Chapter 9. Scalar Initial Value Problem. (IVP)

Consider following (ODE)

$$\begin{array}{ll} \text{(DE)} & \left\{ \begin{array}{r} \dot{u}(t) + a(t)u(t) = f(t), \quad 0 < t \leq T \\ u(0) = u_0 \end{array} \right. \\ \text{(IV)} & \end{array}$$

where $f(t)$ is the source term and $\dot{u}(t) = \dfrac{du}{dt}$.

Here $a(t)$ is bounded and $a(t) \geq 0 \Longrightarrow$ a *parabolic problem*,

while $a(t) > 0 \Longrightarrow$ a *dissipative problem*.

Fundamental solution

Let $A(t) = \displaystyle\int_0^t a(s)ds$, (i.e. $A(t)$ is the *primitive function* of $a(t)$, with $A(0) = 0$), then

$$u(t) = u_0 \cdot e^{-A(t)} + \int_0^t e^{-(A(t)-A(s))} f(s)ds. \tag{1}$$

Proof: The integrating factor is $e^{\int_0^t a(s)ds} = e^{A(t)}$. Multiplying the (DE) by $e^{A(t)}$ and using $a(t) = \dot{A}(t)$, we get

$$\dot{u}(t)e^{A(t)} + \dot{A}(t)e^{A(t)}u(t) = e^{A(t)}f(t), \qquad \text{i.e.} \quad \frac{d}{dt}(u(t)e^{A(t)}) = e^{A(t)}f(t),$$

which integrating over $(0, t)$ gives that

$$\int_0^t \frac{d}{ds}(u(s)e^{A(s)})ds = \int_0^t e^{A(s)}f(s)ds \Longleftrightarrow u(t)e^{A(t)} - u(0)e^{A(0)} = \int_0^t e^{A(s)}f(s)ds.$$

Now since $A(0) = 0$ and $u(0) = u_0$ we get

$$u(t) = u_0 \cdot e^{-A(t)} + \int_0^t e^{-(A(t)-A(s))} f(s)ds.$$

and the proof is complete.

Stability estimates

Using the fundamental solution we can derive the following stability estimates:

(i) If $a(t) \geq \alpha > 0$, then $|u(t)| \leq e^{-\alpha t}|u_0| + \frac{1}{\alpha}(1 - e^{-\alpha t}) \max_{0 \leq s \leq t} |f(s)|$

(ii) If $a(t) \geq 0$ (i.e. $\alpha = 0$ the parabolic case), then

$$|u(t)| \leq |u_0| + \int_0^t |f(s)| ds \text{ or } |u(t)| \leq |u_0| + \|f\|_{L_1}$$

Proof:

(i) For $a(t) \geq 0, \ \forall t > 0$, we have that $A(t) = \int_0^t a(s) ds$ is non-decreasing and $A(t) - A(s) \geq 0, \ \forall t > s$.

Since $a(t) \geq \alpha > 0$ we have $A(t) = \int_0^t a(s) ds \geq \int_0^t \alpha \cdot ds = \alpha t$. Further

$$A(t) - A(s) = \int_s^t a(r)\, dr \geq \alpha(t - s).$$

Thus $e^{-A(t)} \leq e^{-\alpha t}$ and $e^{-(A(t) - A(s))} \leq e^{-\alpha(t-s)}$.

Hence using (1), we get

$$u(t) \leq u_0 \cdot e^{-\alpha t} + \int_0^t e^{-\alpha(t-s)} \max_{0 \leq s \leq t} |f(s)| ds, \tag{2}$$

which after integration gives that

$$|u(t)| \leq e^{-\alpha \cdot t}|u_0| + \max_{0 \leq s \leq t} |f(s)| \left[\frac{1}{\alpha} e^{-\alpha(t-s)}\right]_{s=0}^{s=t}$$

$$|u(t)| \leq e^{-\alpha \cdot}|u_0| + \frac{1}{\alpha}(1 - e^{-\alpha t}) \max_{0 \leq s \leq t} |f(s)| \quad \square$$

(ii) Let $\alpha = 0$ in (2) then $|u(t)| \leq |u_0| + \int_0^t |f(s)| ds$, and the proof is complete.

$\square$

Remark: Recall that we refer to the set of functions where we seek the approximate solution as the *trial space* and the space of functions used for the orthogonality condition, as the *test space*.

9.2. Galerkin finite element methods (FEM) for IVP

cG(1) *Continuous Galerkin of degree 1*:

In this case the trial functions are piecewise linear and continuous while the test functions are piecewise constant and discontinuous, i.e. the trial and test functions are in different spaces.

Note, e.g. the variational formulation: Find a function u in an appropriate space of *trial* functions that satisfies

$$\int_0^1 au'v'dx = \int_0^1 fvdx \quad \text{for all appropriate } \textit{test} \text{ functions } v.$$

dG(0) *Discontinuous Galerkin of degree 0*:

Here both the trial and test functions are piecewise constant and discontinuous (they are in the same space of functions).

gG(q) *Global Galerkin of degree q:*

This method is formulated as follows: Find $U \in P^q(0,T), U(0) = u_0$, such that

$$\int_0^T (\dot{U} + aU)vdt = \int_0^T fvdt, \quad \forall v \in P^q(0,T), \text{ with } \quad v(0) = 0,$$

where $v := \{t, t^2, \ldots, t^q\}$.

cG(q) *Continuous Galerkin of degree q:*

Find $U \in P^q(0,T), U(0) = u_0$, such that

$$\int_0^T (\dot{U} + aU)vdt = \int_0^T fvdt, \quad \forall v \in P^{q-1}(0,T),$$

where now $v := \{1, t, t^2, \ldots, t^{q-1}\}$.

Note the difference between the two test function spaces above.

Example: Consider cG(q) with $q = 1$ then $t^{q-1} = t^0 = 1$ and $v \equiv 1$, thus we have

$$\int_0^T (\dot{U} + aU)vdt = \int_0^T (\dot{U} + aU)dt = U(T) - U(0) + \int_0^T aU(t)dt \tag{3}$$

But $U(t)$ is a linear function through $U(0)$ and *the unknown* quantity $U(T)$, thus

$$U(t) = U(T)\frac{t}{T} + U(0)\frac{T-t}{T}, \tag{4}$$

inserting $U(t)$ in (3) we have

$$U(T) - U(0) + \int_0^T a\Big(U(T)\frac{t}{T} + U(0)\frac{T-t}{T}\Big)\, dt = \int_0^T f\, dt. \tag{5}$$

which gives us $U(T)$ and consequently, through (4) and assuming $U(0)$ is given, $U(t)$. Using this idea we can formulate:

The cG(1) Algorithm for the partition $\mathcal{T}_k$ of $[0,T]$ to subintervals $I_k = (t_{k-1}, t_k]$.

(1) Given $U(0) = U_0$, apply (5) to $(0, t_1]$ and compute $U(t_1)$.
Then authomatically get $U(t), \forall t \in [0, t_1]$ (see (4)).

