Stability of finite difference schemes for hyperbolic initial boundary value problems II

Jean-François Coulombel

Laboratoire Paul Painlevé (UMR CNRS 8524)
CNRS, Université Lille 1
Team Project SIMPAF - INRIA Lille Nord Europe

Nonlinear hyperbolic PDEs, dispersive and transport equations,
Trieste, June 2011

Plan of the second course

1. Continuous hyperbolic boundary value problems : a brief introduction
   - The one-dimensional case
   - The multi-dimensional case
2. Discretized problems : zero initial data
   - The discretized boundary value problem : strong stability
   - The normal modes analysis
   - The Uniform Kreiss-Lopatinskii Condition
   - The main stability result
   - Summary

Plan

Consider the initial boundary value problem :

$$\begin{cases} \partial_t u + A\,\partial_x u = F(t,x)\,, & \text{in } [0,T] \times \mathbb{R}_+\,, \\ B\,u(t,0) = g(t)\,, & \text{in } [0,T]\,, \\ u|_{t=0} = f\,, & \text{on } \mathbb{R}_+\,. \end{cases}$$

Space domain : $\mathbb{R}_+$. Boundary : $\{x = 0\}$.
Linear **system** with **constant coefficients** : $A \in M_N(\mathbb{R})$, $u \in \mathbb{R}^N$, $B \in M_{p,N}(\mathbb{R})$.

General problem

Which boundary conditions B give a well-posed problem ? The solution and its trace should be estimated in the **same** functional spaces as the data.

The method of characteristics

We assume that the operator $\partial_t + A\,\partial_x$ is hyperbolic : A is diagonalizable.
We introduce a set of eigenvalues and eigenvectors : λ_j, r_j, $j = 1, \dots, N$.
The solution u is decomposed on the basis $(r_1, \dots, r_N)$:

$$u(t,x) = \sum_{j=1}^{N} \alpha_j(t,x)\, r_j\,,$$

$$f(x) = \sum_{j=1}^{N} \beta_j(x)\, r_j\,,$$

$$F(t,x) = \sum_{j=1}^{N} F_j(t,x)\, r_j\,.$$

The method of characteristics

The problem reads

$$\begin{cases} \partial_t \alpha_j + \lambda_j \, \partial_x \alpha_j = F_j(t,x) \, , & \text{in } [0,T] \times \mathbb{R}_+ \, , \\ \sum_j \alpha_j(t,0) \, B \, r_j = g(t) \, , & \text{in } [0,T] \, , \\ \alpha_j|_{t=0} = \beta_j \, , & \text{on } \mathbb{R}_+ \, . \end{cases}$$

This system can be solved by the method of characteristics, separating outgoing characteristics from incoming characteristics.

The method of characteristics

Definition

The eigenvalue λ_j corresponds to an outgoing characteristic if $\lambda_j < 0$, and to an incoming characteristic if $\lambda_j > 0$.

The analysis is in two steps. For simplicity, we shall assume that A is invertible. This is the so-called noncharacteristic case.

The method of characteristics

We start with **outgoing** characteristics : $\lambda_j < 0$. Then we find

$$\alpha_j(t,x) = \beta_j(x - \lambda_j\, t) + \int_0^t F_j(s, x - \lambda_j\,(t-s))\,\mathrm{d}s\,,$$

and this formula defines α_j in $\mathbb{R}_+ \times \mathbb{R}_+$ since $\lambda_j < 0$.

In particular, the **trace** $\alpha_j(t,0)$ can be **explicitly** computed from the data :

$$\alpha_j(t,0) = \beta_j(-\lambda_j\, t) + \int_0^t F_j(s, -\lambda_j\,(t-s))\,\mathrm{d}s\,.$$

The method of characteristics

We go on with **incoming** characteristics : $\lambda_j > 0$. We find

$$\alpha_j(t,x) = \beta_j(x - \lambda_j\, t) + \int_0^t F_j(s, x - \lambda_j\,(t-s))\,\mathrm{d}s\,,$$

in the domain $\{(t,x) \in \mathbb{R}_+ \times \mathbb{R}_+ / x \geq \lambda_j\, t\}$, and

$$\alpha_j(t,x) = \alpha_j(t - x/\lambda_j, 0) + \int_{t-x/\lambda_j}^t F_j(s, x - \lambda_j\,(t-s))\,\mathrm{d}s\,,$$

in the domain $\{(t,x) \in \mathbb{R}_+ \times \mathbb{R}_+ / x < \lambda_j\, t\}$.

The trace $\alpha_j(t,0)$ is **not** determined by the data !

The method of characteristics

To determine the solution completely, the boundary condition

$$\sum_{j=1}^{N} \alpha_j(t,0)\, B\, r_j = g(t)\,,$$

should determine the **trace of incoming characteristics** in terms of the source term g and of the trace of outgoing characteristics.

The method of characteristics

To determine the solution completely, the boundary condition

$$\sum_{j=1}^{N} \alpha_j(t,0)\, B\, r_j = g(t)\,,$$

should determine the **trace of incoming characteristics** in terms of the source term g and of the trace of outgoing characteristics.

