Optimality and resonances in a class of compact finite difference schemes of high order

Joackim Bernier

IRMAR (Rennes)

January 31, 2018

Motivation : Understand construction and qualitative properties of compact *finite difference* schemes of high order for *elliptic problems* .

Problem : It is too general ! So we focus on an elementary problem: *homogeneous Dirichlet problem in dimension* 1.

Motivation : Understand construction and qualitative properties of compact *finite difference* schemes of high order for *elliptic problems* .

Problem : It is too general ! So we focus on an elementary problem: *homogeneous Dirichlet problem in dimension* 1.

For a given $f : \mathbb{R} \to \mathbb{C}$, find $u : [0,1] \to \mathbb{C}$ such that

$$\left\{ \begin{array}{l} -u''(x) = f(x), \ \forall x \in]0,1[, \\ u(0) = u(1) = 0. \end{array} \right.$$

Compact finite difference scheme ?

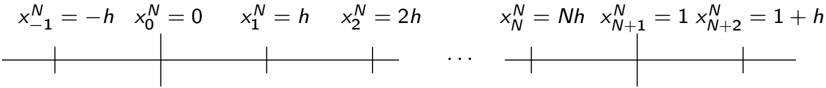

Figure: Regular grid with N points into $]0,1[$.

A compact finite difference scheme is a linear system

$$\mathbf{D}_N \mathbf{u}^N = h^2 \mathbf{S}_N \mathbf{f}^{N,ex}$$

where

- $\mathbf{f}^{N,ex} = (f(x_j^N))_{j \in \mathbb{Z}}$,
- $\mathbf{u}^N \simeq (u(x_j^N))_{j=1\dots N}$ is an approximation of the solution of the Dirichlet problem,
- $\mathbf{D}_N$ and $\mathbf{S}_N$ are matrices.

Examples:

$$\mathbf{D}_N = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} \in \mathscr{L}(\mathbb{C}^N).$$

and (small abuse of notations)

$$\mathbf{S}_N = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix} \in \mathscr{L}(\mathbb{C}^N).$$

$$\mathbf{D}_N = \frac{1}{12}\begin{pmatrix} 29 & -16 & 1 & & & & \\ -16 & 30 & -16 & 1 & & & \\ 1 & -16 & 30 & -16 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -16 & 30 & -16 & 1 \\ & & & 1 & -16 & 30 & -16 \\ & & & & 1 & -16 & 29 \end{pmatrix},$$

and

$$\mathbf{S}_N = \begin{pmatrix} \frac{1}{12} & 1 & & & & & & \\ & & 1 & & & & & \\ & & & 1 & & & & \\ & & & & \ddots & & & \\ & & & & & 1 & & \\ & & & & & & 1 & \\ & & & & & & 1 & \frac{1}{12} \end{pmatrix}.$$

$$\mathbf{D}_N = \frac{1}{20}\begin{pmatrix} 40 & -20 & & & & & & \\ -16 & 34 & -16 & -1 & & & & \\ -1 & -16 & 34 & -16 & -1 & & & \\ & -1 & -16 & 34 & -16 & -1 & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & -1 & -16 & 34 & -16 & -1 \\ & & & & -1 & -16 & 34 & -16 \\ & & & & & & -20 & 40 \end{pmatrix},$$

and

$$\mathbf{S}_N = \frac{1}{60}\begin{pmatrix} 5 & 50 & 5 & & & & & \\ & 8 & 44 & 8 & & & & \\ & & 8 & 44 & 8 & & & \\ & & & \ddots & \ddots & \ddots & & \\ & & & & 8 & 44 & 8 & \\ & & & & & 5 & 50 & 5 \end{pmatrix}.$$

Expectations: convergence of u^N

$$\lim_{N\to\infty} \sup_{j=1,\dots,N} |u_j^N - u(x_j^N)| = 0.$$

In general, we expect there exists $n \in \mathbb{N}^*$, the *order of the scheme* , and $C > 0$ such that

$$|u_j^N - u(x_j^N)| \leq Ch^n, \ \forall N, \forall j.$$

Expectations: convergence of u^N

$$\lim_{N\to\infty} \sup_{j=1,\dots,N} |u_j^N - u(x_j^N)| = 0.$$

In general, we expect there exists $n \in \mathbb{N}^*$, the *order of the scheme* , and $C > 0$ such that

$$|u_j^N - u(x_j^N)| \le Ch^n, \ \forall N, \forall j.$$

Questions:

- In general, are these schemes convergent ?
- Are some of them more efficient than others ?

1. Consistency + Stability $\Rightarrow$ Convergence
2. Construction of the schemes
3. Optimality
4. Resonances

Consistency: (order n) $\forall f \in C^\infty(\mathbb{R}), \exists c > 0, \forall N \in \mathbb{N}^*,$

$$\|\mathbf{D}_N \mathbf{u}^{N,ex} - h^2 \mathbf{S}_N \mathbf{f}^{N,ex}\|_\infty \leq c h^{n+2}.$$

Consistency: (order n) $\forall f \in C^\infty(\mathbb{R}), \exists c > 0, \forall N \in \mathbb{N}^*$,

$$\|\mathbf{D}_N \mathbf{u}^{N,ex} - h^2 \mathbf{S}_N \mathbf{f}^{N,ex}\|_\infty \leq c h^{n+2}.$$