(2) Assume that U is computed in all the successive intervals $(t_{k-1}, t_k]$, $k = 0, 1, n-1$.

(3) Compute $U(t), t \in (t_{n-1}, t_n]$.
This is done through applying (5) to the interval $(t_{n-1}, t_n]$, instead of $(0, T]$: i.e.

$$U(t_n) - U(t_{n-1}) + \int_{t_{n-1}}^{t_n} a\Big(\frac{t - t_{n-1}}{t_n - t_{n-1}} U(t_n) + \frac{t_n - t}{t_n - t_{n-1}} U(t_{n-1})\Big)\, dt = \int_{t_{n-1}}^{t_n} f dt.$$

Now since $U(t_{n-1})$ is known, we can calculate $U(t_n)$ and $U(t), t \in (t_{n-1}, t_n]$ is determined by the n-version of the linear combination formula (4):

$$U(t) = U(t_n)\frac{t}{t_n} + U(t_{n-1})\frac{t_n - t}{t_n},$$

In sequel we shall use the notation $U(t_n) = U_n,\ U(t_{n-1}) = U_{n-1}, \dots$.

Global forms

Continuous Galerkin cG(q): Find $U(t) \in V_k^{(q)}$, such that $U(0) = U_0$ and

$$\int_0^{t_n} (\dot{U} + aU) w dt = \int_0^{t_n} f w dt, \qquad \forall w \in W_k^{(q-1)},$$

$V_k^{(q)} = \{v : v$ continuous piecewise polynomials of degree q on the partition $\mathcal{T}_k\}$,

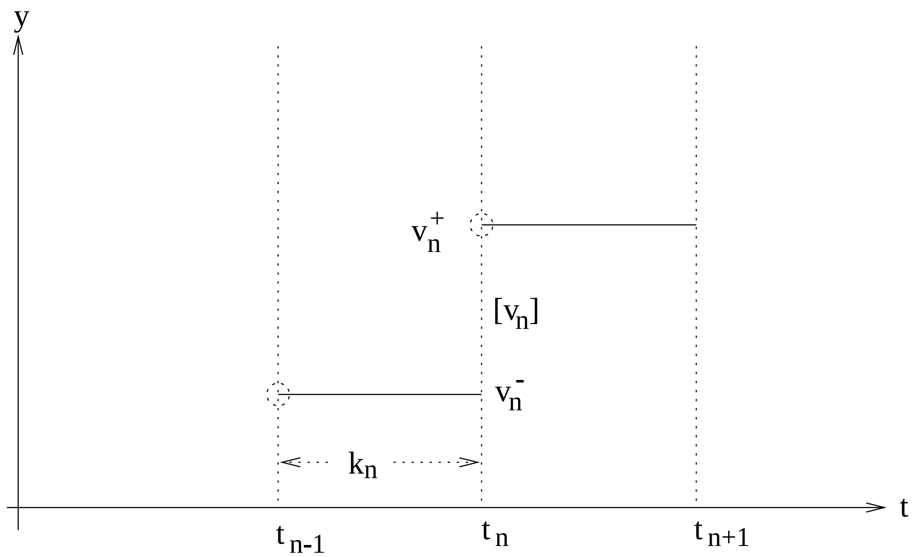
$W_k^{(q-1)} = \{w : w$ discontinuous piecewise polynomials of degree $q-1$ on $\mathcal{T}_k\}$.

Discontinuous Galerkin dG(q): Find $U(t) \in P^q(0,T)$ such that

$$\int_0^T (\dot{U} + aU)v dt + a(U(0) - u(0))v(0) = \int_0^T f v dt, \quad \forall v \in P^q(0,T).$$

This approach gives up the requirement that $U(t)$ satisfies the initial condition. Instead, the intial condition is represented by $(U(0) - u(0))$ with $U(0) \neq u(0)$.

Notation: Let $v^{\pm} n = \lim_{s \to 0^+}$ and $[v_n] = v_n^+ - v_n^-$ is the "jump" in $v(t)$ at time t.



<u>**dG(q):**</u> For $n = 1, \ldots, N$ find $U(t) \in \mathcal{P}^q(t_{n-1}, t_n)$ such that

$$\int_{t_{n-1}}^{t_N} (\dot{U} + aU)v dt + U_{n-1}^+ v_{n-1}^+ = \int_{t_{n-1}}^{t_n} f v dt + U_{n-1}^- v_{n-1}^+, \;\; \forall v \in \mathcal{P}^q(t_{n-1}, t_n). \tag{6}$$

Let $q = 0$, in the case of approximating with piecewise constants $v \equiv 1$ is the only base function and we have $U(t) = U_n = U_{n-1}^+ = U_n^-$ on $I_n = (t_{n-1}, t_n]$, where $U_n = U(t_n)$, ... and $\dot{U} \equiv 0$.

Thus (6) gives the <u>**dG(0)**</u> formulation: For $n = 1, 2, \ldots, N$ find piecewise constants U_n such that

$$\int_{t_{n-1}}^{t_n} aU_n dt + U_n = \int_{t_{n-1}}^{t_n} f dt + U_{n-1}.$$

Finally summing over n in (6), we get the global dG(q) formulation: Find $U(t) \in W_k^{(q)}$, such that

$$\sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} (\dot{U} + aU)w dt + \sum_{n=1}^{N} [U_{n-1}] w_{n-1}^+ = \int_0^{t_N} f w dt, \forall w \in W_k^{(q)}, \;\; U_0^- = u_0.$$

An "a posteriori" error estimate for the cG(1) formulation

The continuous problem:

(7) $$\dot{u}(t) + a(t)u(t) = f(t), \qquad \forall t \in (0, T)$$

The *variational form* for (7) is:

$$\int_0^T (\dot{u} + au)v dt = \int_0^T fv dt, \quad \forall v(t) \in P^q(0, T).$$

Integrating by parts we get

(8) $$u(T)v(T) - u(0)v(0) + \int_0^T u(t)\Big(- \dot{v}(t) + av(t)\Big) dt = \int_0^T fv dt.$$

If we choose v to be the solution of the *dual problem*:

$$-\dot{v} + av = 0, \quad \text{in } (0, T),$$

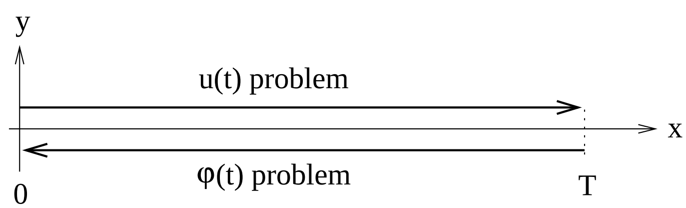
then (8) is equivalent to

$$u(T)v(T) = u(0)v(0) + \int_0^T fv dt, \quad \forall v(t) \in P^q(0, T).$$

Thus we have the following:

