MATHEMATICS OF COMPUTATION
VOLUME 51, NUMBER 184
OCTOBER 1988, PAGES S7–S16

Supplement to Backward Differentiation Approximations of Nonlinear Differential/Algebraic Systems

By Kathryn E. Brenan and Bjorn E. Engquist

A. Convergence Proof for the Index-Three System (Theorem 2).

As we did for the index-two system, we must prove there always exists a set of consistent initial values $(c(t_{k-1}),\ b(t_{k-1}),\ e(t_{k-1}))$ for the linear, index-three DAE system (2.11)-(2.13) corresponding to any set of initial values $(v_{k-1}, w_{k-1}, u_{k-1})$ satisfying (2.3),(2.4). From (2.13) it is clear we should choose $b(t_{k-1}) \in \mathcal{N}(A_{32}(t_{k-1}))$. The consistency condition (2.3) implies we can write $w_{k-1} = w(t_{k-1}) + h^{k+1}c_{1,k-1}$ and $v_{k-1} = v(t_{k-1}) + h^{k+1}c_{2,k-1}$ for some bounded vectors $c_{1,k-1}$ and $c_{2,k-1}$. Since we want our initial values to be consistent with the asymptotic error expansions (2.7),(2.8), we select for $t = t_{k-1}$ and $i = k-1$,

$$b(t) = 0, \qquad p_i = c_{1,i}. \tag{A.1}$$

We must then choose $c(t_{k-1})$ to satisfy the first derivative of the algebraic equation (2.13) in the linear, index-three system: $(t = t_{k-1})$

$$A_{32}(t)A_{21}(t)c(t) = -\frac{1}{(k+1)}A_{32}(t)w^{(k+1)}(t). \tag{A.2}$$

Since the range of $A_{32}(t)$, denoted $\mathbf{R}(A_{32}(t))$, equals $\mathbf{R}(A_{32}(t)A_{21}(t))$ for all $t \in I$, there is always a solution $c(t_{k-1})$ to (A.2). Next we must choose q_{k-1} so that (2.7) is valid for $n = k-1$ (here $t = t_{k-1}$ and $i = k-1$):

$$q_i = -\frac{1}{h}c(t) + c_{2,i}. \tag{A.3}$$

In general, $c(t_{k-1}) \neq 0$ so q_{k-1} will have order $O(1/h)$. The convergence analysis will require us to pick q_{k-1} such that $P_{11}(t_{k-1})q_{k-1}$ is bounded, where $P_{11}(t)$ is a projection, namely $P_{11}(t) = I_p - A_{13}(t)\Gamma(t)A_{32}(t)A_{21}(t)$ and $\Gamma(t) = (A_{32}(t)A_{21}(t)A_{13}(t))^{-1}$. Therefore, if we require $P_{11}(t_{k-1})c(t_{k-1}) = 0$, so that $P_{11}(t_{k-1})q_{k-1}$ is bounded, then from (A.2) it follows that for $t = t_{k-1}$

$$c(t) = -\frac{1}{(k+1)}A_{13}(t)\Gamma(t)A_{32}(t)w^{(k+1)}(t). \tag{A.4}$$

The consistent initial value for $e(t_{k-1})$ is completely determined from the equation obtained by differentiating (2.13) twice and substituting for $c'(t)$ and $b'(t)$:

$$\begin{aligned} e(t) \;=\; & -\Gamma(t)\Big(\big(2A'_{32}A_{21} + A_{32}A'_{21} + A_{32}(A_{21}A_{11} + A_{22}A_{21})\big)c(t) \\ & + \frac{1}{(k+1)}\big(A_{32}A_{21}v^{(k+1)}(t) + (2A'_{32} + A_{32}A_{22})w^{(k+1)}(t) + A_{32}w^{(k+2)}(t)\big)\Big), \end{aligned} \tag{A.5}$$

where all the A_{ij} matrices are evaluated at $t = t_{k-1}$.

Since $P_{11}(t)c(t) = 0$, it follows from (A.1) and (A.3) that there exists a constant η_o such that

$$\|p_i\| = \|c_{1,i}\| \le \eta_o, \tag{A.6}$$

$$\|P_{11}(t)q_i\| = \|P_{11}(t)c_{2,i}\| \le \eta_o \tag{A.7}$$

for $t = t_{k-1}$ and $i = k-1$. Then the solution $(c(t), b(t), e(t))$ to the linear, index-three system (2.11)-(2.13) satisfying the consistency conditions just given for $(c(t_{k-1}), b(t_{k-1}), e(t_{k-1}))$ is unique and smooth for $t \ge t_{k-1}$. Since the k-step BDF requires a numerically consistent starting vector y_{k-1}, we will define $(c(t_i), b(t_i), e(t_i))$ using equations (A.1)-(A.5) when $t = t_i$, and $i = 0, 1, \ldots, k-2$. Let us redefine η_o so that (A.6) and (A.7) are satisfied for all $i = 0, 1, \ldots, k-1$. If the consistency relations are relaxed as stated in the remark following the statement of Theorem 2, consistent initial values $(c(t_{k-1}), b(t_{k-1}), e(t_{k-1}))$ can be determined similarly, although in general equations (A.1)-(A.5) are more complicated (because $b(t_{k-1})$ may not be zero). Then, after substituting equations (2.7),(2.8) for the numerical solution into the BDF equations (2.5),(2.6) and after expanding by Taylor series about the analytic solution at t_n, it follows from (2.11)-(2.13) that the remainders (q_n, p_n, s_n) must satisfy the system

$$S(t_n)\begin{bmatrix} q_n \\ p_n \\ h^2 s_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{k} \gamma_i q_{n-i} \\ \sum_{i=1}^{k} \gamma_i p_{n-i} \\ 0 \end{bmatrix} + \begin{bmatrix} \psi_1(t_n) \\ \psi_2(t_n) \\ \psi_3(t_n) \end{bmatrix}, \tag{A.8}$$

where

$$\begin{aligned}
S(t_n) &= \begin{bmatrix} I_p - \frac{h}{\alpha_o}A_{11}(t_n) & -\frac{h}{\alpha_o}A_{12}(t_n) & -\frac{1}{\alpha_o h}A_{13}(t_n) \\ -\frac{h}{\alpha_o}A_{21}(t_n) & I_q - \frac{h}{\alpha_o}A_{22}(t_n) & 0 \\ 0 & A_{32}(t_n) & 0 \end{bmatrix}, \\
\gamma_i &= -\alpha_i/\alpha_o \text{ where } \alpha_i \ \ (i = 0, 1, \ldots, k) \text{ are the BDF coefficients,} \\
\psi_1(t_n) &= \Big(-hR_1(\xi_n) + h^k c^{(k+1)}(t_n)/(k+1) - h^{k+1}R_{12}(\xi_n) + h^k Q_1(t_n) \\
&\quad + h^{k+1}W_1(q_n, p_n, s_n) + h^{k+2}g_1(q_n, p_n, s_n)\Big)/\alpha_o, && \text{(A.9)} \\
\psi_2(t_n) &= \Big(-h\alpha^* w^{(k+2)}(t_n) + h^2 R_2(\xi_n) + h^k b^{(k+1)}(t_n)/(k+1) - h^{k+1}R_{22}(\xi_n) \\
&\quad + h^k Q_2(t_n) + h^{k+1}W_2(q_n, p_n) + h^{k+2}g_2(q_n, p_n)\Big)/\alpha_o, && \text{(A.10)} \\
\psi_3(t_n) &= -h^{k-1}Q_3(t_n) - h^k W_3(p_n) - h^{k+1}g_3(p_n), && \text{(A.11)}
\end{aligned}$$

and $\alpha^* = \frac{(-1)^k}{(k+2)!}\sum_{i=1}^{k} i^{k+2}\alpha_i$.

The functions $Q_i(t)$ $(i = 1, 2, 3)$ are functions of $c(t)$, $b(t)$, $e(t)$ and partial derivatives of F, G, and H, and are not dependent on q, p, or s. Since $(c(t), b(t), e(t))$ is the solution to the DAE system (2.11)-(2.13), and since the given functions F, G, and H may be assumed to be arbitrarily smooth, it follows that there exists a constant Q such that $\|Q_i(t)\| \le Q$ $(i = 1, 2, 3)$ for all $t \in [t_{k-1}, t_o + T]$. For $k = 1$ the proof of convergence requires not only this boundedness property, but also the smoothness of the $O(h)$ terms in $\psi_2(t)$ and the $O(1)$ term in $\psi_3(t)$, namely $Q_3(t)$. These conditions are valid providing the given functions F, G, and H are sufficiently smooth functions of all their arguments.

