$L^1$ stability theory and rigorous error estimates for some 1D hyperbolic systems of balance laws. (1).

Laurent Gosse (with Debora Amadori)

IAC-CNR (Bari and then Rome)
Plan of the talk

1. Introduction

2. High-order schemes and local truncation error (LTE)
   - Approximation and evolutionary errors
   - Numerical pathologies in standard schemes

3. A trivial example: linearized shallow water + topography
   - Two different types of numerical schemes
   - Numerical investigation of measured errors

4. Controlling global $L^1$ errors of scalar balance laws
   - G. Guerra’s $L^1$ functional (for WFT)
   - Specificities of the Godunov scheme (averaging error)
   - Numerical validation on similarity soln of accretive scalar law
   - Variant of classical LeVeque-Yee’s benchmark [JCP, 1990]
Numerical analysis = approximation + error control

Focus on one-dimensional hyperbolic problems yielding weak (possibly discontinuous) BV-type solns: (splitting for multi-D)

\[
\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = k(x)g(u), \quad u(t = 0, \cdot) = u_0
\]

- numerical scheme: stability (CFL cond’n), entropy consistency \(\Rightarrow\) convergence, \(\|u_{\text{num}}(t, \cdot) - u_{\text{exact}}(t, \cdot)\| \to 0\) as \(\Delta x \to 0\).
- error control: quantify tolerance as accurately as possible,
  \(\|u_{\text{num}}(t, \cdot) - u_{\text{exact}}(t, \cdot)\| \leq E(t, \Delta x^p, \text{parameters, data})\).

90% of literature keeps only grid size \(\Delta x\) and “forgets” the rest. However, LxF has wrong large-time behavior as \(t \gg \frac{1}{\Delta x}\) because \(E_{\text{LxF}} = O(\sqrt{t.\Delta x})\). \(\rightarrow\) time-dependence is crucial.
- if \(k(x) \neq 0\), any dependence on \(k'\) or \(k''\) ? (oscillations)
- weak solns \(\rightarrow p \lesssim 1\), so contrary position w.r.t. 90% (?)
The beautiful constants $C$ (first version, $k(x) \equiv 1$)

The main results of the present work are given by the following two theorems.

**Theorem 1.1.** Let $u_0 \in BV(\mathbb{R}) \cap L^\infty(\mathbb{R}) \cap L^1(\mathbb{R})$, $f \in C^1(\mathbb{R})$, and assume that $g$ satisfies a Lipschitz condition and $g(0) = 0$. Let $S(t)u_0$ denote the unique weak solution of (1.1)-(1.2) satisfying the entropy condition; then the $L^1$ convergence rate of the semi-discrete fractional step algorithm (1.5) is $1/2$. More precisely, for any $t_n = n\Delta t \in [0, T]$, the following estimate holds:

\[
\|S(t_n)u_0 - \left( S_2(\Delta t)S_1(\Delta t) \right)^n u_0 \|_{L^1(\mathbb{R})} \leq C\sqrt{\Delta t},
\]

where $C$ is a constant independent of $\Delta t$. Similar results are valid with the roles of $S_2$ and $S_1$ reversed.
The beautiful constants $C$ (2nd version, $k(x) \equiv 1$)

\begin{equation}
(2.7) \quad v(\cdot, t_n) = [H(\Delta t) E(\Delta t)]^n v_0.
\end{equation}

Note that this approximate solution is defined only at discrete $t$-values. The main result below states that the operator splitting solution, when (2.6) is solved exactly, converges linearly in $\Delta t$ to the entropy solution of (2.4).