Dual problem for (7): Find $\varphi(t)$ such that

(9) $$\begin{cases} -\dot{\varphi}(t) + a(t)\varphi(t) = 0, & t_N > t \geq 0 \\ \varphi(t_N) = e_N, & e_N = u_N - U_N = u(t_N) - U(t_N) \end{cases}$$



Theorem 9.2: For $N = 1, 2, \ldots$ the cG(1) solution $U(t)$ satisfies

$$|e_N| \leq S(t_n) \cdot \max_{[0,t_N]} |k \cdot r(U)|,$$

where $k = k_n = |I_n|$ for $t \in I_n = (t_{n-1}, t_n)$ is the *time step* and $r(U) = \dot{U} + aU - f$ is the residual error. Further, $S(t_N) = \left(\int_0^{t_N} |\dot{\varphi}| dt \right) / e_N$ is the *stability factor*. The stability factor measures the effects of the accumulation of error in the approximation and to give the analysis a quantitative meaning we have to give a quantitative bound of this factor.

$$S(t_n) \leq \begin{cases} e^{\lambda \cdot t_N} & |a(t)| \leq \lambda, \quad \forall t \\ 1 & a(t) \geq 0, \quad \forall t \end{cases}$$

Proof: Let $e(t) = u(t) - U(t)$. By the dual problem: $-\dot{\varphi}(t) + a(t)\varphi(t) = 0$ we can write

$$e_N^2 = e_N^2 + 0 = e_N^2 + \int_0^{t_N} e(-\dot{\varphi} + a\varphi)\, dt,$$

and using partial integration we have

$$\int_0^{t_N} e(-\dot{\varphi} + a(t)\varphi) dt = [-e(t)\varphi(t)]_0^{t_N} + \int_0^{t_N} \dot{e}\varphi\, dt + \int_0^{t_N} ea\varphi\, dt = \{e(0) = 0\} =$$

$$= -\underbrace{e(t_N)}_{=e_N}\underbrace{\varphi(t_N)}_{=e_N} + \int_0^{t_N} (\dot{e} + ae)\varphi\, dt = -e_N^2 + \int_0^{t_N} (\dot{e} + ae)\varphi\, dt.$$

Since

$$\dot{e}(t) + a(t)e(t) = \dot{u}(t) - \dot{U}(t) + a(t)u(t) - a(t)U(t),$$

and $f(t) = \dot{u}(t) + a(t)u(t)$ we can now write

$$\dot{e}(t) + a(t)e(t) = f(t) - \dot{U}(t) - a(t)U(t).$$

Recalling the definition of the residual: $r(U) = \dot{U} + aU - f$ we have that

$$\dot{e}(t) + a(t)e(t) = -r(U),$$

and consequently

$$e_N^2 = e_N^2 + 0 = e_N^2 - e_N^2 - \int_0^{t_N} r(U(t))\varphi(t) dt.$$

Thus we have the *error representation formula*: $e_N^2(t) = -\int_0^{t_N} r(U(t))\varphi(t) dt$. To continue we use the interpolant $\pi_k\varphi$ of φ and write

$$\pi_k\varphi = \frac{1}{k_n} \int_{I_n} \varphi(s) ds \Rightarrow e_N^2 = -\int_0^{t_N} r(U)(\varphi(t) - \pi_k\varphi(t)) dt + \int_0^{t_N} r(U)\pi_k\varphi(t) dt.$$

Now from the discrete variational formulation:

$$\int_0^{t_N} (\dot{U} + aU)\pi_k\varphi(t)dt = \int_0^{t_N} f\pi_k\varphi(t)dt$$

we have the Galerkin orthogonality $\displaystyle\int_0^{t_N} r(U)\pi_k\varphi(t)dt = 0$. Thus the final form of the error representation formula is: $e_N^2 = -\displaystyle\int_0^{t_N} r(U)(\varphi(t) - \pi_k\varphi(t))dt$. Now applying the *interpolation error*

$$\|f - \pi_q f\|_{L_p(a,b)} \leq C_i(b-a)\|f'\|_{L_p(a,b)},$$

to the function φ and the interval I_n, $|I_n| = k_n$ we get $\displaystyle\int_{I_n} |\varphi - \pi_k\varphi|dt \leq k_n \int_{I_n} |\dot{\varphi}|dt$. Thus we may write

$$\int_0^{t_N} |\varphi - \pi_k\varphi|dt = \sum_{n=1}^{N} \int_{I_n} |\varphi - \pi_k\varphi|dt \leq \sum_{n=1}^{N} k_n \int_{I_n} |\dot{\varphi}|dt \tag{10}$$

Let now $|v|_J = \max_{t\in J} |v(t)|$, then using (10) and the final form of the error representation formula we have

$$|e_N|^2 \leq \sum_{n=1}^{N} |r(U)|_{I_n} \cdot k_n \int_{I_n} |\dot{\varphi}|dt \leq \max_{1\leq n\leq N} (k_n|r(U)|_{I_n}) \int_0^{t_N} |\dot{\varphi}|dt.$$

But since $\int_0^{t_N} |\varphi|dt = |e_N| \cdot S(t_N)$, (see the definition of $S(t_N)$), we have that

$$|e_N|^2 \leq |e_N|S(t_N) \max_{[0,t_N]}(k|r(U)|).$$

This completes the proof of the first assertion of the theorem.

To prove the second assertion , we claim that:

(a) $\quad |a(t)| \leq \lambda, \quad 0 \leq t \leq t_N \Rightarrow |\varphi(t)| \leq e^{\lambda\cdot t_N}|e_N|, \quad 0 \leq t \leq t_N$

(b) $\quad |a(t)| \geq 0, \quad \forall t \Rightarrow |\varphi(t)| \leq |e_N|, \quad \forall t \in [0, t_N]$

To prove this claim let $s = t_N - t$, $(t = t_N - s)$ and define $\psi(s) = \varphi(t_N - s)$, then by the chain rule

$$\frac{d\psi}{ds} = \frac{d\psi}{dt} \cdot \frac{dt}{ds} = -\dot{\varphi}(t_N - s).$$

The dual problem is now formulated as follows: find $\varphi(t)$ such that

$$-\dot{\varphi}(t_N - s) + a(t_N - s)\varphi(t_N - s) = 0.$$

The corresponding problem for $\psi(s)$:

$$\begin{cases} \dfrac{d\psi(s)}{ds} + a(t_N - s)\psi(s) = 0, & t_N > s \geq 0 \\ \psi(0) = \varphi(t_N) = e_N, & e_N = u_N - U_N = u(t_N) - U(t_N), \end{cases}$$

has the exact solution $\psi(s) = \psi(0) \cdot e^{-A(t_N)} e^{A(t_N - s)} = e_N \cdot e^{A(t) - A(t_N)}$. Now using the relation $\psi(s) = \varphi(t),\ t_N - s = t$, we get

$$\varphi(t) = e_N \cdot e^{A(t) - A(t_N)}, \quad \text{and} \quad \dot{\varphi}(t) = e_N \cdot a(t) e^{A(t) - A(t_N)}.$$