Weak Consistency: (order n) $\exists l \in \mathbb{N}, \forall f \in C^\infty(\mathbb{R}), \exists c > 0, \forall N \in \mathbb{N}^*$,

$$\left| \left(\mathbf{D}_N \mathbf{u}^{N,ex}\right)_j - h^2 \left(\mathbf{S}_N \mathbf{f}^{N,ex}\right)_j \right| \leq \begin{cases} c h^{n+2} & \text{if } l < j < N+1-l, \\ c h^n & \text{else.} \end{cases}$$

A scheme $(\mathbf{D}_N, \mathbf{S}_N)_N$ (or a sequence of matrix $(\mathbf{D}_N)_N$) is

- stable, if there exists a positive constant $c > 0$ such that for all $N \in \mathbb{N}^*$, we have

$$\forall \mathbf{v} \in \mathbb{C}^N,\ c\|\mathbf{v}\|_\infty \leq h^{-2}\|\mathbf{D}_N\mathbf{v}\|_\infty.$$

A scheme $(\mathbf{D}_N, \mathbf{S}_N)_N$ (or a sequence of matrix $(\mathbf{D}_N)_N$) is

- stable, if there exists a positive constant $c > 0$ such that for all $N \in \mathbb{N}^*$, we have

$$\forall \mathbf{v} \in \mathbb{C}^N,\ c\|\mathbf{v}\|_\infty \leq h^{-2}\|\mathbf{D}_N\mathbf{v}\|_\infty.$$

- strongly stable, if for all $l \in \mathbb{N}$, there exists a positive constant $c > 0$ such that for all $N \in \mathbb{N}^*$,

$$\forall \mathbf{v} \in \mathbb{C}^N,\ c\|\mathbf{v}\|_\infty \leq \sup_{j=1,\dots,N} \begin{cases} h^{-2}\,(\mathbf{D}_N\mathbf{v})_j & \text{if } l < j < N+1-l, \\ (\mathbf{D}_N\mathbf{v})_j & \text{else.} \end{cases}.$$

A scheme $(\mathbf{D}_N, \mathbf{S}_N)_N$ (or a sequence of matrix $(\mathbf{D}_N)_N$) is

- stable, if there exists a positive constant $c > 0$ such that for all $N \in \mathbb{N}^*$, we have

$$\forall \mathbf{v} \in \mathbb{C}^N,\ c\|\mathbf{v}\|_\infty \leq h^{-2}\|\mathbf{D}_N\mathbf{v}\|_\infty.$$

- strongly stable, if for all $l \in \mathbb{N}$, there exists a positive constant $c > 0$ such that for all $N \in \mathbb{N}^*$,

$$\forall \mathbf{v} \in \mathbb{C}^N,\ c\|\mathbf{v}\|_\infty \leq \sup_{j=1,\dots,N} \begin{cases} h^{-2}(\mathbf{D}_N\mathbf{v})_j & \text{if } l < j < N+1-l, \\ (\mathbf{D}_N\mathbf{v})_j & \text{else.} \end{cases}.$$

- stable relatively to a sequence $(\eta_N)_{N\in\mathbb{N}^*}$ of positive numbers, if there exists a positive constant $c > 0$ such that for all $N \in \mathbb{N}^*$, we have

$$\forall \mathbf{v} \in \mathbb{C}^N,\ c\|\mathbf{v}\|_\infty \leq \eta_N\|\mathbf{D}_N\mathbf{v}\|_\infty.$$

Theorem: Lax

- A scheme that is strongly stable and weakly consistent of order n is convergent of order n.
- If $\eta_N h^{n+2} \to_{N\to\infty} 0$ then a scheme that is stable relatively to the sequence $(\eta_N)_{N\in\mathbb{N}^*}$ and consistent of order n is convergent at the rate $\epsilon_N = \eta_N h^{n+2}$.

Proof.

$$\mathbf{D}_N\mathbf{v} = \mathbf{D}_N\left(\mathbf{u}^{N,ex} - \mathbf{u}^N\right) = \mathbf{D}_N\mathbf{u}^{N,ex} - h^2\mathbf{S}_N\mathbf{f}^{N,ex}.$$

□

A couple of *finite difference formulas* $(d, s) \in \left(\mathbb{C}^{(\mathbb{Z})}\right)^2$ is consistent of order n, if

$$\forall u \in C^\infty(\mathbb{R}),\ \sum_{j \in \mathbb{Z}} d_j u(x_j^N) + h^2 s_j u''(x_j^N) = \mathcal{O}(h^{n+2}).$$

Example: $d = 2\mathbb{1}_{\{0\}} - \mathbb{1}_{\{-1,1\}} = "(-1, 2, -1)"$ and $s = \mathbb{1}_{\{0\}} = "(1)"$.

A couple of *finite difference formulas* $(d, s) \in \left(\mathbb{C}^{(\mathbb{Z})}\right)^2$ is consistent of order n, if

$$\forall u \in C^\infty(\mathbb{R}),\ \sum_{j\in\mathbb{Z}} d_j u(x_j^N) + h^2 s_j u''(x_j^N) = \mathcal{O}(h^{n+2}).$$

Example: $d = 2\mathbb{1}_{\{0\}} - \mathbb{1}_{\{-1,1\}} = "(-1, 2, -1)"$ and $s = \mathbb{1}_{\{0\}} = "(1)"$.