The functions W_i $(i = 1, 2, 3)$ are linear functions of their arguments, and for our purposes may be considered as bounded matrix operators. The functions $R_1(\xi_n)$, $R_2(\xi_n)$, $R_{12}(\xi_n)$, and $R_{22}(\xi_n)$ are bounded functions of ξ_n (not necessarily the same ξ_n) where $t_{n-k} \le \xi_n \le t_n$. They are the *remainders* from the k-step BDF formula - i.e., dependent on $u^{(k+3)}(\xi_n)$, $w^{(k+3)}(\xi_n)$, $c^{(k+2)}(\xi_n)$, and $b^{(k+2)}(\xi_n)$, respectively.

All of the remaining terms in the Taylor series expansion are contained in the g_i $(i = 1, 2, 3)$ functions, with the leading terms being quadratic functions of the arguments. For example, the leading term in $g_3(p)$ is $p^T H_{ww}(*)p/2$. All remaining terms in $g_3(p)$ are at least of order $O(h^k)$. Similarly, the leading terms in $g_2(q, p)$ are quadratic functions of q and p, and all remaining terms are $O(h^k)$ at least. These statements are true providing q and p are bounded (which will always be the case). The behavior of $g_1(q, p, s)$ depends on s, since s_n is not bounded until $n \ge 4k - 1$. For $k \ge 2$, the leading terms in $g_1(q, p, s)$ are bounded, quadratic functions of q, p, and s, and any remaining terms in $g_1(q, p, s)$ will contribute at worse an $O(h)$ amount to $\psi_1(t)$ in the case $h^2 s$ is bounded, but hs or s is not. For $k = 1$, we will need to bound $\psi_1(t)$ to $O(h)$ when either $\|hs\|$ or $\|s\|$ is bounded. Since

$$\|h^3 g_1(q, p, s)\| = \|\frac{h}{2}[(hs)^T F_{uu}(*)(hs)^T] + O(h^2)\|$$

when $\|hs\|$ is bounded, and $\|h^3 g_1(q, p, s)\| = O(h^3)$ when $\|s\|$ is bounded, it follows that $\psi_1(t)$ is $O(h)$. In general, we will need to consider only the leading quadratic terms in each g_i $(i = 1, 2, 3)$ for the analysis which follows.

Let us denote the elements of the 3×3 matrix $S^{-1}(t_n)$ by $X_{ij}(t_n)$ $(i, j = 1, 2, 3)$, where the elements are defined below (t-dependence is suppressed):

$$\begin{aligned}
\Gamma &= (A_{32}A_{21}A_{13})^{-1}, && \text{(A.12)} \\
P_{11} &= I_p - A_{13}\Gamma A_{32}A_{21}, && \text{(A.13)} \\
P_{22} &= I_q - A_{21}A_{13}\Gamma A_{32}, && \text{(A.14)} \\
X_{11} &= P_{11} + \frac{h}{\alpha_o}\Big(A_{11} - A_{13}\Gamma A_{32}(A_{22}A_{21} + A_{21}A_{11})\Big)P_{11} + O(h^2), && \text{(A.15)} \\
X_{12} &= \frac{-\alpha_o}{h}A_{13}\Gamma A_{32} - A_{13}\Gamma A_{32}A_{22}P_{22} - P_{11}A_{11}A_{13}\Gamma A_{32} + O(h), && \text{(A.16)} \\
X_{13} &= \frac{\alpha_o}{h}A_{13}\Gamma + P_{11}A_{11}A_{13}\Gamma - A_{13}\Gamma A_{32}A_{22}A_{21}A_{13}\Gamma + O(h), && \text{(A.17)} \\
X_{21} &= \frac{h}{\alpha_o}P_{22}A_{21} + \frac{h^2}{\alpha_o^2}P_{22}(A_{22}A_{21} + A_{21}A_{11})P_{11} + O(h^3), && \text{(A.18)} \\
X_{22} &= P_{22} + \frac{h}{\alpha_o}P_{22}(A_{22}P_{22} + A_{21}A_{11}A_{13}\Gamma A_{32}) + O(h^2), && \text{(A.19)} \\
X_{23} &= A_{21}A_{13}\Gamma + \frac{h}{\alpha_o}P_{22}(A_{22}A_{21} + A_{21}A_{11})A_{13}\Gamma + O(h^2), && \text{(A.20)} \\
X_{31} &= -h\alpha_o\Gamma A_{32}A_{21} - h^2\Gamma A_{32}(A_{22}A_{21} + A_{21}A_{11})P_{11} + O(h^3), && \text{(A.21)} \\
X_{32} &= -\alpha_o^2\Gamma A_{32} + h\alpha_o\Gamma A_{32}(A_{22}A_{21} + A_{21}A_{11})A_{13}\Gamma A_{32} - \alpha_o\Gamma A_{32}A_{22} + O(h^2), && \text{(A.22)} \\
X_{33} &= \alpha_o^2\Gamma - h\alpha_o\Gamma A_{32}(A_{22}A_{21} + A_{21}A_{11})A_{13}\Gamma + O(h^2). && \text{(A.23)}
\end{aligned}$$

These expressions for X_{ij} $(i,j=1,2,3)$ are valid for sufficiently small h. From the relations given in (A.12)-(A.23), it follows there exists a constant K such that

$$\|X_{11}\|,\ h\|X_{12}\|,\ h\|X_{13}\|,\ \frac{1}{h}\|X_{21}\|,\ \|X_{22}\|,\ \|X_{23}\|,\ \frac{1}{h}\|X_{31}\|,\ \|X_{32}\|,\ \|X_{33}\| \ \leq\ K \qquad \text{(A.24)}$$

for all $t \in I$. Define $\mathbf{r}_n = [q_n^T, p_n^T, q_{n-1}^T, p_{n-1}^T, \ldots, q_{n-k+1}^T, p_{n-k+1}^T]^T$ and $\mathbf{s}_n = [s_n^T, s_{n-1}^T, \ldots, s_{n-k+1}^T]^T$. Then system (A.8) can be rewritten in the one-step form as

$$\begin{aligned} \mathbf{r}_n &= F_n \mathbf{r}_{n-1} + f_{1,n}, & \text{(A.25)} \\ h^2 \mathbf{s}_n &= G_n \mathbf{r}_{n-1} + f_{2,n} + h^2 H \mathbf{s}_{n-1}, & \text{(A.26)} \end{aligned}$$

where

$$\begin{aligned} &F_n = \mathrm{CPM}(\gamma_\mu X_n) \qquad (\mu = 1,2,\ldots,k), \qquad \text{(A.27)} \\ &X_n = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}_{|t_n}, \\ &f_{1,n} = [(X_{11}\psi_1 + X_{12}\psi_2 + X_{13}\psi_3)^T,\ (X_{21}\psi_1 + X_{22}\psi_2 + X_{23}\psi_3)^T, 0, \ldots, 0]^T, \\ &G_n = (I_m, 0, \ldots, 0)^T(\gamma_1 U_n, \gamma_2 U_n, \ldots, \gamma_k U_n), \\ &U_n = [X_{31},\ X_{32}]_{|t_n}, \\ &f_{2,n} = [(X_{31}\psi_1 + X_{32}\psi_2 + X_{33}\psi_3)^T, 0, \ldots, 0]^T, \\ &H = \mathrm{CPM}(0_m). \end{aligned}$$

Following a similar approach as was used for the convergence analysis presented in Section 3, the convergence proof here relies on an induction argument as well as the construction of a fixed point iteration. The fixed point iteration scheme will be constructed to solve (A.25) and (A.26), which are implicit in (q_n, p_n, s_n). In particular for $k \geq 2$, we will prove there exists a uniformly bounded solution (q_n, p_n, hs_n) for sufficiently large n. For $k = 1$ we will prove there exists a bounded solution (q_n, p_n, s_n) satisfying (A.25) and (A.26) for $n \geq 3$. Convergence follows then from (2.7),(2.8) since

$$\begin{aligned} \|v_n - v(t_n)\| &\leq h^k\|c(t_n)\| + h^{k+1}\|q_n\|, \\ \|w_n - w(t_n)\| &\leq h^k\|b(t_n)\| + h^{k+1}\|p_n\|, \\ \|u_n - u(t_n)\| &\leq h^k\left(\|e(t_n)\| + \|hs_n\|\right) \end{aligned}$$

are all $O(h^k)$ for either bounded (q_n, p_n, hs_n) or bounded (q_n, p_n, s_n).

Induction Assumption.

