma 3** Let $P_m$ be a polynomial of degree $m$ in one variable. Then the sum $\sum_{i=0}^n P_m(i)$ is a polynomial of degree $(m+1)$ with respect to $n$.

**Proof** Without loss of generality we assume that $P_m(x) = x^m$ and set $F_m(n) = \sum_{i=1}^n P_m^i$. It is then enough to prove that $F_m(n)$ is a polynomial of degree $(m+1)$ with respect to $n$, for each non-negative integer $m$. We proceed by induction on $m$. The cases $m = 0, 1$ are trivial. Assume that $F_m(n)$ is a polynomial of degree $(m+1)$, for arbitrary positive integer $m$. We have the identities

$$(n+1)^{m+2} = \sum_{p=1}^n [(p+1)^{m+2} - p^{m+2}] = \sum_{p=1}^n \sum_{i=0}^{m+1} \binom{m+2}{i} p^i = \sum_{i=0}^{m+1} \binom{m+2}{i} F_i(n),$$

which implies that

$$F_{m+1}(n) = \frac{1}{m+2} (n+1)^{m+2} - \frac{1}{m+2} \sum_{i=0}^{m} \binom{m+2}{i} F_i(n) - \frac{1}{m+2}.$$

According to the induction hypothesis, $\sum_{i=0}^{m} \binom{m+2}{i} F_i(n)$ is a polynomial of degree $(m+2)$ with respect to $n$ and we can then deduce by induction that each $F_m(n)$ is a polynomial of degree $(m+1)$ with respect to $n$.

**Lemma 4** Suppose that $F(t, u) = \lambda u$ and $u_0 = 1$ in the initial value problem (11), where $\lambda$ is a complex number with negative real part ($\lambda \in \mathbb{C}_-$. Then the corresponding approximate solutions from the schemes (22) and (23)-(24) can be written as follows

$$u^{2j+2,n} = \left(\frac{2 + \lambda k}{2 - \lambda k}\right)^{n-j} P_j(n), \text{ for } j = 0, 1, 2, ..., \quad (64)$$

where $P_j(n)$ is a polynomial of degree $j$ with respect to $n$ ($n \geq 2j$).

**Proof** We suppose that $\lambda k \neq -2$, otherwise we trivially have $u^{2j,n+1} = 0$, for $n = j, j+1, ...$. Since $F(t, u) = \lambda u$, we can rewrite (23) as follows

$$u^{2j+2,n+1} = \frac{2 + \lambda k}{2 - \lambda k} u^{2j+2,n} + \frac{2}{2 - \lambda k} \left(kD_- A u^{2j,n+1} - \lambda k \Gamma u^{2j,n+1}\right)$$

where, according to the formula (2) we have

$$kD_- A u^{2j,n} = \sum_{i=1}^j c_{2i+1} k^{2i+1} D_- (D_i D_-)^i u^{2j,n}$$

$$= \sum_{i=1}^j \sum_{m=0}^{2i+1} c_{2i+1} (-1)^m \binom{2i+1}{m} u^{2j,n+i-m},$$

and

$$\sum_{i=1}^j \sum_{m=0}^{2i+1} c_{2i+1} (-1)^m \binom{2i+1}{m} = 0.$$
and
\[ I^2 \hat{u}^{2j,n} = \sum_{i=1}^{j} c_{2i} k^{2i} (D_+ D_-)^i \hat{u}^{2j,n} = \sum_{i=1}^{j} \sum_{m=0}^{2i} (-1)^m c_{2i} \left( \begin{array}{c} 2i \\ m \end{array} \right) \hat{u}^{2j,n+i-m}. \]

