Interactive Theorem Proving 2010

Formal Proof of a Wave Equation Resolution Scheme: the Method Error

Sylvie Boldo,
François Clément,Jean-Christophe Filliâtre, Micaela Mayero,
Guillaume Melquiond and Pierre Weis

INRIA, LRI, CNRS, Université Paris 13

July 11th, 2010

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE | INRIA

centre de recherche SACLAY - ÎLE-DE-FRANCE

Motivations

PDE (Partial Differential Equation) $\Rightarrow$ weather forecast

$\Rightarrow$ nuclear simulation

$\Rightarrow$ optimal control

$\Rightarrow$ ...

Motivations

PDE (Partial Differential Equation) $\Rightarrow$ weather forecast
$\Rightarrow$ nuclear simulation
$\Rightarrow$ optimal control
$\Rightarrow$ ...

Usually too complex to solve by an exact mathematical formula
$\Rightarrow$ approximated by numerical scheme over discrete grids

$\Rightarrow$ mathematical proofs of the convergence of the numerical scheme (we compute something close to the PDE solution if the grids size decreases)

Motivations

PDE (Partial Differential Equation) $\Rightarrow$ weather forecast
$\Rightarrow$ nuclear simulation
$\Rightarrow$ optimal control
$\Rightarrow$ ...

Usually too complex to solve by an exact mathematical formula
$\Rightarrow$ approximated by numerical scheme over discrete grids

$\Rightarrow$ mathematical proofs of the convergence of the numerical scheme (we compute something close to the PDE solution if the grids size decreases)

Let us machine-check this kind of proof! (in Coq)

Outline

Outline

Wave Equation?

Looking for $u : \mathbb{R}^2 \to \mathbb{R}$ regular enough such that:

$$\frac{\partial^2 u(x,t)}{\partial t^2} - c^2 \frac{\partial^2 u(x,t)}{\partial x^2} = s(x,t)$$

with given values for the initial position $u_0(x)$ and the initial velocity $u_1(x)$.

$\Rightarrow$ rope oscillation, sound, radar, oil prospection...

Scheme?

We want $u_j^k \approx u(j\Delta x, k\Delta t)$.

$$\frac{u_j^k - 2u_j^{k-1} + u_j^{k-2}}{\Delta t^2} - c^2 \frac{u_{j+1}^{k-1} - 2u_j^{k-1} + u_{j-1}^{k-1}}{\Delta x^2} = s_j^{k-1}$$

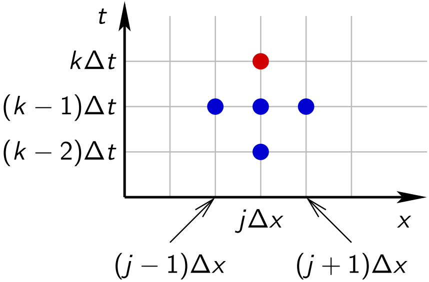
And other horrible formulas to initialize u_j^0 and u_j^1.

Scheme?

We want $u_j^k \approx u(j\Delta x, k\Delta t)$.

$$\frac{u_j^k - 2u_j^{k-1} + u_j^{k-2}}{\Delta t^2} - c^2 \frac{u_{j+1}^{k-1} - 2u_j^{k-1} + u_{j-1}^{k-1}}{\Delta x^2} = s_j^{k-1}$$

And other horrible formulas to initialize u_j^0 and u_j^1.



Three-point scheme: u_j^k depends on u_{j-1}^{k-1}, u_j^{k-1}, u_{j+1}^{k-1} and u_j^{k-2}.

So what?

We measure that u and u_j^k are close when $(\Delta x, \Delta t) \to 0$.

We define $e_j^k \stackrel{\text{def}}{=} \bar{u}_j^k - u_j^k$: convergence error
where $\bar{u}_j^k$ is the value of u at the (j, k) point of the grid.

So what?

We measure that u and u_j^k are close when $(\Delta x, \Delta t) \to 0$.

We define $e_j^k \stackrel{\text{def}}{=} \bar{u}_j^k - u_j^k$: convergence error
where $\bar{u}_j^k$ is the value of u at the (j, k) point of the grid.