For $i = k, k+1, \ldots, 4k-2$,

$$\|q_i\| \leq \eta_1^* \text{ and } \|p_i\| \leq \eta_2^*. \qquad \text{(A.28)}$$

For $k \geq 2$ and $i = k, k+1, \ldots, 3k-2$,

$$\|h^2 s_i\| \leq \eta_3^*. \qquad \text{(A.29)}$$

For $i = 3k-1, 3k, \ldots, 4k-2$,

$$\|hs_i\| \leq \eta_3^{**}. \qquad \text{(A.30)}$$

For $i = 4k-1, 4k, \ldots, n-1$,

$$\begin{aligned} \|q_i\| &\leq \eta_1 e^{(i-3k+1)hL}, & \text{(A.31)} \\ \|p_i\| &\leq \eta_2 e^{(i-3k+1)hL}, & \text{(A.32)} \\ \|hs_i\| &\leq \eta_3 e^{(i-3k+1)hL} \quad \text{for } k \geq 2, & \text{(A.33)} \\ \|s_i\| &\leq \eta_3 e^{ihL} \quad \text{for } k = 1. & \text{(A.34)} \end{aligned}$$

The existence of the constants η_j^* $(j = 1,2,3)$ and η_3^{**} will be verified at the end of the induction proof. It should be noted that (A.29) is not utilized in the convergence proof when $k = 1$, and in fact may not even be true.

It is at this point that the details in the convergence proof for $k = 1$ become significantly more complicated than in the proof for $k \geq 2$. Since the proof for $k \geq 2$ is of more general interest, we will present that case. Key distinctions in the proof for $k = 1$ will be noted later.

Verification of Induction Assumption for $i \geq 4k-1$ and $k \geq 2$.

Now we will prove that there exist constants η_1, η_2, η_3, and L such that relations (A.31)-(A.33) hold uniformly for $i = n$ and $n \geq 4k-1$. Using the induction assumption and the definitions of $\psi_1(t)$, $\psi_2(t)$, and $\psi_3(t)$ given in (A.9)-(A.11), one can derive the following bounds:

$$\begin{aligned} \|\psi_1(t_i)\| &\leq \sigma_{11} h^{k-2} \quad \text{for } i = k, k+1, \ldots, 3k-2 & \text{(A.35)} \\ &\quad \text{where } \sigma_{11} = \tilde{\sigma}_{11}(\eta_3^*) + O(h), \\ \|\psi_1(t_i)\| &\leq \sigma_{12} h \quad \text{for } i = 3k-1, \ldots, n-1, & \text{(A.36)} \\ \|\psi_2(t_i)\| &\leq \sigma_2 h \quad \text{for } i = k, k+1, \ldots, n-1, & \text{(A.37)} \\ \|\psi_3(t_i)\| &\leq \sigma_3 h^{k-1} \quad \text{for } i = k, k+1, \ldots, n-1 & \text{(A.38)} \end{aligned}$$

for some constants σ_{12}, σ_2 and σ_3 independent of $\eta_1^*, \eta_2^*, \eta_3^*, \eta_3^{**}, \eta_1, \eta_2, \eta_3$ and L to $O(h)$ accuracy.

We now proceed with our proof of the existence of the constants η_ℓ $(\ell = 1,2,3)$ and L for $n \geq 4k-1$ by constructing a fixed point iteration. Let us define the fixed point iteration:

$$\begin{aligned} q_n^{(\nu+1)} &= q_n^{(o)} + Z_1\left(q_n^{(\nu)}, p_n^{(\nu)}, hs_n^{(\nu)}\right), \\ p_n^{(\nu+1)} &= p_n^{(o)} + Z_2\left(q_n^{(\nu)}, p_n^{(\nu)}, hs_n^{(\nu)}\right), \\ hs_n^{(\nu+1)} &= hs_n^{(o)} + Z_3\left(q_n^{(\nu)}, p_n^{(\nu)}, hs_n^{(\nu)}\right), \end{aligned}$$

where

$$\begin{aligned} Z_1(q,p,hs) \ = \ & \frac{h^{k+1}}{\alpha_o} X_{11}\left[W_1(q,p,s) + hg_1(q,p,s)\right] + \frac{h^{k+1}}{\alpha_o} X_{12}\left[W_2(q,p) + hg_2(q,p)\right] \\ & - h^k X_{13}\left[W_3(p) + hg_3(p)\right], \end{aligned}$$

$$
\begin{aligned}
Z_2(q,p,hs) &= \frac{h^{k+1}}{\alpha_o}X_{21}\left[W_1(q,p,s)+hg_1(q,p,s)\right]+\frac{h^{k+1}}{\alpha_o}X_{22}\left[W_2(q,p)+hg_2(q,p)\right]\\
&\quad -h^kX_{23}\left[W_3(p)+hg_3(p)\right],\\
Z_3(q,p,hs) &= \frac{h^{k}}{\alpha_o}X_{31}\left[W_1(q,p,s)+hg_1(q,p,s)\right]+\frac{h^{k}}{\alpha_o}X_{32}\left[W_2(q,p)+hg_2(q,p)\right]\\
&\quad -h^{k-1}X_{33}\left[W_3(p)+hg_3(p)\right].
\end{aligned}
$$

For starting approximations select

$$
\begin{aligned}
\mathbf{r}_n^{(o)} &= F_n\mathbf{r}_{n-1}+\tilde{f}_{1,n}, && \text{(A.39)}\\
hs_n^{(o)} &= \frac{1}{h}G_n\mathbf{r}_{n-1}+\frac{1}{h}\tilde{f}_{2,n}+hH\mathbf{s}_{n-1}, && \text{(A.40)}
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{f}_{1,n} &= \left[(\tilde{f}_{1,n}^{(1)})^T,(\tilde{f}_{1,n}^{(2)})^T,0,\ldots,0\right]^T,\\
\tilde{f}_{1,n}^{(i)} &= \left[X_{i1}\tilde{\psi}_1+X_{i2}\tilde{\psi}_2+X_{i3}\tilde{\psi}_3\right]_{|t_n},\quad i=1,2,\\
\tilde{f}_{2,n} &= \left[(X_{31}\tilde{\psi}_1+X_{32}\tilde{\psi}_2+X_{33}\tilde{\psi}_3)^T,0,\ldots,0\right]^T_{|t_n},\\
\tilde{\psi}_1(t_n) &= \left(-hR_1(\xi_n)+h^kc^{(k+1)}(t_n)/(k+1)-h^{k+1}R_{12}(\xi_n)+h^kQ_1(t_n)\right)/\alpha_o,\\
\tilde{\psi}_2(t_n) &= \left(-h\alpha^*w^{(k+2)}(t_n)+h^kb^{(k+1)}(t_n)/(k+1)+h^2R_2(\xi_n)\right.\\
&\quad \left.-h^{k+1}R_{22}(\xi_n)+h^kQ_2(t_n)\right)/\alpha_o,\\
\tilde{\psi}_3(t_n) &= -h^{k-1}Q_3(t_n).
\end{aligned}
$$

Since $\tilde{\psi}_i(t_n)$ $(i=1,2,3)$ are independent of (q_n,p_n,s_n), it follows that

$$
\|\tilde{\psi}_1(t)\|,\ \|\tilde{\psi}_2(t)\|,\ \|\tilde{\psi}_3(t)\| \ \le\ \rho h \qquad \text{(A.41)}
$$

for some constant ρ and $t\in[t_k,t_o+T]$.

It is important to note we are iterating for hs_n, not h^2s_n or s_n. For the implicit function argument, the following three conditions must be satisfied for sufficiently large n (i.e., $n\ge 4k-1$):

I.

$$
\begin{aligned}
\|q_n^{(o)}\| &< \eta_1e^{(n-3k+1)hL}\\
\|p_n^{(o)}\| &< \eta_2e^{(n-3k+1)hL} && \text{(A.42)}\\
\|hs_n^{(o)}\| &< \eta_3e^{(n-3k+1)hL}
\end{aligned}
$$

II.

$$
\|Z(q_n^{(o)},p_n^{(o)},hs_n^{(o)})\|\le\delta/2,\qquad \delta>0,\qquad Z=(Z_1^T,Z_2^T,Z_3^T)^T \qquad \text{(A.43)}
$$

III.

$\|J\|\le 1/2$ for any $q,p,$ and hs such that $\|q-q_n^{(o)}\|\le\delta$, $\|p_n-p_n^{(o)}\|\le\delta$, and $\|hs-hs_n^{(o)}\|\le\delta$ where J is the Jacobian matrix of Z with respect to $q,p,$ and hs. (A.44)

We will show these conditions hold for $\delta=O(h)$ and $n\ge 4k-1$. If condition (A.42) holds, it is straightforward to show conditions (A.43) and (A.44) hold by utilizing the information known about X_{ij} $(i,j=1,2,3)$, W_i and g_i $(i=1,2,3)$. Note that $\|Z_1\|=O(h^{k-1})$, $\|Z_2\|=O(h^k)$, and $\|Z_3\|=O(h^{k-1})$, so for $k\ge 2$ we can select $\delta=O(h)$. Here we have used the facts that hs_n is involved only in the W_1 and g_1 terms, that W_1 is linear, and that g_1 is quadratic in hs_n to $O(h)$ accuracy. To see that condition (A.44) holds for bounded q, p, and hs, one need only observe that taking the partial derivatives of Z with respect to (q_n,p_n,hs_n) does not alter the powers of h.