Conclusion

The problem can be well-posed **only if**

$$\mathbb{R}^p = \mathrm{Span}\big(B\, r_1, \dots, B\, r_q\big)\,,$$

with the convention $\lambda_1, \dots, \lambda_q > 0$, $\lambda_{q+1}, \dots, \lambda_N < 0$.
This implies $p \leq q$.

Conclusion

In the one-dimensional case, the initial boundary value problem is well-posed (existence/uniqueness) **if and only if**

1. The number p of independent **boundary conditions** is equal to the number of **incoming characteristics** $(q = p)$,
2. There holds
$$\mathrm{Ker}\ B \cap \mathrm{Span}\big(r_1, \dots, r_p\big) = \{0\}\,,$$
where $r_1, \dots, r_p$ span the eigenspace of A associated with positive eigenvalues.

Observe that the latter relation is compatible with dimensions ! In particular, there always exists a matrix B for which the problem is well-posed.

Plan

1. Continuous hyperbolic boundary value problems : a brief introduction
   - The one-dimensional case
   - The multi-dimensional case
2. Discretized problems : zero initial data
   - The discretized boundary value problem : strong stability
   - The normal modes analysis
   - The Uniform Kreiss-Lopatinskii Condition
   - The main stability result
   - Summary

Consider the initial boundary value problem :

$$\begin{cases} L(\partial)\, u := \partial_t u + \sum_{j=1}^d A_j\, \partial_{x_j} u = F\,, & \text{in } [0,T] \times \mathbb{R}^d_+\,,\\ B\, u|_{x_d=0} = g\,, & \text{on } [0,T] \times \mathbb{R}^{d-1}\,,\\ u|_{t=0} = f\,, & \text{on } \mathbb{R}^d_+\,. \end{cases}$$

Space domain : half-space $\mathbb{R}^d_+ = \{x_d > 0\}$.
Linear **system** with **constant coefficients**.

General problem

The differential operator $L(\partial)$ is given. Can we find/characterize the boundary conditions B that give a well-posed problem ?

Key points of the analysis

In the non-characteristic case (A_d invertible), the analysis relies on three major assumptions :

1. A **stability** assumption for the Cauchy problem : hyperbolicity.
2. An additional **structural** assumption on the symbol associated with the Cauchy problem (geometric regularity of eigenelements).
3. A **compatibility** condition between the boundary conditions and the hyperbolic system : the Uniform Kreiss-Lopatinskii Condition.

• Assumptions 1 and 2 only involve the operator $L(\partial)$! With these assumptions, one can define a certain vector bundle E over a compact basis Σ (a closed half-sphere).
Kreiss, Sakamoto (1970), Majda-Osher (1975), Métivier (2000), Métivier-Zumbrun (2005)

The bundle E is first defined in the **interior** of Σ, and the difficult part of the job is to **extend continuously** E to the boundary $\partial\Sigma$.

• Assumptions 1 and 2 only involve the operator $L(\partial)$! With these assumptions, one can define a certain vector bundle E over a compact basis Σ (a closed half-sphere).
Kreiss, Sakamoto (1970), Majda-Osher (1975), Métivier (2000), Métivier-Zumbrun (2005)

The bundle E is first defined in the **interior** of Σ, and the difficult part of the job is to **extend continuously** E to the boundary $\partial\Sigma$.

• The formulation of the UKLC is simpler once we have this first result :

UKLC

$$\forall \zeta \in \Sigma \,, \quad E(\zeta) \cap \mathrm{Ker}\; B = \{0\}.$$

In one space dimension, $E(\zeta) = E^+(A)$ for all ζ.

The stability estimate

Theorem (a priori estimate in weighted L^2 spaces)

If the UKLC is satisfied, then for all sufficiently smooth and decaying solution u, there holds :

$$\gamma \iint_{\mathbb{R}\times\mathbb{R}^d_+} \mathrm{e}^{-2\gamma t} |u(t,x)|^2 \, \mathrm{d}t \, \mathrm{d}x + \int_{\mathbb{R}\times\mathbb{R}^{d-1}} \mathrm{e}^{-2\gamma t} |u(t,y,0)|^2 \, \mathrm{d}t \, \mathrm{d}y$$
$$\lesssim \frac{1}{\gamma} \iint_{\mathbb{R}\times\mathbb{R}^d_+} \mathrm{e}^{-2\gamma t} |F(t,x)|^2 \, \mathrm{d}t \, \mathrm{d}x + \int_{\mathbb{R}\times\mathbb{R}^{d-1}} \mathrm{e}^{-2\gamma t} |g(t,y)|^2 \, \mathrm{d}t \, \mathrm{d}y$$

for all $\gamma \geq 1$.

