ERROR ESTIMATES FOR WELL-BALANCED AND TIME-SPLIT SCHEMES ON A DAMPED SEMILINEAR WAVE EQUATION

DEBORA AMADORI AND LAURENT GOSSE

Abstract. *A posteriori* L^1 error estimates (in the sense of [12, 23]) are derived for both well-balanced (WB) and fractional-step (FS) numerical approximations of the unique weak solution of the Cauchy problem for the 1D semilinear damped wave equation. For setting up the WB algorithm, we proceed by rewriting it under the form of an elementary 2×2 system which linear convective structure allows to reduce the Godunov scheme with optimal Courant number (corresponding to $\Delta t = \Delta x$) to a wavefront-tracking algorithm free from any step of projection onto piecewise constant functions. A fundamental difference in the total variation estimates is proved, which partly explains the discrepancy of the FS method when the dissipative (sink) term displays an explicit dependence in the space variable. Numerical tests are performed by means of several exact solutions of the linear damped wave equation.

MSC: 35L71, 65M15, 74J20.

1. Introduction

We consider the one-dimensional damped semilinear wave equation,

$$\partial_{tt} u - \partial_{xx} u + 2k(x) g(\partial_t u) \quad = \quad 0\,, \tag{1}$$

under an assumption on k which is related to scattering problems, in the sense that incoming signals interact and get perturbed by an external phenomenon of bounded extent, which characteristic scale remains small when compared to the entire computational domain:

$$k \in L^1(\mathbb{R})\,, \qquad k(x) \geq 0\,. \tag{2}$$

By introducing the "macroscopic" variables,

$$J = \partial_t u\,, \qquad \rho = -\partial_x u$$

the wave equation (1) is equivalent to the elementary system:

$$\begin{cases} \partial_t \rho + \partial_x J = 0 \\ \partial_t J + \partial_x \rho = -2k(x) g(J)\,. \end{cases} \tag{3}$$

Oppositely, in terms of "microscopic diagonal" variables $f^{\pm}$, defined by

$$\rho = f^+ + f^-\,, \qquad J = f^+ - f^-$$

the system (3) rewrites as a discrete-velocity kinetic model:

$$\begin{cases} \partial_t (f^-) - \partial_x (f^-) = k(x)\, g(f^+ - f^-) \\ \partial_t (f^+) + \partial_x (f^+) = -k(x)\, g(f^+ - f^-)\,. \end{cases} \tag{4}$$

Date: March 18, 2014.

Assume that k satisfies (2) and that

$$g \in C^1(\mathbb{R}), \qquad g(0) = 0, \qquad g \text{ strictly increasing}. \tag{5}$$

A special case of interest is $g(J) = J$, which yields the Goldstein-Taylor model, the linear damped wave equation, or Maxwell-Cattaneo-Vernotte's equation of hyperbolic heat conduction, see [20, 19]. Initial data for (4) can be chosen such that

$$f_0^{\pm} \in BV(\mathbb{R}) \cap L^1(\mathbb{R}). \tag{6}$$

Now we state the main theorem of this paper.

Theorem 1. *Assume* (2) *and* (6), *then the two following properties hold.*

(1) Let $f^{\pm}_{\Delta t}$ stand for the fractional step/wavefront-tracking approximation of (4) and $f^{\pm}$ for its unique solution obtained as the limit of an approximating sequence $f^{\pm}_{\Delta t}$, $\Delta t \to 0$. There exists a $t^* > 0$, see (56), *such that for $t \le t^*$:*

$$\int \left| f^{\pm}_{\Delta t}(t,x) - f^{\pm}(t,x) \right| dx \quad \le \quad \Delta t \left[\mathrm{TV}\,\{f_0^{\pm}\} \, + \, C_1 t \, + \, C_2 t^2 \right], \tag{7}$$

$$C_1 \;=\; 2\, \|k\|_\infty \|g'\|_\infty \mathrm{TV}\,\{f_0^{\pm}\} + 12\, \mathrm{TV}\, k \cdot \|g\|_\infty,$$

$$C_2 \;=\; 2\, \mathrm{TV}\, k \cdot \|g\|_\infty \cdot \|k\|_\infty \cdot \|g'\|_\infty,$$

*where $t = n\Delta t$, $n \in \mathbb{N}$. For $t > t^*$, its error increases at most **linearly**,*

$$\int \left| f^{\pm}_{\Delta t}(t,x) - f^{\pm}(t,x) \right| dx \quad \le \quad \Delta t \left[\mathrm{TV}\,\{f_0^{\pm}\} \, + \, C_1 t \, + \, \tilde{C}_1\,(t - t^*) + C_2 (t^*)^2 \right]$$

$$\tilde{C}_1 \;=\; 8 \|k\|_{L^1} \|g\|_\infty \, \|k\|_\infty \|g'\|_\infty.$$

Moreover, its total variation grows at most linearly in time, see (54), and a maximum principle, see Fig. 5, holds.

(2) There exists a $\delta > 0$ such that, under the smallness restriction

$$\mathrm{TV}\, f_0^{\pm} + \|k\|_{L^1} \quad \le \quad \delta,$$

then for $f^{\pm}_{\Delta x}$ being the well-balanced/wavefront-tracking approximation of (10), the a-posteriori error estimate holds for all $t > 0$: for any $x_1 < x_2$,

$$\int_{x_1}^{x_2} \left| f^{\pm}_{\Delta x}(t,x) - f^{\pm}(t,x) \right| dx \tag{8}$$

$$\le 2\Delta x \left(\mathrm{TV}\,\{f_0^{\pm}; (x_1 - t, x_2 + t)\} + (3\|g\|_\infty + \frac{1}{2}) \|k\|_{L^1(x_1 - t, x_2 + t)} \right),$$

where $f^{\pm}$ stands now for the unique limit of an approximating sequence $f^{\pm}_{\Delta x}$, $\Delta x \to 0$. Its total variation is uniformly bounded with respect to time as written in (39).

For the definitions of $f^{\pm}_{\Delta t}$ and $f^{\pm}_{\Delta x}$, see §3.4 and §2 respectively. One may also include the more usual case of $k(x) = 1 \notin L^1(\mathbb{R})$ by assuming that

$$k = k(x) \in L^1_{loc} \cap BV_{loc}, \quad k \ge 0, \quad f_0^{\pm} \in BV(\mathbb{R}) \text{ with bounded support.}$$

Notice that in case $k(x) = 1$, the problem becomes invariant by x-translations and traveling waves $f^{\pm}(x - st)$, $|s| \neq 0$, connecting asymptotic constant states at $|x| \to +\infty$ may exist. Oppositely, selecting $k(x)$ with compact support eliminates them and time-asymptotic patterns are expected to consist in a standing wave in the vicinity of $x = 0$ and scattering waves exiting the domain of influence of k at velocity ± 1. Let us compare (7), (8) when $k(x) \equiv 1$. By using the slightly improved estimate (41), the error estimate for the well-balanced approximation becomes

$$\int_{x_1}^{x_2} |f^{\pm}_{\Delta x}(t,x) - f^{\pm}(t,x)| dx \;\leq\; 2\,\Delta x\, \mathrm{TV}\,\{f^{\pm}_0; (x_1 - t, x_2 + t)\} \\ + \Delta x\,(6\|g\|_\infty + 1)\,(x_2 - x_1) \;+\; \Delta x\,(2\|g\|_\infty + 1) 4t\,, \tag{9}$$

while the obvious simplifications $\mathrm{TV}\, k = 0$ and $\|k\|_\infty = 1$ can be done in (7).

Three main comments are in order:

- the estimate (8) heavily relies on the Bressan-Liu-Yang L^1-stability theory for weak solutions of hyperbolic systems of conservation laws ([11, 10]). The observation that it yields a-posteriori estimates, similarly as Kružkov theory yielding Kuznetsov a-posteriori estimates [12], was made by Laforest [23]. He derived also rigorous local error indicators, responsible for the increase in time of the L^1 error, called "discrepancies" (see [23], Theorem 3.3). Here, we follow a closely related methodology to study *inhomogeneous* balance laws and highlight which type of source term discretization entails better control on the time-growth of the global L^1-error.
- The WB estimate is independent of $\mathrm{TV}\,(k)$, it only perceives the L^1 norm of k. Hence, assuming that $k \in L^1 \cap BV(\mathbb{R})$ has compact support, say (a,b), there exists an optimal constant of Poincaré's inequality,
$$\|k\|_{L^1(a,b)} \leq \frac{b-a}{2} \mathrm{TV}\,(k).$$
This implies that the existence of a critical time, $\widetilde{t} \in \mathbb{R}^+$, growing with $\sqrt{b-a}$ and the sup-norms of k, g, g' taken on the positively invariant domain of (4), beyond which the error estimate of the split-scheme (7) is inevitably greater that the one of the WB scheme (8).
- By glancing at the estimate (9), one may think that in the case $k \equiv 1$, the overall performance of the WB scheme decreases as its (local) L^1 error appears to be growing linearly with the time t. An explanation could be that, by construction, the WB discretization doesn't perceive moving traveling waves, except for the static ones for which $s = 0$. However, the estimates (7) and (9) still are of quite different natures: in (7), the linear amplification acts on $\Delta x \mathrm{TV}\,(f^{\pm}_0)$ and for small time, $1 + 2t\|k\|_\infty \|g'\|_\infty \simeq \exp(2t\|k\|_\infty \|g'\|_\infty)$. Accordingly, the error amplification depends on both the oscillations of the initial data and a "Gronwall-type" factor (like in [3]). It doesn't seem possible to do the same interpretation for (9) in which the error growth simply results from the widening of the cone of dependence. Moreover, good performances of the WB scheme originally proposed in [19], in terms of consistency with the asymptotic behavior prescribed by [9], were independently shown in [6] (see §7.1.2, the case $\alpha = 0$)

About global existence of solutions to system (3) for BV data with $k \equiv 1$, see for instance [4] where a fractional step approach was used to define approximate solutions. Thanks to the above conditions on g, the increase of total variation at each time step where the source is added can be accurately controlled, hence one obtains strong compactness of sequences of approximations generated by the algorithm. Other relevant references are [14, 25]. A careful treatment of 1D systems of hyperbolic semilinear equations is given in [28]. A general study of the temporal behavior of error estimates appears in [1]. Our framework applies for instance to models endowed with a two-scale discontinuous relaxation parameter, see [13].

The paper is organized as follows. Section 2 focuses on the well-balanced (WB) approximation: §2.1 deals with the non-conservative Riemann problem and §2.2 with general interaction estimates. In §2.3, we give convergence results by means of the BV-bound (39). The error estimate (8) is established in §2.4 by setting up a Bressan-Liu-Yang functional uniformly equivalent with the L^1 norm (following [3, 23]) which decreases in time: then the resulting L^1 error is easily deduced. Section 3 follows by focusing on the fractional-step (time-split) approximation: §3.1 studies the possible increase in total variation due to the space-dependence of the sink term, the BV-bound (54) is deduced in §3.2. A *local truncation error* (LTE) is proved for general space-dependent sink terms, following [5], from which the global L^1 error (7) follows. At last, §4 presents a few numerical results based on several exact solutions given in [24, 21], displaying for instance a bifurcation phenomenon of wave speed. Finally, §5 gives concluding remarks.

2. The Well-Balanced approximation

In this context, the WB approach consists in dealing with the inhomogeneous system (4) by means of a non-conservative homogeneous 3×3 system, which turns out to be equivalent for smooth $a(x)$,

$$\begin{cases} \partial_t f^- - \partial_x f^- - g(f^+ - f^-)\partial_x a &= 0 \\ \partial_t f^+ + \partial_x f^+ + g(f^+ - f^-)\partial_x a &= 0 \\ \partial_t a &= 0 \end{cases} \tag{10}$$

where

$$a = a(x) \doteq \int_{-\infty}^{x} k(y)\, dy \,.$$

From assumption (2) one has that

$$a \in BV(\mathbb{R}) \cap C(\mathbb{R}) \,, \qquad a_x \geq 0 \,. \tag{11}$$

This procedure, which consists in localizing a source term of bounded extent into a countable collection of Dirac masses in order to integrate it inside a Riemann solver by means of an elementary wave which is obviously linearly degenerate, appears to trace back to the paper by Glimm and Sharp [15]. It is extensively used in [17].

The characteristic speeds of system (10) are $\lambda = -1,\, 0,\, 1$ with corresponding eigenvectors $r_- = (0,1,0)^t$, $r_0 = (-g,-g,1)^t$, $r_+ = (1,0,0)^t$. We will call *0-wave curves* those characteristic curves corresponding to $\lambda = 0$. One can easily check that the characteristic curves are straight lines [17, §8.1.3]. Indeed, this is obvious for $r_\pm$, while for r_0 we observe that, along a 0-wave curve, one has $f^+ = f^-$ so that g and hence r_0 are constant. Therefore 0-wave curves are straight lines, not necessarily parallel to each other.

