Energy preserving schemes for nonlinear Hamiltonian systems of wave equations: Application to the vibrating piano string

Juliette Chabassier, Patrick Joly

Energy Preserving Schemes for Nonlinear Hamiltonian Systems of Wave Equations. Application to the Vibrating Piano String.

J. Chabassier[A,*], P. Joly[A]

[a] *POems Team, INRIA Paris - Rocquencourt*
Domaine de Voluceau
Rocquencourt BP 105
78153 Le Chesnay Cedex
France

Abstract

This paper considers a general class of nonlinear systems, "nonlinear Hamiltonian systems of wave equations". The first part of our work focuses on the mathematical study of these systems, showing central properties (energy preservation, stability, hyperbolicity, finite propagation velocity ...). Space discretization is made in a classical way (variational formulation) and time discretization aims at numerical stability using an energy technique. A definition of "preserving schemes" is introduced, and we show that explicit schemes or partially implicit schemes which are preserving according to this definition cannot be built unless the model is trivial. A general energy preserving second order accurate fully implicit scheme is built for any continuous system that fits the nonlinear Hamiltonian systems of wave equations class. The problem of the vibration of a piano string is taken as an example. Nonlinear coupling between longitudinal and transversal modes is modeled in the "geometrically exact model", or approximations of this model. Numerical results are presented.

Key words: Energy preservation, nonlinear Hamiltonian systems of wave equations, finite elements numerical schemes, piano string.

Introduction

We consider in this paper a general class of nonlinear systems, namely nonlinear Hamiltonian systems of wave equations. Our main objective is the construction of energy preserving discretization schemes for such systems. The concrete problem that has motivated this work was to compute the vibrations of a piano string, with the objective to achieve the numerical simulation of a whole concert piano. The full piano model is quite complex and couples the vibrations of the string (a 1D model) with the vibrations of the soundboard (a 2D phenomenon) and with the sound radiation (a 3D phenomenon). Guaranteeing and proving the stability of a numerical method for the coupled problem is not an easy task. Having an energy approach is a very powerful approach to achieve this goal. The Hamiltonian nature of the equations governing the vibrations of the string makes it possible a priori : there is conservation of an energy for the continuous problem. Preserving such a property at the discrete level has two nice consequences : keeping after discretization an important (from both theoretical and physical points of view) property of the exact solution and getting stability results provided that the discrete property has pleasant positivity properties.

The problematic of energy preserving schemes is far from new and has already generated an intensive literature. It appears that results can easily be found in the case of scalar equations (the unknown function takes scalar values), of course in the context of ordinary differential equations ODEs (see for instance [16, 26]) but also of some partial differential equations, particularly semi-linear wave equations (see for instance [6, 11, 12, 23, 25, 31]). The case of systems has been much less investigated and it seems that most results are restricted to very particular systems : see for instance in [27], [15] or [8] in the context of systems of ODE's and [14] (for nonlinear elasticity) or [5] (for nonlinear strings) in the context of systems of PDE's. Finally, the references [3, 4, 17] investigate time FE methods for the N-body problem, nonlinear elastodynamics and are extended to more general problems and higher orders. These very interesting methods rely on the difficult seek of a good quadrature rule and reduce to the previously mentioned methods in particular cases. Our aim in this context was to find a systematic and easily computed energy preserving scheme for any system of PDE's, while keeping a great degree of generality.

*Corresponding author

As said above, we tackle in this paper a rather general class of 1D nonlinear Hamiltonian systems of wave equations, where the unknown function takes values in $\mathbb{R}^N$ for arbitrary $N \geq 1$. The restriction to the 1D case contributes essentially to simplifying the presentation. However, most of our developments can be extended to higher space dimensions. Our article is divided into two parts. The first part (sections 1 and 2) concerns general systems. The second part (section 3) presents the application to the particular system which governs the vibrations of a piano string.