The proof relies on the construction of symbolic symmetrizers, which is based on a suitable block structure reduction. Kreiss, Sakamoto (1970), Majda-Osher (1975)

• The a priori estimate yields well-posedness (unique solvability + estimate) for the boundary value problem with source terms F, g in L^2_γ. The solution and its trace belong to L^2_γ and vanish in $\{t < 0\}$ if the source terms do so (zero initial data).

• If the system is either Friedrichs symmetrizable or hyperbolic with constant multiplicity, one can incorporate **non-zero initial data** in L^2, and obtain semigroup estimates. Rauch (1972), Audiard (2011) $\Rightarrow$ ibvp on $[0, T]$.

• Regularity of the solution for smooth source terms. Rauch-Massey (1974)

• There are also many results of the same kind in the characteristic case.

Plan

Discretizing the equations

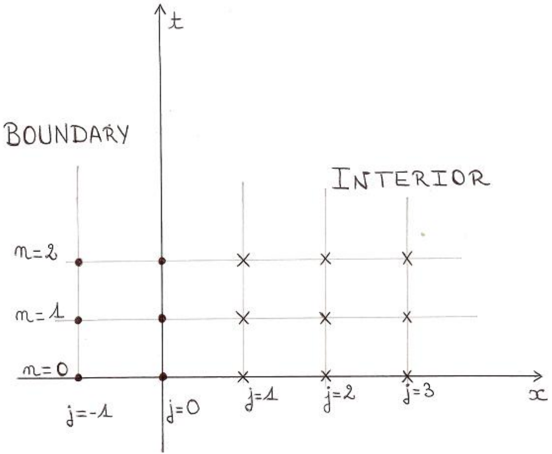
We consider one-dimensional boundary value problems, with **zero initial data**.

We let U_j^n denote the approximation of the solution u on the cell $[n\,\Delta t, (n+1)\,\Delta t[\times[j\,\Delta x, (j+1)\,\Delta x[$, with :

$$n \in \mathbb{N}\,, \quad j \in 1 - r + \mathbb{N}\,.$$

For $1 - r \leq j \leq 0$, the U_j^n's approximate the **trace** of u : **boundary cells**.

The mesh points (boundary and interior domain) :



The numerical scheme

For simplicity only, we consider here a one time step scheme :

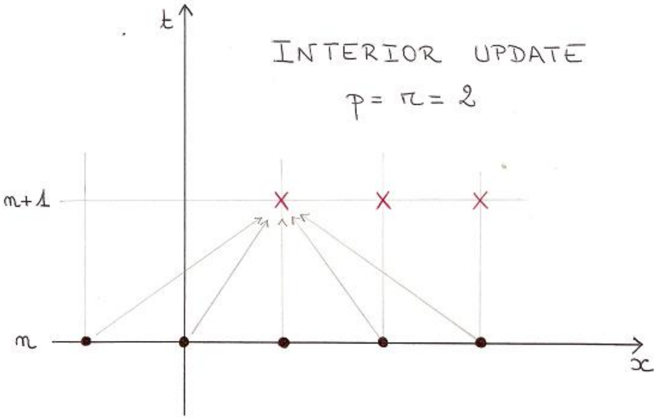
$$\begin{cases} U_j^{n+1} = Q\, U_j^n + \Delta t\, F_j^n\,, & j \geq 1\,, \\ U_j^{n+1} = B_j\, U_1^{n+1} + g_j^{n+1}\,, & j = 1-r, \dots, 0\,, \\ U_j^0 = 0\,, & j \geq 1-r\,, \end{cases}$$

with some operators

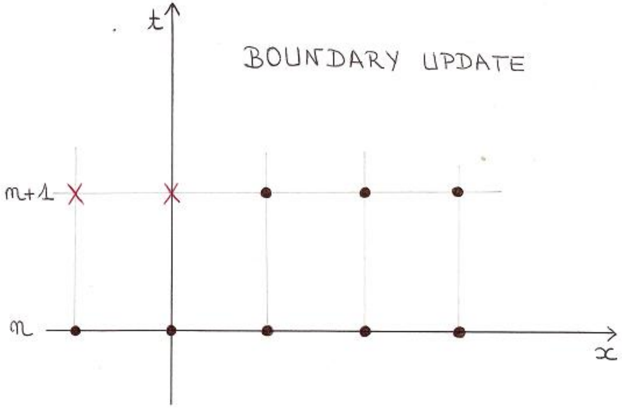
$$Q := \sum_{\ell=-r}^{p} A_\ell\, T^\ell\,, \quad B_j := \sum_{\ell=0}^{q} B_{\ell,j}\, T^\ell\,.$$

It is possible to consider more complicated boundary operators, and the theory works the same.

The numerical scheme in the interior domain :



The numerical scheme on the boundary :



Definition of strong stability

An appropriate notion of stability is the following.