Combining the three last identities we derive the identity
\[ u^{2j+2,n+1} = \frac{2 + \lambda k}{2 - \lambda k} u^{2j+2,n} + \frac{2}{2 - \lambda k} \sum_{i=0}^{2j+1} \alpha_{j,i}(\lambda k) u^{2j,n+1+j-i}, \quad (65) \]
for \( j = 1, 2, ..., \) where \( \alpha_{j,i} \) is affine in \( \lambda k \). Under the condition of the lemma, \( (22) \) matches Crank-Nicolson scheme and we have
\[ u^{2,n} = \left( \frac{2 + \lambda k}{2 - \lambda k} \right)^{n}, \]
that is \( (64) \) is true for \( j = 0 \). Suppose that \( (64) \) holds for arbitrary integer \( j \geq 0 \). From \( (65) \) we have
\[ u^{2j+4,n} = \frac{2 + \lambda k}{2 - \lambda k} u^{2j+4,n-1} + \frac{2}{2 - \lambda k} \sum_{i=0}^{2j+3} \alpha_{j+1,i}(\lambda k) u^{2j+2,n+1+j-i}, \]
with \( n \geq j + 1 \). For \( n + 1 + j - i \geq j \), that is \( n \geq 2j + 2 \), we can substitute each \( u^{2j+2,n+1+j-i} \) by the formula given by the induction hypothesis \( (64) \) and deduce that
\[ u^{2j+4,n} = \frac{2 + \lambda k}{2 - \lambda k} u^{2j+4,n-1} + \left( \frac{2 + \lambda k}{2 - \lambda k} \right)^{n-j-1} Q_j(n), \]
where
\[ Q_j(n) = \frac{2}{2 - \lambda k} \sum_{i=0}^{2j+3} \alpha_{j+1,i}(\lambda k) \left( \frac{2 + \lambda k}{2 - \lambda k} \right)^{j+2-i} P_j(n+1+j-i). \]
It follows that
\[ u^{2j+4,n} = \left( \frac{2 + \lambda k}{2 - \lambda k} \right)^{n-j-1} \left( u^{2j+4,n-1} + \sum_{i=j+2}^{n} Q_j(i) \right). \]

It is clear that \( Q_j(n) \) is a polynomial of degree \( j \) with respect to \( n \) as \( P_j(n) \). Therefore, according to the Lemma \( 3 \), \( \sum_{i=j+2}^{n} Q_j(i) \) is a polynomial of degree \( (j+1) \) with respect to \( n \). Whence,
\[ u^{2j+4,n} = \left( \frac{2 + \lambda k}{2 - \lambda k} \right)^{n-j-1} P_{j+1}(n), \]
for \( n \geq 2j + 2 \),
with
\[ P_{j+1}(n) = u^{2j+4,n+1} + \sum_{i=j+2}^{n} Q_j(i), \]
a polynomial of degree \( (j+1) \) with respect to \( n \). Therefore, we can deduce by induction on \( j \) that the lemma is true for arbitrary non-negative integer \( j \).
Theorem 6 Each deferred correction schemes (22) and (23)-(24) is $A$-stable.

Proof From Lemma 4 we have, for $\text{Re}(\lambda_k) < 0$,

$$\lim_{n \to +\infty} |u^{2j+2,n}| = \lim_{n \to +\infty} \left| \left( \frac{2 + \lambda_k}{2 - \lambda_k} \right)^{n-j} P_j(n) \right| = \lim_{n \to +\infty} \left| P_j(n) |e^{(n-j)\ln|2+\lambda_k|}| \right| = 0,$$

since under the condition $\text{Re}(\lambda_k) < 0$ we have $\left| \frac{2 + \lambda_k}{2 - \lambda_k} \right| < 1$.

Theorem 7 DC schemes for the Forward Euler rule are not $A$-stable, but the Backward Euler rule is.

7 Numerical experiments

For the numerical experiments we choose six classical problems. The first problem is linear and non stiff, but the five others are among the stiffest in [8, 21]. We evaluate the accuracy of $DC_2$, ..., $DC_{10}$, implemented using the Scilab programming language. As stated in Remark 3 we make the codes self-starting by computing some approximate solutions corresponding to $t < t_0 = 0$. For each problem we evaluate the global error (absolute or relative) and the order of accuracy of the five methods using a reference solution computed with $DC_{10}$, for the last five test problems. We use the functions `stiff` (implementing BDF) and `rkf` (implicit Runge-Kutta 4-5) of the solver `ode` from Scilab for a comparison with the DC methods. For each stiff problem we provide the initial step size $k_0$ as prescribed in [8]: «$k_0$ is used to ensure that all interesting initial transients are followed». We recall that if the ODE is of the form $y' = f(y)$, we take $k_0 = 1/|\lambda_{\text{max}}|$ where $\lambda_{\text{max}}$ is the eigenvalue of largest magnitude of the Jacobian matrix ($\partial f/\partial y$) along the solution curve.