We want to bound $\left\| e_h^{k_{\Delta t}(t)} \right\|_{\Delta x}$: the average of the convergence error on all points of the grid at a given time $k_{\Delta t}(t) = \left\lfloor \frac{t}{\Delta t} \right\rfloor \Delta t$.

So what?

We measure that u and u_j^k are close when $(\Delta x, \Delta t) \to 0$.

We define $e_j^k \stackrel{\text{def}}{=} \bar{u}_j^k - u_j^k$: convergence error
where $\bar{u}_j^k$ is the value of u at the (j, k) point of the grid.

We want to bound $\left\| e_h^{k_{\Delta t}(t)} \right\|_{\Delta x}$: the average of the convergence error on all points of the grid at a given time $k_{\Delta t}(t) = \left\lfloor \frac{t}{\Delta t} \right\rfloor \Delta t$.

We want to prove:

$$\left\| e_h^{k_{\Delta t}(t)} \right\|_{\Delta x} = O_{[0, t_{\max}]}(\Delta x^2 + \Delta t^2)$$

Outline

Dot product and finite support

We only consider functions having a finite support:

$$\{f : \mathbb{Z} \to \mathbb{R} \mid \exists a, b \in \mathbb{Z}, \forall i \in \mathbb{Z}, f(i) \neq 0 \Rightarrow a \leq i \leq b\}.$$

Dot product and finite support

We only consider functions having a finite support:

$$\{f : \mathbb{Z} \to \mathbb{R} \mid \exists a, b \in \mathbb{Z}, \forall i \in \mathbb{Z}, f(i) \neq 0 \Rightarrow a \leq i \leq b\}.$$

We have an uninterpreted $\langle ., . \rangle$ such that

$$\forall f\ g\ a\ b, (\forall i, (f(i) \neq 0 \vee g(i) \neq 0) \Rightarrow a \leq i \leq b) \Rightarrow \langle f, g \rangle = \sum_{i=a}^{b} f(i) g(i)$$

Dot product and finite support

We only consider functions having a finite support:

$$\{f : \mathbb{Z} \to \mathbb{R} \mid \exists a, b \in \mathbb{Z}, \forall i \in \mathbb{Z}, f(i) \neq 0 \Rightarrow a \leq i \leq b\}.$$

We have an uninterpreted $\langle ., . \rangle$ such that

$$\forall f\, g\, a\, b, (\forall i, (f(i) \neq 0 \vee g(i) \neq 0) \Rightarrow a \leq i \leq b) \Rightarrow \langle f, g \rangle = \sum_{i=a}^{b} f(i) g(i)$$

(or an Hilbert ε)

Dot product and finite support

We only consider functions having a finite support:

$$\{f : \mathbb{Z} \to \mathbb{R} \mid \exists a, b \in \mathbb{Z}, \forall i \in \mathbb{Z}, f(i) \neq 0 \Rightarrow a \leq i \leq b\}.$$

We have an uninterpreted $\langle ., . \rangle$ such that

$$\forall f\ g\ a\ b, (\forall i, (f(i) \neq 0 \vee g(i) \neq 0) \Rightarrow a \leq i \leq b) \Rightarrow \langle f, g \rangle = \sum_{i=a}^{b} f(i) g(i)$$

(or an Hilbert ε)

Hence $\|f\| \stackrel{\text{def}}{=} \sqrt{\langle f, f \rangle}$.
Hence a predicate *FS* (finite support) with lemmas and a dedicated tactic.

Outline

1. Wave equation resolution scheme?
2. Formal proof: basic blocks
   - Dot product
   - Big O
3. Formal proof: convergence
4. Conclusion & perspectives

Big O = big pain

Usually, the big O uses one variable and $f(\mathbf{x}) = O_{\|\mathbf{x}\| \to 0}(g(\mathbf{x}))$ means

$$\exists \alpha, C > 0, \quad \forall \mathbf{x} \in \mathbb{R}^n, \quad \|\mathbf{x}\| \le \alpha \Rightarrow |f(\mathbf{x})| \le C \cdot |g(\mathbf{x})|.$$

Big O = big pain

Usually, the big O uses one variable and $f(\mathbf{x}) = O_{\|\mathbf{x}\|\to 0}(g(\mathbf{x}))$ means

$$\exists \alpha, C > 0, \quad \forall \mathbf{x} \in \mathbb{R}^n, \quad \|\mathbf{x}\| \leq \alpha \Rightarrow |f(\mathbf{x})| \leq C \cdot |g(\mathbf{x})|.$$