Definition (strong stability), Gustafsson-Kreiss-Sundström (1972)

The numerical scheme is strongly stable if there exists a constant $C > 0$ such that for all $\gamma > 0$ and all $\Delta t \in]0,1]$, there holds :

$$\frac{\gamma}{\gamma\,\Delta t + 1} \sum_{n \geq 0} \sum_{j \geq 1-r} \Delta t\, \Delta x\, \mathrm{e}^{-2\,\gamma\, n\, \Delta t}\, |U_j^n|^2 + \sum_{n \geq 0} \sum_{j=1-r}^{p} \Delta t\, \mathrm{e}^{-2\,\gamma\, n\, \Delta t}\, |U_j^n|^2 \leq C \left(\frac{\gamma\,\Delta t + 1}{\gamma} \|F\|_{\ell^2_\gamma}^2 + \|g\|_{\ell^2_\gamma}^2 \right) .$$

The definition is "compatible" with two obvious asymptotic cases.

- In the limit $\Delta t \to 0$, the estimate **formally** reduces to

$$\gamma \iint_{\mathbb{R}_+ \times \mathbb{R}_+} \mathrm{e}^{-2\,\gamma\,t} |u(t,x)|^2 \,\mathrm{d}t\,\mathrm{d}x + \int_{\mathbb{R}_+} \mathrm{e}^{-2\,\gamma\,t} |u(t,0)|^2 \,\mathrm{d}t$$
$$\leq C \left\{ \frac{1}{\gamma} \iint_{\mathbb{R}_+ \times \mathbb{R}_+} \mathrm{e}^{-2\,\gamma\,t} |F(t,x)|^2 \,\mathrm{d}t\,\mathrm{d}x + \int_{\mathbb{R}_+} \mathrm{e}^{-2\,\gamma\,t} |g(t)|^2 \,\mathrm{d}t \right\} .$$

The definition is "compatible" with two obvious asymptotic cases.

- In the limit $\Delta t \rightarrow 0$, the estimate **formally** reduces to

$$\gamma \iint_{\mathbb{R}_+ \times \mathbb{R}_+} \mathrm{e}^{-2\,\gamma\,t} |u(t,x)|^2 \,\mathrm{d}t\,\mathrm{d}x + \int_{\mathbb{R}_+} \mathrm{e}^{-2\,\gamma\,t} |u(t,0)|^2 \,\mathrm{d}t \\ \le C \left\{ \frac{1}{\gamma} \iint_{\mathbb{R}_+ \times \mathbb{R}_+} \mathrm{e}^{-2\,\gamma\,t} |F(t,x)|^2 \,\mathrm{d}t\,\mathrm{d}x + \int_{\mathbb{R}_+} \mathrm{e}^{-2\,\gamma\,t} |g(t)|^2 \,\mathrm{d}t \right\} .$$

- In the limit $\gamma \rightarrow +\infty$, the estimate **formally** reduces to

$$\frac{1}{\lambda} \sum_{j \ge 1-r} |U_j^1|^2 + \sum_{j=1-r}^{p} |U_j^1|^2 \le C \left\{ \frac{1}{\lambda} \Delta t^2 \sum_{j \ge 1} |F_j^0|^2 + \sum_{j=1-r}^{0} |g_j^1|^2 \right\} ,$$

and this estimate is trivially satisfied (recall $U_j^0 = 0$).

General problem

The finite difference operator Q is given. Can we find/characterize the boundary operators $B_{1-r}, \dots, B_0$ that give a strongly stable scheme ?

Remark : Consistency is "supposed" to be an easier problem. When both strong stability and consistency hold, the numerical scheme converges. Gustafsson (1975). This is in the spirit of "Lax' theorem".

In what follows, we are first going to characterize strong stability in terms of an estimate for the resolvent equation. This characterization makes the definition relevant for practical purposes. Other notions of stability are not so easy to handle.

Plan

The resolvent equation

The normal modes analysis consists in performing a Laplace transform ($\mathcal{Z}$-transform) in the time variable, or equivalently in looking for solutions of the form

$$U_j^n = z^n \, W_j \,, \quad z \in \mathcal{U} \,, \quad (W_j) \in \ell^2 \,,$$

when the source terms F_j^n, g_j^n have the same form.

The resolvent equation

The normal modes analysis consists in performing a Laplace transform ($\mathcal{Z}$-transform) in the time variable, or equivalently in looking for solutions of the form

$$U_j^n = z^n \, W_j \, , \quad z \in \mathcal{U} \, , \quad (W_j) \in \ell^2 \, ,$$

when the source terms F_j^n, g_j^n have the same form.

Observe that such a sequence does not vanish for $n = 0$! Nevertheless, one can forget about the initial condition, perform the transformation and compute the equation satisfied by the sequence (W_j).

The resolvent equation

The sequence (W_j) should satisfy the resolvent equation

$$\begin{cases} W_j - z^{-1}\, Q\, W_j = F_j\,, & j \geq 1\,, \\ W_j - B_j\, W_1 = g_j\,, & j = 1-r, \dots, 0\,, \end{cases}$$

with given source terms (F_j), $g_{1-r}, \dots, g_0$.

The first **crucial** point of the theory is the following characterization.

