Appendices

A Proof of Proposition 1

Recall that $h = t_j - t_{j-1}$ for all j. Straightforward substitutions into the definitions give that $\phi_i \equiv \phi(0) = (0, 1, 0, \ldots, 0)$, $\phi_{i-1} \equiv \phi(-h) = (0, 0, 1, \ldots, 0)$ etc. and hence $\phi_{i-p}^T \phi_{i-q} = \delta_{pq}$, for all $0 \leq p, q \leq s - 1$. Furthermore $\Phi_i \equiv \Phi(0) = (1, 0, 0, \ldots, 0)$ since every component of $\Phi(\omega)$ bar the first is a polynomial of degree s with a factor ω. Finally

$$\Phi_{i+1} \equiv \Phi(h) = \left(1 \quad \int_0^h \ell_0^{0:s-1}(\omega) \, d\omega \quad \cdots \quad \int_0^h \ell_{s-1}^{0:s-1}(\omega) \, d\omega \right)$$

Now by (10) and the standard formulae for Gaussian conditioning, we have

$$\mathbb{E}[y_{i+1}|y_i, f_{i-s+1:i}] =$$

$$\begin{pmatrix} \Phi_{i+1}^T \Phi_i \\ \Phi_{i+1}^T \phi_i \\ \vdots \\ \Phi_{i+1}^T \phi_{i-s+1} \end{pmatrix}^T \underbrace{\begin{pmatrix} \Phi_i^T \Phi_i & \Phi_i^T \phi_i & \cdots & \Phi_i^T \phi_{i-s+1} \\ \phi_i^T \Phi_i & \phi_i^T \phi_i & \cdots & \phi_i^T \phi_{i-s+1} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{i-s+1}^T \Phi_i & \phi_{i-s+1}^T \phi_{i+1} & \cdots & \phi_{i-s+1}^T \phi_{i-s+1} \end{pmatrix}^{-1}}_{\mathbb{I}_{s+1}^{-1}} \begin{pmatrix} y_i \\ f_i \\ \vdots \\ f_{i-s+1} \end{pmatrix}$$

$$= (\Phi_{i+1}^T \Phi_i) y_i + \sum_{k=0}^{s-1} (\Phi_{i+1}^T \phi_{i-k}) f_{i-k}$$

$$= y_i + \sum_{k=0}^{s-1} [\Phi_{i+1}]_{k+2} \cdot f_{i-k} \qquad \begin{pmatrix} \text{where } [\Phi_{i+1}]_{k+2} \text{ denotes the} \\ (k+2)\text{th component of } \Phi_{i+1} \end{pmatrix}$$

$$= y_i + \sum_{k=0}^{s-1} \left[\int_0^h \ell_k^{0:s-1}(\omega) \, d\omega \right] \cdot f_{i-k}$$

$$= y_i + h \sum_{k=0}^{s-1} c_{k,s} f_{i-k} \qquad \text{since } \int_0^h \ell_k^{0:s-1}(\omega) \, d\omega = h c_{k,s}$$

which is equal to the s-step Adams-Bashforth predictor defined by (6) and (7). Next we write

$$\text{Var}[y_{i+1}|y_i, f_{i-s+1:i}] = \Phi_{i+1}^T \Phi_{i+1} - \begin{pmatrix} \Phi_{i+1}^T \Phi_i \\ \Phi_{i+1}^T \phi_i \\ \vdots \\ \Phi_{i+1}^T \phi_{i-s+1} \end{pmatrix}^T \mathbb{I}_{s+1}^{-1} \begin{pmatrix} \Phi_i^T \Phi_{i+1} \\ \phi_i^T \Phi_{i+1} \\ \vdots \\ \phi_{i-s+1}^T \Phi_{i+1} \end{pmatrix}$$

$$= \Phi_{i+1}^T \Phi_{i+1} - \begin{pmatrix} 1 \\ [\Phi_{i+1}]_2 \\ \vdots \\ [\Phi_{i+1}]_{s+1} \end{pmatrix}^T \begin{pmatrix} 1 \\ [\Phi_{i+1}]_2 \\ \vdots \\ [\Phi_{i+1}]_{s+1} \end{pmatrix}$$

$$= \Phi_{i+1}^T \Phi_{i+1} - \Phi_{i+1}^T \Phi_{i+1}$$

$$= 0$$

and the proposition follows.