2.1. The 3×3 Riemann problem and positively invariant domains.

Let

$$U_\ell = (f_\ell^-, f_\ell^+, a_\ell) \,, \qquad U_r = (f_r^-, f_r^+, a_r)$$

be a given a Riemann data for (10). The Riemann problem for system (10) is solved in terms of the three characteristic families, resulting in three waves: the two ± 1–waves, with corresponding speed ± 1, where only $f^\pm$ can change its value;

and the $0-wave$, corresponding to the stationary field of (10), evolving along the stationary equations

$$\partial_x f^{\pm} = -k(x)g(J)\,. \tag{12}$$

Notice that $J = f^+ - f^-$ is constant along stationary solutions.

The intermediate states in the Riemann fan are

$$U_1 = (f_*^-, f_\ell^+, a_\ell)\,, \qquad U_2 = (f_r^-, f_*^+, a_r)\,,$$

while the waves appearing in the solution are as follows: U_ℓ and U_1 are connected by a (-1)–wave of size σ_{-1}, U_1 and U_2 are connected by a 0–wave of size σ_0, and U_2 and U_r are connected by a 1–wave of size σ_1 where

$$\sigma_{-1} \;=\; f_*^- - f_\ell^- \;=\; J_\ell - J_* \;=\; \rho_{*,\ell} - \rho_\ell \tag{13}$$

$$\sigma_0 \;=\; a_r - a_l \tag{14}$$

$$\sigma_1 \;=\; f_r^+ - f_*^+ \;=\; J_r - J_* \;=\; \rho_r - \rho_{*,r}\,. \tag{15}$$

Here the "$*$" denotes the corresponding value related to the 0-wave: more precisely, $(\rho_{*,\ell}, J_*)$ and $(\rho_{*,r}, J_*)$ denote the left and right state long the 0-wave, respectively, in term of the variables (ρ, J). Notice that J is constant across the 0-wave. The states along the 0-wave satisfy a discrete version of (12), that is

$$f_*^+ - f_\ell^+ \;=\; f_r^- - f_*^- \;=\; -g(J_*)(a_r - a_l)\,; \tag{16}$$

this implies (compare with steady equations in (3)):

$$\rho_{*,r} - \rho_{*,\ell} \;=\; -2g(J_*)(a_r - a_l)\,.$$

Remark 1. A more accurate choice for measuring the size of 0-waves, rather than (14), may be $\sigma_0 = g(J_*)(a_r - a_l) = -(\Delta f^{\pm})$ instead. We shall not pursue in this direction hereafter.

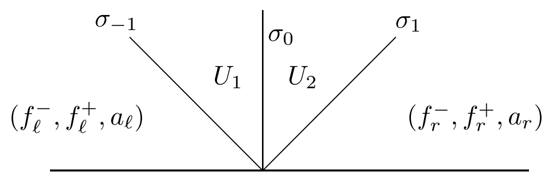

FIGURE 1. The solution to the Riemann problem.

Proposition 1. *Let $m < M$, $\delta := a_r - a_\ell > 0$ and initial states $f_\ell^{\pm}$, $f_r^{\pm} \in [m, M]$. Then one has*

$$m \le f^{\pm}(x,t) \le M \tag{17}$$

and

$$|f_r^+ - f_\ell^+| + |f_r^- - f_\ell^-| - 2C\delta \le |\sigma_{-1}| + |\sigma_1| \le |f_r^+ - f_\ell^+| + |f_r^- - f_\ell^-| + 2C\delta \tag{18}$$

where

$$C \;=\; \max\{|g(M)|, |g(m)|\}\,. \tag{19}$$

Proof. Define the intermediate value J_* implicitly by the equation

$$J_* + g(J_*)\delta \quad = \quad f_\ell^+ - f_r^- \,, \qquad \delta = a_r - a_\ell \geq 0\,. \tag{20}$$

This is well-defined since the map

$$x \mapsto x + g(x)\delta$$

is strictly increasing for $\delta \geq 0$ (recall that $g' > 0$). Hence the values f_*^+, f_*^- are defined by the identity

$$f_*^+ - f_r^- \quad = \quad f_\ell^+ - f_*^- \;=\; J_*\,, \tag{21}$$

and then the intermediate values f_*^-, f_*^+ satisfy (16). The estimates can be proved:

- By denoting $(x, y) = (f_*^+, f_*^-)$, one can easily find that
$$(f_\ell^+ - x)(x - f_r^-) = (f_\ell^+ - y)(y - f_r^-) = J_* \cdot g(J_*)\delta\,. \tag{22}$$
Noticing that $u \cdot g(u) \geq 0$ for all u, we conclude that, if $\delta \geq 0$, the new values $f_*^\pm$ do not leave the interval with extrema f_ℓ^+, f_r^-:
$$m \leq \min\{f_\ell^+, f_r^-\} \leq x,\, y \leq \max\{f_\ell^+, f_r^-\} \leq M\,,$$
therefore (17) is proved.
- Finally, concerning (18), we use (15) and (16) to find that
$$\begin{aligned}
|f_r^+ - f_\ell^+| &\leq |f_r^+ - f_*^+| + |f_*^+ - f_\ell^+| \\
&= |\sigma_1| + |f_*^+ - f_\ell^+| = |\sigma_1| + |g(J_*)|\delta \\
&\leq |\sigma_1| + C\delta \\
&\leq |f_r^+ - f_\ell^+| + |f_*^+ - f_\ell^+| + C\delta \\
&\leq |f_r^+ - f_\ell^+| \,+\, 2C\delta\,,
\end{aligned}$$
with C as in (19), so that
$$|f_r^+ - f_\ell^+| - C\delta \leq \quad |\sigma_1| \quad \leq |f_r^+ - f_\ell^+| \,+\, C\delta\,,$$
An analogous estimate holds for σ_{-1}. Hence we end up with (18).

□

Remark 2. If $\delta = a_r - a_\ell < 0$, assume moreover that

$$(\sup g')|\delta| < 1\,. \tag{23}$$

Then there exists a unique solution to the Riemann problem for (10) with data U_ℓ for $x < 0$, U_r for $x > 0$. In other words, the jump in a should be sufficiently small. Indeed, thanks to (23), the map $\mathbb{R} \ni x \mapsto x + g(x)\delta$ is strictly increasing. Moreover, (17) does not hold necessarily for $\delta < 0$, since (22) does not have the correct sign.

We are now in position to explain the construction of an algorithm able to deliver a Well-Balanced approximation of (3).

Let $a(x)$ satisfy (11) and $f_0^\pm$ satisfy (6). Fix $\Delta x > 0$ and set $x_j = j\Delta x$ for $j \in \mathbb{Z}$. Approximate the initial data $f_0^\pm$ and a with constant values on each (x_j, x_{j+1}), say

$$(f_0^\pm)^{\Delta x}(x) = f_0^\pm(x_j)\,, \qquad a^{\Delta x}(x) = a(x_j) \text{ for } x \in (x_j, x_{j+1})\,.$$

Denote by $f^- = (f^-)^{\Delta x}$ and $f^+ = (f^+)^{\Delta x}$ the approximate solutions naturally defined by a wave-front tracking algorithm. Let $m \leq M$ be constants such that

$$\forall x \in \mathbb{R}, \qquad m \leq f_0^\pm(x) \leq M\,. \tag{24}$$

By means of Prop. 1, since $a_x \geq 0$, the approximate solution remains confined inside the same interval:

$$\forall t > 0, \qquad m \leq f^{\pm}(t,.) \leq M\,. \tag{25}$$

2.2. **General study of interaction patterns.** Now we investigate the interaction between various patterns of waves for the system (10) because the introduction of $a(x)$ yields a nonlinearity. The amplitude of waves is defined at (13)–(15).

Proposition 2. *Let U_ℓ and U_m be connected by a complete Riemann pattern of size $q^-_{\pm 1}$ and q_0. Let U_m and U_r be connected by a single wave as described in the cases below. Finally let $q^+_{\pm 1}$ be the sizes of the ± 1-waves solving the Riemann problem for U_ℓ, U_r (see Figures 2 and 3). Set $C_1 = Lip(g)$, the following holds:*

(a) If U_m and U_r be connected by a -1-wave of size σ_{-1}, then one has

$$|q^+_{-1} - q^-_{-1} - \sigma_{-1}| = |q^+_1 - q^-_1| \quad \leq \quad C_1\, q_0\, |\sigma_{-1}|\,. \tag{26}$$

(b) If U_m and U_r be connected by a 0-wave of size σ_0, then one has

$$|q^+_{-1} - q^-_{-1}| = |q^+_1 - q^-_1| \quad \leq \quad C_1\, |q^-_1|\, \sigma_0\,. \tag{27}$$

(c) If U_m and U_r be connected by a 1-wave of size σ_1, then one has

$$q^+_{-1} = q^-_{-1}\,, \qquad q^+_1 = q^-_1 + \sigma_1\,.$$

Proof. Denote by J^-_*, J^+_* the intermediate values of J in the Riemann problem for (U_ℓ, U_m) and (U_ℓ, U_r) respectively. Then the following identities are valid for the sizes of waves:

$$\begin{cases} q^+_{-1} - q^-_{-1} & = J^-_* - J^+_*\,, \\ q^+_1 - q^-_1 & = (J^-_* - J^+_*) + (J_r - J_m)\,. \end{cases} \tag{28}$$

Indeed, it is sufficient to remind the definitions (13), (15) for the size of the waves; for instance we get $q^+_{-1} - q^-_{-1} = (J_\ell - J^+_*) - (J_\ell - J^-_*)$ and hence the first identity. Similar for the second one.

- In case *(c)* one has $J_r - J_m = \sigma_1$ and $J^-_* = J^+_*$. Hence the thesis simply follows from (28), being $q^+_1 - q^-_1 - \sigma_1 = 0 = q^+_{-1} - q^-_{-1}$.

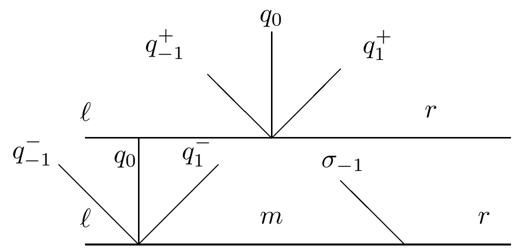

FIGURE 2. Illustration of Case (a).

- About *(a)*, one has that $\sigma_{-1} = J_m - J_r$; moreover, by using (20), the quantities J_*^+, J_*^- satisfy

$$J_*^- + g(J_*^-)q_0 = f_\ell^+ - f_m^-\,, \qquad J_*^+ + g(J_*^+)q_0 = f_\ell^+ - f_r^-\,,$$

so that

$$(J_*^- - J_*^+)\,(1 + g'(\xi)q_0) = \sigma_{-1}\,.$$

Then

$$\begin{aligned} q_1^+ - q_1^- &= \underbrace{(J_r - J_m)}_{=-\sigma_{-1}} + (J_*^- - J_*^+) \\ &= \sigma_{-1}\left(-1 + \frac{1}{1 + g'(\xi)q_0}\right) = -\sigma_{-1}\cdot\frac{g'(\xi)q_0}{1 + g'(\xi)q_0}\,. \end{aligned}$$

Since $1 + g'(\xi)q_0 \geq 1$, then (26) follows:

$$|q_1^+ - q_1^-| \leq |\sigma_{-1}|\,g'(\xi)q_0 \leq C_1\,|\sigma_{-1}|\,q_0\,.$$

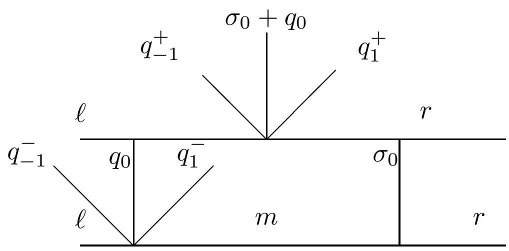

FIGURE 3. Illustration of Case (b).

- It remains to consider *(b)*, where $J_r = J_m$ and hence (28) reduces to

$$q_1^+ - q_1^- = q_{-1}^+ - q_{-1}^- = J_*^- - J_*^+\,. \tag{29}$$

The following identities hold for $J_*^\pm$:

$$J_*^- + g(J_*^-)q_0 = f_\ell^+ - f_m^-\,, \qquad J_*^+ + g(J_*^+)\,(q_0 + \sigma_0) = f_\ell^+ - f_r^-\,.$$

Therefore

$$\left(J_*^- - J_*^+\right)\Big(1 + g'(\xi)q_0\Big) - g(J_*^+)\sigma_0 = f_r^- - f_m^-\,.$$

From (16) we immediately deduce that $f_r^- - f_m^- = -g(J_r)\sigma_0$, so that

$$\begin{aligned} \left(J_*^- - J_*^+\right)(1 + g'(\xi)q_0) &= \sigma_0\left(g(J_*^+) - g(J_r)\right) \\ &= \sigma_0\, g'(\eta)\,\left(J_*^+ - J_r\right) \\ &= \sigma_0\, g'(\eta)\,\left[-(J_*^- - J_*^+) - \underbrace{(J_r - J_*^-)}_{=q_1^-}\right] \\ &= -\sigma_0\, g'(\eta)\,(J_*^- - J_*^+) - \sigma_0\, g'(\eta)\, q_1^-\,. \end{aligned}$$

The previous identity rewrites as

$$\left(J_*^- - J_*^+\right)(1 + g'(\xi)q_0 + \sigma_0\, g'(\eta)) \;=\; -\,\sigma_0\, g'(\eta)\, q_1^- \,.$$

Recalling that g', q_0, σ_0 are all ≥ 0, it follows from the previous identity that

$$|J_*^- - J_*^+| \;\;\leq\;\; \sigma_0\, g'(\eta)\, |q_1^-| \;\leq\; C_1\, \sigma_0\, |q_1^-|$$

and hence, going back to (29), we get (27).