Thus the proof of both claims (a) and (b) are easily followed:

(a) For $|a(t)| \leq \lambda$, we have

$$|\varphi(t)| = |e_N| e^{\int_{t_N}^{t} a(s)ds} \leq |e_N| e^{\max_t |a(t)|(t_N - t)} \leq |e_N| e^{\lambda \cdot t_N}$$

. $\square$

(b) For $|a(t)| \leq 0$, we have $|\varphi(t)| = |e_N| e^{\int_0^{t_N} a(s)ds} \leq |e_N| e^{\min_t a(t)(t - t_N)}$ and since $(t - t_N) < 0$ we get $|\varphi(t)| \leq |e_N|$. $\square$

Further note that for $a(t) \geq 0$ we have

$$\begin{aligned} \int_0^{t_N} |\dot{\varphi(t)}| dt &= |e_N| \int_0^{t_N} a(t) e^{A(t) - A(t_N)} dt = |e_N| \cdot [e^{A(t) - A(t_N)}]_0^{t_N} \\ &= |e_N| \cdot \left(1 - e^{A(0) - A(t_N)}\right) \leq 1, \end{aligned}$$

which gives that $S(t_N) = \dfrac{\int_0^{t_N} |\dot{\varphi(t)}| dt}{|e_N|} \leq 1.$

<u>The case $|a(t)| \leq \lambda$:</u>

We have earlier derived that $\dot{\varphi}(t) = a(t) e_N \cdot e^{A(t) - A(t_N)}$, thus

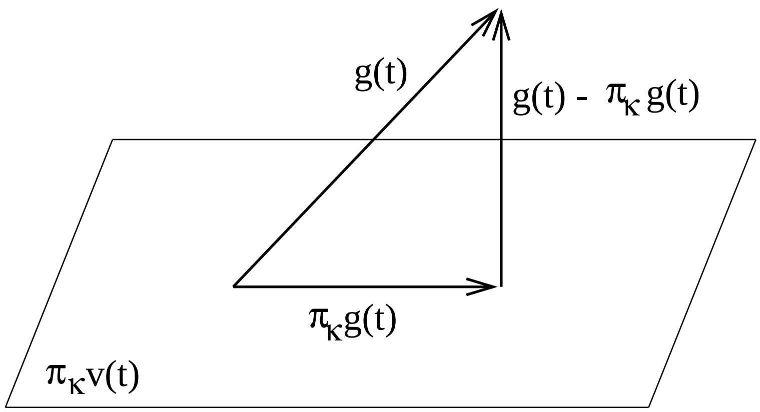
$$|\dot{\varphi}(t)| \leq \lambda |e_N| e^{A(t) - A(t_N)} = \lambda |e_N| e^{\int_{t_N}^{t} a(s)ds} \leq \lambda |e_N| e^{\lambda(t_N - t)}.$$

Integrating over $(0, t_N)$ we get

$$\int_0^{t_N} |\dot{\varphi}(t)| dt \leq |e_N| \int_0^{t_N} \lambda e^{\lambda(t_N - t)} dt = |e_N| \Big[- e^{\lambda(t_N - t)} \Big]_0^{t_N} = |e_N| (-1 + e^{\lambda t_N}),$$

which gives that $S(t_N) \leq (-1 + e^{\lambda \cdot t_N}) \leq e^{\lambda \cdot t_N}$, and completes the proof of the second assertion. $\square$

Convergence order $\mathcal{O}(k^2)$:



Note that $(g(t) - \pi_k g(t)) \perp$ (constants) $\forall g(t)$, and since $\dot{U}(t)$ is constant on I_N we have that $\int_0^{t_N} \dot{U}(\varphi - \pi_k \varphi)dt = 0$, thus

$$\begin{aligned}
e_N^2 &= -\int_0^{t_N} r(U)(\varphi(t) - \pi_k\varphi(t))dt = \int_0^{t_N} (f - aU - \dot{U})(\varphi - \pi_k\varphi)dt \\
&= \int_0^{t_N} (f - aU)(\varphi - \pi_k\varphi)dt - \int_0^{t_N} \dot{U}(\varphi - \pi_k\varphi)dt \\
&= -\int_0^{t_N} (aU - f)(\varphi - \pi_k\varphi)dt.
\end{aligned}$$

Similarly using the fact that $\pi_k(aU - f)$ is a constant

$$\int_0^{t_N} \pi_k(aU - f)(\varphi - \pi_k\varphi)dt = 0$$

and we can write

$$e_N^2 = -\int_0^{t_N} \Big((aU - f) - \pi_k(aU - f)\Big)(\varphi - \pi_k\varphi)dt.$$

Now using Theorem 9.2 and the interpolation error we get

$$\begin{aligned}
|e_N| &\leq S(t_N) \cdot \Big| k|(aU - f) - \pi_k(aU - f)| \Big|_{[0,t_N]} \\
&\leq S(t_N) \cdot \Big| k^2 \frac{d}{dt}(aU - f) \Big|_{[0,t_N]}.
\end{aligned}$$

$\square$

9.3.2. An a posteriori error estimate for the dG(0)

Theorem 9.3. For $N = 1, 2, \ldots$, the dG(0) solution $U(t)$ satisfies

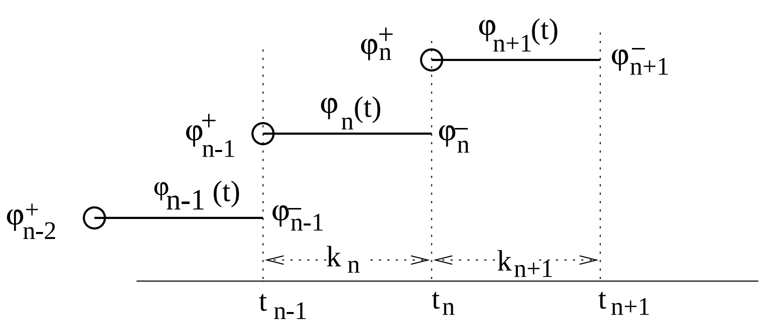
$$|u(t_N) - U_N| \le S(t_N)|kR(U)|_{[0,t_N]}, \quad U_N = U(t_N)$$

where

$$R(U) = \frac{|U_N - U_{N-1}|}{k_n} + |f - aU| \quad \text{for} \quad t_{N-1} < t \le t_N.$$