Here 2 variables: $\mathbf{\Delta x}$ (grid sizes, tends to 0), and $\mathbf{x}$ (time and space).
(Think about Taylor expansions)

Big O = big pain

Usually, the big O uses one variable and $f(\mathbf{x}) = O_{\|\mathbf{x}\| \to 0}(g(\mathbf{x}))$ means

$$\exists \alpha, C > 0, \quad \forall \mathbf{x} \in \mathbb{R}^n, \quad \|\mathbf{x}\| \le \alpha \Rightarrow |f(\mathbf{x})| \le C \cdot |g(\mathbf{x})|.$$

Here 2 variables: $\mathbf{\Delta x}$ (grid sizes, tends to 0), and $\mathbf{x}$ (time and space).
(Think about Taylor expansions)

$$\forall \mathbf{x}, \exists \alpha, C > 0, \quad \forall \mathbf{\Delta x} \in \mathbb{R}^2, \quad \|\mathbf{\Delta x}\| \le \alpha \Rightarrow |f(\mathbf{x}, \mathbf{\Delta x})| \le C \cdot |g(\mathbf{\Delta x})|$$

does not work.

Uniform big O

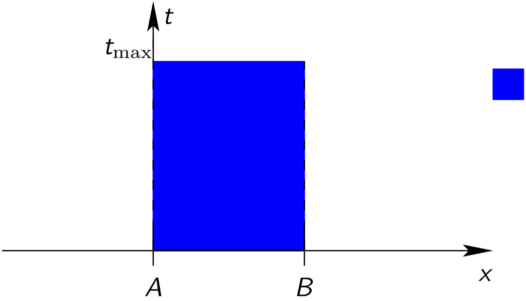
We used a uniform big O:

$$\exists \alpha, C > 0, \quad \forall \mathbf{x}, \mathbf{\Delta x}, \quad \|\mathbf{\Delta x}\| \le \alpha \Rightarrow |f(\mathbf{x}, \mathbf{\Delta x})| \le C \cdot |g(\mathbf{\Delta x})|.$$

where variables $\mathbf{x}$ and $\mathbf{\Delta x}$ are restricted to subsets of $\mathbb{R}^2$.
(for example such that $\Delta t > 0$)
$\Rightarrow$ Taylor expansions

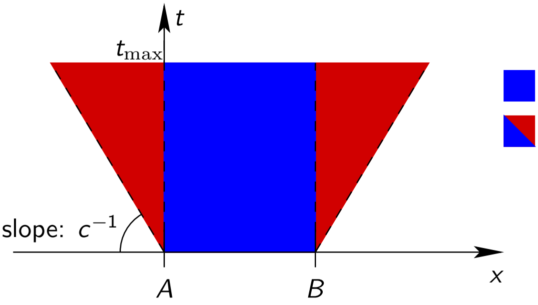
Outline

Finite support



u_0 and u_1 may be nonzero.

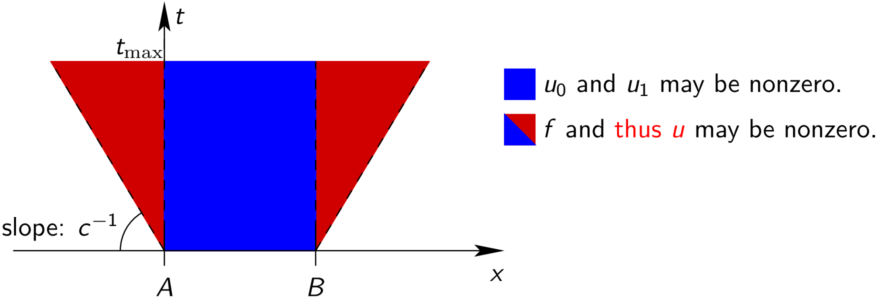
Finite support



u_0 and u_1 may be nonzero.

f may be nonzero.