In section 1, we recall the main properties of 1D nonlinear Hamiltonian systems of wave equations. We insist particularly on the most relevant (for our purpose) properties of (sufficiently) smooth solutions of such systems: energy preservation (leading to H^1 stability), hyperbolicity and finite propagation velocity. Section 2, the main section of the article, is devoted to the discretization schemes. For the space discretization, we use a variational formulation and a Galerkin approximation procedure (Section 2.1). The main difficulties are encountered when looking at the time discretization using finite differences, which is the object of section 2.2. Our desire to preserve a certain energy leads us to introduce a particular class of numerical schemes. We show that this class excludes the explicit scheme (except in the linear case, see Lemma 2.2 of section 2.2.2) as well as partially decoupled implicit schemes (except in some very particular systems, see section 2.2.3). Finally in section 2.2.4, we exhibit inside our class of numerical schemes a fully implicit, second order accurate, energy preserving and unconditionally stable scheme for any nonlinear Hamiltonian system of wave equations, with any number of unknown variables. Note that the implicitness of the scheme is the price to be paid for robustness (obtained via energy conservation).

In the context of the simulation of the piano (section 3), the implicitness of the scheme is by no means a real constraint since the time devoted to the string itself should be a small percentage of the total computational cost while the unconditional stability provides more flexibility and robustness for the coupled model. In Section 3.1 we present the nonlinear vibrating string model introduced in [28], as well as some of its approximations including the one used in [2] and [5]. These models are all nonlinear Hamiltonian systems of wave equations. We give their main mathematical properties in section 3.2. We apply the numerical scheme of section 2.2.4 to this system (section 3.3) and related numerical results are given in section 3.4.

1. Nonlinear Hamiltonian systems of wave equations : general theoretical frame

1.1. General formulation

This paragraph is devoted to 1D nonlinear Hamiltonian systems of wave equations. A function $H : \mathbb{R}^N \to \mathbb{R}$, a potential energy, totally determines the system (N is the size of the system). We shall consider the following Cauchy problem (Ω is a segment of $\mathbb{R}$ or $\mathbb{R}$ itself):

$$\left\{\begin{array}{l} \text{Find } \mathbf{u} = (u_1, \cdots, u_N) : \Omega \times \mathbb{R}^+ \to \mathbb{R}^N, \\ \partial_{tt}^2 \mathbf{u} - \partial_x \left[\nabla H\left(\partial_x \mathbf{u}\right)\right] = 0, \ x \in \Omega, \ t > 0, \\ \mathbf{u}(x,0) = \mathbf{u}_0(x), \quad \partial_t \mathbf{u}(x,0) = \mathbf{u}_1(x), \\ \mathbf{u}(x,t) = 0, \ \forall x \in \partial\Omega. \end{array}\right. \tag{1}$$

Remark 1.1. *The function H is only used through its gradient, hence any $H + \mathcal{L}$ gives the same system of equations as H, with $\mathcal{L}$ a linear function on $\mathbb{R}^N$. Thus, it is not restrictive to assume that $H(0) = 0$ and $\nabla H(0) = 0$.*

The mathematical properties of the system will depend on the properties of H. Here are the main assumptions that we will consider in the following sections:

($\mathcal{H}1$) smoothness : H is of class C^2,

($\mathcal{H}2$) coercivity : $\exists\, K > 0 \quad$ s.t. $\quad H(\mathbf{v}) \geq K\,|\mathbf{v}|^2$,

($\mathcal{H}3$) convexity : H is strictly convex,

($\mathcal{H}4$) There exists $\quad c_+ > 0 \ $ s.t. $\ |\nabla H(\mathbf{v})|^2 \leq 2\, c_+^2\, H(\mathbf{v})$,

($\mathcal{H}5$) There exists $\quad M > 0 \ $ s.t. $\ |\nabla H(\mathbf{v})| \leq M\,(1 + |\mathbf{v}|)$.

Sometimes, these properties will be needed only locally.