Proof: Similar as in cG(1). Note that now the residual error includes jump terms and since dual problem satisfies $-\dot{\varphi}(t) + a(t)\varphi(t) = 0$, we can write

$$\begin{aligned} e_N^2 &= e_N^2 + \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} e(-\dot{\varphi}(t) + a(t)\varphi(t))dt = [PI] = \\ &= e_N^2 + \sum_{n=1}^{N} \Big(\int_{t_{n-1}}^{t_n} (\dot{e} + ae)\varphi(t)dt - [e\varphi]_{t_{n-1}}^{t_n} \Big) \\ &= \Big\{ \dot{e} + ae = \dot{u} - \dot{U} + au - aU = f - aU, \ (U \text{ constant} \Rightarrow \dot{U} = 0) \Big\} \\ &= e_N^2 + \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} (f - aU)\varphi dt - \sum_{n=1}^{N} [e\varphi]_{t_{n-1}}^{t_n}. \end{aligned} \tag{11}$$



We rewrite the last sum as

$$\begin{aligned} \sum_{n=1}^{N} (e\varphi)_{t_{n-1}}^{t_n} &= \sum_{n=1}^{N} \Big(e(t_n^-)\varphi(t_n^-) - e(t_{n-1}^+)\varphi(t_{n-1}^+) \Big) \\ &= \{ g(t_n^-) = g_n^-, g(t_{n-1}^+) = g_{n-1}^+ \} = \\ &= \sum_{n=1}^{N} (e_n^- \varphi_n^- - e_{n-1}^+ \varphi_{n-1}^+) = (e_1^- \varphi_1^- - e_0^+ \varphi_0^+) + (e_2^- \varphi_2^- - e_1^+ e_1^+) + \ldots \\ &\quad + (e_{N-1}^- \varphi_{N-1}^- - e_{N-2}^+ \varphi_{N-2}^+) + (e_N^- \varphi_N^- - e_{N-1}^+ \varphi_{N-1}^+). \end{aligned}$$

To continue we write $\varphi_i^- = (\varphi_i^- - \varphi_i^+ + \varphi_i^+)$, $i = 1, \dots N-1$ then

$$-\sum_{n=1}^{N}(e\varphi)_{t_{n-1}}^{t_n} = -e_N^-\varphi_N^- + e_0^+\varphi_0^+ - e_1^-(\varphi_1^- - \varphi_1^+ + \varphi_1^+) + e_1^+\varphi_1^+$$
$$- e_2^-(\varphi_2^- - +\varphi_2^+ + \varphi_2^+) + e_2^+\varphi_2^+ \dots$$
$$- e_{N-1}^-(\varphi_{N-1}^- - \varphi_{N-1}^+ + \varphi_{N-1}^+) + e_{N-1}^+\varphi_{N-1}^+.$$

We rewrite these terms as

$$\begin{aligned}- e_i^-(\varphi_i^- - \varphi_i^+ + \varphi_i^+) + e_i^+\varphi_i^+ &= -e_i^-\varphi_i^- + -e_i^-\varphi_i^+ - e_i^-\varphi_i^+ + e_i^+\varphi_i^+\\ &= e_i^-(\varphi_i^+ - \varphi_i^-) + \varphi_i^+(e_i^+ - e_i^-) = e_i^-[\varphi_i] + \varphi_i^+[e_i],\end{aligned}$$

where $[g] = g^+ - g^-$ represents the jump. Thus

$$-\sum_{n=1}^{N}(e\varphi)|_{t_{n-1}}^{t_n} = -e_N^2 + e_0^+\varphi_0^+ + \sum_{n=1}^{N-1}[e_n]\varphi_n^+ + \sum_{n=1}^{N-1} e_n^-[\varphi_n].$$

Inserting in (11) we get that

$$\begin{aligned}e_N^2 &= e_N^2 + \sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}(f - aU)\varphi dt - \sum_{n=1}^{N}[e\varphi]_{t_{n-1}}^{t_n}\\
&= e_N^2 + \sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}(f - aU)\varphi dt - e_N^2 + e_0^+\varphi_0^+ + \sum_{n=1}^{N-1}[e_n]\varphi_n^+ + \sum_{n=1}^{N-1}[\varphi_n]e_n^- =\\
&= \{\varphi_n,\ u_n \text{ smooth } \Rightarrow [\varphi_n] = 0,\ [u_n] = 0\}\\
&= e_0^+\varphi_0^+ + \sum_{n-1}^{N}\int_{t_{n-1}}^{t_n}(f - aU)\varphi dt + \sum_{n=1}^{N-1}[e_n]\varphi_n^+ = \{[u_n] = 0 \Rightarrow [e_n] = [-U_n]\}\\
&= \sum_{n=1}^{N}(\int_{t_{n-1}}^{t_n}(f - aU)\varphi dt - [U_{n-1}]\varphi_{n-1}^+) =\\
&= \{\text{Galerkin}\} = \sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}\{(f - aU)(\varphi - \pi_k\varphi) - [U_{n-1}](\varphi - \pi_k\varphi)_{n-1}^+\}dt.\end{aligned}$$

The rest is as in Theorem 9.2. □

Adaptive dG(0):

To guarantee that the dG(0) approximation $U(t)$ satiesfies

$$|e_N| = |u(t_n) - U(t_n)| \le TOL \quad \text{(TOL is a given tolerance)}$$

we seek to determine the time step k_n so that

(12) $$S(t_N)\max_{t\in I_n}|k_nR(U)| = TOL, n = 1, 2, \dots, N.$$

Algorithm:

(i) Compute U_n from U_{n-1} using a predicted step k_n.

Example, $\int_{t_{n-1}}^{t_n} aU_n dt + U_n = \int_{t_{n-1}}^{t_n} f dt + U_{n-1}$

(ii) Compute $|kR(U)|_{I_n} := \max_n |k_n R(U)|$.

$$
\begin{array}{ccc}
\boxed{\begin{array}{l}\text{Is (12) valid}\\ \text{for this } k_n?\end{array}} & \xrightarrow{\text{YES!}} & \boxed{\begin{array}{l}\text{Accept the solution } U_n\\ \text{and go to the next time step}\end{array}} \\
\Big| \ \text{NO!} & & \\
\boxed{\begin{array}{l}\text{Recompute (12) with}\\ \text{a smaller } k_n\end{array}} & &
\end{array}
$$

9.4. A priori error analysis

The discontinuous Galerkin method, dG(0) for $\dot{u} + au = f$, a=constant:

Find $U = U(t),\ t \in I_n$, such that

$$\int_{t_{n-1}}^{t_n} \dot{U} dt + a \int_{t_{n-1}}^{t_n} U dt = \int_{I_n} f dt. \tag{13}$$

Note that $U(t) = U_n$ is constant for $t \in I_n$. Let $U_n = U(t_n),\ U_{n-1} = U(t_{n-1})$ and $k_n = t_n - t_{n-1}$, then

$$\int_{t_{n-1}}^{t_n} \dot{U} dt + a \int_{t_{n-1}}^{t_n} U dt = U(t_n) - U(t_{n-1}) + ak_n U_n = U_n - U_{n-1} + ak_n U_n.$$

Hence with a given initial data $u(0) = u_0$, the equation (13) can be written as

$$\begin{cases} U_n - U_{n-1} + ak_n U_n = \int_{I_n} f dt \quad n = 1, 2, \dots \\ U_0 = u_0. \end{cases} \tag{14}$$

As for the *exact* solution $u(t)$ of $\dot{u} + au = f$, the same procedure would give

$$u(t_n) - u(t_{n-1}) + k_n a u_n(t) = \int_{I_n} f dt + k_n a u_n(t) - a \int_{t_{n-1}}^{t_n} u(t) dt, \tag{15}$$

where we have moved the term $a \int_{t_{n-1}}^{t_n} u(t) dt$ to the right hand side and add $k_n a u_n(t)$ to both sides.