Remark : We only consider *symmetric* finite difference formulas

$$\forall j \in \mathbb{Z},\ d_j = d_{-j} \text{ and } s_j = s_{-j}.$$

Choosing $u \equiv 1$ we get

$$\sum_{j\in\mathbb{Z}} d_j = 0.$$

It is easy to get such couples. For example, choose any $d \in \mathbb{C}^{(\mathbb{Z})}$ symmetric and satisfying

$$\sum_{j\in\mathbb{Z}} d_j = 0$$

then we can get $s \in \mathbb{C}^{(\mathbb{Z})}$ such that (d,s) is consistent of order n solving the linear system

$$(\frac{s_0}{2}, s_1, \dots, s_{\frac{n}{2}-1})((i-1)^{2j-2})_{1\leq i,j\leq \frac{n}{2}} = -\sum_{j>0} d_j \left(\frac{j^2}{2}, \dots, \frac{j^n}{n(n-1)}\right).$$

To design matrices $\mathbf{D}_N$ and $\mathbf{S}_N$ from the formulas d and s, a natural choice would be the following:

$$(\mathbf{D}_N\mathbf{u})_i = \sum_{j\in\mathbb{Z}} d_{i-j}\mathbf{u}_j \text{ and } (\mathbf{S}_N\mathbf{f})_i = \sum_{j\in\mathbb{Z}} s_{i-j}\mathbf{f}_j.$$

To design matrices $\mathbf{D}_N$ and $\mathbf{S}_N$ from the formulas d and s, a natural choice would be the following:

$$(\mathbf{D}_N\mathbf{u})_i = \sum_{j\in\mathbb{Z}} d_{i-j}\mathbf{u}_j \text{ and } (\mathbf{S}_N\mathbf{f})_i = \sum_{j\in\mathbb{Z}} s_{i-j}\mathbf{f}_j.$$

Problem: it does not make sense on the boundary !

If we define *the size of the stencil* of d as

$$\tau(d) = \max\{j \in \mathbb{Z} \mid d_j \neq 0\},$$

then the previous formula involves terms like $\mathbf{u}_{1-\tau(d)}$.

Solution: To introduce formulas (d^i, s^i) consistent of order n (or $n-2$) at a distance i of the boundary.

$$(\mathbf{D}_N\mathbf{u})_i := \begin{cases} \sum_{j>0} d^i_{j-i}\mathbf{u}_j & \text{if } 1 \leq i \leq \tau(d), \\ \sum_{j\in\mathbb{Z}} d_{j-i}\mathbf{u}_j & \text{if } \tau(d) < i < N+1-\tau(d), \\ \sum_{j<N+1} d^{N+1-i}_{-j+i}\mathbf{u}_j & \text{if } N+1-\tau(d) \leq i \leq N+1. \end{cases}$$

$$(\mathbf{S}_N\mathbf{f})_i := \begin{cases} \sum_{j\in\mathbb{Z}} s^i_{j-i}\mathbf{f}_j & \text{if } 1 \leq i \leq \tau(d), \\ \sum_{j\in\mathbb{Z}} s_{j-i}\mathbf{f}_j & \text{if } \tau(d) < i < N+1-\tau(d), \\ \sum_{j\in\mathbb{Z}} s^{N+1-i}_{-j+i}\mathbf{f}_j & \text{if } N+1-\tau(d) \leq i \leq N+1. \end{cases}$$

(d^i, s^i) is consistent of order $\mu \in \{n, n-2\}$ at a distance i of the boundary if

$$\forall u \in C^\infty(\mathbb{R}),\ u(0) = 0 \ \Rightarrow \ \sum_{j>-i} d_j^i u(x_{j+i}^N) + h^2 \sum_{j\in\mathbb{Z}} s_j^i u''(x_{j+i}^N) = \mathcal{O}(h^{\mu+2}), \tag{1}$$

(d^i, s^i) is consistent of order $\mu \in \{n, n-2\}$ at a distance i of the boundary if

$$\forall u \in C^\infty(\mathbb{R}),\ u(0) = 0 \ \Rightarrow \ \sum_{j > -i} d^i_j u(x^N_{j+i}) + h^2 \sum_{j \in \mathbb{Z}} s^i_j u''(x^N_{j+i}) = \mathcal{O}(h^{\mu+2}), \quad (1)$$

Lemma

This scheme is consistent of order n (weakly if (d^i, s^i) consistent of order $n-2$).

(d^i, s^i) is consistent of order $\mu \in \{n, n-2\}$ at a distance i of the boundary if

$$\forall u \in C^\infty(\mathbb{R}),\ u(0) = 0 \ \Rightarrow \ \sum_{j>-i} d^i_j u(x^N_{j+i}) + h^2 \sum_{j\in\mathbb{Z}} s^i_j u''(x^N_{j+i}) = \mathcal{O}(h^{\mu+2}), \tag{1}$$

Lemma

This scheme is consistent of order n (weakly if (d^i, s^i) consistent of order $n-2$).

Question : How to choose (d^i, s^i)?

This choice is *crucial* to hope stability. For example, we could choose $d^i = s^i = 0$ but the scheme would not be stable !

There are methods based on *monotonicity* (i.e. $(\mathbf{D}_N^{-1})_{i,j} \geq 0$) but these methods are not very general and it is not clear if arbitrarily high order schemes may be designed.

Our choice: We keep the relation

$$(\mathbf{D}_N \mathbf{u})_i = \sum_{j \in \mathbb{Z}} d_{i-j} \mathbf{u}_j, \ \forall j = 1, \ldots, N,$$

extending $\mathbf{u}$ as an odd sequence in 0 and $N+1$.