□

Proposition 3. ***(Multiple interaction)*** *Assume that a* 1*–wave, a 0–wave and a* -1*–wave interact. Let* σ_{-1}^-*,* σ_1^- *be the sizes of the incoming waves and* σ_{-1}^+*,* σ_1^+ *be the ones of the waves after interactions. Then*

$$|\sigma_{-1}^+| + |\sigma_1^+| \;\leq\; |\sigma_{-1}^-| + |\sigma_1^-|\,. \tag{30}$$

Besides, for $\delta = a_r - a_\ell$*, one has*

$$\begin{cases} |\sigma_{-1}^+| - |\sigma_{-1}^-| & \leq\; C_1\delta\left(|\sigma_{-1}^-| + |\sigma_1^-|\right) \\ |\sigma_1^+| - |\sigma_1^-| & \leq\; C_1\delta\left(|\sigma_{-1}^-| + |\sigma_1^-|\right)\,. \end{cases} \tag{31}$$

Proof. We proceed by letting interactions occur two at a time, and then collect the result. The same procedure was used in [2].

- Assume first that the $(+1)$–wave interacts with the 0–wave, then two $\pm$ waves of size $\widetilde{\sigma}_{\pm 1}$ will outgo the interaction point. We are in case *(b)* of Prop. 2, where $\sigma_{-1}^- = 0 = \sigma_0$ and (29) reduces to $\widetilde{\sigma}_1 - \sigma_1^- = \widetilde{\sigma}_{-1}$, so that

$$\widetilde{\sigma}_1 - \widetilde{\sigma}_{-1} \;\;=\;\; \sigma_1^-\,. \tag{32}$$

  By equating $\rho_r - \rho_\ell$ before and after the interaction, we find that

$$\widetilde{\sigma}_1 + \widetilde{\sigma}_{-1} - 2g(\widetilde{J})\delta \;\;=\;\; \sigma_1^- - 2g(J_m)\delta\,. \tag{33}$$

  Subtracting (32) from (33), we get

$$2\widetilde{\sigma}_{-1} - 2g(\widetilde{J})\delta \;\;=\;\; -2g(J_m)\delta\,,$$

  so that

$$\begin{aligned} \widetilde{\sigma}_{-1} \;&=\; \left[g(\widetilde{J}) - g(J_m)\right]\delta = g'(\xi)\underbrace{\left(\widetilde{J} - J_m\right)}_{=-\widetilde{\sigma}_1}\delta \\ &=\; -g'(\xi)\widetilde{\sigma}_1\delta\,. \end{aligned}$$

  Hence $\mathrm{sgn}(\widetilde{\sigma}_1) = -\mathrm{sgn}(\widetilde{\sigma}_{-1})$, and using again (32), we find that

$$\mathrm{sgn}(\sigma_1^-) = \mathrm{sgn}(\widetilde{\sigma}_1).$$

  Therefore we have proved that

$$|\widetilde{\sigma}_1| + |\widetilde{\sigma}_{-1}| \;\;=\;\; |\sigma_1^-|\,. \tag{34}$$

- After this interaction, the wave of size $\widetilde{\sigma}_1$ will cross the (-1)–wave of size σ_{-1}^-, clearly without changing size. The interaction between this last wave and the 0–wave will produce two new waves, $\widehat{\sigma}_{\pm 1}$. Analogously as before, they will satisfy

$$|\widehat{\sigma}_1| + |\widehat{\sigma}_{-1}| \;\;=\;\; |\sigma_{-1}^-|\,. \tag{35}$$

Due to the linearity of ±1–waves, no other interaction can occur. The sizes of the outcoming waves σ_{-1}^{+}, σ_1^{+} must satisfy

$$\sigma_{-1}^{+} = \widetilde{\sigma}_{-1} + \widehat{\sigma}_{-1}\,, \qquad \sigma_1^{+} = \widetilde{\sigma}_1 + \widehat{\sigma}_1\,.$$

Therefore, by using (34) and (35), we finally get (30):

$$\begin{aligned} |\sigma_{-1}^{+}| + |\sigma_1^{+}| &\leq |\widetilde{\sigma}_{-1}| + |\widehat{\sigma}_{-1}| + |\widetilde{\sigma}_1| + |\widehat{\sigma}_1| \\ &= |\sigma_{-1}^{-}| + |\sigma_1^{-}|\,. \end{aligned}$$

Finally let us prove (31) for the 1-family, the other one being analogous. From the construction above and Prop. 2, it is easy to deduce that

$$|\widetilde{\sigma}_1 - \sigma_1^{-}| \leq C_1|\sigma_1^{-}|\delta\,, \qquad |\widehat{\sigma}_1| \leq C_1|\sigma_{-1}^{-}|\delta\,.$$

One has

$$|\sigma_1^{+}| - |\sigma_1^{-}| \quad \leq \quad |\widetilde{\sigma}_1| + |\widehat{\sigma}_1| - |\sigma_1^{-}| \;\leq\; |\widetilde{\sigma}_1 - \sigma_1^{-}| + |\widehat{\sigma}_1|\,,$$

therefore we conclude thanks to the above estimates on $|\widetilde{\sigma}_1 - \sigma_1^{-}|$ and on $|\widehat{\sigma}_1|$. □

2.3. **Bounds on total variation for WB approximation.** Previous interaction estimates allow to derive uniform bounds on the total variation for the system (10) (if $a(x) = x$, see also Sect. 8.1.1–2.1 in [17]). As usual, one defines:

$$\begin{aligned} L_{\pm}(t) &= \sum_{(\pm1)-waves} |\Delta f^{+}| + |\Delta f^{-}| \\ L_0(t) &= \sum_{0-waves} |\Delta f^{+}| + |\Delta f^{-}| \end{aligned}$$

and

$$L(t) \;\doteq\; L_{\pm}(t) + L_0(t) \;=\; \mathrm{TV}\, f^{+}(t,\cdot) + \mathrm{TV}\, f^{-}(t,\cdot)\,.$$

Remark 3. *We remark that $L_{\pm}(t)$ coincides with $\mathrm{TV}\, J(t,\cdot)$: indeed, along ±1-waves, it holds that $|\Delta f^{\pm}| = |\Delta J|$; on the other hand, since J is constant along 0-waves, then*

$$L_{\pm}(t) = \sum_{(\pm1)-waves} |\Delta J| \;=\; \mathrm{TV}\, J(t,\cdot)\,. \tag{36}$$

Let us explain how these quantities evolve in time.

- From (30) and (18), one has that $L_{\pm}(t) \leq L_{\pm}(s)$ if $t \geq s$ and thus,

$$\begin{aligned} L_{\pm}(t) &\leq L_{\pm}(0+) \\ &\leq \mathrm{TV}\, f^{+}(0,\cdot) + \mathrm{TV}\, f^{-}(0,\cdot) + 2C\,\mathrm{TV}\, a\,, \end{aligned} \tag{37}$$

  where (18) is used for $L_{\pm}(0+)$. Recall that $C = \max\{|g(M)|, |g(m)|\}$ as in (19).
- By using (16) and (19), a uniform in time estimate for L_0 reads as follows:

$$\begin{aligned} L_0(t) &= 2\sum_j |g(J_*(x_j))|\Delta a(x_j) \\ &\leq 2C\sum_j \Delta a(x_j) \;=\; 2C\,\mathrm{TV}\, a\,. \end{aligned} \tag{38}$$

Recalling that $\mathrm{TV}\, a = \|k\|_{L^1}$, we put together (37), (38) and finally get the following estimate that does not depend on time:

$$\boxed{\mathrm{TV}\, f^+(t,\cdot) + \mathrm{TV}\, f^-(t,\cdot) \;\leq\; \mathrm{TV}\, f^+(0,\cdot) + \mathrm{TV}\, f^-(0,\cdot) \;+\; 4C_0\|k\|_{L^1}\,.} \tag{39}$$

2.4. **Lyapunov functional and linear L^1 error estimate.** Among all the possible interaction patterns, we are especially interested in the ones which occur when following the stability roadmap proposed by Bressan et al. [10]. It consists mainly in introducing a nonlinear functional which tracks the time-evolution of the L^1 distance between two wavefront-tracking approximations by considering their difference as a "transversal Riemann problem" being solved by shock curves only regardless to entropy conditions. Obviously, in the present simplified framework, there is no entropy conditions at all as the system (3) is semilinear. Hence, 2 approximations $f_1^\pm, a(x)$ and $f_2^\pm, b(x)$ being given, at each point t, x, one solves the Riemann problem for (10) with left/right data:

$$f_1^\pm(t,x), a(x), \qquad f_2^\pm(t,x), b(x).$$

Let

$$q_{\pm 1}\,, \quad q_0(x) = b(x) - a(x)$$

stand for the corresponding "transversal wave-strengths", and consider, for instance, that f_1^- has a jump of size σ at the point (t, x_α): see Figure 4. In order to correctly devise the weights involved in the Lyapunov functional, it is necessary to know how the "transversal wave-strengths" evolve according to all the jumps in both $f_1^\pm, a(x)$ and $f_2^\pm, b(x)$.

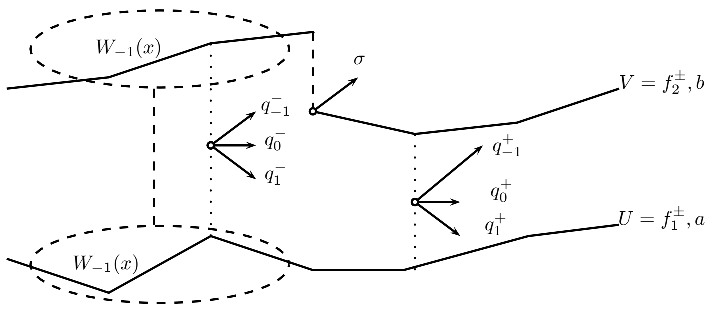

FIGURE 4. Interaction between a "transversal Riemann problem" (left) and a -1-wave resulting in the new Riemann problem (right).

In the sequel, we use all the standard notations by Bressan ([10]); the only exception is that the characteristic families are numbered $-1, 0, 1$ for obvious reasons. Let U, V stand for (f_1^-, f_1^+, a) and (f_2^-, f_2^+, b) respectively. We write σ_i^α for the size a front located at x^α, of the family $i \in \{-1, 0, 1\}$; zero-waves are measured simply by the jump of $a(x)$ or $b(x)$, respectively for U or V (see (14)). Recall that all the σ_0^α are positive, since $a(x)$ and $b(x)$ are assumed to be monotone, non-decreasing. The Lyapunov functional $\Phi[U,V]$ reads, for $x_1 < x_2$ and $t \leq T = (x_2 - x_1)/2$:

$$t \mapsto \Phi[U,V](t) \;=\; \sum_{i=-1}^{1} \int_{x_1+t}^{x_2-t} |q_i(x)| W_i(x)\, dx, \tag{40}$$

where W_i are time-dependent weights, defined as follows:

$$W_i(t,x) \quad = \quad 1+\kappa_1 A_i(t,x)+\kappa_2\big(Q(U)+Q(V)\big)\,, \qquad i=-1\,,\ 0\,,\ 1$$

and

$$\begin{aligned} A_{-1}(x) &= \sum_{x_\alpha<x} \sigma_0^\alpha\,,\\ A_1(x) &= \sum_{x_\alpha>x} \sigma_0^\alpha\,,\\ A_0(t,x) &= \sum_{x_\alpha<x} |\sigma_1^\alpha| + \sum_{x_\alpha>x} |\sigma_{-1}^\alpha|\,. \end{aligned}$$

The sums above extend over all jumps in U and V. An estimate for $A_{\pm1}$ reads:

$$A_{\pm1}(t,x) \quad \leq \quad \mathrm{TV}\,a+\mathrm{TV}\,b\,.$$

On the other hand, an estimate on A_0 goes as follows. Assume that m, M are common bounds on $f_1^\pm(0,\cdot)$ and $f_2^\pm(0,\cdot)$, see (24). By defining

$$\begin{aligned} \mathcal{A}_0 &\doteq \mathrm{TV}\,f_1^-(0,\cdot)+\mathrm{TV}\,f_1^+(0,\cdot)+2C\,\mathrm{TV}\,a\,,\\ \mathcal{B}_0 &\doteq \mathrm{TV}\,f_2^-(0,\cdot)+\mathrm{TV}\,f_2^+(0,\cdot)+2C\,\mathrm{TV}\,b \end{aligned}$$

and recalling (37), one obtains

$$A_0(t,x) \quad \leq \quad L_\pm(t;U)+L_\pm(t;V) \ \leq\ \mathcal{A}_0+\mathcal{B}_0\,.$$