Thus

$$(14) \Leftrightarrow (1+k_n a)U_n(t) = U_{n-1}(t) + \int_{I_n} f\,dt$$

$$(15) \Leftrightarrow (1+k_n a)u_n(t) = u_{n-1}(t) + \int_{I_n} f\,dt + k_n a u_n(t) - a\int_{t_{n-1}}^{t_n} u(t)dt$$

Let now $e_n = u_n - U_n$ and $e_{n-1} = u_{n-1}(t) - U_{n-1}(t)$ then $(15) - (14)$ gives that

$$e_n = (1+k_n a)^{-1}(e_{n-1} + \rho_n) \tag{16}$$

where $\rho_n := k_n a u_n(t) - a\displaystyle\int_{t_{n-1}}^{t_n} u(t)dt$. Thus in order to estimate the error e_n we need an iteration procedure and an estimate of ρ_n:

Lemma 9.1. We have that

$$|\rho_n| \leq \frac{1}{2}|a||k_n|^2 \max_{I_n} |\dot{u}(t)|$$

Proof. Recalling the definition we have that $\rho_n = k_n a u_n(t) - a\int_{t_{n-1}}^{t_n} u(t)dt$. Thus

$$|\rho_n| \leq |a||k_n|\Big|u_n - \frac{1}{|k_n|}\int_{I_n} u dt\Big|.$$

Using a Taylor expansion of the integrand $u(t)$ about t_n:

$$u(t) = u_n + \dot{u}(\xi)(t - t_n), \text{ for some } \xi, t_{n-1} < \xi < t_n$$

yields

$$\begin{aligned}
|\rho_n| &\leq |a||k_n|\Big|u_n - \frac{1}{k_n}\int_{I_n}[u_n + \dot{u}(\xi)(t-t_n)]dt\Big| \\
&\leq |a||k_n|\Big|u_n - \frac{1}{k_n}k_n u_n - \frac{1}{k_n}\dot{u}(\xi)\Big[\frac{(t-t_n)^2}{2}\Big]_{t_{n-1}}^{t_n}\Big| \\
&= |a||k_n|\Big| - \frac{1}{k_n}\dot{u}(\xi)\Big[0 - \frac{k_n^2}{2}\Big] = |a||k_n|\Big| - \frac{1}{k_n}\dot{u}(\xi)\frac{k_n^2}{2}\Big| = |a||k_n|^2\frac{1}{2}|\dot{u}(\xi)|.
\end{aligned}$$

Thus we get the final estimate for ρ_n:

$$|\rho_n| \leq \frac{1}{2}|a||k_n|^2 \max_{I_n} |\dot{u}(t)| \quad \square$$

To simplify the estimate for e_n we split, and gather, the proof of technical details in the following lemma:

Lemma 9.2. For $k_n|a| \le 1/2,\ n \ge 1$ we have

(i) $(1 - k_n|a|)^{-1} \le e^{2k_n|a|}$.

(ii) Let $\tau_n = t_N - t_{n-1}$ then $|e_N| \le \frac{1}{2}\sum_{n=1}^{N}(e^{2|a|\tau_n}|a|k_n) \max_{1\le n\le N} k_n|\dot{u}|I_n$.

(iii) $\sum_{n=1}^{N} e^{2|a|\tau_n}|a|k_n \le e\int_0^{t_N} |a|e^{2|a|\tau}d\tau$.

We postphone the proof of this lemma and first show that using these results we can obtain a bound for the error e_N (our main result) viz,

Theorem 9.4: If $k_n|a| \le \frac{1}{2}, n \ge 1$ then the error of the $dG(0)$ approximation U satisfies

$$|u(t_N) - U(t_N)| = |e_N| \le \frac{e}{4}\Big(e^{2|a|t_N} - 1\Big) \max_{1\le n\le N} k_n|\dot{u}(t)|_{I_n}.$$

Proof: Using the estimates (ii) and (iii) of lemma 9.2 we have that

$$|e_N| \le \frac{1}{2}\sum_{n=1}^{N}(e^{2|a|\tau_n}|a|k_n) \max_{1\le n\le N} k_n|\dot{u}|_{I_n} \le \frac{1}{2}\Big(e\int_0^{t_N} |a|e^{2|a|\tau}d\tau\Big) \max_{1\le n\le N} k_n|\dot{u}|_{I_n} =$$
$$= \frac{1}{2}e\Big[\frac{e^{2|a|\tau}}{2}\Big]_0^{t_N} \cdot \max_{1\le n\le N} k_n|\dot{u}(t)|_{I_n} = \frac{e}{4}\Big(e^{2|a|t_N} - 1\Big) \max_{1\le n\le N} k_n|\dot{u}(t)|_{I_n}.\square$$

Note that the stability constant $\frac{e}{4}\Big(e^{2|a|t_N} - 1\Big)$ may grow depending on $|a|$ and t_N, and then this result may not be satisfactory at all.

Now we return to the proof of our technical results:

Proof of Lemma 9.2:

(i) For $0 \le x \le 1/2$, we have that $1/2 \le 1 - x < 1$ and $0 < 1 - 2x \le 1$. We can now multiply both side of the first claim: $\frac{1}{1-x} < e^{2x}$ by $1 - x \ge 1/2 > 0$ to obtain the equivalent relation

$$f(x) := (1 - x)e^{2x} > 1. \tag{17}$$

Note that $f(0) = 1$ and since $f'(x) = (1 - 2x)e^{2x} > 0$ the relation (17) is valid.