1.2. Energy preservation and H^1 stability

Theorem 1.1. *Any smooth enough solution* $\mathbf{u}$ *of (1) satisfies the energy identity:*

$$\frac{d}{dt}E(t) = 0, \quad \textit{with } E(t) = \int_\Omega \left\{\frac{1}{2}\left|\partial_t \mathbf{u}\right|^2 + H(\partial_x \mathbf{u})\right\} dx. \tag{2}$$

Proof. For completeness, we include the following well known proof. We take the inner product in $\mathbb{R}^N$ of the equation with $\partial_t u$ and integrate over x to obtain :

$$\int_\Omega \left[\partial_{tt}^2 \mathbf{u} - \partial_x \left[\nabla H(\partial_x \mathbf{u})\right]\right] \cdot \partial_t \mathbf{u} = 0.$$

After integration by parts we obtain:

$$\int_\Omega \left[\partial_{tt}^2 \mathbf{u} \cdot \partial_t \mathbf{u} + \nabla H(\partial_x \mathbf{u}) \cdot \partial_{tx}^2 \mathbf{u}\right] = 0,$$

where we recognize

$$\int_\Omega \frac{1}{2}\partial_t\left(|\partial_t \mathbf{u}|^2\right) + \int_\Omega \partial_t\left[H(\partial_x \mathbf{u})\right] = 0.$$

Hence,

$$\frac{d}{dt}\int_\Omega \left\{\frac{1}{2}\left|\partial_t \mathbf{u}\right|^2 + H(\partial_x \mathbf{u})\right\} dx = 0.$$

□

Obviously, (2) yields an upper bound on the H^1-norm of the solution, under the hypothesis ($\mathcal{H}2$).

Corollary 1.1. *Let us assume (H2). Then, there exists $C > 0$ such that*

$$||\mathbf{u}(\cdot,t)||_{H^1} \leq C\, E(0), \qquad \forall t \geq 0.$$

1.3. Hyperbolicity of the system

We begin by recalling here some basic definitions for first order hyperbolic systems (see [13] for more details and mathematical results).

Definition 1.1. HYPERBOLIC SYSTEM
We consider the system of equations

$$\partial_t \mathbf{u} + \partial_x \mathbf{F}(\mathbf{u}) = 0 \tag{3}$$

where $\mathbf{F} : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}^n$ is of class C^1 and we look for $\mathbf{u} = (u_1, \cdots, u_n) : \Omega \times \mathbb{R}^+ \to \mathcal{D}$. We define

$$\mathbf{A}(\mathbf{u}) = \left(\frac{\partial \mathbf{F}_i}{\partial u_j}(\mathbf{u}) \right)_{1 \leq i,j \leq n} \tag{4}$$

the Jacobian matrix of $\mathbf{F}$. The system (3) is said to be hyperbolic if, for any $\mathbf{u} \in \mathcal{D}$, the matrix $\mathbf{A}(\mathbf{u})$ has n real eigenvalues $\mu_1(\mathbf{u}) \leq \ldots \leq \mu_k(\mathbf{u}) \leq \ldots \leq \mu_n(\mathbf{u})$ and p linearly independent corresponding eigenvectors $\mathbf{r}_1(\mathbf{u}), \ldots, \mathbf{r}_k(\mathbf{u}), \ldots, \mathbf{r}_n(\mathbf{u})$, i.e.

$$\mathbf{A}(\mathbf{u})\, \mathbf{r}_k(\mathbf{u}) = \mu_k(\mathbf{u})\, \mathbf{r}_k(\mathbf{u}).$$

If, in addition, the eigenvalues $\mu_k(\mathbf{u})$ are all distinct, the system (3) is called strictly hyperbolic.

The system is said locally (strictly) hyperbolic near $\mathbf{u}_0$ if the appropriate properties are true not for $\mathbf{u} \in \mathcal{D}$ but for $\mathbf{u}$ in a neighborhood of $\mathbf{u}_0$.

With the notations

$$\mathbf{U} \equiv (\mathbf{U}_t, \mathbf{U}_x) := (\partial_t \mathbf{u}, \partial_x \mathbf{u}) \quad \text{and} \quad \mathbf{U}_0(x) = (\mathbf{u}_1, \partial_x \mathbf{u}_0),$$

the system (1) can be written

$$\begin{cases} \text{Find } \mathbf{U} : \Omega \times \mathbb{R}^+ \to \mathbb{R}^{2N}, \\ \partial_t \mathbf{U} + \partial_x F(\mathbf{U}) = 0, \ x \in \Omega, \ t > 0, \\ \mathbf{U}(x,0) = \mathbf{U}_0(x), \ x \in \Omega. \end{cases} \tag{5}$$

where the nonlinear function F is given by

$$\forall\, \mathbf{U} = (\mathbf{U}_t, \mathbf{U}_x) \in \mathbb{R}^N \times \mathbb{R}^N, \quad F(\mathbf{U}) = \begin{pmatrix} -\nabla H(\mathbf{U}_x) \\ -\mathbf{U}_t \end{pmatrix}. \tag{6}$$