B Proof of Proposition 2

We follow the same reasoning as in Proposition 1. Since the additional basis function at the end of ϕ_{i-k}^+ is clearly zero at for all $0 \leq k \leq s - 1$, each inner product of the form $\phi^{+T} \phi^+$, $\Phi^{+T} \phi^+$ and $\phi^{+T} \Phi^+$ is equal to the corresponding inner product $\phi^T \phi$, $\Phi^T \phi$ and $\phi^T \Phi$ as no additional

contribution from the new extended basis arises. It therefore suffices to check only the terms of the form $\Phi^{+T}\Phi$.

Integrating the additional basis function gives a polynomial of degree $s+1$ with a constant factor ω. Evaluating this at $t_i = 0$ means that the additional term is also 0 in Φ_i. Therefore $\Phi_{i+1}^{+T}\Phi_i^+ = \Phi_{i+1}^T\Phi_i$ and $\Phi_i^{+T}\Phi_i^+ = \Phi_i^T\Phi_i$. It follows that the expression for $\mathbb{E}[y_{i+1}|y_i, f_{i-s+1:i}]$ is exactly the same as when using the unaugmented basis function set.

The argument in the previous paragraph means we can immediately write down that

$$\mathrm{Var}[y_{i+1}|y_i, f_{i-s+1:i}] = \Phi_{i+1}^{+T}\Phi_{i+1}^+ - \Phi_{i+1}^T\Phi_{i+1}$$

Since the first $s+1$ components of Φ_{i+1}^{+T} are equal to the $s+1$ components of Φ_{i+1}^T, this expression reduces to the contribution of the augmented basis element. Therefore

$$\mathrm{Var}[y_{i+1}|y_i, f_{i-s+1:i}] = \left(\alpha h^s \int_0^h \ell_{-1}^{-1:s-1}(\omega)\,\mathrm{d}\omega\right)^2$$

$$= \left(\alpha h^{s+1}\beta_{-1,s+1}^{AM}\right)^2$$

The Adams-Moulton coefficient $\beta_{-1,s+1}^{AM}$ is equal to the local truncation error constant for the s-step Adams-Bashforth method [12] and the proposition follows.

C Extension to Adams-Moulton

We collect here the straightforward modifications required to the constructions in the main paper to produce implicit Adams-Moulton methods instead of explicit Adams-Bashforth versions.

The telescopic decomposition (5) becomes

$$p(y_{1:N}, f_{0:N}|y_0) = \prod_{i=0}^N p(f_i|y_i) \times \prod_{i=0}^{N-1} p(y_{i+1}|y_i, f_{\max(0,i-s+1):i+1}) \tag{14}$$

where it is particularly to be noted that f_N is no longer superfluous.

The Lagrange interpolation resulting in the the Adams-Moulton method is

$$Q_i(\omega) = \sum_{j=-1}^{s-1} \ell_j^{-1:s-1}(\omega)f_{i-j} \qquad \ell_j^{-1:s-1}(\omega) = \prod_{\substack{k=-1 \\ k \neq j}}^{s-1} \frac{\omega - t_{i-k}}{t_{i-j} - t_{i-k}}, \tag{15}$$

the analogous vectors of basis polynomials to (8) and (9) are

$$\psi(\omega) = \begin{pmatrix} 0 & \ell_{-1}^{-1:s-1}(\omega) & \ell_0^{-1:s-1}(\omega) & \ell_1^{-1:s-1}(\omega) & \ldots & \ell_{s-1}^{-1:s-1}(\omega) \end{pmatrix}^T \tag{16}$$

$$\Psi(\omega) = \int \psi(\omega)\,\mathrm{d}\omega = \begin{pmatrix} 1 & \int \ell_{-1}^{-1:s-1}(\omega)\,\mathrm{d}\omega & \ldots & \int \ell_{s-1}^{-1:s-1}(\omega)\,\mathrm{d}\omega \end{pmatrix}^T \tag{17}$$

and the iterator is defined by

$$y_{i+1} - y_i \approx \int_{t_i}^{t_{i+1}} Q_i(\omega)\,\mathrm{d}\omega = h \sum_{j=-1}^{s-1} \beta_{j,s+1}^{AM} f_{i-j} \tag{18}$$

with $\beta_{j,s+1}^{AM} \equiv h^{-1} \int_0^h \ell_j^{-1:s-1}(\omega)\,\mathrm{d}\omega$ are the Adams-Moulton coefficients.