Our choice: We keep the relation

$$(\mathsf{D}_N \mathbf{u})_i = \sum_{j \in \mathbb{Z}} d_{i-j} \mathbf{u}_j, \ \forall j = 1, \dots, N,$$

extending $\mathbf{u}$ as an odd sequence in 0 and $N+1$.
With the previous formalism we get

$$d^i_j = d_j - d_{2i+j}, \quad i = 1, \dots, \tau(d), \quad j \in \mathbb{Z}$$

Our choice: We keep the relation

$$(\mathbf{D}_N \mathbf{u})_i = \sum_{j \in \mathbb{Z}} d_{i-j} \mathbf{u}_j, \ \forall j = 1, \ldots, N,$$

extending $\mathbf{u}$ as an odd sequence in 0 and $N+1$.
With the previous formalism we get

$$d_j^i = d_j - d_{2i+j}, \quad i = 1, \ldots, \tau(d), \quad j \in \mathbb{Z}$$

Lemma

There exists s^i such that (d^i, s^i) is consistent of order $\mu \in \{n, n-2\}$ at a distance i of the boundary.

Sketch of proof: Let $u \in C^\infty(\mathbb{R})$ such that $u(0) = 0$

$$\sum_{j>-i} d_j^i u(x_{i+j}^N) = \sum_{j\in\mathbb{Z}} d_j u(x_{j+i}^N) - \sum_{j<-i} d_j u(x_{j+i}^N) - \sum_{j>-i} d_{j+2i} u(x_{j+i}^N)$$
$$= -h^2 \sum_{j\in\mathbb{Z}} s_j u''(x_{j+i}^N) - \sum_{j>0} d_{i+j} \left(u(x_j^N) + u(x_{-j}^N) \right) + \mathcal{O}(h^{n+2}).$$

However, we have

$$u(x_j^N) + u(x_{-j}^N) = u(x_j^N) + u(x_{-j}^N) - 2u(0) = -h^2 \sum_{l\in\mathbb{Z}} b_l^j u''(x_l^N) + \mathcal{O}(h^{\mu+2}),$$

where b^j is obtained solving the linear system

$$(\frac{b_0^j}{2}, s_1, \dots, b_{\frac{\mu}{2}-1}^j)((i-1)^{2j-2})_{1\leq i,j\leq \frac{\mu}{2}} = -\left(\frac{j^2}{2}, \dots, \frac{j^\mu}{\mu(\mu-1)}\right).$$

Question : Why such a choice on the boundary ?

Question : Why such a choice on the boundary ?
We have constructed

$$\begin{array}{ccc} \mathscr{S}_{\mathbb{C}} & \to & M_N(\mathbb{C}) \\ d & \mapsto & \mathsf{D}_N(d) \end{array}$$

with $\mathscr{S}_{\mathbb{C}}$ the space of symmetric complex valued formulas.

Question : Why such a choice on the boundary ?
We have constructed

$$\begin{array}{ccc} \mathscr{S}_{\mathbb{C}} & \to & M_N(\mathbb{C}) \\ d & \mapsto & \mathsf{D}_N(d) \end{array}$$

with $\mathscr{S}_{\mathbb{C}}$ the space of symmetric complex valued formulas. If we equip $\mathscr{S}_{\mathbb{C}}$ of its structure of *algebra* for the *convolution* , it is a morphism of algebra !

Question : Why such a choice on the boundary ?
We have constructed

$$\begin{array}{rcl} \mathscr{S}_{\mathbb{C}} & \to & M_N(\mathbb{C}) \\ d & \mapsto & \mathsf{D}_N(d) \end{array}$$

with $\mathscr{S}_{\mathbb{C}}$ the space of symmetric complex valued formulas. If we equip $\mathscr{S}_{\mathbb{C}}$ of its structure of *algebra* for the *convolution* , it is a morphism of algebra !

However, if $a = 2\mathbb{1}_{\{0\}} - \mathbb{1}_{\{-1,1\}} = "(-1, 2, -1)"$ then

$$\begin{array}{rcl} \mathbb{C}[X] & \to & \mathscr{S}_{\mathbb{C}} \\ P & \mapsto & P(a) \end{array} \text{ is an isomorphism of algebra.}$$

Question : Why such a choice on the boundary ?
We have constructed

$$\begin{array}{ccc} \mathscr{S}_{\mathbb{C}} & \to & M_N(\mathbb{C}) \\ d & \mapsto & \mathbf{D}_N(d) \end{array}$$

with $\mathscr{S}_{\mathbb{C}}$ the space of symmetric complex valued formulas. If we equip $\mathscr{S}_{\mathbb{C}}$ of its structure of *algebra* for the *convolution* , it is a morphism of algebra !

However, if $a = 2\mathbb{1}_{\{0\}} - \mathbb{1}_{\{-1,1\}} = "(-1, 2, -1)"$ then

$$\begin{array}{ccc} \mathbb{C}[X] & \to & \mathscr{S}_{\mathbb{C}} \\ P & \mapsto & P(a) \end{array} \quad \text{is an isomorphism of algebra.}$$

Conclusion: if $d = P(a)$ then

$$\mathbf{D}_N(d) = \mathbf{D}_N(P(a)) = P(\mathbf{D}_N(a)) =: P(\mathbf{A}_N).$$

$\mathbf{A}_N$ is the square matrix defined by

$$\mathbf{A}_N = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} \in \mathscr{L}(\mathbb{C}^N).$$

It is a well known matrix whose spectral decomposition is given by

$$\mathbf{A}_N \mathbf{e}_k^N = 4\sin^2\left(\frac{\pi}{2}kh\right)\mathbf{e}_k^N, \quad \text{with } \mathbf{e}_k^N := (\sin(\pi khj))_{j=1,\dots,N}.$$

$\mathbf{A}_N$ is the square matrix defined by

$$\mathbf{A}_N = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} \in \mathscr{L}(\mathbb{C}^N).$$

It is a well known matrix whose spectral decomposition is given by

$$\mathbf{A}_N \mathbf{e}_k^N = 4\sin^2\left(\frac{\pi}{2}kh\right)\mathbf{e}_k^N, \quad \text{with } \mathbf{e}_k^N := (\sin(\pi khj))_{j=1,\dots,N}.$$

Conclusion: To get convergence, we just have to study stability of matrices of the type $P(\mathbf{A}_N)$!