Characterization of strong stability

Theorem, Gustafsson-Kreiss-Sundström (1972)

The numerical scheme is strongly stable **if and only if** there exists a constant $C > 0$ such that **for all** $z \in \mathcal{U}$, for all $(F_j) \in \ell^2$ and for all vectors $g_{1-r}, \dots, g_0 \in \mathbb{C}^N$, the resolvent equation has a unique solution $(W_j) \in \ell^2$ and this solution satisfies

$$\frac{|z|-1}{|z|} \sum_{j \geq 1-r} |W_j|^2 + \sum_{j=1-r}^{p} |W_j|^2 \leq C \left\{ \frac{|z|}{|z|-1} \sum_{j \geq 1} |F_j|^2 + \sum_{j=1-r}^{0} |g_j|^2 \right\} .$$

Characterization of strong stability

Theorem, Gustafsson-Kreiss-Sundström (1972)

The numerical scheme is strongly stable **if and only if** there exists a constant $C > 0$ such that **for all** $z \in \mathcal{U}$, for all $(F_j) \in \ell^2$ and for all vectors $g_{1-r}, \dots, g_0 \in \mathbb{C}^N$, the resolvent equation has a unique solution $(W_j) \in \ell^2$ and this solution satisfies

$$\frac{|z|-1}{|z|} \sum_{j \geq 1-r} |W_j|^2 + \sum_{j=1-r}^{p} |W_j|^2 \leq C \left\{ \frac{|z|}{|z|-1} \sum_{j \geq 1} |F_j|^2 + \sum_{j=1-r}^{0} |g_j|^2 \right\} .$$

This result reduces the problem of one space dimension, to the price of introducing a complex parameter.

The Godunov-Ryabenkii condition

An immediate consequence of the Theorem is the following

Corollary (Godunov-Ryabenkii condition)

If the numerical scheme is strongly stable, then for all $z \in \mathcal{U}$, the only sequence $(W_j) \in \ell^2$ solution to

$$\begin{cases} W_j - z^{-1}\, Q\, W_j = 0\,, & j \geq 1\,, \\ W_j - B_j\, W_1 = 0\,, & j = 1 - r, \dots, 0\,, \end{cases}$$

is zero.

This property is called the Godunov-Ryabenkii condition. It is a **necessary** condition for strong stability but unfortunately it is not a **sufficient** condition.

Some words on the proof of the Theorem. We first assume that the numerical scheme is strongly stable.

Let us consider $(W_j) \in \ell^2$ and $z \in \mathcal{U}$. We test the definition of strong stability on the sequence of solutions :

$$\forall\, j \geq 1-r\,, \quad \forall\, n \geq 0\,, \quad U_j^n(\nu) := \begin{cases} z^n\, W_j/\sqrt{\nu}\,, & \text{if } 1 \leq n \leq \nu, \\ 0\,, & \text{otherwise.} \end{cases}$$

For a fixed ν, $(U_j^n(\nu))$ is a solution with appropriate source terms. When ν is large, the solution behaves more and more as a pure normal mode.

The proof relies on computing the limits, as ν tends to $+\infty$, of the sums

$$\sum_{n\geq 0} \sum_{j\geq 1} \mathrm{e}^{-2\,\gamma\,(n+1)}\,|F_j^n(\nu)|^2\,,$$

$$\sum_{n\geq 1} \sum_{j=1-r}^{0} \mathrm{e}^{-2\,\gamma\,n}\,|g_j^n(\nu)|^2\,,$$

with $\gamma := \ln|z| > 0$.

These sums with respect to n are transformed into integrals by applying Plancherel's Theorem. Then one applies more or less standard arguments of convolution theory (in the spirit of Féjer's Theorem).

This provides an **a priori estimate** for solutions of the resolvent equation :

$$\frac{|z|-1}{|z|} \sum_{j \geq 1-r} |W_j|^2 + \sum_{j=1-r}^{p} |W_j|^2$$
$$\leq C \left\{ \frac{|z|}{|z|-1} \sum_{j \geq 1} |(L(z)\,W)_j|^2 + \sum_{j=1-r}^{0} |(L(z)\,W)_j|^2 \right\},$$

with

$$(L(z)\,W)_j := \begin{cases} W_j - z^{-1}\,Q\,W_j\,, & \text{if } j \geq 1, \\ W_j - B_j\,W_1\,, & \text{if } 1-r \leq j \leq 0. \end{cases}$$

The conclusion follows from an observation and an abstract argument of functional analysis.

The conclusion follows from an observation and an abstract argument of functional analysis.

Lemma

There exists $R_0 \geq 1$ such that for all $z \in \mathbb{C}$ with $|z| > R_0$, the operator $L(z)$ is an isomorphism on ℓ^2.

Lemma

Let E be a Banach space, and let $\mathcal{T}$ denote a nonempty connected set. Let $\mathcal{L}$ be a continuous function on $\mathcal{T}$ with values in the space of bounded operators on E. Assume moreover that the two following conditions are satisfied :

- there exists a constant $M > 0$ such that for all $t \in \mathcal{T}$ and for all $x \in E$, we have $\|x\|_E \leq M \, \|\mathcal{L}(t)\, x\|_E$,
- there exists some $t_0 \in \mathcal{T}$ such that $\mathcal{L}(t_0)$ is an isomorphism.