**Theorem 2.1.** Let $u = u(x, t)$ be the entropy solution of (2.4) and $v = v(x, t_n)$ be the operator splitting solution (2.7). There exists a positive constant $K_0$, depending on $TV(u_0),\ TV(v_0),\ f,\ g,\ T_0$, such that for $T = N \Delta t \leq T_0$,

\begin{equation}
(2.8) \quad \|u(\cdot, T) - v(\cdot, T)\| \leq K_0(\|u_0 - v_0\| + \Delta t).
\end{equation}

Hence, there is a finite constant $K_0 = e^{CT_0}$, independent of $\Delta t$, such that

\[ \|u(\cdot, T) - v(\cdot, T)\| \leq K_0(\|u_0 - v_0\| + \Delta t), \]

which concludes the proof of Theorem 2.1.
Let a Cauchy problem for a given partial differential operator $\mathcal{L}$ be,

$$\partial_t u = \mathcal{L} u, \quad u(t = 0, \cdot) = u_0 \quad \text{(C)}.$$  

For $\Delta x = 2^{-J}$ fixed, a finite-differences/volumes approximation of $\mathcal{L}$ acting on $\Delta x \cdot \mathbb{Z}$ is denoted $\mathcal{L}_J$, so (C) reduces to an (infinite) differential system (Method of Lines), with $\tilde{u}(t, \cdot) \in \ell^\infty(\mathbb{Z})$, say:

$$\frac{d}{dt} \tilde{u} = \mathcal{L}_J \tilde{u}, \quad \tilde{u}(t = 0, \cdot) = \mathcal{P}_J u_0. \quad \text{(MOL)}$$

Assuming smoothness for Taylor dvts, how does the error inherent to (MOL) behave with respect to e.g. time (or other parameters)?
One “triangulates” $u(t, \cdot) - \tilde{u}(t, \cdot)$ by inserting $P_J u(t, \cdot)$, 

$$
    u - \tilde{u} = (\text{Id} - P_J)u + (P_J u - \tilde{u}) := a_J + e_J,
$$

where $a_J$ is purely an approximation error. On the contrary, $e_J(t)$ is an evolutionary error, which may accumulate in time, 

$$
    \frac{d}{dt} e_J = \frac{d}{dt} P_J u - \frac{d}{dt} \tilde{u} = P_J L u - L_J \tilde{u}. \quad (D)
$$

Triangulating, $\frac{d e_J}{dt} = (P_J L u - L_J P_J u) + (L_J P_J u - L_J \tilde{u})$,

$$
    \frac{d}{dt} e_J + (L_J \tilde{u} - L_J P_J u) = (P_J L u - L_J P_J u) := L.T.E.,
$$

and by substituting $\tilde{u}$ by $P_J u - e_J$, we get finally:

$$
    \frac{d}{dt} e_J + [L_J (P_J u - e_J) - L_J P_J u] = L.T.E., \text{ (Local Truncation Error)}.
$$

Hence, the L.T.E. is a source term inside the differential equation (D) governing the scheme’s evolutionary error.
Dissipative PDE with smooth solns

If the approximation $\mathcal{L}_J$ is linear, then (D) simplifies into,

$$\forall t > 0, \quad \frac{d}{dt} e_J(t) = \mathcal{L}_J e_J(t) + \tau_u(t),$$

where $\tau_u(t)$ stands for the L.T.E. related to ($x$-derivatives of) the exact solution $u(t, \cdot)$ to (C) at time $t$. Duhamel’s principle yields:

$$e_J(t) = \exp(t \cdot \mathcal{L}_J) \left( e_J(t = 0) + \int_0^t \exp(-s \cdot \mathcal{L}_J) \tau_u(s) ds \right).$$

Quantities like $\exp(t \cdot \mathcal{L}_J)$ are estimated by “logarithmic norms”.

In case both (C) and its (consistent) discrete approximation $\mathcal{L}_J$, are dissipative (“contractive”, “strongly stable”), the source term induces most of the error $e_J$; on the contrary, like for strictly hyperbolic systems of conservation laws, then both $\mathcal{L}_J$ and the L.T.E. can contribute to the increase of the evolutionary error.
“2nd order” (SOR) reconstruction is vulnerable to wave-interactions (accretive in $BV$) in systems of C.L.; for instance a 1-shock (first-order) interacts with a (formally second-order) 2-rarefaction. A 2-rarefaction comes out (mixing 1rst and 2nd order data): what is its (formal) accuracy?

