F. Chaplais

# A finite element method for solving $n^{th}$ order differential equations

*The method presented here uses both the Strang and Fix conditions and the regularity of the finite elements. By keeping a $n^{th}$ order formulation, we reduce the number of equation variables at the expense of a greater regularity requirement on the atoms. We also show that an $n^{th}$ order theory of characeristics can be applied to the variations of the solutions.*

## 1. Assumptions and problem statement

Let $f$ a $C^n$ mapping from $\mathbb{R}^{n+1}$ into $\mathbb{R}$, Lipshitz with respect to its first $n$ arguments, and define, for a regular function $x$, $F(x,s) \stackrel{\text{def}}{=} f\left(x(s),\ldots,x^{(n-1)}(s),s\right)$. Given a polynomial $p$ of degree $\leq n-1$, we wish to compute an approximate solution of the $n^{th}$ order integral equation:

$$x(t) = p(t) + \int_0^t dt_1 \int_0^{t_1} \ldots \int_0^{t_{n-1}} F(x,t_n)dt_n \stackrel{\text{def}}{=} p(t) + I_t^n\left(F(x,t_n)\right) \tag{1}$$

To do so, we use a (small) step $\delta$ and two compactly supported functions $\varphi$ and $\varphi^*$ that define a biorthogonal system $\varphi_k(t) \stackrel{\text{def}}{=} \varphi(t/\delta - k)$, $k \in \mathbb{Z}$, and $\varphi_k^*(t) \stackrel{\text{def}}{=} (1/\delta)\varphi^*(t/\delta - k)$, $k \in \mathbb{Z}$. We define the projector $\Pi_\delta$ with $\Pi_\delta x = \sum_{k\in\mathbb{Z}} \varphi_k < x, \varphi_k^* >$. If $E \stackrel{\text{def}}{=} \text{Span}\{\varphi_k, k \in \mathbb{Z}\}$ includes all polynomials of degree $< r$, then $[\Pi_\delta x]^{(q)}$ approximates $x^{(q)}$ with an error of order $\delta^{r-q}$ if $x$ is of class $C^r$ ([2]). Hence the idea to approximate equation (1) by

$$z = p + \Pi_\delta\left[I_\bullet^n\left(F(z,t_n)\right)\right] \ , \ z \in E \tag{2}$$

which makes sense if $r \geq n$ and if $\varphi$ is of class $C^n$. We shall use the two following results:

L e m m a 1. *There exist $d$ and $a$ such that, for $x$ and $y$ in $E$, $|F(x(t),t) - F(y(t),t)| \leq a \sup_{|s-t|\leq d\delta} |x(s) - y(s)|$.*

L e m m a 2. *Assume that $|x(t)| \leq \Delta(|t|)$ with $\Delta$ nondecreasing. Then there exists $A$ and $\gamma$ such that $|\Pi_\delta I_t^n x| \leq A I^n_{|t|+\gamma\delta}\left[\Delta(|\bullet|)\right]$.*

## 2. Existence, unicity and regularity of the solutions of (2)

L e m m a 3. *Assume $\delta < (Aa)^{-1/n}[n/(e(\gamma+d))]$, and let $\mu_1$ and $\mu_2$ the nonnegative solutions of $Aa = \mu^n e^{-\mu(d+\gamma)\delta}$. Let $x$ and $y$ such that $|x(t) - y(t)| \leq M e^{\mu|t|}$, $\mu_1 < \mu < \mu_2$. Then there exists $K \in [0,1[$ such that*

$$|\Pi_\delta I_t^n\left[F(x) - F(y)\right]| \leq K M e^{\mu|t|} \tag{3}$$

P r o o f. This comes from the previous lemmas and the study of $h(x) = x^n e^{-x}$.

R e m a r k We shall assume the previous assumptions to be verified. If we have $\mu \leq n/(\delta(\gamma+d))$ then we can in fact use $\tilde{\mu} = n/(\delta(\gamma+d))$ instead, which leads to $K = Aa[(e\delta(d+\gamma))/n]^n$.

T h e o r e m 1. *Let $\mu_2$ as defined above, and $E_{\mu_2}$ the space of functions $x \in E$ which satisfy $|x(t)| \leq Me^{\mu|t|}$, for some $M \geq 0$ and $\mu < \mu_2$. Assume that $F(p,s)$ belongs to $E_{\mu_2}$. Then (2) has a unique solution $z$ in $E_{\mu_2}$; it is such that $|z - p| \leq Me^{\mu|t|}/(1-K)$. Moreover, if $g$ is a mapping similar to $f$ with $(|F(x,s) - G(x,s)| \leq me^{\mu|t|}\ \forall x)$ and $(|F(p,s) - G(p,s)| \leq m(ae^{\mu(|s|+d\delta)})/(1-K) \stackrel{\text{def}}{=} B(m))$, then the difference between the two solutions is bounded by $B(m)$.*

P r o o f. The unicity comes from the contraction property (3). The existence is proved constructively by using the fixed point algorithm $z_{n+1} = p + \Pi_\delta I_\bullet^n F(z_n)$ with $z_0 = p$. The regular perturbation result is obtained by starting the algorithm with $p$ and showing that the bound remains valid.