The Gaussian process prior resulting in AM is

$$
\begin{pmatrix} y_{i+1} \\ y_i \\ f_{i+1} \\ f_i \\ f_{i-1} \\ \vdots \\ f_{i-s+1} \end{pmatrix} = \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \Psi_{i+1}^T \Psi_{i+1} & \Psi_{i+1}^T \Psi_i & \Psi_{i+1}^T \psi_{i+1} & \cdots & \Psi_{i+1}^T \psi_{i-s+1} \\ \Psi_i^T \Psi_{i+1} & \Psi_i^T \Psi_i & \Psi_i^T \psi_{i+1} & \cdots & \Psi_i^T \psi_{i-s+1} \\ \psi_{i+1}^T \Psi_{i+1} & \psi_{i+1}^T \Psi_i & \psi_{i+1}^T \psi_{i+1} & \cdots & \psi_{i+1}^T \psi_{i-s+1} \\ \psi_i^T \Psi_{i+1} & \psi_i^T \Psi_i & \psi_i^T \psi_{i+1} & \cdots & \psi_i^T \psi_{i-s+1} \\ \psi_{i-1}^T \Psi_{i+1} & \psi_{i-1}^T \Psi_i & \psi_{i-1}^T \psi_{i+1} & \cdots & \psi_{i-1}^T \psi_{i-s+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_{i-s+1}^T \Psi_{i+1} & \psi_{i-s+1}^T \Psi_i & \psi_{i-s+1}^T \psi_{i+1} & \cdots & \psi_{i-s+1}^T \psi_{i-s+1} \end{pmatrix} \right] \tag{19}
$$

D Adams-Moulton integrator with $s = 4$

The conditional distribution of interest is $p(y_{i+1}|y_i, f_{i+1}, f_i, f_{i-1}, f_{i-2}) \equiv p(y_{i+1}|y_i, f_{i+1:i-2})$. In the deterministic case the vectors of basis functions become

$$
\psi(\omega)_{s=4} = \begin{pmatrix} 0 & \frac{\omega(\omega+h)(\omega+2h)}{-6h^3} & \frac{(\omega-h)(\omega+h)(\omega+2h)}{2h^3} & \frac{\omega(\omega-h)(\omega+2h)}{-2h^3} & \frac{\omega(\omega-h)(\omega+h)}{6h^3} \end{pmatrix} \tag{20}
$$

$$
\Psi(\omega)_{s=4} = \begin{pmatrix} 1 & \frac{\omega^2(2h+\omega)^2}{24h^3} & \frac{\omega(3\omega^3+8h\omega^2-6h^2\omega-2fh^3)}{-24h^3} & \frac{\omega^2(3\omega^2+4h\omega-12h^2)}{24h^3} & \frac{\omega^2(\omega^2-2h^2)}{-24h^3} \end{pmatrix} \tag{21}
$$

and the resulting calculations give

$$
\mathbb{E}(y_{i+1}|y_i, f_{i+1:i-2}) = y_i + h\left(\frac{3}{8} f_{i+1} + \frac{19}{24} f_i - \frac{5}{24} f_{i-1} + \frac{1}{24} f_{i-2} \right)
$$

$$
\mathrm{Var}(y_{i+1}|y_i, f_{i+1:i-2}) = 0
$$

The probabilistic version is

$$
\psi^+(\omega)_{s=4} = \begin{pmatrix} \cdots \psi(\omega)_{s=4} \cdots & \frac{\alpha\omega(\omega-h)(\omega+h)(\omega+2h)}{24} \end{pmatrix} \tag{22}
$$

$$
\Psi^+(\omega)_{s=4} = \begin{pmatrix} \cdots \Psi(\omega)_{s=4} \cdots & \frac{\alpha\omega^2(6\omega^3+15\omega^2h-10\omega h^2-30h^3)}{720} \end{pmatrix} \tag{23}
$$

and further calculation shows that

$$
\mathbb{E}(y_{i+1}|y_i, f_{i-1:i+2}) = y_i + h\left(\frac{3}{8} f_{i+1} + \frac{19}{24} f_i - \frac{5}{24} f_{i-1} + \frac{1}{24} f_{i-2} \right) \tag{24}
$$

$$
\mathrm{Var}[y_{i+1}|y_i, f_{i+1:i-2}] = \left(\frac{19h^5\alpha}{720} \right)^2 \tag{25}
$$

Remark

Proofs analogous to those of Propositions 1 and 2, for the Adams-Moulton case, follow the same line of reasoning as for the Adams-Bashforth case.