Theorem 1.2. *Local hyperbolicity of the system (1) is equivalent to local convexity of H.*

PROOF. The Jacobian of F is given by

$$\forall\, \mathbf{U} = (\mathbf{U}_t, \mathbf{U}_x) \in \mathbb{R}^N \times \mathbb{R}^N,$$

$$A(\mathbf{U}) = DF(\mathbf{U}) = \begin{pmatrix} 0 & -D^2H(\mathbf{U}_x) \\ -I & 0 \end{pmatrix}. \tag{7}$$

where $D^2H(\mathbf{U}_x)$ refers to the Hessian matrix of H. The eigenvalue problem for $A(\mathbf{U})$:

$$\text{Find } \big(\mathbf{Z}(\mathbf{U}) = (\mathbf{Z}_t(\mathbf{U}), \mathbf{Z}_x(\mathbf{U})\big) \neq 0 \in \mathbb{C}^{2N} \text{ and } \mu(\mathbf{U}) \in \mathbb{C},$$

$$DF(\mathbf{U})\, \mathbf{Z}(\mathbf{U}) = \mu(\mathbf{U})\, \mathbf{Z}(\mathbf{U}).$$

is equivalent to

$$\begin{cases} D^2H(\mathbf{U}_x)\, \mathbf{Z}_x(\mathbf{U}) = \mu^2(\mathbf{U})\, \mathbf{Z}_x(\mathbf{U}), \\ \mathbf{Z}_t(\mathbf{U}) = -\mu(\mathbf{U})\, \mathbf{Z}_x(\mathbf{U}). \end{cases}$$

Thus it is clear that the local hyperbolicity of (1) relies on the positivity of the eigenvalues of $D^2H(\mathbf{U}_x)$, i.e. on the local convexity of H.

□

1.4. Finite propagation velocity

Theorem 1.3. *We assume (H4). Then, any smooth enough solution $\mathbf{u}$ of (1) propagates with a velocity lower than C.*

PROOF. Let us assume hypothesis (H4). We can notice that this property induces that H is positive. We will use an energy technique. Let $V > 0$, to be specified later, and $a \in \mathbb{R}$ such that the initial data have their support in $]-\infty, a[$. We have for all $t > 0$:

$$\int_{a+Vt}^{+\infty} \Big(\partial_{tt}^2 \mathbf{u} - \partial_x \left[\nabla H(\partial_x \mathbf{u}) \right] \Big) \cdot \partial_t \mathbf{u}\, dx = 0,$$

that is to say, after integration by parts,

$$\Big| \int_{a+Vt}^{+\infty} \partial_t \Big(\frac{1}{2} \left| \partial_t \mathbf{u} \right|^2 \Big)\, dx + \int_{a+Vt}^{+\infty} \left[\nabla H(\partial_x \mathbf{u}) \right] \cdot \partial_{xt}^2 \mathbf{u}\, dx - \left[\nabla H(\partial_x \mathbf{u}) \cdot \partial_t \mathbf{u} \right](a+Vt, t) = 0,$$

which, with the energy density

$$\mathbf{e} = \frac{1}{2} \left| \partial_t \mathbf{u} \right|^2 + H(\partial_x \mathbf{u})$$

can be written

$$\int_{a+Vt}^{+\infty} \frac{\partial \mathbf{e}}{\partial t}\, dx - \left[\nabla H(\partial_x \mathbf{u}) \cdot \partial_t \mathbf{u} \right](a+Vt, t) = 0$$

or, after derivation under the integral,

$$\frac{d}{dt} \int_{a+Vt}^{+\infty} \mathbf{e}\, dx + \Phi(a+Vt, t) = 0,$$

where

$$\Phi := V\, \mathbf{e} - \left[\nabla H(\partial_x \mathbf{u}) \cdot \partial_t \mathbf{u} \right].$$