For solutions $u = (u_1, \cdot \cdot \cdot, u_d) : [0, T] \to \mathbb{R}^d$, $1 \leq d \leq 4$, with large magnitude we calculate the absolute error of the approximate solutions $\{u_{2p,n}^{i}\}_{0 \leq n \leq N}$, $1 \leq p \leq 5$, with the norm

$$||u_{i}^{2p} - u_{i}|| = \max_{0 \leq n \leq N} |u_{i}^{2p,n} - u_{i}(t_n)|,$$

while for solutions with small magnitude we calculate the relative error

$$||u_{i}^{2p} - u_{i}|| = \max_{1 \leq n \leq N} \left| \frac{u_{i}^{2p,n} - u_{i}(t_n)}{u_{i}(t_n)} \right|,$$

For very large $N$ we extract solutions at $4 \times 10^6$ discrete times evenly spread over the interval $[0, T]$. 
7.1 Modified oscillatory initial value problem

\[ u' = 2u \cos(t), \quad u(0) = 1, \quad T = 10^6. \]  

The exact solution is \( u(t) = e^{2 \sin(t)}. \) Table 3 gives the absolute error and the order of accuracy computed using \( k = 1/8 \) and \( k = 1/16. \) The solvers \( \text{rkf} \) and \( \text{stiff} \) are used with tolerances \( \text{atol} = \text{rtol} = 10^{-10}. \)

\[ \text{Table 3: Absolute error for the modified oscillator} \]

| \( k \) | DC2  | DC4  | DC6  | DC8  | DC10 | \( \text{rkf} \) | \( \text{stiff} \) |
|--------|------|------|------|------|------|----------------|----------------|
| 0.25   | 0.25 | 1.86e-3 | 3.27e-4 | 1.35e-4 | 3.11e-5 | 9.0793       | 3.05e-2        |
| 0.125  | 6.16e-2 | 1.18e-4 | 4.61e-6 | 6.18e-7 | 3.61e-8 | 8.32e-2       | 3.05e-2        |
| 0.0625 | 1.53e-2 | 7.44e-6 | 7.03e-8 | 2.50e-9 | 3.75e-11 | 4.71e-3       | 3.05e-2        |
| Order  | 2.01 | 3.99 | 6.03 | 7.94 | 9.90 | 4.14 | -              |

7.2 Krogh [8][22]

\[ y' = -By + U^T(z_1^2/2 - z_2^2/2, z_1z_2, z_1^2, z_2^2)^t, \quad y(0) = (0, -2, -1, -1)^t, \quad T = 1000, \]  

where \( z = Uy \) and

\[
U = \frac{1}{2} \begin{bmatrix}
-1 & 1 & 1 & 1 \\
1 & -1 & 1 & 1 \\
1 & 1 & -1 & 1 \\
1 & 1 & 1 & -1
\end{bmatrix}, \quad B = U \begin{bmatrix}
-10 & -10 & 0 & 0 \\
10 & -10 & 0 & 0 \\
0 & 0 & 1000 & 0 \\
0 & 0 & 0 & 0.0001
\end{bmatrix} U. 
\]

For this problem \( k_0 = 10^{-3} \) [8]. A reference solution is computed with DC10 with the time step \( k = 2 \times 10^{-6}. \) The solver \( \text{rkf} \) and \( \text{stiff} \) are used with \( \text{atol} = \text{rtol} = 10^{-10}. \) Table 4 gives the absolute errors (these are uniform in the four components of the approximate solutions) and the order of accuracy. We use the values in bold to calculate the order of convergence taking into account the fact that the error stagnates near \( 6 \times 10^{-9}. \) The error at the final time \( T = 1000 \) for each component of the approximate solution is equal to \( 1.6 \times 10^{-9} \) for any scheme DC2,...,DC10, when \( k \leq 1/1000. \)