The proof that the starting iterates are bounded as in (A.42) involves straightforward algebraic manipulations. The details in the algebra are omitted here. However, the essential steps are given below. The proof for $k=2$ requires more exploitation of the structure of the block matrices than when $k\ge 3$. In both cases, the proof relies on some projection matrices to annihilate the effects of $1/h$ factors, as is demonstrated repeatedly in certain matrix product relations.

Since equations (A.25) and (A.26) hold when n is replaced by i and $i=k,k+1,\ldots,n-1$, we can rewrite equations (A.39),(A.40) as

$$
\begin{aligned}
\mathbf{r}_n^{(o)} &= \left(\prod_{i=k}^{n}F_i\right)\mathbf{r}_{k-1}+\sum_{i=k}^{n-1}\left(\prod_{j=i+1}^{n}F_j\right)f_{1,i}+\tilde{f}_{1,n}, && \text{(A.45)}\\
hs_n^{(o)} &= \sum_{\ell=0}^{k-1}\left(\frac{1}{h}H^\ell G_{n-\ell}\left(\prod_{i=k}^{n-\ell-1}F_i\right)\mathbf{r}_{k-1}+\sum_{i=k}^{n-\ell-2}\left(\prod_{j=i+1}^{n-\ell-1}F_j\right)f_{1,i}+f_{1.n-\ell-1}\right)\\
&\quad +\frac{1}{h}\sum_{\ell=1}^{k-1}H^\ell f_{2,n-\ell}+\frac{1}{h}\tilde{f}_{2,n}. && \text{(A.46)}
\end{aligned}
$$

First we will bound $\mathbf{r}_n^{(o)}$, starting with the first term in (A.45) involving $\mathbf{r}_{k-1}$. Recall that F_i is a $\ell k\times\ell k$ block companion matrix, with blocks X_i containing a subblock $X_{12}(t_i)$ of order $O(1/h)$. Since $n\ge 4k-1$, the product $\prod_{i=k}^{n}F_i$ involves $2k$ or more factors of F_i, so all blocks are sums of terms with at least two factors of X_i $(i=k,\ldots,n)$. We will need the following lemma proven in [1]:

Lemma 1 . *There exist constants K^* $(K^*\ge 1)$ and $\mathcal{E}$ such that*

$$
\|\prod_{j=i+1}^{n}F_j\|\le K^*e^{(n-i)h\mathcal{E}}
$$

for all $n-i\ge 2k$ where F_j is defined in (A.27).

The proof of this lemma is technical but straightforward, involving the application of a stability theorem for matrix products ([16], [26]) as well as a theorem verifying the existence of smooth similarity tranformations [13]. One trick in the proof involves partitioning the product into groups of $2k$ factors of F_j because such groups can be shown to be bounded while individual factors F_j are not. The eigenvalues and eigenvectors of a group must be analyzed in detail. The details of this analysis can be found in [1] (see Lemma 6.14). Therefore, the product $\prod_{i=k}^{n}F_i$ in (A.45) is bounded, but the starting vector $\mathbf{r}_{k-1}$ contains $O(1/h)$ terms q_j, $j = 0, 1, \ldots, k-1$. Recall the starting values were chosen to satisfy (A.6),(A.7). Now using the structure of the product $\prod_{i=k}^{n} F_i$, one can show the q_j terms always appear as $P_{11}(t_i)q_j$ which are $O(1)$. All the terms involving p_j are also clearly $O(1)$. Therefore, there exist constants $\eta_o^{(\ell)}, \ell = 1, 2, \ldots, 2k$, such that

$$\| \left(\prod_{i=k}^{3k+\ell-2} F_i \right) \mathbf{r}_{k-1} \| \leq \eta_o^{(\ell)}. \tag{A.47}$$

For $n \geq 5k - 1$, it follows that

$$\| \left(\prod_{i=k}^{n} F_i \right) \mathbf{r}_{k-1} \| \leq K^* e^{(n-3k+1)h\mathcal{E}} \eta_o^{(1)},$$

since $\| \prod_{i=3k}^{n} F_i \| \leq K^* e^{(n-3k+1)h\mathcal{E}}$ from Lemma 1. We summarize this bound as

$$\| \left(\prod_{i=k}^{n} F_i \right) \mathbf{r}_{k-1} \| \leq \eta_o^* K^* e^{(n-3k+1)h\mathcal{E}}$$

for $n \geq 3k - 1$ and $\eta_o^* = \max_{\ell=1,\ldots,2k} \left(\eta_o^{(\ell)} \right)$. Next, in the bounding of $\mathbf{r}_n^{(o)}$, we break the terms $\sum_{i=k}^{n-1} \left(\prod_{j=i+1}^{n} F_j \right) f_{1,i}$ into two parts:

$$\sum_{i=k}^{n-1} \left(\prod_{j=i+1}^{n} F_j \right) \left(R_i \left[\psi_1(t_i)^T, \psi_2(t_i)^T, 0, \ldots, 0 \right]^T + \left[(X_{13}(t_i)\psi_3(t_i))^T, (X_{23}(t_i)\psi_3(t_i))^T, 0, \ldots, 0 \right]^T \right), \tag{A.48}$$

where $R_i =$ diag $(X_i, 0, \ldots, 0)$. In bounding both parts, it will be necessary to consider the case $k = 2$ separately from $k \geq 3$. Consider the terms involving ψ_1 and ψ_2. It is possible to show that there exists a constant N_1 such that (see Lemma 6.15 in [1])

$$\| \left(\prod_{j=i+1}^{n} F_j \right) R_i \| \leq N_1 K^* e^{nh\mathcal{E}} \qquad \text{for } n - i \geq k.$$

For $n - i < k$, any $O(1/h)$ terms in $\left(\prod_{j=i+1}^{n} F_j \right) R_i$ will multiply the $\psi_2(t_i)$ term, which is always $O(h)$ by (A.37). Thus, there exists a constant Δ_1 independent of η_1, η_2, η_3 or L such that

$$\| \left(\prod_{j=i+1}^{n} F_j \right) R_i \left[\psi_1(t_i)^T, \psi_2(t_i)^T, 0, \ldots, 0 \right]^T \| \leq \Delta_1$$

for $i = n - k + 1, \ldots, n - 1$. Then, using relations (A.35)-(A.38) it follows that

$$\begin{aligned} \Lambda &= \| \sum_{i=k}^{n-1} \left(\prod_{j=i+1}^{n} F_j \right) R_i \left[\psi_1(t_i)^T, \psi_2(t_i)^T, 0, \ldots, 0 \right]^T \| \\ &\leq [(2k-1)h + (n - 4k + 2)h] N_1 K^* e^{nh\mathcal{E}} \max(\sigma_{11}, \sigma_{12}, \sigma_2) + (k-1)\Delta_1 \end{aligned}$$

providing $n \geq 4k - 1$ and $k \geq 3$. If $k = 2$, the terms $\sum_{i=k}^{3k-2} \cdots$ must be bounded more carefully since $\|\psi_1(t_i)\| \leq \sigma_{11}$ for $i = k, k+1, \ldots, 3k-2$. Still, there are a finite number of such terms (each $O(1)$), so we can summarize this result for $k \geq 2$ as $\Lambda \leq \Delta_1^*$ for some constant Δ_1^* independent of η_1, η_2, η_3 and L to $O(h)$ accuracy.

Next, we bound the terms in (A.48) involving $\psi_3(t_i)$. The case when $k \geq 3$ is fairly straightforward, requiring the use of (A.38) and some information concerning the structure of the matrices (i.e, the location of $O(1/h)$ elements). If $k = 2$, the terms must be broken into two parts:

$$\sum_{i=k}^{n-2k} \cdots + \sum_{i=n-2k+1}^{n-1} \cdots.$$

The finite number of terms is easily bounded by noting any $O(1/h)$ terms in $\prod_{j=i+1}^{n} F_j$ will multiply the $O(h)$ term $X_{23}(t_i)\psi_3(t_i)$. The remaining terms $(i = k, \ldots, n-2k)$ involve the product $\prod_{j=i+1}^{n} F_j$ containing $2k$ or more factors of F_j. Therefore, these terms will contain products involving two or more factors of X_j, such as

$$X_{i+j+\ell} X_{i+j} \begin{bmatrix} X_{13,i} \\ X_{23,i} \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix},$$

where A_1 is $O(1)$ and A_2 is $O(h)$ when $\ell \neq j$. This last fact follows by direct computation using relations (A.12)-(A.23). Similar relations hold for $\left(\prod_{j\in\sigma} X_j \right) \left[(X_{13,i})^T, \, (X_{23,i})^T \right]^T$, where σ contains two or more indices. Then, as for $k \geq 3$, these terms can be bounded, and in summary, there exists a Δ_2^* such that for $k \geq 2$ and $n \geq 4k - 1$

$$\| \sum_{i=k}^{n-1} \left(\prod_{j=i+1}^{n} F_j \right) \left[(X_{13}(t_i)\psi_3(t_i))^T, (X_{23}(t_i)\psi_3(t_i))^T, 0, \ldots, 0 \right]^T \| \leq \Delta_2^*,$$

where Δ_2^* is independent of η_1, η_2, η_3 and L to $O(h)$ accuracy.