If $l = \tau(d) - 1$ and $m = \tau(s)$ the matrices we have constructed are

$$\mathbf{D}_N(d) = \begin{pmatrix} d_{\mathbf{0}} & \cdots & d_{l+\mathbf{1}} & & & \\ \vdots & \ddots & & \ddots & & \\ & & \ddots & & \ddots & \\ d_{l+\mathbf{1}} & & & & & \\ & \ddots & & \ddots & & d_{l+\mathbf{1}} \\ & & \ddots & & \ddots & \vdots \\ & & & d_{l+\mathbf{1}} & \cdots & d_{\mathbf{0}} \end{pmatrix} - \begin{pmatrix} d_{\mathbf{2}} & \cdots & d_{l+\mathbf{1}} & & & \\ \vdots & \iddots & & & & \\ d_{l+\mathbf{1}} & & & & & \\ & & & & & d_{l+\mathbf{1}} \\ & & & & \iddots & \vdots \\ & & & d_{l+\mathbf{1}} & \cdots & d_{\mathbf{2}} \end{pmatrix},$$

$$\mathbf{S}_N = \begin{pmatrix} s_m & \cdots & s_{\mathbf{0}} & \cdots & s_m & & & & \\ & \ddots & & \ddots & & \ddots & & & \\ & & \ddots & & \ddots & & \ddots & & \\ & & & \ddots & & \ddots & & \ddots & \\ & & & & \ddots & & \ddots & & \ddots \\ & & & & s_m & \cdots & s_{\mathbf{0}} & \cdots & s_m \end{pmatrix} + \begin{pmatrix} \mathbf{B}_\mu^+ & & \\ & \mathbf{0}_{N-\mu+\mathbf{2},N-\mu+\mathbf{2}} & \\ & & \mathbf{B}_\mu^- \end{pmatrix}.$$

To get efficient schemes, we want to **minimise l and m** but to **conserve** the consistency **order.**

We can look for symmetric formulas d, s as polynomial of a

$$d = P(a) \text{ and } s = Q(a).$$

Considering formulas of consistency order larger than or equal 2, necessarily we have $P(0) = 0$. Consequently, it can be factorized

$$P = XR.$$

We can look for symmetric formulas d, s as polynomial of a

$$d = P(a) \text{ and } s = Q(a).$$

Considering formulas of consistency order larger than or equal 2, necessarily we have $P(0) = 0$. Consequently, it can be factorized

$$P = XR.$$

We remark that

$$\tau(d) = 1 + \deg R \text{ and } \tau(s) = \deg Q.$$

We can look for symmetric formulas d, s as polynomial of a

$$d = P(a) \text{ and } s = Q(a).$$

Considering formulas of consistency order larger than or equal 2, necessarily we have $P(0) = 0$. Consequently, it can be factorized

$$P = XR.$$

We remark that

$$\tau(d) = 1 + \deg R \text{ and } \tau(s) = \deg Q.$$

Lemma

(d, s) is consistent of order $n = 2k$ if and only if

$$R(X) = C(X)Q(X) \mod X^k.$$

where

$$C(X) := 4\left(\frac{\arcsin(\frac{\sqrt{X}}{2})}{\sqrt{X}}\right)^2 = 2\sum_{n\in\mathbb{N}} \frac{X^n}{(n+1)^2 C_{2n+2}^{n+1}}.$$

Looking for a solution of $R(X) = C(X)Q(X) \mod X^k$ with $\deg R \leq l$ and $\deg Q \leq m$ we get a linear system with

- k equations,
- $l + m + 2$ unknowns.

Looking for a solution of $R(X) = C(X)Q(X) \mod X^k$ with $\deg R \leq l$ and $\deg Q \leq m$ we get a linear system with

- k equations,
- $l + m + 2$ unknowns.

So, if $l + m + 1 \geq k$ there is at least one non trivial solution.

Looking for a solution of $R(X) = C(X)Q(X) \mod X^k$ with $\deg R \leq l$ and $\deg Q \leq m$ we get a linear system with

- k equations,
- $l + m + 2$ unknowns.

So, if $l + m + 1 \geq k$ there is at least one non trivial solution.

Lemma

For any fixed, l, m, k such that $l + m + 1 = k$, this solution is unique up to multiplication by a constant.

We denote it $R_{l,m}$ and $Q_{l,m}$ (choosing $R_{l,m}(0) = 1$).

Sketch of proof: It is a classical problem of *Padé approximant.* Indeed, we are looking for a rational approximation $\frac{R}{Q}$ of C in $X = 0$.