Then $\mathcal{L}(t)$ is an isomorphism **for all** $t \in \mathcal{T}$.

We now assume that the resolvent equation is uniquely solvable with a good estimate.
We consider some source terms (F_j^n), (g_j^n) with compact support, and we let (U_j^n) denote the solution to the numerical scheme.

We now assume that the resolvent equation is uniquely solvable with a good estimate.
We consider some source terms (F_j^n), (g_j^n) with compact support, and we let (U_j^n) denote the solution to the numerical scheme.

Very crude estimates yield

$$\sum_{n \geq s+1} \sum_{j \geq 1-r} \mathrm{e}^{-2\gamma n} |U_j^n|^2 < +\infty \,,$$

for γ large enough. We can thus define the Laplace transform of (U_j^n) for every fixed j. These are holomorphic functions on a half-plane $\{\mathrm{Re}\, \tau > \gamma_0\}$.

The sequence $(\widehat{U}_j(\tau))$ is a solution to the resolvent equation for Re τ large enough, and we would like it to solve the resolvent equation for all Re $\tau > 0$. Is it well-defined ?

The sequence $(\widehat{U}_j(\tau))$ is a solution to the resolvent equation for Re τ large enough, and we would like it to solve the resolvent equation for all Re $\tau > 0$. Is it well-defined ?

The conclusion follows from :

1. the unique continuation principle for holomorphic functions ($\widehat{U}_j$ coincides with a function W_j that is holomorphic on the half-plane Re $\tau > 0$),
2. the Paley-Wiener Theorem (W_j is the Laplace transform of some function),
3. Plancherel's Theorem (which makes the link between estimates for the numerical scheme and estimates for the resolvent equation).

A technical refinement

For technical reasons, we shall need the following refined version.

Theorem

The numerical scheme is strongly stable **if and only if** for all $R \geq 2$, there exists a constant $C_R > 0$ such that for all $z \in \mathcal{U}$ with $|z| \leq R$, for all $(F_j) \in \ell^2$ and for all vectors $g_{1-r}, \dots, g_0 \in \mathbb{C}^N$, the resolvent equation has a unique solution $(W_j) \in \ell^2$ and this solution satisfies

$$\frac{|z|-1}{|z|} \sum_{j \geq 1-r} |W_j|^2 + \sum_{j=1-r}^{p} |W_j|^2 \leq C_R \left\{ \frac{|z|}{|z|-1} \sum_{j \geq 1} |F_j|^2 + \sum_{j=1-r}^{0} |g_j|^2 \right\} .$$

Plan

We are going to give a quantitative version of the Godunov-Ryabenkii condition, meaning an estimate that will be satisfied for strongly stable schemes.

We are going to give a quantitative version of the Godunov-Ryabenkii condition, meaning an estimate that will be satisfied for strongly stable schemes.

We rewrite the resolvent equation as an induction of the first order. We start from

$$\begin{cases} W_j - z^{-1}\, Q\, W_j = F_j\,, & j \geq 1\,, \\ W_j - B_j\, W_1 = g_j\,, & j = 1-r, \dots, 0\,, \end{cases}$$

which we rewrite as

$$\begin{cases} \sum_{\ell=-r}^{p} \mathbb{A}_\ell(z)\, W_{j+\ell} = F_j\,, & j \geq 1\,, \\ W_j - B_j\, W_1 = g_j\,, & j = 1-r, \dots, 0\,. \end{cases}$$

The matrices $\mathbb{A}_\ell$ depend holomorphically on $z \neq 0$.
From now on, we make the following assumption.

Assumption (Noncharacteristic discrete boundary)

The matrices $\mathbb{A}_{-r}(z)$ and $\mathbb{A}_p(z)$ are invertible for all $z \in \overline{\mathcal{U}}$.

The matrices $\mathbb{A}_\ell$ depend holomorphically on $z \neq 0$.
From now on, we make the following assumption.

Assumption (Noncharacteristic discrete boundary)

The matrices $\mathbb{A}_{-r}(z)$ and $\mathbb{A}_p(z)$ are invertible for all $z \in \overline{\mathcal{U}}$.

We can thus introduce the matrix

$$\mathbb{M}(z) := \begin{pmatrix} -\mathbb{A}_p(z)^{-1}\,\mathbb{A}_{p-1}(z) & \dots & \dots & -\mathbb{A}_p(z)^{-1}\,\mathbb{A}_{-r}(z) \\ I & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & I & 0 \end{pmatrix}.$$

The resolvent equation equivalently reads

$$\begin{cases} \mathcal{W}_{j+1} = \mathbb{M}(z)\, \mathcal{W}_j + \mathcal{F}_j\,, \quad j \geq 1\,, \\ \mathbb{B}\, \mathcal{W}_1 = \mathcal{G}\,, \end{cases}$$

with $\mathcal{W}_j := (W_{j+p-1}, \dots, W_{j-r}) \in \mathbb{C}^{N\,(p+r)}$, $j \geq 1$, and new source terms.