The remaining terms $\tilde{f}_{1,n}$ in $\mathbf{r}_n^{(o)}$ can be bounded easily using the relations (A.41):

$$\|\tilde{f}_{1,n}\| \leq K \max(\rho h + 2\rho, O(h)) \leq \Delta_3^*$$

for some constant Δ_3^*. Finally, for $n \geq 4k - 1$ we have shown

$$\|\mathbf{r}_n^{(o)}\| \leq \eta_o^* K^* e^{(n-3k+1)h\mathcal{E}} + \Delta_1^* + \Delta_2^* + \Delta_3^*.$$

Since this bound is independent of η_1, η_2, η_3 and L, at least to $O(h)$ accuracy, there exists a constant η and $L = \mathcal{E}$ such that

$$\|\mathbf{r}_n^{(o)}\| < \eta e^{(n-3k+1)h\mathcal{E}} \qquad \text{for } n \geq 4k - 1.$$

Let $\eta_1 = \eta$ and $\eta_2 = \eta$. Then, for $n \geq 4k - 1$

$$\|q_n^{(o)}\| \leq \|\mathbf{r}_n^{(o)}\| < \eta_1 e^{(n-3k+1)h\mathcal{E}}$$

and

$$\|p_n^{(o)}\| \leq \|\mathbf{r}_n^{(o)}\| < \eta_2 e^{(n-3k+1)h\mathcal{E}}.$$

The procedure for bounding $hs_n^{(o)}$, defined in (A.46), is very similar to the previous analysis for $\mathbf{r}_n^{(o)}$, although additional properties of matrix products are utilized [3]. We will give only a summary of the intermediate results required for the estimate. The terms involving the initial remainder $\mathbf{r}_{k-1}$ can be bounded with the use of (A.47) and the following properties:

1. $\|G_{n-\ell}\prod_{i=3k}^{n-\ell-1} F_i\| = O(h)$ for all $n-\ell \geq 5k$
2. $\|G_{n-\ell}\left(\prod_{i=k}^{n-\ell-1} F_i\right)\mathbf{r}_{k-1}\| = O(h)$ for $n-\ell \geq 3k$.

Properties (1) and (2) can be easily proven by noting the following fact:

$$[X_{31},\ X_{32}]|_{t_{i+j+s}} X_{i+j} X_i = [A_1,\ A_2],$$

where A_1 is $O(h^2)$ and A_2 is $O(h)$ for $s \neq j$. This fact is proven by direct computation of the matrix products using relations (A.12)-(A.23) and Taylor series expansion about a common t value. It follows then that

$$\|\sum_{\ell=0}^{k-1}\frac{1}{h}H^{\ell}G_{n-\ell}\left(\prod_{i=k}^{n-\ell-1}F_i\right)\mathbf{r}_{k-1}\| \leq \Delta_4^*$$

for $n \geq 4k-1$ and for some constant Δ_4^*.

As we did earlier for $\mathbf{r}_n^{(o)}$, the terms involving $f_{1,i}$ are split into two pieces, and the cases $k=2$ and $k \geq 3$ are done separately. Using the structure of the matrices and relations (A.35)-(A.38), it can be shown that the first part is bounded for $k \geq 2$ as

$$\|\sum_{\ell=0}^{k-1}\frac{1}{h}H^{\ell}\sum_{i=k}^{n-\ell-2}G_{n-\ell}\left(\prod_{j=i+1}^{n-\ell-1}F_j\right)R_i\left[\psi_1(t_i)^T,\psi_2(t_i)^T,0,\ldots,0\right]^T\| \leq \Delta_5^*,$$

where Δ_5^* is independent of η_1, η_2, η_3 and L to $O(h)$ accuracy. Bounding the second part involving $\psi_3(t_i)$ is straightforward when $k \geq 3$, but when $k=2$ we will need the fact that there exists a constant $\tilde{N}$ such that

$$\|H^{\ell}G_{n-\ell}\prod_{j=i+1}^{n-\ell-1}F_j\| \leq \tilde{N}h^2$$

for $n-\ell-i-1 \geq 2k$ (see Lemma 6.12 in [1]). Then, we split this part further,

$$\sum_{i=k}^{n-\ell-2k-1}\cdots \text{ and } \sum_{i=n-\ell-2k}^{n-2-\ell}\cdots,$$

where the first sum is bounded as

$$\|\sum_{\ell=0}^{k-1}\frac{1}{h}H^{\ell}\sum_{i=k}^{n-\ell-2k-1}G_{n-\ell}\prod_{j=i+1}^{n-\ell-1}F_j\left[(X_{13}(t_i)\psi_3(t_i))^T,(X_{23}(t_i)\psi_3(t_i))^T,0,\ldots,0\right]^T\|$$
$$\leq k(n-3k)h\tilde{N}K\sigma_3 h^{k-2}.$$

Each term in the finite sum $\sum_{i=n-\ell-2k}^{n-\ell-2}\cdots$ involves products of X_{ij} which can be bounded to $O(1)$ by direct computation. Hence, for $k=2$ as well as for $k \geq 3$, we have

$$\|\sum_{\ell=0}^{k-1}\frac{1}{h}H^{\ell}G_{n-\ell}\sum_{i=k}^{n-\ell-2}\left(\prod_{j=i+1}^{n-\ell-1}F_j\right)\left[X_{13}(t_i)^T,X_{23}(t_i)^T,0,\cdots,0\right]^T\psi_3(t_i)\| \leq \Delta_6^*$$

for some constant Δ_6^* independent of η_1, η_2, η_3 and L to $O(h)$ accuracy.

Next, using relations (A.24), (A.35)-(A.38) and (A.41), we have

$$\|\sum_{\ell=1}^{k-1}\frac{1}{h}H^{\ell}f_{2,n-\ell}+\frac{1}{h}\tilde{f}_{2,n}\| \leq \Delta_7^*$$

for some constant Δ_7^* and $k \geq 2$. Finally, it is possible to prove that there exists a constant Δ_8^* such that

$$\|\sum_{\ell=0}^{k-1}\frac{1}{h}H^{\ell}G_{n-\ell}f_{1,n-\ell-1}\| \leq \Delta_8^* \qquad \text{for } k \geq 2.$$

This bound can be obtained by examining matrix products such as $X_{31}X_{13}$ and $X_{32}X_{23}$, and by applying the bounds (A.35)-(A.38). To summarize the final bound for $hs_n^{(o)}$, we have

$$\|hs_n^{(o)}\| \leq \sum_{i=4}^{8}\Delta_i^* \qquad \text{for } n \geq 4k-1,$$

where the constants Δ_i^* are independent of η_1, η_2, η_3 and L to $O(h)$ accuracy. Hence, we can select a constant η_3 such that for all $n \geq 4k-1$

$$\|hs_n^{(o)}\| \leq \|hs_n^{(o)}\| \leq \sum_{i=4}^{8}\Delta_i^* < \eta_3 e^{(n-3k+1)h\mathcal{E}}.$$

We have proven there exist constants η_1, η_2, η_3 and $\mathcal{E}$ such that the starting iterates are bounded uniformly providing $n \geq 4k-1$, as in conditions (A.42). It follows immediately that conditions (A.43) and (A.44) are satisfied. Hence, the fixed point iteration converges to a solution (q_n, p_n, hs_n) of equations (A.25) and (A.26). In addition,

$$\begin{aligned}\|q_n - q_n^{(o)}\| &= \|Z_1(q_n,p_n,hs_n)\| \leq \delta,\\ \|p_n - p_n^{(o)}\| &= \|Z_2(q_n,p_n,hs_n)\| \leq \delta,\\ \|hs_n - hs_n^{(o)}\| &= \|Z_3(q_n,p_n,hs_n)\| \leq \delta,\end{aligned}$$

where $\delta = O(h)$ for $k \geq 2$. Therefore, for sufficiently small h, we have for $n \geq 4k-1$

$$\begin{aligned}\|q_n\| &\leq \|q_n^{(o)}\| + \delta &&\leq \eta_1 e^{(n-3k+1)h\mathcal{E}},\\ \|p_n\| &\leq \|p_n^{(o)}\| + \delta &&\leq \eta_2 e^{(n-3k+1)h\mathcal{E}},\\ \|hs_n\| &\leq \|hs_n^{(o)}\| + \delta &&\leq \eta_3 e^{(n-3k+1)h\mathcal{E}}.\end{aligned}$$

To complete the induction argument, we must still verify the starting assumptions (A.28) - (A.30). However, presuming for the moment that they do hold, then q_n, p_n and hs_n are uniformly bounded for $n \geq 4k-1$. Hence, the k-step BDF method produces a numerical solution which converges to $O(h^k)$ accuracy to a solution of system (1.8)-(1.10) for $n \geq 4k-1$.