E Expansion of backward difference coefficient approximation for $s = 3$

From (13), we have for $s = 3$

$$
\begin{aligned}
y_{i+1} &= y_i + h\left(\frac{23}{12} f_i - \frac{4}{3} f_{i-1} + \frac{5}{12} f_{i-2} \right) - \frac{3}{8} h^4 y^{(4)}(t_{i+1}) + O(h^5) \\
&= y_i + h\left(\frac{23}{12} f_i - \frac{4}{3} f_{i-1} + \frac{5}{12} f_{i-2} \right) - \frac{3}{8} h^4 f'''(t_{i+1}) + O(h^5) \\
&= y_i + h\left(\frac{23}{12} f_i - \frac{4}{3} f_{i-1} + \frac{5}{12} f_{i-2} \right) - \frac{3}{8} h^4 \left[\frac{-f_i + 3f_{i-1} - 3f_{i-2} + f_{i-3}}{h^3} + O(h) \right] + O(h^5) \\
&= y_i + h\underbrace{\left(\frac{55}{24} f_i - \frac{59}{24} f_{i-1} + \frac{37}{24} f_{i-2} - \frac{3}{8} f_{i-3} \right)}_{\text{AB4}} + O(h^5)
\end{aligned}
$$

F   Proof of Theorem 3

Proposition 2 implies that our integrator can be written as

$$y_{i+1} = y_i + h \sum_{j=0}^{s-1} \beta_{j,s}^{AB} f(y_{i-j}, t_{i-j}) + \xi_i \tag{26}$$

where y_i denotes the numerical solution at iteration i, and $\xi_i \in \mathbb{R}^d$ is a Gaussian random variable satisfying $\mathbb{E}|\xi_i \xi_i^T| = Q h^{2s+2}$ for some fixed $d \times d$ matrix Q. We denote the true solution of the ODE (1) at iteration i by $Y_i \equiv y(t_i)$ and we have that

$$Y_{i+1} = Y_i + h \sum_{j=0}^{s-1} \beta_{j,s}^{AB} f(Y_{i-j}, t_{i-j}) + \tau_i \tag{27}$$

where by construction the local truncation error $\tau_i = O(h^{s+1})$. If we now subtract (26) from (27) and denote the accumulated error at iteration i by $E_i = Y_i - y_i$, we have

$$E_{i+1} = E_i + \Delta \phi_i + \tau_i - \xi_i$$

where

$$\Delta \phi_i := h \sum_{j=0}^{s-1} \beta_{j,s}^{AB} \Delta f_{i-j}, \qquad \Delta f_{i-j} := f(Y_{i-j}, t_{i-j}) - f(y_{i-j}, t_{i-j})$$

We will rearrange this s-step recursion to give an equivalent one-step recursion in an higher-dimensional space. In particular, using the trivial identities $E_{i-1} = E_{i-1}, \cdots, E_{i-s+1} = E_{i-s+1}$ we obtain

$$\underbrace{\begin{pmatrix} E_{i+1} \\ E_i \\ \vdots \\ E_{i-s+2} \end{pmatrix}}_{=: \, \mathcal{E}_{i+1}} = \underbrace{\begin{pmatrix} \mathbb{I}_d & 0 & \cdots & 0 \\ \mathbb{I}_d & 0 & \cdots & 0 \\ & \ddots & \ddots & \\ 0 & & \mathbb{I}_d & 0 \end{pmatrix}}_{=: \, \mathcal{A}} \underbrace{\begin{pmatrix} E_i \\ E_{i-1} \\ \vdots \\ E_{i-s+1} \end{pmatrix}}_{=: \, \mathcal{E}_i} + \underbrace{\begin{pmatrix} \Delta \phi_i \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{=: \, \Delta \Phi_i} + \underbrace{\begin{pmatrix} \tau_i \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{=: \, \mathcal{T}_i} - \underbrace{\begin{pmatrix} \xi_i \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{=: \, \Xi_i}$$

or in compact form,

$$\mathcal{E}_{i+1} = \mathcal{A}\mathcal{E}_i + \Delta \Phi_i + \mathcal{T}_i - \Xi_i, \qquad i = s-1, \ldots, N-1, \qquad N = T/h \tag{28}$$