The new boundary conditions apply to the vector $\mathcal{W}_1$ for $q < p$.

The resolvent equation equivalently reads

$$\begin{cases} \mathcal{W}_{j+1} = \mathbb{M}(z)\,\mathcal{W}_j + \mathcal{F}_j\,, \quad j \geq 1\,, \\ \mathbb{B}\,\mathcal{W}_1 = \mathcal{G}\,, \end{cases}$$

with $\mathcal{W}_j := (W_{j+p-1}, \dots, W_{j-r}) \in \mathbb{C}^{N\,(p+r)}$, $j \geq 1$, and new source terms.

The new boundary conditions apply to the vector $\mathcal{W}_1$ for $q < p$.

In the case $q \geq p$, there is another equivalent formulation.

The characterization of strong stability transposes to this equivalent form of the resolvent equation.

Characterization of strong stability

Theorem (Characterization of strong stability for $q < p$)

The numerical scheme is strongly stable **if and only if** for all $R \geq 2$, there exists a constant $C_R > 0$ such that for all $z \in \mathcal{U}$ with $|z| \leq R$, for all $(\mathcal{F}_j) \in \ell^2$ and for all $\mathcal{G} \in \mathbb{C}^{Nr}$, the **new** resolvent equation has a unique solution $(W_j) \in \ell^2$ and this solution satisfies

$$\frac{|z|-1}{|z|} \sum_{j \geq 1-r} |\mathcal{W}_j|^2 + |\mathcal{W}_1|^2 \leq C_R \left\{ \frac{|z|}{|z|-1} \sum_{j \geq 1} |\mathcal{F}_j|^2 + |\mathcal{G}|^2 \right\} .$$

Stable eigenvalues

We take a closer look at the case $\mathcal{F}_j = 0$. Then ℓ^2 solutions to

$$\mathcal{W}_{j+1} = \mathbb{M}(z)\,\mathcal{W}_j\,,$$

correspond to eigenvalues of $\mathbb{M}(z)$ in the unit disk. They are uniquely determined by $\mathcal{W}_1$.

Stable eigenvalues

We take a closer look at the case $\mathcal{F}_j = 0$. Then ℓ^2 solutions to

$$\mathcal{W}_{j+1} = \mathbb{M}(z)\,\mathcal{W}_j\,,$$

correspond to eigenvalues of $\mathbb{M}(z)$ in the unit disk. They are uniquely determined by $\mathcal{W}_1$.

Lemma, Kreiss (1968)

Let the discretization of the Cauchy problem be stable. Then for $z \in \mathcal{U}$, $\mathbb{M}(z)$ has **no eigenvalue** on $\mathbb{S}^1$. Its stable eigenspace associated with eigenvalues in the unit disk is denoted $E^s(z)$; it depends holomorphically on $z \in \mathcal{U}$ and its dimension equals $N\,r$.

The UKLC

Proposition

Let the discretization of the Cauchy problem be stable. If the numerical scheme is strongly stable, then for all $R \geq 2$, there exists a constant $C_R > 0$ such that for all $z \in \mathcal{U}$ with $|z| \leq R$, there holds

$$\forall \mathcal{W} \in E^s(z) , \quad |\mathcal{W}| \leq C_R \, |\mathbb{B} \, \mathcal{W}| \, .$$

In other words, the mapping

$$\Phi(z) \, : \, \mathcal{W} \in E^s(z) \longmapsto \mathbb{B} \, \mathcal{W} \in \mathbb{C}^{Nr} \, ,$$

is an isomorphism for all $z \in \mathcal{U}$. Moreover for all $R \geq 2$, the inverse $\Phi(z)^{-1}$ is uniformly bounded with respect to $z \in \mathcal{U}$, $|z| \leq R$.

Observe that the Godunov-Ryabenkii condition asks that $\Phi(z)$ be an isomorphism for all $z \in \mathcal{U}$, but there is no control of the norm of the inverse mapping as z approaches the boundary of $\mathcal{U}$. This norm may explode !

This is a main gap, and it it the analogue of the gap between the Lopatinskii condition and the **uniform** Lopatinskii condition for PDEs.

The Godunov-Ryabenkii condition with a uniform control of the inverse mapping will be called the Uniform Kreiss-Lopatinskii Condition. It is a **necessary** condition for strong stability, and it can be checked on some specific examples.

Plan

Assumption (ℓ^2-stability for the discretized Cauchy problem)

The amplification matrix $\widehat{Q}(\kappa)$ is uniformly power bounded for $\kappa = e^{i\xi} \in \mathbb{S}^1$.

Gustafsson-Kreiss-Sundström (1972) : characterization of strong stability by the UKLC under **strong structural assumptions** on the eigenelements of $\widehat{Q}$ (either regular contact point with the circle, or unitary behavior as in the leap-frog scheme).