Verification of Induction Assumption for $i \leq 4k-2$.

The proof of convergence of the k-step BDF method depends on the induction assumption, and in particular on the behavior of q_i, p_i and $h^2 s_i$ during the first steps. The proof that the induction assumptions (A.28) and (A.29) hold is very similar to the proof just given above. That is, it is necessary to construct a fixed point iteration using equations (A.25) and (A.26), but now we iterate for q_i, p_i, and $h^2 s_i$, and not for hs_i.

The starting iterates selected in (A.39),(A.40) may be used except $q_i^{(o)}$ must be altered slightly to include the $O(1)$ term $(h^2 s_i^{(o)})^T F_{uu}(*)(h^2 s_i^{(o)})/2$, at least when $k = 2$. This term is added because the behavior of the $h^{k+2} g_1(q, p, s)$ term in $\psi_1(t)$ will have order $O(h^{k-2})$ when $h^2 s$ is bounded, but hs is not.

The assumptions (A.28) and (A.29) are first shown to hold for $i = k$, and then the result follows for later steps ($i = k+1, \ldots$) in the same fashion. In bounding the initial iterates for the fixed point iteration, we will use the same properties of the starting values and projection matrices used in the above induction argument. In fact, the consistency relation (2.4) for the starting values is vital in the proof that $q_i^{(o)}$ is bounded. Bounding the initial iterates is much simpler here because there are only finite sums involved. This argument verifies that q_n, p_n, and $h^2 s_n$ are bounded for any *finite* number of steps.

The induction assumption (A.30) must still be verified before the convergence proof is complete. From (A.28) we know q_n and p_n are bounded for $n = k, k+1, \ldots, 4k-2$. Therefore, we can construct a fixed point iteration just for hs_n, $n = 3k-1, \ldots, 4k-2$, using the difference equation (A.26):

$$hs_n^{(o)} = \frac{1}{h} X_{31}(t_n) \left(\sum_{i=1}^{k} \gamma_i q_{n-i} + \hat{\psi}_1(t_n) \right) + \frac{1}{h} X_{32}(t_n) \left(\sum_{i=1}^{k} \gamma_i p_{n-i} + \psi_2(t_n) \right) + \frac{1}{h} X_{33}(t_n) \psi_3(t_n),$$

where $\hat{\psi}_1(t_n) = \psi_1(t_n) -$ terms involving s_n and

$$hs_n^{(\nu+1)} = hs_n^{(o)} + Z(hs_n^{(\nu)}),$$

where

$$Z(hs_n^{(\nu)}) = \frac{1}{h} X_{31}(t_n) \left(\frac{h^{k+1}}{\alpha_o} \tilde{W}_1(s_n^{(\nu)}) + \frac{h^{k+2}}{\alpha_o} \tilde{g}_1(q_n, p_n, s_n^{(\nu)}) \right),$$

and where $\tilde{W}_1$ and $\tilde{g}_1$ are the terms in W_1 and g_1 involving s_n, respectively. The proof that this fixed point iteration converges is straightforward. The key to bounding $hs_n^{(o)}$ lies in the fact that since $n \geq 3k-1$, the terms involving q_{n-i} ($i = 1, 2, \ldots, k$) do not reference any of the starting $O(1/h)$ values of q. The other important step is noting that $A_{32}(t_j) p_j = \psi_3(t_j)$ for $j = k, k+1, \ldots$ and $X_{32}(t_n) = -\alpha_o^2 \Gamma(t_n) A_{32}(t_n) + O(h)$. Therefore, since $\|\psi_3(t)\| = O(h^{k-1})$, it follows that $\| \sum_{i=1}^{k} \gamma_i X_{32}(t_n) p_{n-i}/h \| = O(h^{k-2})$. The induction assumption (A.30) is therefore true when $n = 3k-1$, and it follows when $n = 3k, \ldots, 4k-2$ in a similar fashion. Now that the induction assumption has been verified for the first few steps, the induction and fixed point argument presented earlier for $n \geq 4k-1$ is complete.

Remark. We have shown the numerical solution $\mathbf{u}_n = (u_n, u_{n-1}, \ldots, u_{n-k+1})$ is globally $O(h^k)$ accurate for all $n \geq 3k-1$. Equivalently, after $k+1$ steps from t_{k-1}, the BDF method produces a numerical solution $\mathbf{u}_{2k}$ at t_{2k} which has accuracy $O(h^k)$. The numerical solution expressed in *long* vector notation first has $O(h^k)$ accuracy in all components when $n = 3k-1$.

Convergence Proof for k = 1 (Theorem 2).

While the convergence proof for $k = 1$ is very similar to the one for $k \geq 2$ just presented and thus will not be given in detail, there are some additional difficulties that deserve to be noted. In some respect, these difficulties tend to make the convergence proof for the Backward Euler formula even more interesting. The primary difficulty starts with the fact that $\|h^2 s_1\|$ may not be bounded, so $\psi_1(t_1)$ may not be bounded as $h \to 0$. Therefore, the proof must be modified to avoid the use of the induction assumption (A.29). Instead, the proof can be altered to use the facts that $\|X_{11}(t_1)\psi_1(t_1)\| \leq \Delta_{11}$ and $\|X_{21}(t_1)\psi_1(t_1)\| \leq \Delta_{21} h$ for some constants Δ_{11} and Δ_{21} which depend only on η_o and η_1^* to $O(h)$ accuracy. The fixed point argument must also be modified to iterate for (p_n, q_n, s_n) instead of (q_n, p_n, hs_n). As a result, the smoothness of the leading error terms follows immediately from the convergence proof for $k = 1$ and $n \geq 3$. The starting iterates for $q_n^{(o)}$ and $s_n^{(o)}$ in the fixed point iteration must be altered to include some terms which are now $O(1)$:

$$\begin{aligned} q_n^{(o)} &= X_{11}(t_n) \left(q_{n-1} + \tilde{\psi}_1(t_n) \right) + X_{12}(t_n) \left(p_{n-1} + \tilde{\psi}_2(t_n) \right) + X_{13}(t_n) \left(-Q_3(t_n) - hW_3(p_n^{(o)}) \right), \\ s_n^{(o)} &= \frac{1}{h^2} \Big[X_{31}(t_n) \left(q_{n-1} + \tilde{\psi}_1(t_n) \right) + X_{32}(t_n) \left(p_{n-1} + \tilde{\psi}_2(t_n) + h^2 W_2(q_n^{(o)}, p_n^{(o)}) \right) \\ &\quad + X_{33}(t_n) \left(-Q_3(t_n) - hW_3 \left(p_n^{(o)} - hX_{23}(t_n) W_3(p_n^{(o)}) \right) - h^2 g_3(p_n^{(o)}) \right) \Big]. \end{aligned}$$

Note that the initial iterates must be chosen in the order $p_n^{(o)}$, $q_n^{(o)}$, and then $s_n^{(o)}$, and the definitions of Z_1, Z_2, Z_3 must be modified accordingly. As before, the main difficulty in proving the fixed point iteration converges is to show the initial iterates are uniformly bounded. This proof requires numerous cancellation properties of matrix products and sums of the X_{ij} matrices, most of which can be verified by straightforward (although tedious) algebraic manipulation and expansion by Taylor series.

As in the proof for $k \geq 2$, the expressions for the initial iterates are rewritten as follows:

$$\begin{aligned} \begin{bmatrix} q_n^{(o)} \\ p_n^{(o)} \end{bmatrix} &= \prod_{i=1}^{n} X_i \begin{bmatrix} q_o \\ p_o \end{bmatrix} + \sum_{i=1}^{n-1} \left(\prod_{j=i}^{n} X_j \begin{bmatrix} \psi_1(t_i) \\ \psi_2(t_i) \end{bmatrix} + \prod_{j=i+1}^{n} X_j \begin{bmatrix} X_{13}(t_i) \\ X_{23}(t_i) \end{bmatrix} \psi_3(t_i) \right) \\ &\quad + X_n \begin{bmatrix} \tilde{\psi}_1(t_n) \\ \tilde{\psi}_2(t_n) \end{bmatrix} - \begin{bmatrix} X_{13}(t_n) \\ X_{23}(t_n) \end{bmatrix} Q_3(t_n) - \begin{bmatrix} hX_{13}(t_n) W_3(p_n^{(o)}) \\ 0 \end{bmatrix}, \\ s_n^{(o)} &= \frac{1}{h^2} [X_{31}(t_n),\ X_{32}(t_n)] \Bigg(\prod_{i=1}^{n-1} X_i \begin{bmatrix} q_o \\ p_o \end{bmatrix} + \sum_{i=1}^{n-1} \prod_{j=i}^{n-1} X_j \begin{bmatrix} \psi_1(t_i) \\ \psi_2(t_i) \end{bmatrix} \\ &\quad + \sum_{i=1}^{n-2} \prod_{j=i+1}^{n-1} X_j \begin{bmatrix} X_{13}(t_i) \\ X_{23}(t_i) \end{bmatrix} \psi_3(t_i) \\ &\quad + \begin{bmatrix} X_{13}(t_{n-1}) \\ X_{23}(t_{n-1}) \end{bmatrix} \psi_3(t_{n-1}) + \begin{bmatrix} \tilde{\psi}_1(t_n) \\ \tilde{\psi}_2(t_n) + h^2 W_2(q_n^{(o)}, p_n^{(o)}) \end{bmatrix} \Bigg) \\ &\quad + \frac{1}{h^2} X_{33}(t_n) \left(-Q_3(t_n) - hW_3 \left(p_n^{(o)} - hX_{23}(t_n) W_3(p_n^{(o)}) \right) - h^2 g_3(p_n^{(o)}) \right). \end{aligned} \quad (A.49)$$