Here the constants κ_1, κ_2 are positive, to be chosen, and $Q(U)$, $Q(V)$ stand for the interaction potential between ±1–waves and 0–waves that show up in U, V respectively:

$$Q(U)(t) \quad = \quad \sum_\beta \sigma_0^\beta \left[\sum_{\alpha,\ x_\alpha<x_\beta} |\sigma_1^\alpha| + \sum_{\alpha,\ x_\alpha>x_\beta} |\sigma_{-1}^\alpha|\right]$$

where the sum runs over all jumps of U in (x_1+t, x_2-t). Hence

$$\begin{aligned} Q(U)(t) &\leq \mathrm{TV}\,\{a\}\,L_\pm(t;U)\\ &\leq \mathrm{TV}\,\{a\}\,L_\pm(0+,U)\\ &\leq \mathrm{TV}\,\{a\}\,\mathcal{A}_0\,. \end{aligned}$$

The situation is analogous for V:

$$Q(V)(t) \quad \leq \quad \mathrm{TV}\,\{b\}\,\mathcal{B}_0\,.$$

We estimate the sum of the Q as follows:

$$Q(U)+Q(V) \quad \leq \quad \mathrm{TV}\,\{a\}\,\mathcal{A}_0\ +\ \mathrm{TV}\,\{b\}\,\mathcal{B}_0\,.$$

In order to ensure that these weights are uniformly bounded, one must deal with the bounds:

$$\begin{aligned} W_{\pm1}(t,x) &\leq 1+\kappa_1\left(\mathrm{TV}\,a+\mathrm{TV}\,b\right)+\kappa_2\left(\mathrm{TV}\,\{a\}\,\mathcal{A}_0\ +\ \mathrm{TV}\,\{b\}\,\mathcal{B}_0\right)\,,\\ W_0(t,x) &\leq 1+\kappa_1\left(\mathcal{A}_0+\mathcal{B}_0\right)+\kappa_2\left(\mathrm{TV}\,\{a\}\,\mathcal{A}_0\ +\ \mathrm{TV}\,\{b\}\,\mathcal{B}_0\right)\,. \end{aligned}$$

Hence, once that the constant values κ_1 and κ_2 are determined, it is necessary to restrict both the total variation of initial data $f_0^\pm$ and the strength of the source term. More precisely there exists $\delta>0$ such that, if

$$\mathrm{TV}\,a\,,\quad \mathrm{TV}\,b\,,\quad \mathrm{TV}\,f_1^\pm(0,\cdot)\,,\quad \mathrm{TV}\,f_2^\pm(0,\cdot)<\delta\,,$$

then the weights satisfy $1 \leq W_1(t,x) \leq 2$.

Let us present the main steps of the analysis:

(1) We first quantify the relation between $\Phi[U,V](t)$ and the L^1 difference between the two approximate solutions. Define

$$I(t) = \int_{x_1+t}^{x_2-t} |f_1^+(t,x) - f_2^+(t,x)| + |f_1^-(t,x) - f_2^-(t,x)| dx.$$

Recalling (18) and using $W_{\pm 1} \geq 1$, one gets

$$\begin{aligned} I(t) &\leq \int_{x_1+t}^{x_2-t} |q_1| W_1 + |q_{-1}| W_{-1} + 2C|a-b|\, dx \\ &\leq \Phi[U,V](t) + (2C-1) \int_{x_1+t}^{x_2-t} |a-b|\, dx \end{aligned}$$

and also, always taking advantage of (18),

$$\begin{aligned} \Phi[U,V](t) &\leq 2 \sum_{i=-1,1} \int_{x_1+t}^{x_2-t} |q_i| dx \;+\; 2 \int_{x_1+t}^{x_2-t} |a-b| dx \\ &\leq 2I(t) \;+\; (4C+2) \int_{x_1+t}^{x_2-t} |a-b|\, dx\,. \end{aligned}$$

Altogether, assuming that $t \mapsto \Phi[U,V](t)$ decreases, it comes that:

$$\begin{aligned} I(t) &\leq \Phi[U,V](t) + (2C-1) \int_{x_1+t}^{x_2-t} |a-b|\, dx \\ &\leq \Phi[U,V](0) + (2C-1) \int_{x_1+t}^{x_2-t} |a-b|\, dx \\ &\leq 2I(0) \;+\; (4C+2) \int_{x_1}^{x_2} |a-b|\, dx \;+\; (2C-1) \int_{x_1+t}^{x_2-t} |a-b|\, dx\,. \quad (41) \end{aligned}$$

(2) Then, we want to prove that Φ does not increase in time. To see this, let's start considering interaction times:
   - When a wave front leaves the left boundary, of family $k = -1$, the weights W_i change continuously in L^1_{loc}. If the leaving wave-front is of family $k = 0$, the weight W_{-1} possibly decreases.
   - When the interaction between two waves $+1$, -1 occurs, their size do not change across interaction, so the functional does not change.
   - Now we consider the case of a ± 1 wave-front interacting with a 0-wave; the presence of reflected waves induces a possible increase in the weights W_i, which is controlled by means of (31) in Prop. 3.
     However, thanks to the presence of both the interaction potentials $Q(U)$ and $Q(V)$, the possible increase of W_i is compensated with their corresponding decay. Hence, for κ_2 big enough, the overall functional decreases.

(3) Now, following Bressan (see [10, p.155]), outside interaction times it is convenient to write the time-derivative of Φ as follows:

$$\begin{aligned}\frac{d\Phi[U,V]}{dt} &= \sum_{i=-1}^{1} |q_i(x)|W_i(x)(-1+\lambda_i)\big|_{x=x_1+t} \\ &\quad + \sum_{i=-1}^{1} |q_i(x)|W_i(x)(-1-\lambda_i)\big|_{x=x_2-t} + \sum_{\alpha}\sum_{i=-1}^{1} E_{\alpha,i}\,,\end{aligned}$$

being

$$\begin{aligned}E_{\alpha,i} &= |q_i^{\alpha+}|W_i^{\alpha+}(\lambda_i^{\alpha+}-\dot{x}^{\alpha}) - |q_i^{\alpha-}|W_i^{\alpha-}(\lambda_i^{\alpha-}-\dot{x}^{\alpha}) \\ &= \left[|q_i^{\alpha+}|W_i^{\alpha+} - |q_i^{\alpha-}|W_i^{\alpha-}\right](\lambda_i^{\alpha}-\dot{x}^{\alpha})\end{aligned}$$

where we used that the λ_i's are constant. Thanks to the linear structure of families ± 1, lots of simplification occur. For instance, if $i = k_\alpha$ then the corresponding speeds coincide $\lambda_i^\alpha = \dot{x}^\alpha$ thus $E_{\alpha,i} = 0$. Since $|\lambda_i| \le 1$, the contribution from the boundaries is non-positive and then:

$$\frac{d\Phi[U,V]}{dt} \le \sum_{\alpha}\sum_{i=-1}^{1} E_{\alpha,i}\,.$$

We formerly assumed that $t \mapsto \Phi[U,V](t)$ decreases, the next lemma proves this:

Lemma 1. *Let U, V be two approximate solutions, generated by the Well-Balanced algorithm, from initial data $U_0 = \big(f_1^{\pm}(t=0,\cdot), a(\cdot)\big)$, $V_0 = \big(f_2^{\pm}(t=0,\cdot), b(\cdot)\big)$ endowed with sufficiently small total variation so that the corresponding weights satisfy the uniform bound $1 \le W_i(t,x) \le 2$, $i \in \{0, \pm 1\}$.*

If $\kappa_1 \ge 4C_1$ then one has, outside interaction times:

$$\frac{d\Phi[U,V]}{dt} \le 0.$$

Proof. We will analyze the jumps that occur in the $V = (f_2^{\pm}, b)$ vector of unknowns; the analysis for the jumps in U is completely similar (see also [10, p.160]). Such a framework exactly meets with the interaction estimates given in Prop. 2. Accordingly, let $k_\alpha \in \{\pm 1, 0\}$ denote the characteristic family of the jump present at the abscissa x_α. To carry on, one distinguishes between each value of k_α:

- if $k_\alpha = -1 = \dot{x}^\alpha$, an easy computation shows that $E_{-1} = 0$ and that

$$q_0^+ = q_0^-\,, \qquad W_1^+ = W_1^-\,, \qquad W_0^+ - W_0^- = -\kappa_1|\sigma_{-1}|$$

  and hence

$$\begin{aligned}E_0 &= |q_0^-|\left\{W_0^+ - W_0^-\right\} = -\kappa_1|\sigma_{-1}||q_0^-|\,,\\ E_1 &= 2\left\{|q_1^+| - |q_1^-|\right\}W_1^-\,.\end{aligned}$$

  Moreover it follows from Proposition 2 that $|q_1^+| \le |q_1^-| + C_1|q_0^-||\sigma_{-1}|$.

  Therefore, recalling that the weights are supposed to be smaller that 2, one gets

$$\begin{aligned}\sum_{i=-1}^{1} E_i = E_0 + E_1 &\le -\kappa_1|q_0^-||\sigma_{-1}| + 2W_1^- C_1|q_0^-||\sigma_{-1}| \\ &\le |q_0^-||\sigma_{-1}|\,(-\kappa_1 + 4C_1) \le 0\,.\end{aligned}$$

- if $k_\alpha = 1 = \dot{x}^\alpha$, this is the simple Case (c), and

$$\begin{aligned}\sum_{i=-1}^{1} E_i &= E_{-1} + E_0 \\ &= -2\left\{|q_{-1}^+|W_{-1}^+ - |q_{-1}^-|W_{-1}^-\right\} - \left\{|q_0^+|W_0^+ - |q_0^-|W_0^-\right\}.\end{aligned}$$

Here q_0, q_{-1}, W_{-1} do not change, while

$$W_0^+ - W_0^- = +\kappa_1 \sigma_1 .$$

Hence one gets a negative sign for every $\kappa_1 > 0$:

$$\sum_{i=-1}^{1} E_i = -|q_0|\left\{W_0^+ - W_0^-\right\} = -\kappa_1 |q_0| \sigma_1 \leq 0 .$$

- if $k_\alpha = 0 = \dot{x}^\alpha$, this is Case (b), depicted in Fig. 3, with $\dot{x} = \lambda_0 = 0$ and thus $E_0 = 0$.

$$\begin{aligned}\sum_{i=-1}^{1} E_i &= E_{-1} + E_1 \\ &= -\left\{|q_{-1}^+|W_{-1}^+ - |q_{-1}^-|W_{-1}^-\right\} + \left\{|q_1^+|W_1^+ - |q_1^-|W_1^-\right\}.\end{aligned}$$

The weights $W_i^\pm$, $i = \pm 1$ jump as follows:

$$W_{-1}^+ - W_{-1}^- = +\kappa_1|\sigma_0| \geq 0 , \qquad W_1^+ - W_1^- = -\kappa_1|\sigma_0|.$$

Hence, by means of (27), we find that

$$\begin{aligned}E_{-1} &= -|q_{-1}^+|\left\{W_{-1}^+ - W_{-1}^-\right\} - W_{-1}^-\left\{|q_{-1}^+| - |q_{-1}^-|\right\} \\ &\leq -W_{-1}^-\left\{|q_{-1}^+| - |q_{-1}^-|\right\} \\ &\leq +2\left||q_{-1}^-| - |q_{-1}^+|\right| \\ &\leq 2|q_{-1}^- - q_{-1}^+| \\ &\leq 2C_1\, \sigma_0\, |q_1^-|\end{aligned}$$

while, in a quite similar way,

$$\begin{aligned}E_1 &= |q_1^-|(W_1^+ - W_1^-) + (|q_1^+| - |q_1^-|)W_1^+ \\ &\leq -\kappa_1\sigma_0|q_1^-| + 2\left|q_1^+ - q_1^-\right| \\ &\leq -\kappa_1\sigma_0|q_1^-| + 2C_1\sigma_0|q_1^-| \\ &\leq \sigma_0|q_1^-|(2C_1 - \kappa_1)\end{aligned}$$

At this point, having $\kappa_1 \geq 4C_1$ again ensures $E_{-1} + E_1 \leq 0$.

□

Since we now have the time-decay of $\Phi[U, V]$ at hand, by just selecting

$$b = P^{\Delta x} a, \qquad \partial_x a(x) = k(x),$$

and $V(t=0,\cdot) = P^{\Delta x}U(t=0,\cdot)$, one obtains that the L^1 error of the WB scheme at time $t>0$ is bounded by a uniform constant times the initial error:

$$\begin{aligned}
&\int_{x_1+t}^{x_2-t} |f^{\pm}_{\Delta x}(t,x) - f^{\pm}(t,x)|dx \\
&\quad \leq\ 2\int_{x_1}^{x_2} |f^{\pm}_{\Delta x}(0,x) - f^{\pm}(0,x)|dx + (6C+1)\Delta x \mathrm{TV}\,(a) \\
&\quad \leq\ 2\Delta x \left(\mathrm{TV}\,(f_0^+) + \mathrm{TV}\,(f_0^-) + (3C+\frac{1}{2}) \int_{x_1}^{x_2} k(x)dx \right)
\end{aligned}$$

This concludes the proof of (8) in Theorem 1.