Wave-curves deformation coming from local projections → artificial viscosity: creates intermediate points which don’t belong to shock-curve, so numerical approx. of a Riemann soln containing only a (big) 1-shock (say !) will inevitably contain small spurious 2-waves → issue of “slowly moving shocks”.
Numerical viscosity points which don’t belong to shock-curve (Jin-Liu, Arora-Roe JCP) and Engquist-Sjogreen (SINUM, 2000).
A trivial example: linearized shallow water + topography

Two different types of numerical schemes

**Linearized shallow water equations in 1D**

Linearization of SW eqns with topography around a state $\bar{\rho} > 0$, $\bar{u} = 0$ in diagonal variables, for $a(x) \in C_\infty(\mathbb{R})$ and $k(x) = \partial_x a$,

$$\partial_t u - \partial_x u = \partial_x a \quad (1)$$

Use conservative form: ($\rightarrow$ covariant deriv. on Riem. manif.)

$$\partial_t u - \partial_x (u + a) = 0 \quad (2)$$

Let $\Delta x > 0$, $\Delta t = \nu \Delta x$ with $0 < \nu \leq 1$ : discretize $(1)$ or $(2)$?

(FS) $u(t^{n+1}, x_j) - u(t^n, x_j) = \frac{u(t^n, x_j) - u(t^n, x_{j-1})}{\Delta x} - k(x_j) = R_j^n$

(WB) $u(t^{n+1}, x_j) - u(t^n, x_j) = \frac{(u + a)(t^n, x_j) - (u + a)(t^n, x_{j-1})}{\Delta x} = \tilde{R}_j^n$.

$R_j^n, \tilde{R}_j^n$ “local truncation error” (cf. Taylor) in each comp’al. cell.
Linearized shallow water equations in 1D

Pointwise error \( E(t) = \|u^{\Delta t} - u\|_{\infty} \) with smooth \( u \) is retrieved by means of bounds on sums \( \sum_n \Delta t \cdot |R^n_j| \) and Taylor developments:

For (FS): \( R^n_j = \Delta x O(\partial_{xx} u) + \Delta t O(\partial_{xx} a) \), but \( \partial_{xx} u \) grows with \( t \),

\[
E(t) \leq E(0) + \Delta t \cdot t \cdot \|a\|_{C^2} + \Delta x \cdot t \cdot (1 - \nu) (\|u(0, \cdot)\|_{C^2} + t \|a\|_{C^3}) + O(1) t^2 (\Delta x)^2
\]

Nonlinear in \( t \) ! → may grow rather fast, so \( \Delta x \) necessary to get \( E(t) \leq \tau \) (\( \tau > 0 \) a given tolerance) can shrink excessively rapidly.

For (WB): Pure transport, so \( \tilde{R}^n_j = \Delta x O(\partial_{xx}(u + a)) + O(\Delta x^2) \)

\[
E(t) \leq E(0) + \Delta x \cdot \frac{t}{2} (1 - \nu) \|u(0, \cdot) + a\|_{C^2} + C \cdot t (\Delta x)^2.
\]

N.B.: this last estimate improves when \( \Delta t / \Delta x = \nu = 1 \).
$L^1$ stability and error estimates

A trivial example: linearized shallow water + topography

Numerical investigation of measured errors

Time-evolution of measured $L^\infty$ errors of well-balanced (WB, blue) and centered source (FS, black) methods for smooth fcts $u$, $a$:

Courant numbers $\nu = 0.9, 0.7, 0.5, 0.3, 0.1$ (left) and $\nu = 1$ (right)
L1 stability and error estimates
Controlling global L1 errors of scalar balance laws

General outline: stability theory $\rightarrow$ rigorous error

- Strictly hyperbolic $n \times n$ system of balance laws, non-resonant $(0 \notin \text{spec}(\nabla f))$, $k(x) \in L^1(\mathbb{R})$ (waves exit interaction area):
  \[
  \partial_t u + \partial_x f(u) = k(x)g(u), \quad k(x)(L \cdot \nabla g \cdot R) \not\leq 0.
  \]
- Set $a_x = k$ and consider augmented system of $(n + 1)$ eqns,
  \[
  \partial_t u + \partial_x f(u) - g(u)\partial_x a = 0 \quad \partial_t a = 0. \quad (*)
  \]
  Non-resonance $\rightarrow$ strict hyperbolicity of $(*)$ $\rightarrow$ scattering.