Following classical theory, this result is a corollary of the fact that $C(-4X)$ is a Stieltjes transform of a positive function $\rho : (0, 1) \to \mathbb{R}$,

$$C(-4X) = \int_0^1 \frac{\rho(s)}{1 + sX} \, \mathrm{d}s \, .$$

Theorem: D. Karp and E. Prilepkina (2012)

$$\text{If} \begin{cases} 1 \leq \alpha_1 \leq \cdots \leq \alpha_q, \\ 0 < \beta_1 \leq \cdots \leq \beta_q, \\ \forall k \in [\![1, q]\!], \ \sum_{j=1}^{k} \alpha_j \leq \sum_{j=1}^{k} \beta_j \end{cases}$$

then ${}_{q+1}F_q\left[\begin{smallmatrix} 1, \alpha_1, \ldots, \alpha_q \\ \beta_1, \ldots, \beta_q \end{smallmatrix}; -X\right]$ is a Stieltjes transform of a positive function $\rho : (0,1) \to \mathbb{R}$.

$${}_pF_q\left[\begin{matrix} \alpha_1, \ldots, \alpha_p \\ \beta_1, \ldots, \beta_q \end{matrix}; X\right] := \sum_{k \in \mathbb{N}} \frac{(\alpha_1)_k \ldots (\alpha_p)_k}{(\beta)_1 \ldots (\beta_q)_k} \frac{X^k}{k!} \text{ with } (\gamma)_k = \prod_{j=0}^{k-1} \gamma + j.$$

Theorem: D. Karp and E. Prilepkina (2012)

$$\text{If} \begin{cases} 1 \leq \alpha_1 \leq \cdots \leq \alpha_q, \\ 0 < \beta_1 \leq \cdots \leq \beta_q, \\ \forall k \in [\![1, q]\!], \ \sum_{j=1}^{k} \alpha_j \leq \sum_{j=1}^{k} \beta_j \end{cases}$$

then ${}_{q+1}F_q\left[{1,\alpha_1,\dots,\alpha_q \atop \beta_1,\dots,\beta_q}; -X\right]$ is a Stieltjes transform of a positive function $\rho : (0,1) \to \mathbb{R}$.

$${}_pF_q\left[{\alpha_1,\dots,\alpha_p \atop \beta_1,\dots,\beta_q}; X\right] := \sum_{k\in\mathbb{N}} \frac{(\alpha_1)_k \dots (\alpha_p)_k}{(\beta)_1 \dots (\beta_q)_k} \frac{X^k}{k!} \text{ with } (\gamma)_k = \prod_{j=0}^{k-1} \gamma + j.$$

$$C(X) = {}_3F_2\left[{1,1,1 \atop \frac{3}{2},2}; \frac{X}{4}\right].$$

Corollary: Zero points of $R_{l,m}$ are localised in $(4, \infty)$.

Theorem

Let $P \in \mathbb{C}[X]$ be a polynomial such that

$$P(0) = 0, \; P'(0) \neq 0 \text{ and } \forall x \in]0, 4], \; P(x) \neq 0.$$

Then the sequence of matrices $(P(\mathbf{A}_N))_{N \in \mathbb{N}^*}$ is strongly stable.

Application: $P = XR_{l,m}$.

Corollary: Zero points of $R_{l,m}$ are localised in $(4, \infty)$.

Theorem

Let $P \in \mathbb{C}[X]$ be a polynomial such that

$$P(0) = 0, \; P'(0) \neq 0 \text{ and } \forall x \in]0, 4], \; P(x) \neq 0.$$

Then the sequence of matrices $(P(\mathbf{A}_N))_{N \in \mathbb{N}^*}$ is strongly stable.

Application: $P = XR_{l,m}$.

Conclusion: If we want a scheme of order $n = 2k$ with $l = \tau(d) + 1$ and $m = \tau(m)$ as small as possible, we have to choose $l + m + 1 = k$. In this case there exists an unique normalized scheme given by $d = (XR_{l,m})(a)$ and $s = Q_{l,m}(a)$. Furthermore, this scheme is strongly stable and so convergent of order n.

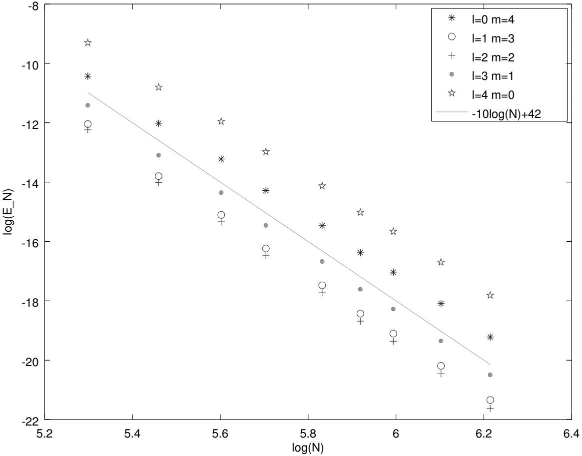

Figure: Convergence curves, with $u(x) = x(1-x)e^{4\cos(41x)}$ and $E_N := \|\mathbf{u}^N - \mathbf{u}^{N,ex}\|_\infty$, $N \in \{200, 235, 271, 300, 341, 372, 401, 447, 500\}$, for the optimal schemes, with $n = 10$, $\mu = 8$, $d = d^{l,m}$ and $s = s^{l,m}$.

Sketch of proof: This property is well known for $\mathbf{A}_N$, so we just have to prove that if R does not vanish on $[0,4]$ then

$$\sup_{N\in\mathbb{N}^*} \|R(\mathbf{A}_N)^{-1}\|_\infty < \infty.$$

Sketch of proof: This property is well known for $\mathbf{A}_N$, so we just have to prove that if R does not vanish on $[0,4]$ then

$$\sup_{N\in\mathbb{N}^*} \|R(\mathbf{A}_N)^{-1}\|_\infty < \infty.$$

Consider the Fourier transform of $R^{-1}(4\sin^2(\frac{\theta}{2}))$

$$R^{-1}(4\sin^2(\frac{\theta}{2})) = \sum_{n\in\mathbb{N}} a_n \cos(n\theta), \ (a_n) \in \ell^1.$$