Finite support

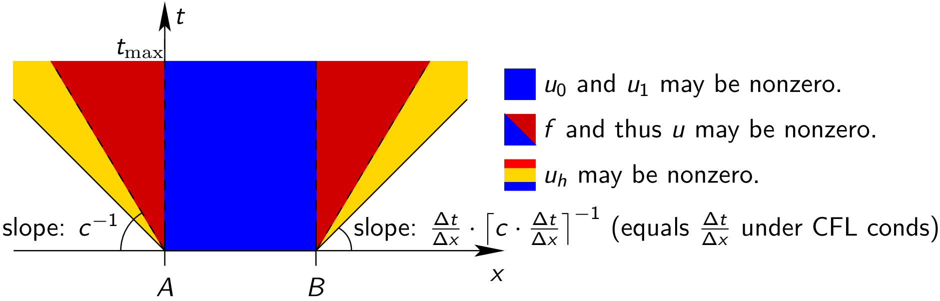


One axiom about the exact solution of the PDE:

$$x \notin [A - c \cdot t, B + c \cdot t] \quad \Rightarrow \quad u(x,t) = 0$$

(mathematically proved using d'Alembert's formula)

Finite support



One axiom about the exact solution of the PDE:

$$x \notin [A - c \cdot t, B + c \cdot t] \quad \Rightarrow \quad u(x,t) = 0$$

(mathematically proved using d'Alembert's formula)

Proof idea 1/3: consistency

The truncation error is defined as how much the exact solution solves the numerical scheme:

$$\varepsilon_j^{k-1} = \frac{\bar{u}_j^k - 2\bar{u}_j^{k-1} + \bar{u}_j^{k-2}}{\Delta t^2} - c^2 \frac{\bar{u}_{j+1}^{k-1} - 2\bar{u}_j^{k-1} + \bar{u}_{j-1}^{k-1}}{\Delta x^2} - s_j^{k-1}$$

Proof idea 1/3: consistency

The truncation error is defined as how much the exact solution solves the numerical scheme:

$$\varepsilon_j^{k-1} = \frac{\bar{u}_j^k - 2\bar{u}_j^{k-1} + \bar{u}_j^{k-2}}{\Delta t^2} - c^2 \frac{\bar{u}_{j+1}^{k-1} - 2\bar{u}_j^{k-1} + \bar{u}_{j-1}^{k-1}}{\Delta x^2} - s_j^{k-1}$$

The consistency is the boundedness of the truncation error:

$$\left\| \varepsilon_h^{k_{\Delta t}(t)} \right\|_{\Delta x} = O_{[0,t_{\max}]}(\Delta x^2 + \Delta t^2)$$

By Taylor series and many computations.

Proof idea 2/3: stability

We define a discrete energy by

$$E_h(c)(u_h)^{k+\frac{1}{2}} \stackrel{\text{def}}{=} \frac{1}{2}\left\|\frac{u_h^{k+1}-u_h^k}{\Delta t}\right\|_{\Delta x}^2 + \frac{1}{2}\left\langle u_h^k, u_h^{k+1}\right\rangle_{A_h(c)}$$

kinetic energy potential energy

$\left\langle v_h, w_h\right\rangle_{A_h(c)} \stackrel{\text{def}}{=} \left\langle A_h(c)\, v_h, w_h\right\rangle_{\Delta x}$ and $\left(A_h(c)\, v_h\right)_j \stackrel{\text{def}}{=} -\, c^2 \frac{v_{j+1}-2v_j+v_{j-1}}{\Delta x^2}$.

Proof idea 2/3: stability

We define a discrete energy by

$$E_h(c)(u_h)^{k+\frac{1}{2}} \stackrel{\text{def}}{=} \frac{1}{2}\left\|\frac{u_h^{k+1}-u_h^k}{\Delta t}\right\|_{\Delta x}^2 + \frac{1}{2}\left\langle u_h^k, u_h^{k+1}\right\rangle_{A_h(c)}$$

kinetic energy potential energy

$\langle v_h, w_h\rangle_{A_h(c)} \stackrel{\text{def}}{=} \langle A_h(c)\, v_h, w_h\rangle_{\Delta x}$ and $(A_h(c)\, v_h)_j \stackrel{\text{def}}{=} -c^2\frac{v_{j+1}-2v_j+v_{j-1}}{\Delta x^2}$.

Note that this energy is constant if $f = 0$.
We prove an overestimation and an underestimation of this energy.
$\Rightarrow u_h$ does not diverge.