- Numerical scheme for $(*)$: $(u^{\Delta x}, a^{\Delta x})$. The source is naturally treated as a stationary wave at each $x = x_j$ in Riemann solver.

- Error estimates for $(u, a) - (u^{\Delta x}, a^{\Delta x})$, via $L^1$ stab. theory for $(*)$. Such theory (strict hyper.) $\rightarrow$ [Bressan-Liu-Yang, 1999].
Accretive 1D scalar balance law

Nonlinear problem : weak solutions $\rightarrow$ Taylor doesn’t work.

$$\partial_t u + \partial_x f(u) = k(x)g(u) \quad (\ast)$$


- $g$ smooth, $k \in L^1 \cap BV(\mathbb{R})$ ($\rightarrow$ weaker is possible)
- ($\ast$) can be recasted as a **homogeneous 2 $\times$ 2 Temple system** (a ”scalar system”, cf. [Isaacson & Temple SIAP, G. MCOM]) by setting $\partial_x a = k(x)$ and consequently $\partial_t a \equiv 0$:

$$\begin{cases}
\partial_t u + \partial_x f(u) - g(u) \partial_x a = 0, \\
\partial_t a = 0.
\end{cases}$$

- Charact. speeds $\{0, f'(u)\}$ : **Strictly hyperbolic** if $f' > 0$.
- Riemann Inv. : $\{a, w(a, u) = \phi^{-1}(\phi(u) - a)\}$, with $\phi' = f'/g$. 
G. Guerra’s stability result [JDE ’04] for WFT approx’s

Lyapunov functional for WFT approx. of (*) with small param. $\delta$.

- For any 2 partitions $\delta_1 \geq 0, \delta_2$ and 2 WFT $U_1 = (b, v), U_2 = (a, u)$, the functional $\Phi[U_1, U_2](t)$ is uniformly equivalent to the $L^1$ distance and ”almost decaying”:

$$|\Phi(t) - \Phi(s)| \leq |t - s| \exp(\kappa TV(a))[O(\delta_1) + O(\delta_2)],$$

”$O$” is for TV of Riemann inv., $f, g$ and non-resonance $g/f'$.

- $k(x) \in L^1 \cap C^0(\mathbb{R}) : WB\text{-}WFT \rightarrow$ Kruzkov entropy solution.

Denote $p = a - b, \omega = \phi^{-1}(\phi(v) + p), W_1, W_2 \simeq$ jumps in R.I.

$$\Phi[U_1, U_2](t) = \int_{\mathbb{R}} W_2(x)[W_1(t, x)|p(x)| + |u - \omega|(t, x)]dx$$
Introduce the general methodology!

Consider \( U_1 = (u, a), \ U_2 = (v, b) \), WB-WFT approximations of:

\[
\partial_t u + \partial_x f(u) - g(u)\partial_x a = 0 \quad \partial_t a = 0.
\]

Key tool:

Lyapunov functional \( t \mapsto \Phi[U_1, U_2](t) \) unif. equivalent to \( L^1 \) norm:

\[
O(1)\left(\|u(t) - v(t)\|_{L^1} + \|a - b\|_{L^1}\right) \leq \Phi[U_1, U_2](t) \leq \Phi[U_1, U_2](0) + O(t) \\
\leq O(1)\left(\|u(0) - v(0)\|_{L^1} + \|a - b\|_{L^1}\right) + t \cdot (O(\delta_1) + O(\delta_2)).
\]

Here both the \( O(1) \) are indep. on \( t \), and possibly made explicit.