7.3 Roberson (1966) [8][17][20]

\[
\begin{align*}
y_1' &= -0.04y_1 + 10^4 y_2 y_3 \\
y_2' &= 0.04y_1 - 10^4 y_2 y_3 - 3.10^7 y_2^2 \\
y_3' &= 3.10^7 y_2^2 \\
y(0) &= (1, 0, 0)^t, \quad T = 10^4.
\end{align*}
\]
Table 4 Absolute error for Krogh

| k      | DC2   | DC4   | DC6   | DC8   | DC10  | rkf  | stiff |
|--------|-------|-------|-------|-------|-------|------|-------|
| 1.00e-3| 0.017 | 7.49e-3 | 1.05e-3 | 8.84e-4 | 4.64e-4 | 5.54e-5 | 7.66e-8 |
| 4.00e-4| 1.82e-3 | 8.15e-5 | 2.86e-6 | 1.06e-7 | 6.09e-9 | 1.72e-6 | 7.65e-8 |
| 2.50e-4| 9.66e-5 | 1.90e-5 | 3.01e-7 | 6.09e-9 | 6.09e-9 | 1.98e-7 | 7.66e-8 |
| 1.388e-4| 2.88e-6 | 1.67e-6 | 7.82e-9 | 6.09e-9 | 6.08e-9 | 9.39e-9 | 7.66e-8 |
| 3.125e-5| 1.49e-5 | 6.02e-9 | 6.02e-9 | 6.02e-9 | 6.02e-9 | 6.02e-9 | 7.66e-8 |

Order 2.03 4.05 6.21 8.57 12.26 4.39 –

This is one of the three problems considered as stiffest in [20]. The authors in [20] suggest varying the final time $T$ up to $10^{11}$. We are limited to $T = 10^4$ to guarantee a good reference solution. The Jacobian ($\partial f/\partial y$) along the solution curve in $[0, 10^4]$ computed shows that $k_0$ decreases when the final time $T$ increases. This means that one would need smaller step size to compute solution when $T$ is larger as suggested. The reference solution is computed with DC10 with time step $k = 1.67 \times 10^{-5}$. We computed $k_0 \simeq 1.12 \times 10^{-4}$, from the reference solution. The maximal magnitude of the approximate solution is about 1 in the first and third component and about $5.78 \times 10^{-5}$ in the second component. Table 5 gives the relative errors and the order of accuracy. DC10 converges with order 11.36 in the second and third components for $k = 6.25 \times 10^{-4}$ to $k = 3.12 \times 10^{-4}$. The solver stiff and rkf are used with atol = $10^{-16}$ and rtol = $10^{-12}$. The solver rkf turns out to be inefficient for $k = 0.01$ and $k = 1.56 \times 10^{-4}$, consequently we drop computing its corresponding errors.

Table 5 Relative error for the Roberson problem

| k     | DC2   | DC4   | DC6   | DC8   | DC10  | rkf  | stiff |
|-------|-------|-------|-------|-------|-------|------|-------|
| 0.01  | 5.77e-6 | 3.29e-7 | 1.48e-7 | 1.23e-7 | 1.56e-7 | –    | 8.41e-11 |
| 6.25e-4| 4.77e-3 | 6.63e-4 | 2.47e-5 | 1.36e-5 | 8.21e-5 | 5.4073 | 2.90e-11 |
| 3.12e-4| 7.26e-4 | 3.76e-4 | 1.40e-5 | 7.74e-5 | 4.66e-5 | 0.3394 | 5.64e-11 |
| 1.56e-4| 3.47e-11 | 2.84e-11 | 2.76e-11 | 2.73e-11 | 2.76e-11 | 9.68e-13 | 7.93e-11 |
| Order 1.97 4.05 7.35 8.14 9.18 – –
### 7.4 Klopfenstein (1970) [8] Problem D6

\[
\begin{align*}
y_1' & = -y_1 + 10^8 y_3(1 - y_1) \\
y_2' & = -10y_2 + 3 \times 10^7 y_3(1 - y_2) \\
y_3' & = -y_1' - y_2' \\
y(0) & = (1, 0, 0)^T, T = 1.
\end{align*}
\] (69)