Proof idea 3/3: convergence

The convergence error is solution of the same discrete scheme with inputs

$$u_{0,j} = 0, \qquad u_{1,j} = \frac{e_j^1}{\Delta t}, \qquad \text{and} \qquad s_j^k = \varepsilon_j^{k+1}.$$

+ proofs about the initializations.

Proof idea 3/3: convergence

The convergence error is solution of the same discrete scheme with inputs

$$u_{0,j} = 0, \qquad u_{1,j} = \frac{e_j^1}{\Delta t}, \qquad \text{and} \qquad s_j^k = \varepsilon_j^{k+1}.$$

$+$ proofs about the initializations.

All these proofs require the existence of ζ and ξ in $]0,1[$ with $\zeta \leq 1-\xi$ and we require that $\zeta \leq \frac{c\Delta t}{\Delta x} \leq 1-\xi$ (CFL conditions).

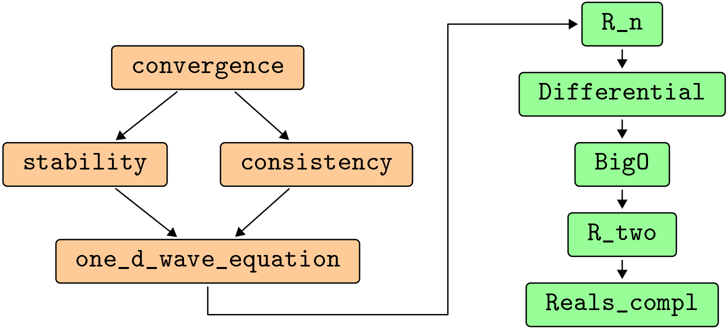
Convergence

We proved that:

$$\left\| e_h^{k_{\Delta t}(t)} \right\|_{\Delta x} = O_{\substack{t \in [0, t_{\max}] \\ (\Delta x, \Delta t) \to 0 \\ 0 < \Delta x \,\wedge\, 0 < \Delta t \,\wedge \\ \zeta \le c\frac{\Delta t}{\Delta x} \le 1 - \xi}} (\Delta x^2 + \Delta t^2).$$

Outline

Conclusion



4500 lines of Coq (half dedicated, half re-usable)
$\approx$ as long as a detailed paper proof

Conclusion

- synergy applied mathematicians / logicians

Conclusion

- synergy applied mathematicians / logicians
- filling the gaps of pen&paper proofs

Conclusion

- synergy applied mathematicians / logicians
- filling the gaps of pen&paper proofs
- 1 axiom: finite support of the exact solution
  (+1 ε operator)

Perspectives

- re-use the proofs with reflections (the rope has two ends).

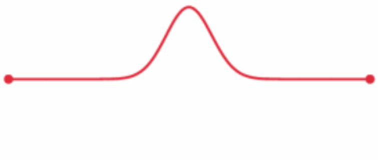
Perspectives

- re-use the proofs with reflections (the rope has two ends).

- link this to the C program
  $\Rightarrow$ full proof of the program (with already done floating-point proof)

Perspectives

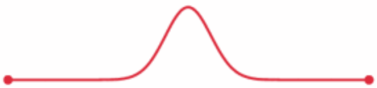
- re-use the proofs with reflections (the rope has two ends).



- link this to the C program
  $\Rightarrow$ full proof of the program (with already done floating-point proof)
- extract the C and α of the big O (done)

Perspectives

- re-use the proofs with reflections (the rope has two ends).



- link this to the C program
  $\Rightarrow$ full proof of the program (with already done floating-point proof)
- extract the C and α of the big O (done)
- prove Lax equivalence for as many schemes as possible:
  consistency $\Rightarrow$ (stability $\Leftrightarrow$ convergence)

Perspectives

- re-use the proofs with reflections (the rope has two ends).

- link this to the C program
  $\Rightarrow$ full proof of the program (with already done floating-point proof)
- extract the C and α of the big O (done)
- prove Lax equivalence for as many schemes as possible:
  consistency $\Rightarrow$ (stability $\Leftrightarrow$ convergence)
- other schemes for the same PDE
- other PDEs
- ODEs

Thank you for your attention

F∮st

http://fost.saclay.inria.fr