Let \( v(0) = u^{\Delta x}(0), \ b = a^{\Delta x} \) be suitably defined (piecewise constant, with jumps at the points \( j\Delta x \)), the decay of \( \Phi \) yields ...
Controlling global $L^1$ errors of scalar balance laws

G. Guerra’s $L^1$ functional (for WFT)

Error WFT-WB : non-resonance yields no Gronwall

Send $\delta_1 \to 0$, with $\delta_2 > 0$, → in $L^1$ norm (possibly local):

$$\| u(t) - v(t) \| \leq C \exp(\kappa \| k \|) \left[ \| u_0 - v_0 \| + \| a - b \| + \mathcal{O}(\delta_2) t \right].$$

The strict hyperbolicity ”killed” the exponential $L^1$-error amplification in time (cf. Langseth, Tveito & Winther) and changed it into a linear one. Improved accuracy without any reference to stabilization onto steady-state regimes.

Steady-state means $TV(w) = 0 \Rightarrow \frac{d\phi}{dt} \leq 0$. (WB property)

Weaker result for strictly hyperbolic $n \times n$ systems [ARMA] for $TV(u_0), TV(v_0), \| k \|$ small enough (Glimm interaction potential) + decay of positive waves [G., Goatin PAMS].

Observe that $\| a - b \|_{L^1} = \mathcal{O}(\Delta x) \| k \|_{L^1}$, $k'(x)$ plays no role.
From WFT to a Well-Balanced Godunov scheme

Time-marching process: Exact Riemann step $E$ followed by local projection step $P^\Delta x : u^\Delta x(t^n, \cdot) = (P^\Delta x \circ E(\Delta t))^n P^\Delta x u_0$,

\[
\begin{align*}
\partial_t u + \partial_x f(u) - g(u) \partial_x a &= 0, \\
\partial_t a &= 0,
\end{align*}
\]

$u(0, \cdot) = u_o \in BV(\mathbb{R})$

Self-similar Riemann step for “scalar system” solves exactly $(\ast)$.
- $\Delta x = L \Delta t$, $L = \max f'(u)$
- At $t = 0$ set $u^0_j = u_0(j \Delta x +)$, $a^\Delta x(x) = a(j \Delta x)$.

Set $t_n = n \Delta t$, $x_{j + 1/2} = (j + 1/2) \Delta x$. By non-resonance,
- In each open cell $(t_n, t_{n+1}) \times (x_{j-1/2}, x_{j+1/2})$, $u^\Delta t$ is solution of $\partial_t u + \partial_x f(u) = 0$, at $x = x_{j-1/2}$ a stationary wave appears;
- At $t = t_{n+1}$, projection step: average over cell $(x_{j-1/2}, x_{j+1/2})$. These steps make Lyapunov functional $\Phi$ grow $\rightarrow$ infinite sum.
Main error estimate (cf. [AG, JDE (2013)])

Let $f' > 0$ and $N = \sup_{x,u} k(x) g'(u) > 0$, $\forall t > 0$ and $x_1 < x_2$,

$$
\int_{x_1}^{x_2} |u^{\Delta t}(t, x) - u(t, x)| \, dx \leq \min \{E_1, E_2\}
$$

$$
E_1(\Delta x, t) = C_1 \Delta x + C_2 t,
$$

$$
E_2(\Delta x, t) = \sqrt{\Delta x} A(t) + \Delta x B(t).
$$

For small $t$: $A(t) \sim \sqrt{t}$, $B(t) \sim \text{const.} > 0$;

For large $t$: $A(t), B(t) \sim e^{Nt}$

The estimate $E_2(t)$ results of Kuznetsov’s method + Gronwall lemma: it is convenient for small values of $t \Delta x$. The estimate $E_1(t)$ is good for large ones ($C_2$ comes from averaging, $P^{\Delta x}$).
First error term: $E_1$

$L^1$ stability for "scalar systems" [Guerra, JDE (2004)]: $\delta_1, \delta_2 \rightarrow 0$, 

$$E_1(\Delta x, t) = C e^{\kappa TV\{a\}} \left[ (TV\{u_0\} + 1)\Delta x + TV\{w_0\} \right]$$

Consequence of decreasing nonlinear functional for non-resonant "scalar system": it decreases during the exact Riemann steps, but it remotely increases at local projection steps. So $E_1$ depends on time only because of averages ($\rightarrow$ Godunov), but not of source $g$.