These starting iterates can then be bounded uniformly for all $n \geq 3$. This bounding process requires a careful, but straightforward examination of all the terms, in particular those terms involving q_o and $\psi_1(t_1)$. Terms involving q_o can generally be bounded with the help of (A.7). Note that the terms with $\psi_1(t_1)$ are actually either $X_{11}(t_1)\psi_1(t_1)$ or $X \quad (t_1)\psi_1(t_1)$, which can be bounded. In the bounding of $q_n^{(o)}$ it is necessary to use the fact that the $O(1/h)$ terms in

$$X_{13}(t_n)\left(Q_3(t_n) + hW_3(p_n^{(o)})\right) - \left(X_{11}(t_n)X_{13}(t_{n-1}) + X_{12}(t_n)X_{23}(t_{n-1})\right)\psi_3(t_{n-1})$$

cancel because $Q_3(t)$ is smooth and

$$X_{11}(t_n)X_{13}(t_{n-1}) + X_{12}(t_n)X_{23}(t_{n-1}) = -\frac{\alpha_o}{h}A_{13}(t_n)\Gamma(t_n) + O(1).$$

The bounding of $s_n^{(o)}$ is even more complicated, requiring smoothness of $\tilde{\psi}_2(t_n)$ and $\tilde{\psi}_3(t_n)$ and matrix product properties such as

$$\|[X_{31}(t_n),\ X_{32}(t_n)]X_{n-1}X_{n-2}\| \leq Nh^2,$$
$$\|[X_{31}(t_n),\ X_{32}(t_n)]X_{n-1}X_{n-2}X_{n-3}\| \leq Nh^2$$

for some constant N, and

$$\|[X_{31}(t_3),\ X_{32}(t_3)]X_2X_1 \begin{bmatrix} q_o \\ p_o \end{bmatrix} \| \leq \beta_1 h^2$$

for some constant β_1 dependent on η_o. The smoothness of $\tilde{\psi}_2(t)$ can be used to obtain a cancellation of the $O(1/h)$ terms between the $\psi_2(t_{n-1})$ and the $\tilde{\psi}_2(t_n)$ term in (A.49). In particular, there exists a constant β_2 which depends on η_1^*, η_2^*, η_1, η_2 and $\mathcal{E}$, but not on η_3 to $O(h)$ accuracy such that for $n \geq 2$

$$\|\left(X_{31}(t_n)X_{12}(t_{n-1}) + X_{32}(t_n)X_{22}(t_{n-1})\right)\psi_2(t_{n-1}) + X_{32}(t_n)\left(\tilde{\psi}_2(t_n) + h^2W_2(q_n^{(o)}, p_n^{(o)})\right)\| \leq \beta_2 h^2.$$

The bounding of $s_n^{(o)}$ will also require a cancellation of terms involving $\psi_3(t_i)$ at $i = n$, $n-1$, and $n-2$:

$$\|\frac{1}{h^2}X_{33}(t_n)\left[-Q_3(t_n) - hW_3\left(p_n^{(o)} - hX_{23}(t_n)W_3(p_n^{(o)})\right) - hg_3(p_n^{(o)})\right]$$
$$+\frac{1}{h^2}[X_{31}(t_n),\ X_{32}(t_n)]\left(\begin{bmatrix} X_{13}(t_{n-1}) \\ X_{23}(t_{n-1}) \end{bmatrix}\psi_3(t_{n-1}) + X_{n-1}\begin{bmatrix} X_{13}(t_{n-2}) \\ X_{23}(t_{n-2}) \end{bmatrix}\psi_3(t_{n-2})\right)\| \quad < \quad \infty.$$

Matrix product properties such as

$$\|[X_{31}(t_n),\ X_{32}(t_n)]X_{n-1}X_{n-2}\begin{bmatrix} X_{13}(t_{n-3}) \\ X_{23}(t_{n-3}) \end{bmatrix}\| = O(h^2)$$

will also be utilized. For more details of the bounding of the starting iterates, see [3]. It should be clear from the examples given above that the convergence proof for $k = 1$ requires additional matrix product and cancellation of terms properties in order to resolve the difficulties associated with the behavior of s_1.

The induction assumptions (A.28) and (A.30) must also be verified before the convergence argument is complete. Again a different approach must be taken than the one used when $k \geq 2$. In particular, the argument must not require any knowledge of the behavior of $h^2 s_1$, but instead must rely on the behavior of

$$X_{11}(t_i)\psi_1(t_i) \quad = \quad X_{11}(t_i)\left(I_p - hA_{11}(t_i)\right)q_i - hX_{11}(t_i)A_{12}(t_i)p_i - X_{11}(t_i)q_{i-1},$$
$$X_{21}(t_i)\psi_1(t_i) \quad = \quad X_{21}(t_i)\left(I_p - hA_{11}(t_i)\right)q_i - hX_{21}(t_i)A_{12}(t_i)p_i - X_{21}(t_i)q_{i-1}.$$

These difficulties can be resolved fully, allowing the starting induction assumptions (A.28) and (A.30) to be verified. Hence the Backward Euler solution does converge to a solution of the index-three system as $n \to \infty$, $h \to 0$ for $t_n \in [t_2, t_o + T]$.

Note that we have shown $\|hs_n\| \leq \eta_3^{**}$ for $n = 2$ and $\|s_n\| \leq \eta_3 e^{nh\mathcal{E}}$ for all $n \geq 3$. Therefore, convergence of the numerical solution for the algebraic variables u is not obtained until the second Backward Euler step (i.e., $n = 2$). The fact that we can not bound either s_n or hs_n on the first step is not a deficiency in the proof; convergence is not in general obtained on the first step, and an $O(1)$ error in the numerical solution will be apparent in the algebraic variables. This $O(1)$ error will be present even if the initial values used to start the numerical method are consistent (i.e., contain no errors) with a true solution of the system.

While we have proven the numerical solution converges to a true solution of the system, we have not proven it converges to a solution consistent with the given initial values for the algebraic variables. Because of the $O(1)$ error in the algebraic variables at the end of the first step, and since the convergence analysis does not depend on past values of u, but only on the current time level (i.e., on u_n at t_n), it is possible for the numerical solution to jump to a different solution curve. This behavior has been observed in practice [2]. It will not occur if the algebraic variables appear only linearly in the system.

Remark. If the initial values do not satisfy (2.4), then convergence is not obtained until $n \geq 3$. At the end of the first step, the consistency condition (2.4) is satisfied for $t = t_1$. The method now requires two more steps, starting from *consistent* initial values, before convergence is obtained in the algebraic variables.

Proof of Corollary 2.