For the subsequent calculations it will be necessary to find a scalar product inducing a matrix norm such that the norm of the matrix $\mathcal{A}$ is less or equal to 1. This is possible if the eigenvalues of the Frobenius matrix $\mathcal{A}$ lie inside the unit circle on the complex plane and are simple if their modulus is equal to 1. It is easy to show that the eigenvalues of $\mathcal{A}$ are roots of the characteristic polynomial associated with the deterministic integrator (7). Since we have assumed that the deterministic integrator is convergent, $\mathcal{A}$ does have the claimed property, since it is equivalent to the root condition in Dahlquist's equivalence theorem [12]. Thus there exists a non-singular matrix Λ with a block structure like $\mathcal{A}$ such that $||\Lambda^{-1}\mathcal{A}\,\Lambda||_2 \leq 1$. We can therefore choose a scalar product for $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{ds}$ as

$$\langle \mathcal{X}, \mathcal{Y} \rangle_* := \langle \Lambda^{-1} \mathcal{X}, \Lambda^{-1} \mathcal{Y} \rangle_2$$

and then have $|\cdot|_*$ and $||\cdot||_*$ as the induced vector and matrix norms respectively, with $||\mathcal{A}||_* = ||\Lambda^{-1}\mathcal{A}\,\Lambda||_2 \leq 1$ as required. We also have

$$\langle \mathcal{X}, \mathcal{Y} \rangle_* = \mathcal{X}^T \Lambda^{-T} \Lambda^{-1} \mathcal{Y} = \mathcal{X}^T \Lambda^* \mathcal{Y} \text{ with } \Lambda^* = \Lambda^{-T} \Lambda^{-1} = (\lambda_{ij}^* \otimes \mathbb{I}_d)_{1 \leq i,j \leq s} \tag{29}$$

Due to the equivalence of norms there exist constants $c^*, c_* > 0$ such that

$$|\mathcal{X}|_2^2 \leq c^* |\mathcal{X}|_*^2 \quad \text{and} \quad |\mathcal{X}|_*^2 \leq c_* |\mathcal{X}|_\infty^2 \quad \text{for all } \mathcal{X} \in \mathbb{R}^{ds}, \tag{30}$$

where $|\mathcal{X}|_2^2 = \sum_{j=1,\ldots,s} |x_j|^2$ and $|\mathcal{X}|_\infty = \max_{j=1,\ldots,s} |x_j|$ for $\mathcal{X} = (x_1^T, \cdots, x_s^T)^T$, $x_j \in \mathbb{R}^d$.

For the particular vectors $\tilde{\mathcal{X}} = (x^T, 0, \cdots, 0)^T$ and $\tilde{\mathcal{Y}} = (y^T, 0, \cdots, 0)^T$ with $\tilde{\mathcal{X}}, \tilde{\mathcal{Y}} \in \mathbb{R}^{ds}$ and $x, y \in \mathbb{R}^d$, one has

$$\langle \tilde{\mathcal{X}}, \tilde{\mathcal{Y}} \rangle_* = \lambda_{11}^* \langle x, y \rangle_2 = \lambda_{11}^* x^T y, \tag{31}$$

where λ_{11}^* is as in (29). Applying the norm $|\cdot|_*^2$ to (28) and taking expectations gives

$$
\begin{aligned}
\mathbb{E}|\mathcal{E}_{i+1}|_*^2 &= \mathbb{E}|\mathcal{A}\mathcal{E}_i + \Delta\Phi_i + \mathcal{T}_i - \Xi_i|_*^2 \\
&= \mathbb{E}|\mathcal{A}\mathcal{E}_i + \Delta\Phi_i + \mathcal{T}_i|_*^2 + O(h^{2s+2}) \\
&= \mathbb{E}|\mathcal{A}\mathcal{E}_i + \Delta\Phi_i|_*^2 + 2\mathbb{E}\langle h^{1/2}(\mathcal{A}\mathcal{E}_i + \Delta\Phi_i), \mathcal{T}_i h^{-1/2}\rangle_* + \mathbb{E}|\mathcal{T}_i|_*^2 + O(h^{2s+2}) \\
&= \mathbb{E}|\mathcal{A}\mathcal{E}_i + \Delta\Phi_i|_*^2 + 2\mathbb{E}\langle h^{1/2}(\mathcal{A}\mathcal{E}_i + \Delta\Phi_i), \mathcal{T}_i h^{-1/2}\rangle_* + O(h^{2s+2}) \qquad (32)
\end{aligned}
$$

We now consider the term $|\mathcal{A}\mathcal{E}_i + \Delta\Phi_i|_*^2$ and expand it as

$$
|\mathcal{A}\mathcal{E}_i + \Delta\Phi_i|_*^2 = \underbrace{|\mathcal{A}\mathcal{E}_i|_*^2}_{A} + \underbrace{|\Delta\Phi_i|_*^2}_{B} + \underbrace{2\langle \mathcal{A}\mathcal{E}_i, \Delta\Phi_i\rangle_*}_{C}
$$

For term A we immediately have $|\mathcal{A}\mathcal{E}_i|_*^2 \le |\mathcal{E}_i|_*^2$ by construction of the norm $|\cdot|_*^2$.