The comments from Shampine [17] indicate that this problem is not in its original version, but we prefer this version for a good reference solution since the original one may have solution with magnitude about \(10^{-16}\). For this problem \(k_0 = 3.3 \times 10^{-8}\) [8]. A reference solution is computed with DC10 with time step \(k = 10^{-9}\). The solution for this problem has small magnitude. For example, the first two components have a maximal magnitude of order 1, but the third component has a magnitude of about \(10^{-8}\). Table 6 gives the relative errors and the orders of accuracy. Very high order convergence is observed for \(k = 5 \times 10^{-8}\) to \(k = 2.5 \times 10^{-5}\) for DC4, ..., DC10. We use the solver stiff with tolerances \(atol = 10^{-18}\) and \(rtol = 10^{-13}\) while \(rkf\) is run with \(atol = rtol = 10^{-13}\).

| \(k\)  | DC2   | DC4   | DC6   | DC8   | DC10 | \(rkf\) | stiff |
|-------|-------|-------|-------|-------|------|--------|-------|
| 0.01  | 2.81e-5 | 3.32e-8 | 1.43e-8 | 1.43e-8 | 1.43e-8 | -      | 4.5e-13 |
| 2.5e-8| 1.10e-13 | 1.10e-13 | 1.10e-6 | 1.10e-13 | 1.10e-13 | 1.21e-13 | 5.4e-13 |
| 1.25e-8| 2.79e-13 | 2.79e-13 | 2.79e-13 | 2.82e-13 | 2.83e-13 | -      | 5.3e-13 |

Order 1.88883 3.49358 4.80098 8.48155 9.77595 –

### 7.5 Oregonator [20]

\[
\begin{align*}
y_1' & = 77.27(y_2 + y_1(1 - 8.375 \times 10^{-6} y_1 - y_2)) \\
y_2' & = \frac{1}{77.27}(y_3 - (1 + y_1)y_2) \\
y_3' & = 0.161(y_1 - y_3) \\
y(0) & = (1, 2, 3)^T, T = 360.
\end{align*}
\] (70)
This is one of the three stiffest problems in [20]. The reference solution is computed with DC10 with time step $k = 3.6 \times 10^{-7}$. We compute $k_0 \simeq 7.33 \times 10^{-6}$, the eigenvalue of largest magnitude of the Jacobian is achieved for $y'(37)$. The solution for this problem has large magnitude in the three components. The magnitude of the solution varies in $[1, 117845.8]$, $[0.003, 1768.7]$ and $[1.005, 31263.85]$, respectively, for the first, second and third component. Table 7 gives the absolute errors and orders of accuracy. We use the solvers stiff and rkf with $atol = 10^{-12}$ and $rtol = 10^{-12}$.

Table 7 Absolute error for the Oregonator

| $k$   | DC2   | DC4   | DC6   | DC8   | DC10  | rkf   | stiff |
|-------|-------|-------|-------|-------|-------|-------|-------|
| 3.6e-3| 2255.06 | 69.2541 | 2.00839 | 0.35983 | 0.22683 | 1.0669 | 3.53e-3 |
|       | 0.42362 | 0.01432 | 6.25e-4 | 6.59e-5 | 4.24e-5 | 1.76e-4 | 6.02e-7 |
|       | 18.4070 | 0.52556 | 1.00e-2 | 1.73e-4 | 7.84e-5 | 1.89e-3 | 2.88e-5 |
| 1.80e-3| 563.229 | 4.21482 | 3.75e-2 | 2.42e-3 | 5.20e-4 | 8.85e-2 | 1.72e-3 |
|       | 0.10578 | 8.69e-4 | 1.06e-5 | 4.75e-7 | 1.05e-7 | 1.81e-5 | 2.90e-7 |
|       | 4.59668 | 3.18e-2 | 1.51e-4 | 8.81e-8 | 3.20e-7 | 3.89e-5 | 1.41e-5 |
| 1.28e-3| 287.302 | 1.09068 | 3.03e-3 | 1.82e-4 | 2.14e-5 | 1.09e-2 | 5.63e-3 |
|       | 5.39e-2 | 2.08e-4 | 6.18e-7 | 3.45e-8 | 3.27e-9 | 2.16e-6 | 9.74e-7 |
|       | 2.34470 | 8.26e-3 | 1.99e-5 | 6.26e-8 | 5.83e-8 | 4.98e-6 | 4.60e-5 |
| Order | 2.0006 | 4.0175 | 7.4767 | 7.6943 | 9.4849 | 6.22 | - |
| 3.00e-5| 1.56e-2 | 7.82e-5 | 7.83e-5 | 7.83e-5 | 6.20e-5 | 5.21e-4 | 5.52e-3 |
|       | 2.93e-5 | 1.26e-8 | 1.26e-8 | 1.26e-8 | 1.20e-8 | 8.76e-8 | 9.54e-7 |
|       | 1.27e-3 | 6.40e-7 | 6.41e-7 | 6.41e-7 | 6.15e-7 | 3.78e-7 | 4.51e-5 |