Sketch of proof: This property is well known for $\mathbf{A}_N$, so we just have to prove that if R does not vanish on $[0,4]$ then

$$\sup_{N\in\mathbb{N}^*} \|R(\mathbf{A}_N)^{-1}\|_\infty < \infty.$$

Consider the Fourier transform of $R^{-1}(4\sin^2(\frac{\theta}{2}))$

$$R^{-1}(4\sin^2(\frac{\theta}{2})) = \sum_{n\in\mathbb{N}} a_n \cos(n\theta),\ (a_n) \in \ell^1.$$

Then observe that

$$\mathbf{D}_N(\frac{\mathbb{1}_{\{-n,n\}}}{2})\mathbf{e}_k^N = \cos(n\pi kh)\mathbf{e}_k^N.$$

Sketch of proof: This property is well known for $\mathbf{A}_N$, so we just have to prove that if R does not vanish on $[0,4]$ then

$$\sup_{N\in\mathbb{N}^*} \|R(\mathbf{A}_N)^{-1}\|_\infty < \infty.$$

Consider the Fourier transform of $R^{-1}(4\sin^2(\frac{\theta}{2}))$

$$R^{-1}(4\sin^2(\frac{\theta}{2})) = \sum_{n\in\mathbb{N}} a_n \cos(n\theta),\ (a_n) \in \ell^1.$$

Then observe that

$$\mathbf{D}_N(\frac{\mathbb{1}_{\{-n,n\}}}{2})\mathbf{e}_k^N = \cos(n\pi kh)\mathbf{e}_k^N.$$

Consequently, we have

$$R(\mathbf{A}_N)^{-1} = \sum_{n\in\mathbb{N}} a_n \mathbf{D}_N(\frac{\mathbb{1}_{\{-n,n\}}}{2}).$$

However, we have obviously $\|\mathbf{D}_N(\frac{\mathbb{1}_{\{-n,n\}}}{2})\|_\infty = 1$, so we get

$$\|R(\mathbf{A}_N)^{-1}\|_\infty \leq \|(a_n)\|_{\ell^1}.$$

Consider a scheme where $\mathbf{D}_N(d) = P(\mathbf{A}_N)$. If we don't take care in the construction of d, P could vanish into $(0, 4)$.

However, we have

$$\mathrm{Sp}_{\mathbb{C}}\ \mathbf{D}_N(d) = \left\{ P\left(4\sin^2\left(\frac{\pi k h}{2}\right)\right) \mid k = 1, \ldots, N \right\}.$$

So $\mathbf{D}_N(d)$ could have very small eigenvalues. (stability $\simeq$ control of $h^2 \|\mathbf{D}_N^{-1}\|$)

Consider a scheme where $\mathbf{D}_N(d) = P(\mathbf{A}_N)$. If we don't take care in the construction of d, P could vanish into $(0,4)$.

However, we have

$$\mathrm{Sp}_{\mathbb{C}}\ \mathbf{D}_N(d) = \left\{ P\left(4\sin^2\left(\frac{\pi kh}{2}\right)\right) \mid k = 1,\dots,N \right\}.$$

So $\mathbf{D}_N(d)$ could have very small eigenvalues. (stability $\simeq$ control of $h^2 \|\mathbf{D}_N^{-1}\|$)

Questions:

- In general, does P really have zero points into $(0,4)$?
- Are small eigenvalues of $\mathbf{D}_N(d)$ large enough to allow stability ?

Theorem

For almost all symmetric $\mathbb{C}$ valued formula d such that

$$\sum_{j\in\mathbb{Z}} d_j = 0,$$

P does not vanish in $]0,4]$, $P(0) = 0$ and $P'(0) \neq 0$.

Corollary: Almost all complex valued scheme (consistent of order larger than or equal to 2) is strongly stable !

Theorem

For almost all symmetric $\mathbb{C}$ valued formula d such that

$$\sum_{j\in\mathbb{Z}} d_j = 0,$$

P does not vanish in $]0,4]$, $P(0) = 0$ and $P'(0) \neq 0$.

Corollary: Almost all complex valued scheme (consistent of order larger than or equal to 2) is strongly stable !

Problem: In practice, we consider real valued formulas. And for any $l > 0$,

$$\mathrm{Int}\{R \in \mathbb{R}_l[X] \mid \exists x_0 \in (0,4),\ R(x_0) = 0\} \neq \emptyset.$$

It can not be a null set.

A numerical experiment:

- Fix $l \geq 1$ and $n \in 2\mathbb{N}^*$,
- Choose randomly a formula d of sum 0 such that $\tau(d) \leq l + 1$,
- Determine s to get a couple (d, s) consistent of order n,
- Assemble the matrices $\mathbf{D}_N(d)$ and $\mathbf{S}_N$,
- Choose a test function $u \in C^\infty(\mathbb{R})$ such that $u(0) = u(1) = 0$ and define $f = -u''$,
- Plot the convergence curves.