- $w_0 = w_0(x)$ denotes the Riemann invariant at time $t = 0$
- $C$ and $\kappa$ are independent of both time and $\Delta x$.
- The $L^1$, $L^\infty$ norms and TV above are referred to the interval $(x_1 - Lt, x_2)$. In particular, $TV\ a = \|k\|_{L^1}$.

Increases at most **linearly** with $t$ and doesn’t see coeff. oscillations ($\rightarrow k'(x)$). Nice “large-time/coarse-grid” estimate is specific WB.
Second error term : $E_2$

Obtained by means of the entropy dissipation [Kuznetsov (1976)]

$$E_2(\Delta x, t) = \sqrt{\Delta x} A(t) + \Delta x B(t)$$

with coeffts, ($\rightarrow A(t)$ sees the $x$-derivative $k'(x))$, $\tilde{L} = L + 1$ and,

$$A(t)^2 = h(t) \left[ \text{TV} \{w_0\} + \|k\|_{L^1} \right] \left[ \tilde{L} \cdot \text{TV} \{u_0\} e^{Nt} + \text{TV} \{k\} \|g\|_{\infty} h(t) \right]$$

$$B(t) = \text{TV} \{u_0\} e^{Nt} + \|k\|_{\infty} h(t) \left[ \text{TV} \{w_0\} + \|k\|_{L^1} \right]$$

and the Gronwall Lemma brings an exponential term,

$$h(t) = \frac{e^{Nt} - 1}{N}, \quad N = \sup_{x,u} k(x)g'(u) > 0.$$

$\rightarrow$ The term $E_2$ dominates for small $t$ as $\Delta x \to 0$ (very fine grids).
Controlling global $L^1$ errors of scalar balance laws
Numerical validation on similarity soln of accretive scalar law

**N-wave for** $\partial_t v + \partial_x (v^4/4) = v$, $k'(x) \equiv 0$.

Exact similarity soln [Ha, Kim, JCP 2006] : TS’s numerical viscosity reacts with accretive splitted source and shifts the shock location $\rightarrow$ exponential growth. WB less viscous $\rightarrow$ linear growth.
Controlling global $L^1$ errors of scalar balance laws

Variant of classical LeVeque-Yee's benchmark [JCP, 1990]

“Speedup” Riccati source term, $u_0(x) = 2 - 3Y(x)$

\[
\partial_t u + \partial_x u^2 = + k(x)(1 + u)(2 - u), \quad k(x) = 2 \left(1 + \sin\left(\frac{\pi x}{10}\right)\right).
\]
Controlling global $L^1$ errors of scalar balance laws

Variant of classical LeVeque-Yee's benchmark [JCP, 1990]

“Slowdown” Riccati source term, $u_0(x) = 2 - 3 Y(x)$

$$\partial_t u + \partial_x \frac{u^2}{2} = -k(x)(1 + u)(2 - u), \quad k(x) = 2 \left( \alpha + \sin \left( \frac{\beta \pi x}{10} \right) \right).$$
Controlling global $L^1$ errors of scalar balance laws

Variant of classical LeVeque-Yee’s benchmark [JCP, 1990]

“Slowdown” Riccati source term, $u_0(x) = 2 - 3Y(x)$

$$\partial_t u + \partial_x \frac{u^2}{2} = -k(x)(1 + u)(2 - u),\ k(x) = 2\left(\alpha + \sin\left(\frac{\beta \pi x}{10}\right)\right).$$