Goldberg-Tadmor (1978, ...) : convenient formulations of the UKLC using simultaneous diagonalization of the matrices (scalar problems).

Michelson (1983) : characterization of strong stability by the UKLC in **multi-d** under a strong **dissipation assumption** on $\widehat{Q}$.

In these works, the dissipation assumption “seems” only to be **technical** (sufficient, but not necessary). Moreover, it is not completely satisfactory in view of applications. In multi-d, dissipation is not so common.

In some sense, ℓ^2-stability should be “close to” sufficient in order to characterize strong stability (this is what the theory for PDEs tells us).

Main goal of the study

Characterize strong stability by the UKLC under the **most general** structural assumptions on the eigenelements of $\widehat{Q}$. In particular :

$\longrightarrow$ Do not use simultaneous diagonalization of the matrices A_ℓ.
$\longrightarrow$ Preliminary work for the multi-d analysis.

The resolvent equation

We recall that the resolvent equation is written as

$$\begin{cases} \mathcal{W}_{j+1} = \mathbb{M}(z)\,\mathcal{W}_j + \mathcal{F}_j\,, \quad j \geq 1\,, \\ \mathbb{B}\,\mathcal{W}_1 = \mathcal{G}\,, \end{cases}$$

where $\mathbb{M}$ is **holomorphic** on a neighborhood of $\{|z| \geq 1\}$.

The stable eigenspace of $\mathbb{M}(z)$, associated with eigenvalues in the unit disk, is denoted $E^s(z)$. It depends holomorphically on $z \in \mathcal{U}$ and its dimension equals $N\,r$. In other words, E^s defines a holomorphic bundle over the open set $\mathcal{U}$.

Extending the stable subspace

Assumption (geometric regularity of eigenelements)

The operator Q is geometrically regular.

Extending the stable subspace

Assumption (geometric regularity of eigenelements)

The operator Q is geometrically regular.

Theorem, C. (2009)

We assume that the numerical boundary is noncharacteristic, and that the discretization of the Cauchy problem is stable and geometrically regular. Then the stable bundle E^s **extends continuously** to the circle $\{|z| = 1\}$, and the extended bundle is continuous on $\{|z| \geq 1\}$.

The main stability result

Theorem, C. (2009)

Under the same assumptions as in the previous Theorem, the numerical scheme is strongly stable **if and only if** there holds
$E^s(z) \cap \text{ Ker } \mathbb{B}(z) = \{0\}$ for all $|z| \geq 1$.

The latter condition is an equivalent -and more convenient- formulation of the UKLC.

This result is in the same spirit as the work by Métivier-Zumbrun (2004-2005) on PDEs, and gives an optimal generalization of the works by Gustafsson-Kreiss-Sundström (1972).

Outline of the proof : construction of K-symmetrizers, as in Métivier-Zumbrun (2004). **Both theorems follow from only one argument** !

Outline of the proof : construction of K-symmetrizers, as in Métivier-Zumbrun (2004). **Both theorems follow from only one argument** !

Diagonalize $\mathbb{M}(z)$ with an appropriate discrete block structure, and construct a symmetrizer for each block.

According to the behavior of the eigenvalues of $\widehat{Q}(\kappa)$, the blocks of $\mathbb{M}(z)$ have various behaviors (size, sign of the coefficients etc.).

In the continuous case, the spectrum of the symbol $A(\xi)$ is included in $\mathbb{R}$ $\longrightarrow$ very good localization, and therefore **precise information** on the block structure.

In the discrete case, the spectrum of $\widehat{Q}(\mathrm{e}^{i\xi})$ is included in the closed unit disk, a **thick** region. **Less precise information** because the scheme can produce parabolic, dispersive etc. behavior at high frequencies. (This behavior is independent of the consistency analysis.)

In the continuous case, the spectrum of the symbol $A(\xi)$ is included in $\mathbb{R}$ $\longrightarrow$ very good localization, and therefore **precise information** on the block structure.

In the discrete case, the spectrum of $\widehat{Q}(\mathrm{e}^{i\xi})$ is included in the closed unit disk, a **thick** region. **Less precise information** because the scheme can produce parabolic, dispersive etc. behavior at high frequencies. (This behavior is independent of the consistency analysis.)

Classification of the blocks according to their size and to their dissipation index. Construction of **new symmetrizers**. Unfortunately Kreiss' construction **does not work** here ! The analysis in Gustafsson-Kreiss-Sundström (1972) only considers two possible cases.

Plan

- **Two main assumptions** on the discretized Cauchy problem : stability and geometric regularity of eigenelements (stability is equivalent to the von Neumann condition).
- Precise definition of the **discrete block structure condition**.
- We **have not used** any simultaneous diagonalization property on the matrices of the numerical scheme !
- The analysis works in exactly the same way for multi-steps schemes (e.g., leap-frog).

Awfully pessimistic conclusion (not entirely true)

Numerical schemes in 1d are worse than PDEs in multi-d...