## 3. Application to the solving of ODEs

Theorem 2. *The solution $z$ of (2) approximates the solution $x$ of (1) at the order $r-n+1$ with respect to $\delta$.*

Proof. We first notice that $F(x(t),t)-F((\Pi_\delta x)(t),t)$ is of order $r-n+1$ with respect to $\delta$, and has an exponential behaviour with respect to time. On the other hand, $\Pi_\delta x$ satisfies an equation of the type (2) with the dynamics $G(z,t)=F(z,t)+F(x(t),t)-F((\Pi_\delta x)(t),t)$. Note that $G(z,t)-F(z,t)$ does not depend on $z$. Provided that the step $\delta$ is small enough with respect to the Lipschitz constant of $f$, this proves that $z-\Pi_\delta x$, and hence, $z-x$, is of order $r-n+1$.

Let us turn now to the numerical solving of (2). We assume $\delta < D(a)$ and $\mu \le n/(\delta(\gamma+d))$. Then $K$ is of order $n$ with respect to $\delta$; this shows that only a finite number $N$ of fixed point iterations, determined by the ratio $r/n$, is needed to get an optimal precision on the solution of (2), that is, one that is consistent with the Strang and Fix conditions.

To do these computations, we recall ([1],[3]) that there exists a family of $n$ biorthogonal systems with generators $\varphi^{(d)}$ and $\varphi^{*(d)}$ such that $d/dt\left(\sum_k x_k\varphi_k^{(d)}\right)=\sum_k(x_k-x_{k-1})\varphi_k^{(d+1)}$ with $\varphi^{(0)}=\varphi$ and $\varphi^{*(0)}=\varphi^*$. Denoting by $\Delta$ the difference operator $(\Delta x)_k \stackrel{\text{def}}{=} x_k-x_{k-1}$, the fixed point agorithm $z_{n+1}=p+\Pi_\delta I^n_\bullet F(z_n)$ implies

$$(\Delta^n z_{j+1})_k = <\varphi_k^{*(n)}, F(z_j,t)> \stackrel{\text{def}}{=} F_{j,k} \tag{4}$$

Because the finite elements are compactly supported, the computation of $F_{j,k}$ involves only the knowledge of $z_{j,l} \stackrel{\text{def}}{=} < z_j,\varphi_l^* >$, for $k-a\le l\le k+b$ and some fixed integers $a$ and $b$. This suggests the following algorithm:

- compute the $z_{j,k}$, $1\le j\le N$, $(N-n)b\le k\le n-1+(N-n)b$, starting from $z_0=p$ and using the integral formulation of the fixed point algorithm
- given $z_{j,k}$, $1\le j\le N$, $p+(N-n)b\le k\le p+n-1+(N-n)b$, increase $p$ by one by using the N $n^{th}$ order discrete recursions (4).

We see that, while (2) links all of the coordinates of $z$, computing only a finite number of fixed point iterates $z_j$s makes it possible to design an algorithm that is causal with respect to the index $k$ of the coordinates $z_{j,k}$ of the $z_j$s.

## 4. Sliding along the solutions

An interesting feature of the integral formulation (2) is that it allows to study the variation of $z$ with repect to $p$ and the "initial" time.

Theorem 3. *Let us define $I^n_{a,b}x=\int_a^b dt_1\int_a^{t_1}\dots\int_a^{t_{n-1}}x(t_n)dt_n$ and $z(a_0,\dots,a_{n-1},t_0)$ the solution of $z(t)=\sum_{k=0}^{k=n-1}a_k\left[(t-t_0)^{n-1-k}/(n-1-k)!\right]+\Pi_\delta I^n_{t_0,t}F(z,t_n)$. Then we have the following characteristic equation*

$$\frac{\partial z}{\partial t_0}+\frac{\partial z}{\partial a_0}F(z,t)(t_0)+\sum_{i=1}^{i=n-1}\frac{\partial z}{\partial a_i}a_{i-1}=0 \tag{5}$$

Proof. The reader will check that $I^n_{t_0,t}x=I^n_{t_0+h,t}x+\sum_{k=0}^{k=n-1}\left[(t-t_0)^{n-k-1}/(n-k-1)!\right]I^1_{t_0,t}I^k_{t_0+h,\bullet}x$. Defining $b_i(h)=\sum_{k=0}^{k=i}\left(h^{i-k}/(i-k)!\right)\left[a_k+I^1_{t_0,t_0+h}I^k_{t_0+h,\bullet}F(z)\right]$, we see then that

$$z(a_0,\dots,a_{n-1},t_0)=z(b_0(h),\dots,b_{n-1}(h),t_0+h)\ \forall h \tag{6}$$

## 5. References


1 Cieselski Z. & Fiegiel T.: Spline bases in classical function spaces on compact manifolds, part II, Studia Math. **LXXVCI** (1983), 95-136.

2 Fix G. & Strang G.: Fourier Analysis of the Finite element method in Ritz-Galerkin theory, Studia Appl. Math. **48** (1969), 265-273.

3 Lemarie-Rieusset P.G.: Analyses multi-resolutions non orthogonales, commutation entre projecteurs et dérivation et ondelettes vecteurs à divergence nulle, Revista Matemática Iberoamericana, **Vol. 8 No 2** (1992), 221-237.



*Addresses:* Prof. F. Chaplais, CAS-Ecole des Mines de Paris, 35 rue Saint-Honoré, 77305 Fontainebleau Cedex, France. e-mail: chaplais@cas.ensmp.fr.