For technical simplicity, we first prove this corollary under the assumption that the difference equations are all solved exactly at each step. We know from the difference equations (A.8) that

$$A_{32}(t_n)p_n = \psi_3(t_n) \qquad \text{for} \quad n \geq k. \tag{A.50}$$

Next we use the difference equations (A.8) to derive a relation for s_n. Namely, the first block difference equation is

$$\left(I_p - \frac{h}{\alpha_o}A_{11}(t_n)\right) q_n - \frac{h}{\alpha_o}A_{12}(t_n)p_n - \frac{h}{\alpha_o}A_{13}(t_n)s_n = \sum_{i=1}^{k}\gamma_i q_{n-i} + \psi_1(t_n). \tag{A.51}$$

Multiply (A.51) by $A_{32}(t_n)A_{21}(t_n)$ and invert the coefficient matrix of s_n to obtain

$$s_n = -\frac{\alpha_o}{h}\Gamma(t_n)A_{32}(t_n)A_{21}(t_n)\left(\sum_{i=1}^{k}\gamma_i q_{n-i} + \psi_1(t_n) - \left(I_p - \frac{h}{\alpha_o}A_{11}(t_n)\right) q_n + \frac{h}{\alpha_o}A_{12}(t_n)p_n\right). \tag{A.52}$$

For bounded q_n, p_n, and hs_n we know $\psi_1(t_n)$ is order $O(h)$. Thus, in order to prove s_n is bounded, we need only to prove the terms

$$-\frac{\alpha_o}{h}\Gamma(t_n)A_{32}(t_n)A_{21}(t_n)\left(\sum_{i=1}^{k}\gamma_i q_{n-i} - q_n\right) \tag{A.53}$$

are $O(1)$. The second block difference equation in (A.8) is

$$-\frac{h}{\alpha_o}A_{21}(t_n)q_n + \left(I_q - \frac{h}{\alpha_o}A_{22}(t_n)\right) p_n = \sum_{i=1}^{k}\gamma_i p_{n-i} + \psi_2(t_n).$$

Then,

$$A_{21}(t_{n-i})q_{n-i} = \frac{\alpha_o}{h}\left(\left(I_q - \frac{h}{\alpha_o}A_{22}(t_{n-i})\right) p_{n-i} - \sum_{j=1}^{k}\gamma_j p_{n-i-j} - \psi_2(t_{n-i})\right)$$

for $i = 1, 2, \ldots, k$ and $n \geq 2k$, and hence

$$\begin{aligned} A_{21}(t_n)\left(q_n - \sum_{i=1}^{k}\gamma_i q_{n-i}\right) &= \left(\frac{\alpha_o}{h}I_q - A_{22}(t_n)\right)\left(p_n - \sum_{i=1}^{k}\gamma_i p_{n-i}\right) - \\ &\frac{\alpha_o}{h}\left(\sum_{j=1}^{k}\gamma_j(p_{n-j} - \sum_{i=1}^{k}\gamma_i p_{n-i-j}) + \left(\psi_2(t_n) - \sum_{i=1}^{k}\gamma_i\psi_2(t_{n-i})\right)\right) + O(h) \end{aligned} \tag{A.54}$$

for $n \geq 2k$. Multiply (A.54) by $A_{32}(t_n)$ and rearrange the terms to obtain

$$\begin{aligned} A_{32}(t_n)A_{21}(t_n)\left(q_n - \sum_{i=1}^{k}\gamma_i q_{n-i}\right) &= \frac{\alpha_o}{h}A_{32}(t_n)\left(p_n - 2\sum_{i=1}^{k}\gamma_i p_{n-i} + \sum_{j=1}^{k}\gamma_j\sum_{i=1}^{k}\gamma_i p_{n-i-j}\right) \\ &- A_{32}(t_n)\left[A_{22}(t_n)\left(p_n - \sum_{i=1}^{k}\gamma_i p_{n-i}\right) + \frac{\alpha_o}{h}\left(\psi_2(t_n) - \sum_{i=1}^{k}\gamma_i\psi_2(t_{n-i})\right)\right] + O(h). \end{aligned} \tag{A.55}$$

From equation (A.25) it follows that $p_n = \sum_{i=1}^{k}\gamma_i X_{22}(t_n)p_{n-i} + O(h)$ for $n \geq 2k$ and for q_n, p_n, hs_n bounded ($k \geq 2$). Then,

$$\begin{aligned} p_n - \sum_{i=1}^{k}\gamma_i p_{n-i} &= \sum_{i=1}^{k} -\gamma_i A_{21}(t_n)A_{13}(t_n)\Gamma(t_n)A_{32}(t_n)p_{n-i} + O(h) \\ &= \sum_{i=1}^{k} -\gamma_i A_{21}(t_n)A_{13}(t_n)\Gamma(t_n)\psi_3(t_{n-i}) + O(h) \\ &= O(h^{k-1}) + O(h). \end{aligned} \tag{A.56}$$

By definition of $\psi_2(t)$ in (A.10), $\psi_2(t_n) = h\Psi(t_n) + O(h^2)$ where $\Psi(t_n)$ is a smooth function, namely $\Psi(t) = -\alpha^* w^{(k+2)}(t)/\alpha_o$. Therefore,

$$\begin{aligned} \psi_2(t_n) - \sum_{i=1}^{k}\gamma_i\psi_2(t_{n-i}) &= h\left(\Psi(t_n) - \sum_{i=1}^{k}\gamma_i\Psi(t_{n-i})\right) + O(h^2) \\ &= h\Psi(t_n)(1 - \sum_{i=1}^{k}\gamma_i) + O(h^2) = O(h^2), \end{aligned} \tag{A.57}$$

since $\sum_{i=1}^{k}\gamma_i = 1$. We are trying to prove the terms given in (A.53) are $O(1)$, or equivalently those given by (A.55) are $O(h)$. After applying (A.56) and (A.57), there are still some terms left in (A.55) which must be shown to be $O(h)$, namely

$$\frac{\alpha_o}{h}A_{32}(t_n)\left(p_n - 2\sum_{i=1}^{k}\gamma_i p_{n-i} + \sum_{j=1}^{k}\gamma_j\sum_{i=1}^{k}\gamma_i p_{n-i-j}\right). \tag{A.58}$$

Using the consistency relations (A.50), the terms in (A.58) can be rewritten as

$$\begin{aligned} &\frac{\alpha_o}{h}\left(\psi_3(t_n) - 2\sum_{i=1}^{k}\gamma_i\psi_3(t_{n-i}) + \sum_{j=1}^{k}\gamma_j\sum_{i=1}^{k}\gamma_i\psi_3(t_{n-i-j}) - \right. \\ &\left. hA'_{32}(t_n)\left(2\sum_{i=1}^{k}i\gamma_i p_{n-i} - \sum_{j=1}^{k}\gamma_j\sum_{i=1}^{k}\gamma_i(i+j)p_{n-i-j}\right)\right) + O(h). \end{aligned} \tag{A.59}$$

Note

$$\psi_3(t_n) - \sum_{i=1}^{k}\gamma_i\psi_3(t_{n-i}) = h^{k-1}\left(-Q_3(t_n) + \sum_{i=1}^{k}\gamma_i Q_3(t_{n-i})\right) + O(h^k) = O(h^k)$$

and

$$-\sum_{i=1}^{k}\gamma_i\psi_3(t_{n-i}) + \sum_{j=1}^{k}\gamma_j\sum_{i=1}^{k}\gamma_i\psi_3(t_{n-i-j}) = -\sum_{i=1}^{k}\gamma_i\left(\psi_3(t_{n-i}) - \sum_{j=1}^{k}\gamma_j\psi_3(t_{n-i-j})\right) = O(h^k),$$

since $Q_3(t)$ is a smooth function of t. Except for the terms involving $A'_{32}(t)$ in (A.59), we have shown all the terms to be $O(h)$:

$$\begin{aligned} -\alpha_o A'_{32}(t_n)\left(2\sum_{i=1}^{k}i\gamma_i p_{n-i} - \sum_{j=1}^{k}\gamma_j\sum_{i=1}^{k}\gamma_i(i+j)p_{n-i-j}\right) &= \\ -\alpha_o A'_{32}(t_n)\left(\sum_{i=1}^{k}i\gamma_i(p_{n-i} - \sum_{j=1}^{k}\gamma_j p_{n-i-j}) + \sum_{j=1}^{k}j\gamma_j(p_{n-j} - \sum_{i=1}^{k}\gamma_i p_{n-i-j})\right) &= O(h) \end{aligned}$$

from (A.56) providing $n \geq 3k$. Hence, we have shown the terms in (A.55) are $O(h)$, and equivalently those in (A.53) are $O(1)$. It follows from (A.52) that s_n is bounded. Since the bounds for q_n, p_n, and hs_n are uniform for sufficiently large n, we can bound s_n uniformly. Therefore, the principal leading error term in u_n is $h^k e(t_n)$ for $n \geq 3k$.

This proof can be generalized in a straightforward way to include residual terms representing the effect of not solving the difference equations exactly. Specifically, if one introduces residual terms of order $O(h^{k+2})$, $O(h^{k+3})$, and $O(h^{k+3})$, respectively in the difference equations corresponding to v', w', and the algebraic equations, then one can modify the definitions of ψ_1, ψ_2, and ψ_3 to include corresponding terms of order $O(h)$, $O(h^2)$, and $O(h^2)$. The proof then follows as before. Moreover, if the starting values $\mathbf{y}_{k-1}$ also satisfy the algebraic equations to $O(h^{k+3})$ accuracy, then the smoothness of the leading error term $h^k e(t)$ for u follows for $n \geq 2k$ and $k \geq 2$. To establish this result for $k = 1$, a more careful analysis utilizing the smoothness of $\psi_3(t)$ is required.

Remark. If the result in Corollary 2 were to hold for the numerical solution expressed in *long* vector notation, one would have to require that $n \geq 4k - 1$. Specifically, the principal leading error term in $\mathbf{u}_n$ is $h^k(e(t_n), e(t_{n-1}), \ldots, e(t_{n-k+1}))$ for all $n \geq 4k - 1$.