(ii) Recall that $e_n = (1+k_na)^{-1}(e_{n-1}+\rho_n)$. To deal with the coefficient $(1+k_na)^{-1}$ first we note that $(1 + k_na)^{-1} \le (1 - k_na)^{-1}$ if $a \ge 0$. Thus $(1 + k_n|a|)^{-1} \le$

$(1-k_n|a|)^{-1}$, $a \in \mathbb{R}$. Further the assumption $k_n|a| \le \frac{1}{2}$, $n \ge 1$, combined with (i), implies that $(1-k_n|a|)^{-1} \le e^{2k_n|a|}$, $n \ge 1$. Thus we have

$$\text{(18)} \quad |e_N| \le \frac{1}{1-k_N|a|}|e_{N-1}| + \frac{1}{1-k_N|a|}|\rho_N| \le |e_{N-1}| \cdot e^{2k_N|a|} + |\rho_N| \cdot e^{2k_N|a|}.$$

Changing, e.g. N to $N-1$ we get

$$|e_{N-1}| \le |e_{N-2}| \cdot e^{2k_{N-1}|a|} + |\rho_{N-1}| \cdot e^{2k_{N-1}|a|} = e^{2k_{N-1}|a|}\Big(|e_{N-2}| + |\rho_{N-1}|\Big),$$

which inserting in (18) gives that

$$\text{(19)} \qquad |e_N| \le e^{2k_N|a|}e^{2k_{N-1}|a|}\Big(|e_{N-2}| + |\rho_{N-1}|\Big) + |\rho_N| \cdot e^{2k_N|a|}.$$

Similarly we have $|e_{N-2}| \le e^{2k_{N-2}|a|}\Big(|e_{N-3}| + |\rho_{N-2}|\Big)$. Now iterating (19) and using the fact that $\underline{e_0 = 0}$ we get,

$$\begin{aligned} |e_N| \le & e^{2k_N|a|}e^{2k_{N-1}|a|}e^{2k_{N-2}|a|}|e_{N-3}| + e^{2k_N|a|}e^{2k_{N-1}|a|}e^{2k_{N-2}|a|}|\rho_{N-2}| \\ & + e^{2k_N|a|}e^{2k_{N-1}|a|}|\rho_{N-1}| + |\rho_N| \cdot e^{2k_N|a|} \le \cdots \le \\ \le & e^{2|a|\sum_{n=1}^N k_n}|e_0| + \sum_{n=1}^N e^{2|a|\sum_{m=n}^N k_m}|\rho_n| \\ = & \sum_{n=1}^N e^{2|a|\sum_{m=n}^N k_m}|\rho_n|. \end{aligned}$$

Invoking by Lemma 9.1: $|\rho_n| \le \frac{1}{2}|a||k_n|^2 \max_{I_n} |\dot{u}(t)|$ we have

$$|e_N| \le \sum_{n=1}^N e^{2|a|\sum_{m=n}^N k_m}\frac{1}{2}|a||k_n|^2 \max_{I_n} |\dot{u}(t)|.$$

Note that

$$\sum_{m=n}^N k_m = (t_n - t_{n-1}) + (t_{n+1} - t_n) + (t_{n+2} - t_{n+1}) + \ldots + (t_N - t_{N-1}) = t_N - t_{n-1}.$$

Hence we have shown the assertion (ii) of the lemma:

$$|e_N| \le \sum_{n=1}^N e^{2|a|(t_N - t_{n-1})}\frac{1}{2}|a||k_n|^2 \max_{I_n} |\dot{u}(t)| = \frac{1}{2}\sum_{n=1}^N (e^{2|a|\tau_n}|a|k_n) \max_{1 \le n \le N} k_n|\dot{u}|_{I_n} \quad \square$$

(iii) To prove this part we note that

$$\tau_n = t_N - t_{n-1} = (t_N - t_n) + (t_n - t_{n-1}) = \tau_{n+1} + k_n,$$

and since $|a|k_n \le 1/2$ we have $2|a|\tau_n = 2|a|\tau_{n+1} + 2|a|k_n \le 2|a|\tau_{n+1} + 1$. Further for $\tau_{n+1} \le \tau \le \tau_n$, we can write

$$
\begin{aligned}
e^{2|a|\tau_n} \cdot k_n &= \int_{\tau_{n+1}}^{\tau_n} e^{2|a|\tau_n} d\tau \le \int_{\tau_{n+1}}^{\tau_n} e^{(2|a|\tau_{n+1}+1)} d\tau \\
&= \int_{\tau_{n+1}}^{\tau_n} e^1 \cdot e^{2|a|\tau_{n+1}} d\tau \le e \int_{\tau_{n+1}}^{\tau_n} e^{2|a|\tau} d\tau.
\end{aligned} \tag{20}
$$

Multiplying (20) by $|a|$ and summing over n we get

$$
\begin{aligned}
\sum_{n=1}^{N} e^{2|a|\tau_n} |a| k_n &\le e\Big(\sum_{n=1}^{N} \int_{\tau_{n+1}}^{\tau_n} e^{2|a|\tau} d\tau\Big)|a| \\
&= e \int_{\tau_{N+1}}^{\tau_1} e^{2|a|\tau} |a| d\tau = e \int_0^{t_N} |a| e^{2|a|\tau} d\tau,
\end{aligned}
$$

which is the desired result and the proof is complete. $\square$

Parabolic case, $(a(t) \ge 0)$.

Theorem 9.5: Consider the $dG(0)$ approximation U for $\dot{u} + au = f, \underline{a(t) \ge 0}$. Assume that $k_j |a|_{I_j} \le \frac{1}{2}$ $\forall j$, then we have the error estimates

$$
|u(t_N) - U_N| \le \begin{cases} 3e^{2\lambda t_N} \max\limits_{0 \le t \le t_N} |k\dot{u}| & \text{if } |a(t)| \le \lambda \\ 3 \max\limits_{0 \le t \le t_N} |k\dot{u}| & \text{if } a(t) \ge 0 \end{cases}
$$

Proof:

Let $e = u - U = (u - \pi_k u) + (\pi_k u - U) := \tilde{e} + \bar{e}$, where $\tilde{e}$ is the interpolation error with $\pi_k u$ being the L_2-projection into $W_k^{(0)}$.

To estimate $\bar{e}$, we use *the discrete dual problem (DDP)*:

Find $\Phi \in W_k^{(0)}$, such that for $n = N, N-1, \ldots, 1$.

$$
(DDP) \qquad \begin{cases} \displaystyle\int_{t_{n-1}}^{t_n} (-\dot{\Phi} + a(t)\Phi) v\, dt - [\Phi_n] v_n = 0, \quad \forall v \in W_k^{(0)} \\ \Phi_N^+ = \Phi_{N+1} = (\pi_k u - U)_N :\equiv \bar{e}_N \end{cases}
$$

Let now $v = e$, then

$$
|\bar{e}_N|^2 = \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} (-\dot{\Phi} + a(t)\Phi)\bar{e}\, dt - \sum_{n=1}^{N-1} [\Phi_n]\bar{e}_n + \Phi_N \bar{e}_N. \tag{21}
$$

We now use $\bar{e} = (\pi_k u - U) = (\pi_k u - u + u - U)$ and write (21) as

$$\begin{aligned}|e_N|^2 = &\sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} (-\dot{\Phi} + a(t)\Phi)(\pi_k u - u + u - U)dt \\ &- \sum_{n=1}^{N-1} [\Phi_n](\pi_k u - u + u - U)_n + \Phi_N(\pi_k u - u + u - U)_N.\end{aligned}$$

Using Galerkin orthogonality we replace u by U. Therfore the total contributiuon from the terms with the factor $u - U$ is identical to zero. Thus, using the fact that $\dot{\Phi} = 0$ on each subinterval, we have the error representation formula:

$$\begin{aligned}|e_N|^2 &= \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} (-\dot{\Phi} + a(t)\Phi)(\pi_k u - u)dt - \sum_{n=1}^{N-1} [\Phi_n](\pi_k u - u)_n + \Phi_N(\pi_k u - u)_N \\ &= \int_0^{t_N} (a(t)\Phi)(u - \pi_k u)\, dt + \sum_{n=1}^{N-1} [\Phi_n](u - \pi_k u)_n - \Phi_N(u - \pi_k u)_N.\end{aligned}$$

To continue we need the following results:

Lemma 9.3. If $|a(t)| \le \lambda, \forall t \in (0, t_N)$ and $k_j |a|_{I_j} \le \frac{1}{2}, j = 1, 2, \ldots, N$, then the solution of the discrete dual problem satisfies

(i) $|\Phi_n| \le e^{2\lambda(t_N - t_{n-1})} |\bar{e}_N|$.