7.6 van der Pol oscillator [8][17][20]

\[
\begin{align*}
    y_1' &= y_2 \\
    y_2' &= \mu(1 - y_1^2)y_2 - y_1 \\
    y_1(0) &= 2, \quad y_2(0) = 0, \quad T = 3000, \quad \mu = 1000.
\end{align*}
\]

(71)

This problem was initially proposed for $T = 1$ and $\mu = 5$ in [8]. The actual version results from a suggestion by Shampine [17]. The authors in [20] has a rescaled form of the van der Pol’s equation which is considered as one of their stiffest problem investigated. The reference solution is computed with DC10 with time step $k = 2.50 \times 10^{-6}$. We compute $k_0 = 3.33 \times 10^{-6}$. The magnitude of the solution varies in $[-2, 2.000073]$ and $[-1323.04, 1231.35]$, respectively, for the first and second components. Table 8 gives the absolute errors and orders of accuracy. Since the errors for DC2 and DC4 did not reach the region of asymptotic convergence for the time steps attempted, we drop computing
their order of accuracy. For $atol = rtol = 10^{-10}$ and the step size $k = 5 \times 10^{-5}$, the absolute errors computed with \texttt{rKF} are $8.78 \times 10^{-3}$ and $8.92$, respectively, for the first and second components of the solution while the absolute errors computed with the \texttt{stiff} solver are $6.89 \times 10^{-2}$ and $60.91$. When we force $atol = 10^{-14}$ and $rtol = 10^{-24}$, the \texttt{stiff} solver gives an absolute error of $7.44 \times 10^{-7}$ and $8.17 \times 10^{-4}$, respectively, for the first and second components.

### Table 8 Absolute error for the van der Pol’s equation

| $k$     | DC2        | DC4        | DC6        | DC8        | DC10       |
|---------|------------|------------|------------|------------|------------|
| $5.00e-5$ | 3.016      | 3.007      | 2.999      | 1.561      | $1.79e-2$  |
|         | 1330.989   | 1333.20    | 1322.93    | 1293.96    | $19.077$   |
| $2.50e-5$ | 2.9982     | 2.9856     | 1.8354     | $7.08e-3$  | $6.14e-4$  |
|         | 1329.52    | 1322.69    | 1084.20    | $7.3741$   | $0.6205$   |
| $1.5e-5$  | 2.976      | 2.944      | 0.108      | $1.15e-4$  | $2.24e-4$  |
|         | 1333.26    | 1330.34    | 113.71     | $0.1075$   | $0.24838$  |
| $7.50e-6$ | 2.870      | 2.694      | $1.60e-3$  | 1.95e-6    | 1.98e-5    |
|         | 1327.40    | 1286.55    | $1.63$     | 1.88e-3    | 2.00e-2    |
| $3.75e-6$ | 2.177      | 2.138      | $2.37e-5$  | 7.74e-7    | 7.39e-7    |
|         | 1251.54    | 1163.63    | $2.72e-2$  | 8.31e-4    | 8.22e-4    |
| $2.50e-6$ | 1.3742     | 0.9420     | 1.47e-6    | 1.44e-7    | 0          |
|         | 1147.32    | 689.075    | 2.54e-3    | 1.60e-4    | 0          |
| Order    | –          | –          | 6.07       | 8.07       | 4.9        |