3. The Fractional Step approximation

A general Fractional Step setup proceeds by first, fixing $\Delta t > 0$, then computing iteratively an approximation $(f^-, f^+) := (f^-, f^+)^{\Delta t}$ of system (4) with $f_0^{\pm}$ satisfying (6) and $k(x)$ satisfying (2). Accordingly,

(1) on $[0,\Delta t)$ the $f^{\pm}$ are given by the exact solution of the linear problem

$$\partial_t f^- - \partial_x f^- = 0\,, \qquad \partial_t f^+ + \partial_x f^+ = 0\,. \tag{42}$$

The group operator $\mathcal{S}$ giving the solution writes as

$$\mathcal{S}_t\left(f_0^-, f_0^+\right)(x) = \left(f_0^-(x+t)\,,\, f_0^+(x-t)\right)\,, \qquad \left(f_0^-, f_0^+\right) \in L^1(\mathbb{R})\,.$$

(2) At time $t = \Delta t$ the solution is updated by the source term:

$$\left(f^-, f^+\right)(\Delta t+, x) = O_{\Delta t}\big(\left(f^-, f^+\right)(\Delta t-, x); k(x)\big)\,, \tag{43}$$

where O_t denotes the solution operator of the ordinary differential equation

$$\begin{cases} y' = kg(z-y) \\ z' = -kg(z-y)\,, \end{cases} \tag{44}$$

and a constant value $k>0$. In other words,

$$t \mapsto O_t\big(\,(y_0, z_0)\,; k\big) \in \mathbb{R}^2 \tag{45}$$

satisfies (44) with initial data (y_0, z_0) for $t=0$.

In the time strip $(\Delta t, 2\Delta t)$ the solution is extended as in step **(1)** by taking $(f^-, f^+)(\Delta t+, \cdot)$ as initial data. The procedure is then repeated inductively.

Let $\mathcal{O}_t$ be the group associated to (42) and (44) respectively; more precisely, recalling (45), for (u,v) as in (6) and k now as in (2) we have:

$$\forall x \in \mathbb{R}, \qquad \mathcal{O}_t\left((u,v); k\right)(x) \quad = \quad O_t\left((u(x), v(x)); k(x)\right)$$

We denote by $t \mapsto (\mathcal{U}^{\Delta t})_t$ the fractional step operator. It rewrites:

$$\begin{aligned}
(\mathcal{U}^{\Delta t})_0 &= Id\,, \\
(\mathcal{U}^{\Delta t})_t &= \mathcal{S}_{t-j\Delta t} \circ (\mathcal{U}^{\Delta t})_{j\Delta t}\,, \qquad j\Delta t < t < (j+1)\Delta t\,,\ \ j \geq 0,
\end{aligned}$$

while at fractional steps $t = j\Delta t\,,\ j \geq 1$ we let the space-dependent source act,

$$(\mathcal{U}^{\Delta t})_{j\Delta t} \quad = \quad \mathcal{O}_{\Delta t}\left((\mathcal{U}^{\Delta t})_{j\Delta t-}; k(x)\right)\,,$$

where $(\mathcal{U}^{\Delta t})_{j\Delta t-} = \mathcal{S}_{\Delta t} \circ (\mathcal{U}^{\Delta t})_{(j-1)\Delta t}\,.$

3.1. **Action of the time-ODE** (44) **on the total variation.** Despite the fact that the right-hand side of the simple system (4) appears at first glance to be dissipative, it involves a dependence in the space variable through the smooth function $k(x)$ which, somewhat counter-intuitively, can increase the space variation of the time-splitting approximation. There is indeed an accretive effect, weaker than in [3], which can increase the total variation of the approximate solution. In terms of the variables $\rho = y + z$ and $J = z - y$, the system (44) turns into

$$\rho' = 0\,, \qquad J' = -2kg(J)\,. \tag{46}$$

Recalling assumption (5) on g we notice that $|J|' = -2k|g(J)| < 0$ (dissipative effect for x fixed). Now, since $g'(0) > 0$, we get

$$\begin{aligned} \rho(t) &= \rho(0) = y_0 + z_0\,, \\ |J(t)| &\leq |J(0)|\mathrm{e}^{-\alpha t}\,, \quad \alpha \sim 2kg'(0)\,. \end{aligned}$$

Being $y = (\rho - J)/2$ and $z = (\rho + J)/2$, we deduce that

$$y(t)\,,\ z(t) \to \frac{\rho_0}{2} = \frac{y_0 + z_0}{2}\,, \qquad t \to \infty\,.$$

Lemma 2. *Let g satisfy* (5) *and consider the system* (44)*:*

(1) For $k > 0$, let $(y_1, z_1)(t)$ and $(y_2, z_2)(t)$ be two solutions of (44). *Then*

$$\frac{d}{dt}\{|y_1(t) - y_2(t)| + |z_1(t) - z_2(t)|\} \quad\leq\quad 0\,. \tag{47}$$

(2) Take $0 < k_1 < k_2$, and let $(y_1, z_1)(t)$ and $(y_2, z_2)(t)$ be two solutions of (44) *corresponding to $k = k_1$ or k_2, respectively. Assume also that $(y_1, z_1)(0) = (y_2, z_2)(0)$. Then*

$$|y_1(t) - y_2(t)| + |z_1(t) - z_2(t)| \quad\leq\quad 2C(k_2 - k_1)t\,, \tag{48}$$

where

$$C = \sup_{J\in[a,b]} |g(J)|\,, \qquad [a,b] = \begin{cases} [0, J(0)] & \text{if } J(0) > 0\,, \\ [J(0), 0] & \text{otherwise.} \end{cases} \tag{49}$$

Proof. (1) One easily computes that

$$\begin{aligned} &\frac{d}{dt}\{|y_1(t) - y_2(t)| + |z_1(t) - z_2(t)|\} \\ &= k\,[g(z_1 - y_1) - g(z_2 - y_2)]\,\{\mathrm{sgn}(y_1 - y_2) - \mathrm{sgn}(z_1 - z_2)\} \\ &= kg'(\xi)\ [(z_1 - z_2) - (y_1 - y_2)]\mid \{\mathrm{sgn}(y_1 - y_2) - \mathrm{sgn}(z_1 - z_2)\} \\ &= kg'(\xi)\ [-|y_1 - y_2| - |z_1 - z_2| \\ &\qquad\qquad + (z_1 - z_2)\mathrm{sgn}(y_1 - y_2) + (y_1 - y_2)\mathrm{sgn}(z_1 - z_2)] \\ &\leq 0\,. \end{aligned}$$

(2) Denoting by $(Y(t), Z(t))$ the integral curve of (44) with initial data $(y_1, z_1)(0)$ and $k = 1$, one sees that the space-dependence of $k(x)$ leads to:

$$(y_1, z_1)(t) = (Y(k_1 t), Z(k_1 t))\,, \qquad (y_2, z_2)(t) = (Y(k_2 t), Z(k_2 t))\,,$$

hence, setting $J = Z - Y$

$$\begin{aligned} |y_1(t) - y_2(t)| + |z_1(t) - z_2(t)| &\leq 2\int_{k_1 t}^{k_2 t} |g(J(s))|\, ds \\ &\leq 2C\ (k_2 - k_1)t \end{aligned}$$

where C as in (49). □

In the following lemma we illustrate the dependence of the solution on the variations of the parameter k. We will use the notation $|(\alpha, \beta)|_1 = |\alpha| + |\beta|$.

Lemma 3. *Under the assumptions of Lemma 2, there hold:*

(1) Let $U_\ell = (f_\ell^-, f_\ell^+)$, $U_r = (f_r^-, f_r^+)$, $k_\ell > 0$ and $k_r > 0$ be given. Then

$$|O_t\big(U_\ell; k_\ell\big) - O_t\big(U_r; k_r\big)|_1 \quad \leq \quad |U_\ell - U_r|_1 \, + \, 2C|k_\ell - k_r|t \tag{50}$$

with C given by (49).

(2) Assume moreover that $f_\ell^- = f_r^-$. Denote by $y_\ell(t)$, $y_r(t)$ the first component of $O_t\big(U_\ell; k_\ell\big)$, $O_t\big(U_r; k_r\big)$ respectively, they satisfy

$$\begin{aligned} &|y_\ell(t) - y_r(t))| \\ &\quad \leq t\, \left\{ C|k_\ell - k_r|(1+t) \, + \, \max\{k_\ell, k_r\}\|g'\|_\infty\, |f_\ell^+ - f_r^+| \right\}. \end{aligned} \tag{51}$$

Proof. *(1)* Recalling (47) and (48) in Lemma 2, we find that

$$\begin{aligned} &|O_t\big(U_\ell; k_\ell\big) - O_t\big(U_r; k_r\big)|_1 \\ &\quad \leq |O_t\big(U_\ell; k_\ell\big) - O_t\big(U_r; k_\ell\big)|_1 \, + \, |O_t\big(U_r; k_\ell\big) - O_t\big(U_r; k_r\big)|_1 \\ &\quad \leq |U_\ell - U_r|_1 \, + \, 2C|k_r - k_\ell|t\,. \end{aligned}$$

(2) Let $z_\ell(t)$, $z_r(t)$ be the second components of $O_t\big(U_\ell; k_\ell\big)$, $O_t\big(U_r; k_r\big)$ respectively:

$$y_\ell(t) - y_r(t)) \;=\; (k_\ell - k_r)\int_0^t g(z_\ell - y_\ell)\, d\tau + k_r \int_0^t \left[g(z_\ell - y_\ell) - g(z_r - y_r)\right]\,,$$

so that

$$|y_\ell(t) - y_r(t))| \;\leq\; |k_\ell - k_r|\, C\, t + \max\{k_\ell, k_r\}(\sup g')\int_0^t |z_\ell - z_r| + |y_\ell - y_r|\, d\tau\,.$$

To estimate this last integral, we use (50) and get

$$\int_0^t |z_\ell - z_r| + |y_\ell - y_r|\, d\tau \;\leq\; |f_\ell^+ - f_r^+|\, t \, + \, C|k_r - k_\ell|t^2\,.$$

Combining together the last two estimates, we end up with (51). □

3.2. **Bounds on total variation for FS approximation.** Here we seek for BV bounds of the Fractional Step approximation defined at the beginning of this Section. The quantity

$$L(t) \;=\; \mathrm{TV}\, f^+(t, \cdot) + \mathrm{TV}\, f^-(t, \cdot)$$

is constant between time steps. On the other hand, at each time step we estimate the possible increase of L by means of estimate (50) in Lemma 3:

$$L(\Delta t+) \;\leq\; L(\Delta t-) + 2C\Delta t\, \mathrm{TV}\,\{k\} \tag{52}$$

where C depends on the L^∞ norm of J. An estimate on $\|J(t, \cdot)\|_{L^\infty}$ goes as follows:

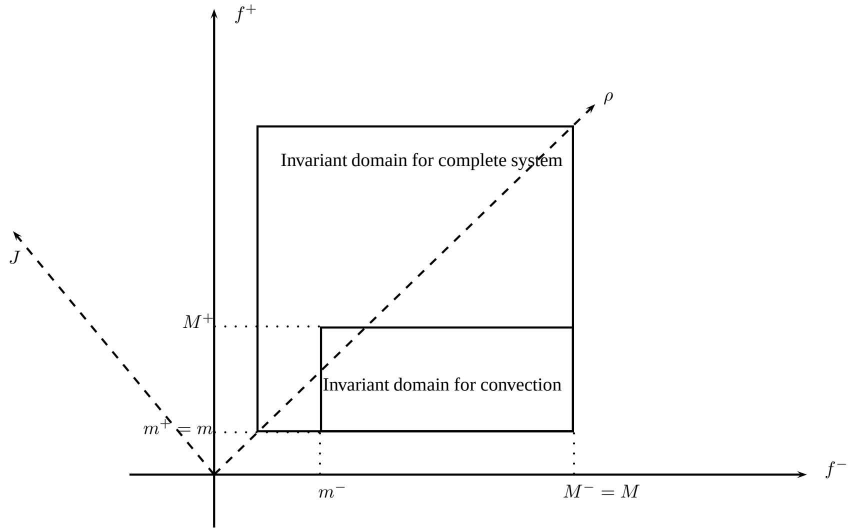

FIGURE 5. Invariant domain for both homogeneous and complete system

- For the linear system (42), each rectangle $[a, b] \times [c, d]$ is invariant.
- On the other hand, recalling (46), those square regions having a diagonal along the line $f^+ = f^-$ ($J = 0$) are invariant domains for the ODE (44). This is because $k \geq 0$ and $\operatorname{sgn} g(J) = \operatorname{sgn} J$.