(ii) $\displaystyle\sum_{n=1}^{N-1} |[\Phi_n]| \le e^{2\lambda t_N} |\bar{e}_N|$.

(iii) $\displaystyle\sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} a(t)|\Phi_n| dt \le e^{2\lambda t_N} |\bar{e}_N|$.

(iv) If $a(t) \ge 0$ then $\quad \text{Max}\Big(|\Phi_n|, \sum_{n=1}^{N-1} |[\Phi_n]|, \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} a(t)|\Phi_n|\, dt\Big) \le |\bar{e}_N|$.

Proof: We show the last estimate (iv), (the proofs of (i)-(iii) are similar to that of the stability factor in Theorem 9.2).

Consider the discrete dual problem with $v \equiv 1$:

$$(DDP) \qquad \begin{cases} \displaystyle\int_{t_{n-1}}^{t_n} (-\dot{\Phi} + a(t)\Phi) v dt - [\Phi_n] v_n = 0, \forall v \in W_k^{(0)} \\ \Phi_{N+1} = (\pi_k u - U)_N :\equiv \bar{e}_N. \end{cases}$$

For dG(0) this becomes

$$(DDP)\quad \begin{cases} -\Phi_{n+1}+\Phi_n+\Phi_n\int_{t_{n-1}}^{t_n}a(t)=0, & n=N,N-1,\ldots,1\\ \Phi_{N+1}=\bar{e}_N, \qquad \Phi_n=\Phi|_{I_n}. \end{cases}$$

Iterating we get

$$\Phi_n=\prod_{j=n}^{N}\Big(1+\int_{I_j}a(t)dt\Big)^{-1}\Phi_{N+1} \tag{22}$$

For $a(t)\geq 0$ we have $\Big(1+\int_{I_j}a(t)dt\Big)^{-1}\leq 1$, thus (22) implies that

$$|\Phi_n|\leq\Phi_{N+1}=|\bar{e}_N|. \tag{23}$$

Further we have using (22) that

$$\Phi_{n-1}=\prod_{j=n-1}^{N}\Big(1+\int_{I_j}a(t)dt\Big)^{-1}\Phi_{N+1}=\Big(1+\int_{I_{n-1}}a(t)dt\Big)^{-1}\Phi_n\leq\Phi_n$$

which implies that

$$[\Phi_n]=\Phi_n^+-\Phi_n^-=\Phi_{n+1}-\Phi_n\geq 0.$$

Thus

$$\begin{aligned}\sum_{n=1}^{N}|[\Phi_n]|&=\Phi_{N+1}-\Phi_N+\Phi_N-\Phi_{N-1}+\ldots+\Phi_2-\Phi_1\\ &=\Phi_{N+1}-\Phi_1\leq\Phi_{N+1}\leq|\bar{e}_N|.\end{aligned} \tag{24}$$

Finally in the discrete equation:

$$\int_{t_{n-1}}^{t_n}(-\dot{\Phi}+a(t)\Phi)v\,dt-[\Phi_n]v_n=0, \qquad \forall v\in W_k^{(0)} \tag{25}$$

we have $v\equiv 1$ and $\dot{\Phi}\equiv 0$ for the dG(0). Hence (25) can be rewritten as

$$\int_{t_{n-1}}^{t_n}a(t)\Phi_n dt=[\Phi_n],$$

which gives, summing over n, that

$$\sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}a(t)\Phi_n dt\leq\sum_{n=1}^{N}[\Phi_n]\leq|\bar{e}_n|. \tag{26}$$

Combining (23), (24), and (26) the proof of (iv) is now complete.

As we mentioned earlier (i)-(iii) are proved in a similar way as Theorem 9.2(a). We omit the details. $\square$

Quadrature of f: (Assume that a=constant)

Then the error representation formula, combining $dG(0)$, with the quadrature role for f is as follows:

$$e_N^2 = \sum_{n=1}^{N} \Big(\int_{t_{n-1}}^{t_n} (f - aU)(\varphi - \pi_k\varphi)dt - [U_{n-1}](\varphi - \pi_k\varphi)_{n-1}^{+} \\ + \underbrace{\int_{t_{n-1}}^{t_n} f\pi_k\varphi dt - (\overline{f\pi_k\varphi})_n k_n}_{\text{quadrature error}} \Big)$$

where $\bar{g}_n = \begin{cases} g(t_n) & \text{for-the-endpoint-rule} \\ g(t_{(n-1/2)}) & \text{for-the-midpoint-rule} \end{cases}$

<u>Definition</u>:

If φ is the solution of the dual problem

$$\begin{cases} -\dot{\varphi} + a\varphi = 0, & \text{for} \quad t_N > t \geq 0 \\ \varphi(t_N) = e_N \end{cases}$$

then $\tilde{S}(t_N) = \dfrac{\int_0^{t_N} |\varphi| dt}{|e_N|}$ is called the *weak stability factor*.

<u>Note.</u> $\pi_k\varphi$ is piecewise constant and

$$\int_{I_n} |\pi_k\varphi(t)|dt \leq \int_{I_n} |\varphi(t)|dt.$$

We have the following relations between the two stability factors:

$$\tilde{S}(t_N) \leq t_N(1 + S(t_N)).$$

If $a > 0$ is sufficiently small, then $\tilde{S}(t_N) >> S(t_N)$.

Theorem 9.7: (The modified a posteriori estimate for dG(0)),

The $dG(0)$ approximation $U(t)$ computed using quadrature on terms involving f satisfies for $N = 1, 2, \ldots$

$$|u(t_n) - U_n| \leq S(t_n)|kR(U)|_{(0,t_N)} + \tilde{S}(t_N)C_j|k^j f^{(j)}|_{(0,t_N)},$$

where

$$R(U) = \frac{|U_n - U_{n-1}|}{k_n} + |f - aU|, \qquad \text{on} \;\; I_n$$

and $j = 1$ for the rectangle rule, $j = 2$ for the midpoint rule, $C_1 = 1$, $C_2 = \frac{1}{2}$, $f^{(1)} = \dot{f}$ and $f^{(2)} = \ddot{f}$.