For term B we have that

$$
\begin{aligned}
|\Delta\Phi_i|_*^2 &= \lambda_{11}^*|\Delta\phi_i|^2 && \text{from (31)} \\
&= \lambda_{11}^*\left|h\sum_{j=0}^{s-1}\beta_{j,s}^{AB}\Delta f_{i-j}\right|^2 \\
&\le \lambda_{11}^* s h^2 \sum_{j=0}^{s-1}\left|\beta_{j,s}^{AB}\Delta f_{i-j}\right|^2 && \text{by Cauchy-Schwarz} \\
&\le \lambda_{11}^* s h^2 L_f^2 \sum_{j=0}^{s-1}(\beta_{j,s}^{AB})^2|E_{i-j}|^2 && \text{since } f \text{ is Lipschitz} \\
&\le \lambda_{11}^* s h^2 L_f^2 C_\beta^2 \sum_{j=0}^{s-1}|E_{i-j}|^2 && \text{where } C_\beta^2 = \max_{j=0,\dots,s-1}\beta_{j,s}^{AB} \\
&\le \lambda_{11}^* s h^2 L_f^2 C_\beta^2 c^* |\mathcal{E}_i|_*^2 && \text{from (30)} \\
&=: \Gamma^2 h^2 |\mathcal{E}_i|_*^2 && \text{where } \Gamma^2 = \lambda_{11}^* s L_f^2 C_\beta^2 c^*
\end{aligned}
$$

For term C we have $2\langle \mathcal{A}\mathcal{E}_i, \Delta\Phi_i\rangle_* \le 2|\mathcal{A}\mathcal{E}_i|_*|\Delta\Phi_i|_* \le 2\Gamma h|\mathcal{E}_i|_*^2$ and it follows that

$$
|\mathcal{A}\mathcal{E}_i + \Delta\Phi_i|_*^2 \le (1 + O(h))|\mathcal{E}_i|_*^2
$$

Then from (32) we have

$$
\begin{aligned}
\mathbb{E}|\mathcal{E}_{i+1}|_*^2 &= \mathbb{E}|\mathcal{A}\mathcal{E}_i + \Delta\Phi_i|_*^2 + 2\mathbb{E}\langle h^{1/2}(\mathcal{A}\mathcal{E}_i + \Delta\Phi_i), \mathcal{T}_i h^{-1/2}\rangle_* + O(h^{2s+2}) \\
&\le (1 + O(h))\mathbb{E}|\mathcal{E}_i|_*^2 + 2h\mathbb{E}|\mathcal{A}\mathcal{E}_i + \Delta\Phi_i|_*^2 + 2h^{-1}\mathbb{E}|\mathcal{T}_i|_*^2 + O(h^{2s+2}) \\
&\le (1 + O(h))\mathbb{E}|\mathcal{E}_i|_*^2 + O(h^{2s+1}) + O(h^{2s+2}) \qquad (33)
\end{aligned}
$$

Then by applying the Gronwall inequality we have (for different K in each line)

$$
\max_{0\le kh\le T}\mathbb{E}|\mathcal{E}_k|_*^2 \le K(T)h^{2s}
$$

and since $\mathcal{E}_k = (E_k, E_{k-1}, \cdots, E_{k-s+1})$ we conclude that

$$
\max_{0\le kh\le T}\mathbb{E}|E_k|^2 \le K(T)h^{2s}
$$

Note that in (33), the $O(h^{2s+2})$ term derived from the introduced perturbations ξ_i is of one higher order than the $O(h^{2s+1})$ term representing the truncation error in the deterministic solver. This observation implies that a noise vector satisfying $\mathbb{E}|\xi_i\xi_i^T| = Qh^{2s+1}$ would also give rise to an integrator of order s.

Remark

An analogous proof for the Adams-Moulton case follows with straightforward modifications.