### 8 General observation

The six problems chosen for these numerical experiments are representative of many tests that we ran to assess the efficiency of the DC methods presented in this paper. The numerical experiments show the strong stability of the DC methods and their quick convergence on stiff and non-stiff problems even for step sizes that are not necessarily small. The expected order of accuracy of each DC scheme is achieved even for the Van der Pol system, which result from a second order equation (not necessarily stiff) transformed to a very stiff first order system of ODEs. Even if the best precision of DC4 on each of the problems shown is not investigated, it is clear that this method compares favorably with the solvers \texttt{rKF} and \texttt{stiff} (the solver \texttt{stiff} use BDF up to order 5). The precision obtained with DC4 is in many cases better than for these two adaptive methods run with the maximal possible precision(i.e. with the smallest $atol$ and $rtol$).
Acknowledgements The authors would like to acknowledge the financial support of the Discovery Grant Program of the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. Byrne, G.D., Hindmarsh, A.C.: Stiff ode solvers: A review of current and coming attractions. Journal of Computational physics 70(1), 1–62 (1987)
2. Carter, H.: Calcul Differentiel. Hermann (1967)
3. Chung, T.: Computational fluid dynamics, 2nd edn. Cambridge university press (2010)
4. Collatz, L.: The Numerical Treatment of Differential Equation, 3rd edn. Springer-Verlag Berlin Heidelberg (1960)
5. Dahlquist, G., Bjöck, A.: Numerical methods in scientific computing, volume I, vol. 8. SIAM (2008)
6. Daniel, J.W., Pereyra, V., Schumaker, L.L.: Iterated deferred corrections for initial value problems. Wisconsin Univ Madison Mathematics Research center (1967)
7. Dutt, A., Greengard, L., Rokhlin, V.: Spectral deferred correction methods for ordinary differential equations. BIT Numerical Mathematics 40(2), 241–266 (2000)
8. Enright, W.H., Hull, T., Lindberg, B.: Comparing numerical methods for stiff systems of ode’s. BIT Numerical Mathematics 15, 1–48 (1975)
9. Gustafsson Bertil, W.K.: Deferred Correction methods for initial value problems. BIT 41(5), 986–995 (2001)
10. Hansen, A.C., Strain, J.: On the order of deferred correction. Applied numerical mathematics 61(8), 961–973 (2011)
11. Hildebrand, F.B.: Introduction to numerical analysis. International Series in Pure and Applied Mathematics. McGraw-Hill, New York (1974)
12. Isaacson, E., Keller, H.B.: Analysis of numerical methods. Wiley (1966)
13. Kushnir, D., Rokhlin, V.: A highly accurate solver for stiff ordinary differential equations. SIAM Journal on Scientific Computing 34(3), A1296–A1315 (2012)
14. Nguyen-Ba, T., Giordano, T., Vaillancourt, R.: Three-stage Hermite–Birkhoff solver of order 8 and 9 with variable step size for stiff odes. Calcolo 52(3), 371–405 (2015)
15. Pereyra, V.: Iterated deferred corrections for nonlinear operator equations. Numerische Mathematik 10(4), 316–323 (1967)
16. Quarteroni, A., Sacco, R., Saleri, F.: Numerical Mathematics, vol. 37, 2 edn. Springer Science & Business Media (2010)
17. Shampine, L.F.: Evaluation of a test set for stiff ode solvers. ACM Transactions on Mathematical Software (TOMS) 7, 409–420 (1981)
18. Spijker, M.N.: Stiffness in numerical initial-value problems. Journal of Computational and Applied Mathematics 72(2), 393–406 (1996)
19. Thomée, V.: Galerkin finite element methods for parabolic problems, Springer Series in Computational Mathematics, vol. 25. Springer-Verlag, Berlin, (2006)
20. Wanner, G., Hairer, E.: Solving ordinary differential equations II. Stiff and Differential-Algebraic Problems, vol. 14. Springer-Verlag Berlin Heidelberg (1991)
21. Wanner, G.: Dahlquist’s classical papers on stability theory. BIT Numerical Mathematics 46(3) (2006)
22. Williams, J., de Hoog, F.: A class of A-stable advanced multistep methods. Mathematics of Computation 28(125), 163–177 (1974)