Now, if m and M are the constant values defined in (24), then the square region $D \doteq [m, M] \times [m, M]$ is invariant for both the linear system and the ODE, therefore it holds that $(f^-, f^+)(x, t) \in [m, M] \times [m, M] = D$. Hence a global bound for J is found to be

$$-(M - m) \leq J(t, x) \leq M - m \,.$$

Therefore obtains the following estimate valid for $t \geq 0$:

$$\begin{aligned} L(t+) &\leq L(0) + 2C_0\, t\, \mathrm{TV}\, \{k\} \,, \\ C_0 &= \sup_{|J| \leq M-m} |g(J)| \,. \end{aligned} \tag{53}$$

Thanks to the apriori bounds on $\|f^\pm\|_\infty$ and on $\mathrm{TV}\, f^\pm$, independent on $\Delta t = \Delta x$, one can apply Helly's theorem for a given time interval $[0, T]$ and obtain a subsequence that converge strongly in L^1_{loc} to a solution of the Cauchy problem for (4), (6). In the limit the following estimate holds:

$$\boxed{\mathrm{TV}\, f^+(t, \cdot) + \mathrm{TV}\, f^-(t, \cdot) \leq \mathrm{TV}\, f_0^+ + \mathrm{TV}\, f_0^- \; + 2C_0 \mathrm{TV}\, \{k\}\, t \,.} \tag{54}$$

Remark 4. *The estimate* (54)*, being itself a consequence of* (50)*, is at the heart of the matter. Indeed it shows that, despite the maximum principle holds on the amplitude of time-splitting approximate solutions (as a consequence of the invariant domain displayed on Fig. 5), their corresponding total variation in space can grow linearly in time because of the local variations of $k(x)$, in sharp contrast with the one generated by the WB process* (39)*. This accretive effect shows that one must be careful when considering the choice of one or another numerical scheme, especially when space-dependent source terms are involved.*

However notice that, in the limit $\Delta x \to 0$, both estimates (54) *and* (39) *hold, because of uniqueness. Therefore we can take the minimum, as follows:*

$$\boxed{\mathrm{TV}\, f^+(t,\cdot) + \mathrm{TV}\, f^-(t,\cdot) \le \mathrm{TV}\, f_0^+ + \mathrm{TV}\, f_0^- \ + 2C_0 \min\{\mathrm{TV}\,\{k\}\, t, 2\|k\|_{L^1}\}\,.} \tag{55}$$

Therefore there exists a time t^*, with

$$\mathrm{TV}\,\{k\} t^* \quad = \quad 2\|k\|_{L^1} \tag{56}$$

such that for $t > t^*$ the BV-bound becomes constant in time.

3.3. **Local Truncation Errors.** Let $U(t) = U(t;k) = (f^-, f^+)(t,\cdot)$ be the exact solution to system (4) for some initial data $U(0) = (f_0^-, f_0^+) = U_0$ satisfying (6) and for $k \in BV(\mathbb{R})$ that satisfies (2). Now we state the so called *local truncation error* estimate.

In the following, the L^1–norm is intended as follows: for instance $\|U(t)\|_{L^1} = \|f^-(\cdot,t)\|_{L^1} + \|f^+(\cdot,t)\|_{L^1}$. Notice that, with this norm, the group $\mathcal{S}$ is an isometry in L^1. Similarly, if $U = (f^-, f^+)$, we denote by $\mathrm{TV}\, U$ the sum $(\mathrm{TV}\, f^- + \mathrm{TV}\, f^+)$.

Lemma 4. *There exists a time $\tau > 0$ sufficiently small, such that*

$$\|U(\Delta t;k) - \mathcal{O}_{\Delta t}\left(\mathcal{S}_{\Delta t}\, U_0;k\right)\|_{L^1} \le \widetilde{C}\cdot \Delta t^2\,, \qquad \Delta t \le \tau \tag{57}$$

where

$$\widetilde{C} = \widetilde{C}(U_0,k) = 2\|k\|_\infty \|g'\|_\infty \mathrm{TV}\, U_0 + 10\|g\|_\infty\, \mathrm{TV}\, k\,.$$

Proof. Let $\Delta t > 0$ be fixed. We will proceed with a similar technique to the one used in [5, Lemma 5.1]. The exact solution $U(\Delta t;k)$ results as the limit of the fractional step procedure for a sequence of space-time meshes $s_j = \Delta t/2^j \to 0$. For the moment we omit the dependence on k.

Then let us fix $s > 0$ and denote by $U^s(t)$ the approximate solution related to $U_0 = (f_0^-, f_0^+)$ and k. Let $j_0 \in \mathbb{N}$ such that $j_0 s \le \Delta t < (j_0+1)s$, and define for $1 \le j \le j_0$:

$$\phi_j^+ \quad = \quad \|U^s(js) - \mathcal{O}_{js}\left(\mathcal{S}_{js}\, U_0\right)\|_{L^1}$$

$$\phi_j^- \quad = \quad \|U^s(js-) - \mathcal{O}_{(j-1)s}\left(\mathcal{S}_{js}\, U_0\right)\|_{L^1}$$

For $j = 0$ the term ϕ_0^+ makes sense at it is $= 0$. Hence we can write

$$\phi_{j_0}^+ \quad = \quad \sum_{j=1}^{j_0} [\phi_j^+ - \phi_j^-] \ + \ \sum_{j=1}^{j_0} [\phi_j^- - \phi_{j-1}^+]\,.$$

One can easily check that $\phi_j^+ - \phi_j^- \le 0$, since

$$\begin{aligned} \phi_j^+ \quad &= \quad \|\mathcal{O}_s\left(U^s(js-)\right) - \mathcal{O}_s \circ \mathcal{O}_{(j-1)s}\left(\mathcal{S}_{js}\, U_0\right)\|_{L^1} \\ &\le \quad \|U^s(js-) - \mathcal{O}_{(j-1)s}\left(\mathcal{S}_{js}\, U_0\right)\|_{L^1}\,. \end{aligned}$$

Here we used the L^1 contractivity of the operator $\mathcal{O}_s$ that results from (47).

Thus it only remains to estimate $\phi_j^- - \phi_{j-1}^+$. Using that $\mathcal{S}$ is an isometry, we first notice that

$$\begin{aligned} \phi_{j-1}^+ \quad &= \quad \|U^s((j-1)s) - \mathcal{O}_{(j-1)s}\left(\mathcal{S}_{(j-1)s}\, U_0\right)\|_{L^1} \\ &= \quad \|\underbrace{\mathcal{S}_s U^s((j-1)s)}_{=U^s(js-)} - \mathcal{S}_s \mathcal{O}_{(j-1)s}\left(\mathcal{S}_{(j-1)s}\, U_0\right)\|_{L^1} \end{aligned}$$

Therefore, using triangular inequality, we get

$$\begin{aligned}\phi_j^- - \phi_{j-1}^+ &\leq \|\mathcal{O}_{(j-1)s}\left(\mathcal{S}_{js}\, U_0\right) - \mathcal{S}_s\mathcal{O}_{(j-1)s}\left(\mathcal{S}_{(j-1)s}\, U_0\right)\|_{L^1}\\ &= \|\mathcal{O}_{(j-1)s}\, \mathcal{S}_s(V) - \mathcal{S}_s\, \mathcal{O}_{(j-1)s}(V)\|_{L^1}\,,\end{aligned}$$

where $V = \mathcal{S}_{(j-1)s}\, U_0$.

Therefore we are led to estimate a commutator, for a generic $V \in L^1(\mathbb{R};\mathbb{R}^2)$:

$$\|\mathcal{O}_t\left(\mathcal{S}_s(V);k\right) - \mathcal{S}_s\left(\mathcal{O}_t(V;k)\right)\|_{L^1}\,, \qquad 0 < s \leq t \leq \Delta t\,.$$

In more details, we have (see Figure 6)

$$\begin{aligned}\mathcal{O}_t\left(\mathcal{S}_s(V);k\right)(x) &= O_t\left(\mathcal{S}_s(V)(x);k(x)\right)\\ &= O_t\left[\left(V^-(x+s), V^+(x-s)\right); k(x)\right]\,,\end{aligned}$$

and (see Figure 7)

$$\mathcal{S}_s\left(\mathcal{O}_t(V;k)\right)(x) = \left[\left(\mathcal{O}_t\left(V;k\right)(x+s)\right)^-, \left(\mathcal{O}_t\left(V;k\right)(x-s)\right)^+\right]\,.$$

Notice that, in the "-" component, the two flows differ for the k (computed at $k(x)$ and at $k(x+s)$ respectively) and for the initial data V^+. A similar consideration holds for the "+" component.

Recalling (51), we find that

$$\begin{aligned}&|\mathcal{O}_t\left(\mathcal{S}_s(V);k\right)(x) - \mathcal{S}_s\left(\mathcal{O}_t(V;k)\right)(x)|_1\\ &\quad\leq t\left\{\|k\|_\infty\|g'\|_\infty|V^-(x+s) - V^-(x-s)| + |V^+(x+s) - V^+(x-s)|\right.\\ &\qquad\left. + 2C\left\{|k(x) - k(x+s)| + |k(x) - k(x-s)|\right\}(1+t)\right\}\,.\end{aligned}$$

Now we integrate in x and use inequalities such as $\int_{\mathbb{R}} |k(x) - k(x+s)|\, dx \leq s\mathrm{TV}\, k$; moreover we assume that $\Delta t \leq 1$. Recalling the definition of C_0 in (53), we obtain

$$\begin{aligned}&\|\mathcal{O}_t\left(\mathcal{S}_s(V);k\right) - \mathcal{S}_s\left(\mathcal{O}_t(V;k)\right)\|_{L^1}\\ &\quad\leq 2st\left\{\|k\|_\infty\|g'\|_\infty\left(\mathrm{TV}\, V^- + \mathrm{TV}\, V^+\right) + 4C_0\,\mathrm{TV}\, k\right\}\,.\end{aligned}$$

By taking $V = \mathcal{S}_{(j-1)s}\, U_0$, we have that $\mathrm{TV}\, V = \mathrm{TV}\, U_0$ and then

$$\phi_j^- - \phi_{j-1}^+ \leq 2s\Delta t\left\{\|k\|_\infty\|g'\|_\infty\mathrm{TV}\, U_0 + 4C_0\,\mathrm{TV}\, k\right\}\,.$$

In conclusion we have

$$\phi_{j_0}^+ \leq \sum_{j=1}^{j_0}\left[\phi_j^- - \phi_{j-1}^+\right] \leq 2(\Delta t)^2\left\{\|k\|_\infty\|g'\|_\infty\mathrm{TV}\, U_0 + 4C_0\,\mathrm{TV}\, k\right\}\,.$$

□

3.4. **Global Truncation Errors for a fractional step numerical algorithm.**

In this subsection we first restrict the general procedure outlined in (42) and (43) to a more practical numerical splitting algorithm which can be rigorously analyzed. Then a *local truncation error* (LTE) between the exact solution of system (4) and this aforementioned approximation is proved, for a small time step $\Delta t > 0$. Based on this, we will provide a global estimate holding for any positive time $T > 0$.

A practical algorithm reads:

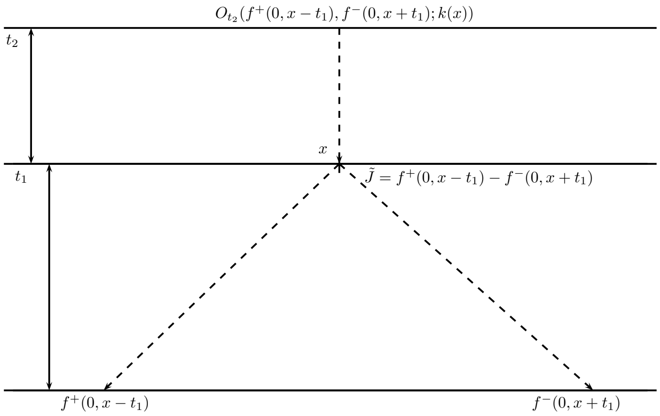

FIGURE 6. First commutator term: $\mathcal{O}_{t_2} \circ \mathcal{S}_{t_1}$.

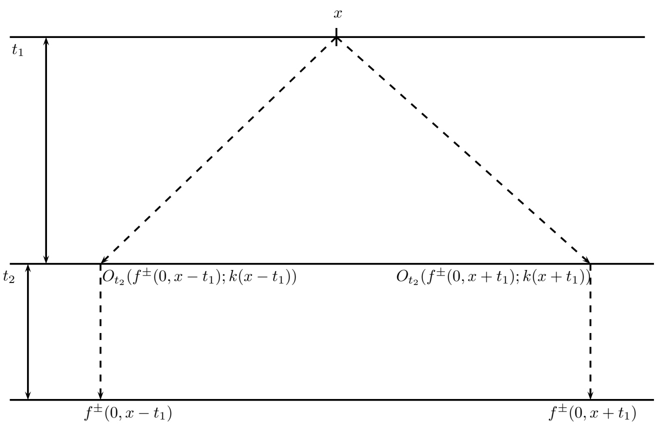

FIGURE 7. Second commutator term: $\mathcal{S}_{t_1} \circ \mathcal{O}_{t_2}$.

- Choose $\Delta x = \Delta t$ and set $x_j = j\Delta x$ for $j \in \mathbb{Z}$. Approximate the initial data $f_0^\pm$ and k with constant values on each (x_j, x_{j+1}), say

$$(f_0^\pm)^{\Delta t}(x) = f_0^\pm(x_j)\,, \qquad k^{\Delta t}(x) = k(x_j) \text{ for } x \in (x_j, x_{j+1})\,.$$

- In correspondence to $(f_0^\pm)^{\Delta t}$ and $k^{\Delta t}$, the approximate solution $f^- = (f^-)^{\Delta t}$ and $f^+ = (f^+)^{\Delta t}$ is defined with the procedure at the begin of this Section. Since $\Delta t = \Delta x$ and the wave speeds are ± 1, the function in (43) is possibly discontinuous only at the points $\{x_j\}_{j\in\mathbb{Z}}$.