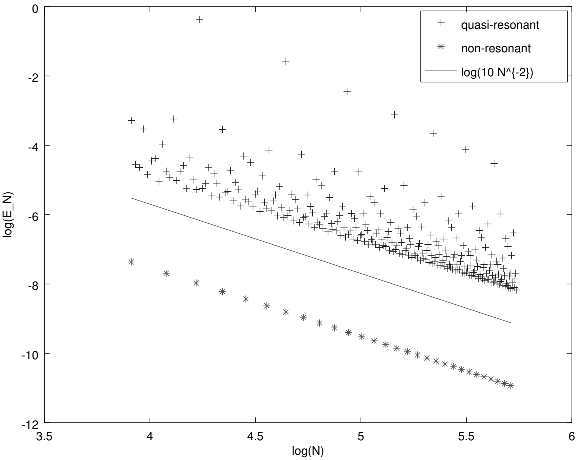

Figure: Convergence curves with $u(x) = x(1-x)e^{2x}$, $n = 2$ and $E_N := \|\mathbf{u}^N - \mathbf{u}^{N,ex}\|_\infty$. For the non-resonant scheme $d = 2\mathbb{1}_{\{0\}} - \mathbb{1}_{\{-1,1\}}$ and for the quasi-resonant scheme $d = (2-6z)\mathbb{1}_{\{0\}} + (4z-1)\mathbb{1}_{\{-1,1\}} - z\mathbb{1}_{\{-2,2\}}$ with $z = 0.358946420670826$.

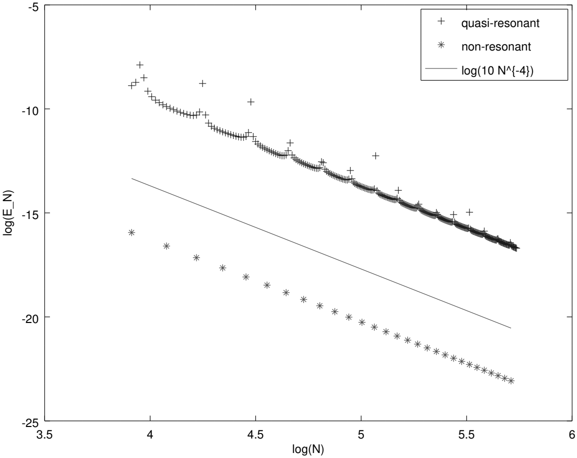

Figure: Convergence curves with $u(x) = x(1-x)e^{2x}$, $n = 4$ and $E_N := \|\mathbf{u}^N - \mathbf{u}^{N,ex}\|_\infty$. For the non-resonant scheme $d = 2\mathbb{1}_{\{0\}} - \mathbb{1}_{\{-1,1\}}$ and for the quasi-resonant scheme $d = (2-6z)\mathbb{1}_{\{0\}} + (4z-1)\mathbb{1}_{\{-1,1\}} - z\mathbb{1}_{\{-2,2\}}$ with $z = 32.12121212$.

Observations:

- Two kinds of behaviours.
- Resonant case $\iff$ P vanishes into $]0,4]$
- In the resonant case the scheme seems convergent with the good rate.

Theorem

Let $P \in \mathbb{C}[X]$ be a polynomial and let Λ be the set of the roots of P in $[0,4]$ and assume that P satisfies the following assumptions:

i) $0 \in \Lambda$,

ii) $4 \notin \Lambda$,

iii) the roots of P in $[0,4]$ are simple,

iv) $\exists \delta : \mathbb{N}^* \to \mathbb{R}_+^*$,

$$\forall \lambda \in \Lambda, \forall q \in \mathbb{N}^*, \forall 1 \leq p \leq q-1, \quad 0 < \delta_q \leq \left| \lambda - 4\sin^2\left(\frac{\pi}{2}\frac{p}{q}\right) \right|.$$

Then the sequence of finite difference matrices $(P(\mathbf{A}_N))_{N\in\mathbb{N}^*}$ is stable relatively to the sequence $\eta_N = \frac{1}{\delta_{N+1}}$

Remark: If $\delta_{N+1} = h^2$ then $P(\mathbf{A}_N)$ is stable.

Theorem Khinchin's

Let $(\nu_q)_q$ be a sequence of positive real numbers such that the series $\sum \nu_q$ converges. Then, for almost all $\alpha \in \mathbb{R}$, there exists a constant $c > 0$ such that for all $p, q \in \mathbb{Z} \times \mathbb{N}^*$, one has

$$|\alpha - \frac{p}{q}| \geq c\frac{\nu_q}{q}.$$

Theorem Khinchin's

Let $(\nu_q)_q$ be a sequence of positive real numbers such that the series $\sum \nu_q$ converges. Then, for almost all $\alpha \in \mathbb{R}$, there exists a constant $c > 0$ such that for all $p, q \in \mathbb{Z} \times \mathbb{N}^*$, one has

$$|\alpha - \frac{p}{q}| \geq c\frac{\nu_q}{q}.$$

"Corollary"

If $\delta_q = \frac{\nu_q}{q}$ and $q\nu_q$ is bounded then for all $l \geq 1$ and almost all $P \in X\mathbb{R}_l[X]$, if Λ is the set of the roots of P into $[0,4]$ then there exists $C > 0$ s.t.

i) $0 \in \Lambda$,

ii) $4 \notin \Lambda$,

iii) the roots of P in $[0,4]$ are simple,

iv) $\exists \delta : \mathbb{N}^* \to \mathbb{R}_+^*$,

$$\forall \lambda \in \Lambda, \forall q \in \mathbb{N}^*, \forall 1 \leq p \leq q-1, \quad 0 < \frac{\nu_q}{q} \leq C\left|\lambda - 4\sin^2\left(\frac{\pi}{2}\frac{p}{q}\right)\right|.$$

In practice we can choose $\nu_q = \frac{1}{q \log^2 q}$.

Conclusion: Almost all real valued consistent scheme of order $n \geq 2$ is convergent at the rate $h^n \log^2(h)$.

In practice we can choose $\nu_q = \frac{1}{q \log^2 q}$.

Conclusion: Almost all real valued consistent scheme of order $n \geq 2$ is convergent at the rate $h^n \log^2(h)$.

Thank you for your attention