By the choice of the approximate initial data and $k^{\Delta t}$, one has that $\mathrm{TV}\, k^{\Delta t} \leq \mathrm{TV}\, k$ and that $\mathrm{TV}\,(f_0^\pm)^{\Delta t} \leq \mathrm{TV}\,(f_0^\pm)$ so that the estimates for the total variation holds uniformly in Δx as in the previous section. Hence we can again apply Helly's theorem and get a subsequence converging strongly to the exact solution.

In the following we obtain a global error estimate for our scheme, that corresponds to part *(1)* of Theorem 1.

Lemma 5. *Let $T > 0$. Assume that $T = N\Delta t > 0$ for some $N \in \mathbb{N}$, with $0 < \Delta t \leq \tau$ as in Lemma 4. Let t^* as in (56). If $t^* > T - \Delta t$, then*

$$\begin{aligned} \|U(T) \quad &- \quad U^{\Delta t}(T)\|_{L^1} \\ &\leq \quad \Delta t \,\mathrm{TV}\, U_0 + \Delta t\, T\Big\{2\|k\|_\infty \|g'\|_\infty \mathrm{TV}\, U_0 \;+\; 12 C_0 \,\mathrm{TV}\, k\Big\} \\ &\qquad + 2\Delta t\, T^2 C_0 \|k\|_\infty \|g'\|_\infty \mathrm{TV}\, k \end{aligned} \tag{58}$$

while, if $t^* \leq T - \Delta t$, then

$$\begin{aligned} \|U(T) \quad &- \quad U^{\Delta t}(T)\|_{L^1} \\ &\leq \quad \Delta t \,\mathrm{TV}\, U_0 + \Delta t\, T\Big\{2\|k\|_\infty \|g'\|_\infty \mathrm{TV}\, U_0 \;+\; 12 C_0 \,\mathrm{TV}\, k\Big\} \\ &\qquad + \; 2\Delta t \cdot (t^*)^2 C_0 \|k\|_\infty \|g'\|_\infty \mathrm{TV}\, k \\ &\qquad\qquad + \; 8\Delta t\, (T - t^*)\, C_0 \|k\|_\infty \|g'\|_\infty \|k\|_{L^1}\,. \end{aligned} \tag{59}$$

Proof. For convenience of the reader, the proof is divided into 3 steps.

Step 1. We start by extending (57) to take into account of different initial data and different k. Given $V_0 \in L^1$ and $k^{\Delta t}$ as defined above, we claim that

$$\|U(\Delta t; k) - \mathcal{O}_{\Delta t}\left(\mathcal{S}_{\Delta t}\, V_0; k^{\Delta t}\right)\|_{L^1} \quad \leq \quad \|U_0 - V_0\|_{L^1} + C_2(\Delta t)^2 \tag{60}$$

being

$$C_2 = C_2(U_0, k) \quad = \quad 2\|k\|_\infty \|g'\|_\infty \mathrm{TV}\, U_0 \;+\; 12 C_0 \,\mathrm{TV}\, k \tag{61}$$

with $C_0 = \|g\|_\infty$ as in (54).

Indeed, by using the triangular inequality we obtain

$$\|U(\Delta t; k) - \mathcal{O}_{\Delta t}\left(\mathcal{S}_{\Delta t}\, V_0; k^{\Delta t}\right)\|_{L^1} \leq \|U(\Delta t; k) - \mathcal{O}_{\Delta t}\left(\mathcal{S}_{\Delta t}\, U_0); k\right)\|_{L^1} \tag{62}$$

$$+\|\mathcal{O}_{\Delta t}\left(\mathcal{S}_{\Delta t}\, U_0; k\right) - \mathcal{O}_{\Delta t}\left(\mathcal{S}_{\Delta t}\, V_0; k^{\Delta t}\right)\|_{L^1}\,. \tag{63}$$

The term on the right hand side in (62) is handled by means of (57). The term in (63) is estimated with the help of (50) in Lemma 3; indeed, it consists in estimating in L^1 the stability of the ODE flow with respect to a slight variation of both initial data and parameter k:

$$\begin{aligned} &\|\mathcal{O}_{\Delta t}\left(\mathcal{S}_{\Delta t}\, U_0; k\right) - \mathcal{O}_{\Delta t}\left(\mathcal{S}_{\Delta t}\, V_0; k^{\Delta t}\right)\|_{L^1} \\ &\qquad \leq \|\mathcal{S}_{\Delta t}\,(U_0 - V_0)\|_{L^1} + 2C_0 \Delta t \|k - k^{\Delta t}\|_{L^1} \\ &\qquad \leq \|U_0 - V_0\|_{L^1} + 2C_0 \Delta t \cdot \Delta x \,\mathrm{TV}\, k\,. \end{aligned}$$

By combining the estimates for (62) and (63), we end up with the inequality (60).

Step 2. Next, denote by t_n the time $t = n\Delta t$. Recalling that

$$U^{\Delta t}(t_n) = \mathcal{O}_{\Delta t}\left(\mathcal{S}_{\Delta t}(U^{\Delta t}(t_{n-1})), k^{\Delta t}\right)\,,$$

the global error of the fractional step procedure can be controlled by means of (60):

$$\|U(t_{n+1}) - U^{\Delta t}(t_{n+1})\|_{L^1} \quad \leq \quad \|U(t_n) - U^{\Delta t}(t_n)\|_{L^1} + C_2(U(t_n), k)(\Delta t)^2 \tag{64}$$

where C_2 is estimated by means of (55), since $U(t_n)$ is the exact solution at time $t = t_n$.

More precisely, let t^* be given as in Remark 4. We claim that

$$C_2(U(t_n),k) \quad\le\quad 2\|k\|_\infty\|g'\|_\infty \mathrm{TV}\,U_0 \;+\; 12C_0\,\mathrm{TV}\,k \\ \qquad\qquad + \begin{cases} 4C_0\|k\|_\infty\|g'\|_\infty\mathrm{TV}\,k\cdot t_n & \text{if } t_n\le t^* \\ 8C_0\|k\|_\infty\|g'\|_\infty\|k\|_{L^1} & \text{if } t_n > t^*\,. \end{cases}$$

Indeed, from (55) we have

$$\mathrm{TV}\,U(t_n) \quad\le\quad \mathrm{TV}\,U_0 + \begin{cases} 2C_0 t_n \mathrm{TV}\,k & \text{if } t_n\le t^* \\ 4C_0\|k\|_{L^1} & \text{if } t_n > t^*\,. \end{cases}$$

It is then enough to substitute in the definition of C_2, (61), to prove the claimed estimate.

Step 3. By setting

$$E_n = \|U(t_n) - U^{\Delta t}(t_n)\|_{L^1}\,,$$

we rewrite (64) as

$$E_{n+1} \quad\le\quad E_n + C_2(U(t_n),k)(\Delta t)^2$$

and therefore

$$E_N \quad\le\quad E_0 \;+\; (\Delta t)^2 \sum_{n=0}^{N-1} C_2(U(t_n),k)\,.$$

Let n^* be such that $n^*\Delta t\le t^*$ and $(n^*+1)\Delta t > t^*$. Assume that $t^*\le T-\Delta t$, so that $n^*+1\le N$. Recalling that $N\Delta t = T$, we compute

$$\Delta t \sum_{n=0}^{N-1} C_2(U(t_n),k) \quad\le\quad T\Big\{2\|k\|_\infty\|g'\|_\infty\mathrm{TV}\,U_0 \;+\; 12C_0\,\mathrm{TV}\,k\Big\} \\ +\; \Delta t\sum_{n=1}^{n^*-1} 4C_0\|k\|_\infty\|g'\|_\infty\mathrm{TV}\,k\cdot t_n \;+\; \Delta t\sum_{n=n^*}^{N-1} 8C_0\|k\|_\infty\|g'\|_\infty\|k\|_{L^1}\,. \quad (65)$$

We easily compute that $\sum_{n=1}^{n^*-1} t_n = \Delta t\sum_{n=1}^{n^*-1} n = \Delta t\frac{(n^*-1)n^*}{2} \le \frac{t^* n^*}{2}$. The terms in (65) are estimated by

$$2t^*n^*C_0\|k\|_\infty\|g'\|_\infty\mathrm{TV}\,k \;+\; 8\,(T-t^*)\,C_0\|k\|_\infty\|g'\|_\infty\|k\|_{L^1}\,.$$

We are now ready to complete the estimate on E_N: since $E_0\le\Delta t\,\mathrm{TV}\,U_0$, we finally get

$$E_N \;\le\; \Delta t\,\mathrm{TV}\,U_0 + \Delta t\,T\left\{2\|k\|_\infty\|g'\|_\infty\mathrm{TV}\,U_0 \;+\; 12C_0\,\mathrm{TV}\,k\right\} \\ +\; 2\Delta t\,(t^*)^2C_0\|k\|_\infty\|g'\|_\infty\mathrm{TV}\,k \;+\; 8\Delta t\,(T-t^*)\,C_0\|k\|_\infty\|g'\|_\infty\|k\|_{L^1}\,.$$

that is exactly (59). Notice that for the case under consideration, that is for $T\ge t^*+\Delta t$, the estimate increases at a linear rate, given by

$$\Delta t\left\{2\|k\|_\infty\|g'\|_\infty\mathrm{TV}\,U_0 \;+\; 12C_0\,\mathrm{TV}\,k + 8\,C_0\|k\|_\infty\|g'\|_\infty\|k\|_{L^1}\right\}.$$

Finally, in the simpler case of $t^* > T-\Delta t$, hence $n^*+1>N$, the same argument as above leads to

$$E_N \quad\le\quad \Delta t\,\mathrm{TV}\,U_0 + \Delta t\,T\Big\{2\|k\|_\infty\|g'\|_\infty\mathrm{TV}\,U_0 \;+\; 12C_0\,\mathrm{TV}\,k\Big\} \\ +\; 2\,\Delta t\,T^2C_0\|k\|_\infty\|g'\|_\infty\mathrm{TV}\,k\,,$$

that gives (58). The proof of the Lemma is complete. □

4. Numerical results for $k'(x) \equiv 0$ and $g(J) = J$.

In this simple context, an exact solution for the equation (1) on the whole real line is given in *e.g.* [25]: however, it involves the integral of a modified Bessel function. Thus we prefer to turn to more specific problems, for which simpler exact solutions can be derived. Observe that in this context where $k(x) \equiv K > 0$, the equivalence between (3) and (1) is even simpler. By differentiating the first conservation law (the continuity equation) with respect to t, and then the second balance law with respect to x, it comes

$$\partial_{tt}\rho - \partial_{xx}\rho - 2K\partial_x J = 0.$$

It remains to use $\partial_t \rho = -\partial_x J$ in order to recover the usual form of the linear damped wave equation. Notice that, in contrast with (1), we have now that $u(t,x) = \rho(t,x)$.

4.1. **Bifurcation of plane waves speed.** The present benchmark is inspired by the analysis proposed by McCartin [24]: roughly speaking, it consists in seeking plane-wave solutions of (1) with $k(x) \equiv 1$ under the form $\rho(t,x) = \exp(i(\omega t - \xi x))$. By defining the wave-speed $a(\xi) = \frac{\omega}{\xi}$ with $\alpha = 1$, $\beta = 2$, McCartin explains that a bifurcation phenomenon occurs when crossing the critical value $\xi^* = 1$, namely

- for $\xi < 1$, the plane wave is static, $a(\xi) = 0$, but decays exponentially at two different rates,
- for $\xi > 1$, the plane wave propagates, and decays at the rate $\exp(-t)$.

The first benchmark consists in setting up the system (4) in the domain $x \in [0, 2\pi/\xi)$ with periodic boundary conditions on each side. By prescribing the analytical solutions computed in [24], it is possible to derive at each time-step $t^n = n\Delta t$ a *measured relative L^1 error* by computing:

$$E_{rel}(t) = \frac{1}{\overline{|\rho|}(t)} \big(\|\rho(t,.) - \rho_{exact}(t,.)\|_{L^1} + \|J(t,.) - J_{exact}(t,.)\|_{L^1} \big).$$

Indeed, a measured error with respect to an exact solution is made even more significant if it is "normalized" according to the average value of the modulus of the unknown, denoted here by $\overline{|\rho|}(t)$. All the numerical results were obtained by fixing $\Delta t = \Delta x$ and 2^7 grid points. On Figure 8, one sees the outcome of both the WB and the time-splitting scheme when computing a standing wave corresponding to $\xi = 0.99$. The relative error of the WB scheme is smaller, and doesn't oscillate like the one coming from the fractional step algorithm. Oppositely, setting up a wave-

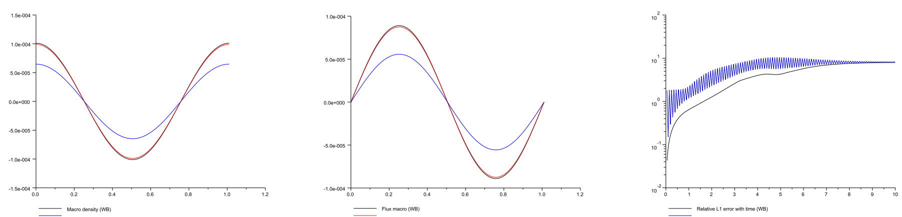

Figure 8. $\xi = 0.99$: Density (ρ, left), flux (J, middle) and relative error.

number $\xi > 1$ yields a propagating periodic signal. Therefore, it is necessary to implement periodic boundary conditions on the edges of the computational domain. Such "inhomogeneous boundary conditions" fit very naturally in the well-balanced framework because the source terms are treated at the interfaces: see *e.g.* Figure 2.1 in [16] or Remark 10.1 (page 199) in [17]. By carefully inspecting the macroscopic flux $J(t = 10, x)$ generated by the fractional step scheme, one sees that some error appears on both the borders, and the greater deviation with the WB approximation (see Figure 9, right) is likely to come from that. One may wonder if setting up

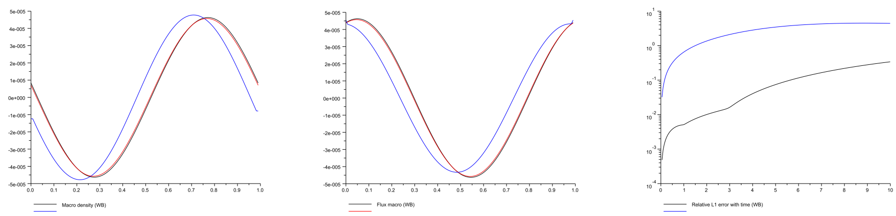

FIGURE 9. $\xi = 1.01$: Density (ρ, left), flux (J, middle) and relative error.

the well-known second-order in time "Strang-splitting" can improve the picture: on Figure 10, we display the time-evolution of (measured) relative L^1 errors with both the same benchmarks and computational grids. What appears is that such a second-order splitting decreases notably the error in the case where the wave-number $\xi = 0.99 < 1$, *i.e.* for the "standing wave" setup. , in the opposite situation $\xi = 1.01 > 1$, the relative L^1 error of the splitting scheme is still way bigger than the one generated by the well-balanced scheme.

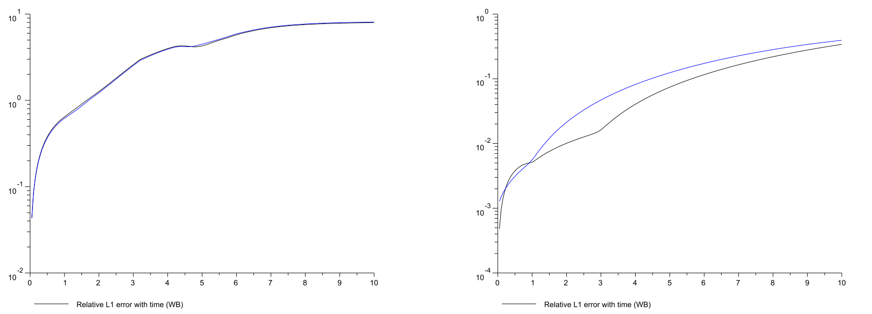

FIGURE 10. Relative L^1 errors for both well-balanced and Strang-splitting schemes w.r.t. time: $\xi = 0.99$ (left), $\xi = 1.01$ (right).

4.2. **Initial-boundary-value problem without positively invariant domain.** Now we focus on the numerical investigation of the Example 1, page 238 of [21], which involves the slightly modified damped wave equation,

$$\partial_{tt} u - \partial_{xx} u + 3\partial_t u = 0, \qquad x \in (0, 1),$$

with both initial data $u(t=0,x) = \partial_t u(t=0,x) = \exp(2x)$ and boundary data $u(t,x=0) = \exp(t)$, $u(t,x=1) = \exp(2+t)$. The exact solution reads $u(t,x) = \exp(t+2x)$ and its exponential growth in time implies that the relative L^1 error $t \mapsto E_{rel}(t)$ is expected to decrease. As before, we set up both the well-balanced and time-splitting schemes with $\Delta t = \Delta x$ in order to measure the relative L^1 errors against the exact solution. The only change is that both a coarse (2^7 points) and a fine (2^8 points) computational grids are considered: see Figure 11. Now, it can

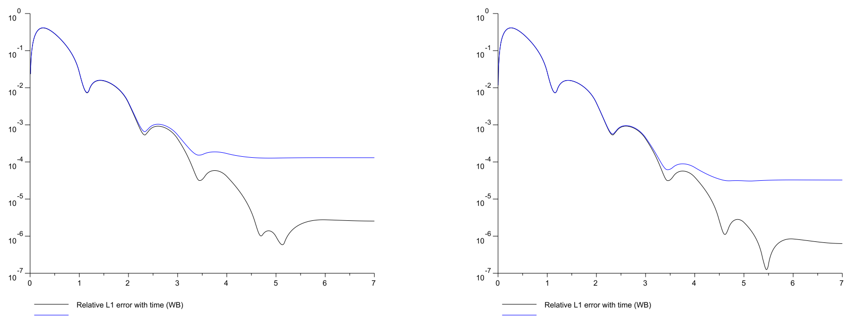

FIGURE 11. Relative L^1 errors of both WB and simple splitting schemes w.r.t. time: 2^7 (left), 2^8 grid points (right).

be interesting to see whether the use of second-order Strang-splitting still strongly improves the computational results like in the former test-case. On Figure 12, the same experiment is conducted with this better splitting algorithm: what emerges is only a marginal improvement for both the coarse and the fine computational grids.

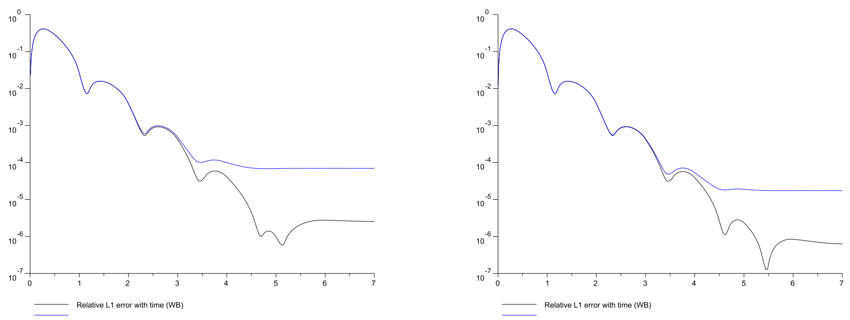

FIGURE 12. Relative L^1 errors of WB and Strang-split schemes w.r.t. time: 2^7 (left), 2^8 grid points (right).

5. Conclusion and outlook

The fundamental estimate for the WB algorithm is really (39) which states that its total variation in x depends on $\|k\|_{L^1}$ and doesn't grow as times goes by. In sharp contrast, for the fractional step, only (54) holds, meaning that now, the total variation may depend on $TV(k)$, which is supposedly bigger. As both numerical schemes are endowed with positively invariant domains, these estimates imply that some oscillations are more likely to develop in the FS approximation as soon as $\partial_x k \not\equiv 0$. Such features manifest themselves in both the error bounds given in Theorem 1. Another aspect of both these estimates, besides being local in space, is their *a posteriori* character: all the quantities showing up in the error bounds depend only on g, the initial data and the approximate solution. In particular, there is no mention of the exact solution: this shares a lot of similarities with the simpler framework of [18]. In particular it may allow for the development of *adaptive algorithms* [7, 8, 22, 27] for such inhomogeneous semilinear systems based on rigorous a posteriori local indicators of the form presented in this paper.

References

[1] S. Abarbanel, A. Ditkowski, B. Gustafsson. On Error Bounds of Finite Difference Approximations to Partial Differential Equations–Temporal Behavior and Rate of Convergence *J. Scient. Comput.* **15** (2000), 79–116

[2] D. Amadori, A. Corli. Glimm estimates for a model of multiphase flow, Preprint 2012

[3] D. Amadori, L. Gosse. Transient L^1 error estimates for well-balanced schemes on non-resonant scalar balance laws, *J. Differential Equations* **255** (2013), 469–502

[4] D. Amadori, G. Guerra. Global BV solutions and relaxation limit for a system of conservation laws, *Proc. Roy. Soc. Edinburgh Sect. A* **131** (2001), 1–26

[5] D. Amadori, G. Guerra. Uniqueness and continuous dependence for systems of balance laws with dissipation, *Nonlinear Anal. TMA* **49** (2002), 987–1014

[6] D. Aregba-Driollet, M. Briani, R. Natalini. Time Asymptotic High Order schemes for dissipative BGK hyperbolic systems, Preprint (2012)

[7] Ch. Arvanitis, Th. Katsaounis, Ch. Makridakis. Adaptive finite element relaxation schemes for hyperbolic conservation laws, *Math. Model. Numer. Anal.* **35** (2001), 17–33

[8] Ch. Arvanitis, Ch. Makridakis, A.E. Tzavaras. Stability and convergence of a class of finite element schemes for hyperbolic systems of conservation laws, *SIAM J. Numer. Analysis* **42** (2004), 1357–1393

[9] S. Bianchini, B. Hanouzet, R. Natalini. Asymptotic behavior of smooth solutions for partially dissipative hyperbolic systems with a convex entropy, *Comm. Pure Appl. Math.* **60** (2007), 1559–1622

[10] A. Bressan. **Hyperbolic Systems of Conservation Laws – The one-dimensional Cauchy problem**, Oxford Lecture Series in Mathematics and its Applications **20** Oxford University Press, Oxford, 2000

[11] A. Bressan, T.-P. Liu, T. Yang. L^1 stability estimates for $n \times n$ conservation laws. *Arch. Ration. Mech. Anal.* **149** (1999), no. 1, 1–22

[12] B. Cockburn, H. Gau. A posteriori error estimates for general numerical schemes for conservations laws, *Mat. Apl. Comput.* **14** (1995), 37–47

[13] F. Coquel, S. Jin, J.-G. Liu, Li Wang, Well-posedness and singular limit of a semilinear hyperbolic relaxation system with a two-scale discontinuous relaxation rate. Preprint 2013

[14] R.M. Colombo, G. Guerra, Differential equations in metric spaces with applications. *Discrete Contin. Dyn. Syst. Series A* **23** (2009), 733–753

[15] J. Glimm, D. H. Sharp. An S-matrix theory for classical nonlinear physics, *Found. Phys.* **16** (1986), 125–141

[16] L. Gosse. Maxwellian decay for well-balanced approximations of a super-characteristic chemotaxis model, *SIAM J. Scient. Comput.* **34** (2012), A520–A545

[17] L. Gosse. **Computing Qualitatively Correct Approximations of Balance Laws**, SIMAI Springer Series, Vol. 2, Springer (2013)

[18] L. Gosse, Ch Makridakis. Two a Posteriori Error Estimates for One-Dimensional Scalar Conservation Laws, *SIAM J. Numer. Anal.* **30** (2000), 964–988

[19] L. Gosse, G. Toscani. An asymptotic-preserving well-balanced scheme for the hyperbolic heat equations *C. R. Acad. Sci Paris,Ser. I.* **334** (2002), 337–342

[20] M. Kac. A stochastic model related to the telegrapher's equation. *Rocky. Mountain J. Math.* **4** (1974), 497–509

[21] D. Kaya. An application of the decomposition method for second order wave equations, *International Journal of Computer Mathematics* **75** (2000), 235–245

[22] H. Kim, M. Laforest, D. Yoon. An adaptive version of Glimm's scheme, *Acta Mathematica Scientia*, special edition for James Glimm's 75th anniversary, **30**(2) (2010), 428–446

[23] M. Laforest, A Posteriori Error Estimate for Front-Tracking: System of Conservation Laws, *SIAM J. Math. Anal.* **35** (2004), 1347–1370

[24] B.J. McCartin. Exponential Fitting of the Damped Wave Equation, *Applied Mathematical Sciences* **4** (2010), 1899–1930

[25] P. Marcati, K. Nishihara. The $L^p - L^q$ estimates of solutions to one-dimensional damped wave equations and their application to the compressible flow through porous media, *J. Differential Equations* **191** (2003), 445–469

[26] M. Ohlberger. A review of a posteriori error control and adaptivity for approximations of nonlinear conservation laws. *Int. J. for Numer. Meth. in Fluids* **59** (2009), 333–354

[27] G. Puppo, M. Semplice. Numerical entropy and adaptivity for finite volume schemes. *Commun. Comput. Phys.* **10** (2011), 1132–1160

[28] J. Rauch, M. Reed. Jump Discontinuities of Semilinear, Strictly Hyperbolic Systems in Two Variables: Creation and Propagation. *Commun. Math. Phys.* **81** (1981) 203–227

DISIM, Università degli Studi dell'Aquila, 67100 L'Aquila, Italy

E-mail address: amadori@univaq.it

IAC–CNR "Mauro Picone" (sezione di Roma) - Via dei Taurini, 19 - 00185 Rome, Italy

E-mail address: l.gosse@ba.iac.cnr.it