Using the hypothesis on the initial data, we have

$$\forall t > 0, \quad \int_{a+Vt}^{+\infty} \mathbf{e}(x,t)\, dx = -\int_0^t \Phi(a+Vs, s)\, ds.$$

Now we choose V large enough such that Φ is positive. This is possible since

$$\left| \nabla H(\partial_x \mathbf{u}) \cdot \partial_t \mathbf{u} \right| \leq \frac{V}{2} \left| \partial_t \mathbf{u} \right|^2 + \frac{1}{2V} \left| \nabla H(\partial_x \mathbf{u}) \right|^2$$

Consequently, using $(\mathcal{H}4)$

$$\Phi \geq V H(\partial_x \mathbf{u}) - \frac{1}{2V}\left|\nabla H(\partial_x \mathbf{u})\right|^2 \geq \left(V - \frac{c_+^2}{V}\right) H(\partial_x \mathbf{u})$$

because of the hypothesis $(\mathcal{H}4)$. If we choose $V = c_+$, the function Φ is positive, and we have

$$\mathbf{e}(x,t) = 0 \quad \text{for} \quad x > a + c_+ t, \quad t > 0.$$

Since H is positive, we have

$$\mathbf{u}(x,t) = 0 \quad \text{for} \quad x > a + c_+ t, \quad t > 0.$$

This result shows that the propagation velocity of the solution of the Cauchy problem (1) is bounded above by:

$$c_+ = \left(\frac{1}{2} \sup_{(\mathbf{u}_x)} R(\mathbf{u}_x)\right)^{\frac{1}{2}}, \quad R(\mathbf{u}_x) := \frac{\left|\nabla H(\mathbf{u}_x)\right|^2}{H(\mathbf{u}_x)}. \tag{8}$$

□

2. Finite element energy preserving numerical schemes for nonlinear Hamiltonian systems of wave equations

Each time one wishes to discretize in space and time an evolution problem whose solution satisfies the conservation of an energy, as in the case of the systems (1) but more generally of many mechanical models, it is a natural idea to try to construct numerical schemes that preserve rigorously a discrete energy that is equivalent of the continuous energy. As we shall see immediately in the next paragraph, in the case of (1), the use of variational techniques (such as the finite element method) for the space semi-discretization ensures "by construction" the conservation of a positive semi-discrete energy. The difficulties really occur when the time discretization is concerned. This essential issue will be the object of section 2.2.

2.1. Spatial semi discretization

2.1.1. Variational formulation

Let us consider the following system of partial differential equations:

$$\begin{cases} \partial_t^2 \mathbf{u} - \partial_x \left[\nabla H(\partial_x \mathbf{u})\right] = 0, \\ \mathbf{u}(x,t) = 0, \quad \forall\, t > 0,\ \forall\, x \in \partial\Omega. \end{cases} \tag{9}$$

Even though all of the following could probably be generalized to a more general context, we shall assume that the function H satisfies the coercivity property $(\mathcal{H}2)$ and the additional assumption $(\mathcal{H}5)$:

$$\exists\, M > 0 \text{ such that } |\nabla H(\mathbf{v})| \leq M\,(1 + |\mathbf{v}|),\ \forall \mathbf{v} \in \mathbb{R}^N. \tag{10}$$

In this case, according to the continuous energy identity, we expect that the solution $\mathbf{u}$ satisfies

$$\mathbf{u} \in C^0(\mathbb{R}^+; H_0^1(\Omega)^N) \tag{11}$$

which implies, because of $(\mathcal{H}5)$

$$H(\partial_x \mathbf{u}) \in L^\infty(\mathbb{R}^+; L^2(\Omega)^N). \tag{12}$$

In this framework, we can write a variational formulation in space of (1) in the space:

$$\mathcal{V} = \left(H_0^1(\Omega)\right)^N. \tag{13}$$

Let us take the inner product (in $\mathbb{R}^N$) of (1) by $\mathbf{v} \in \mathcal{V}$ and integrate over space the resulting equality. We get after integration by parts, since the boundary terms vanish,

$$\frac{d^2}{dt^2}\Big(\int_\Omega \mathbf{u} \cdot \mathbf{v}\Big) + \int_\Omega \nabla H(\partial_x \mathbf{u}) \cdot \partial_x \mathbf{v} = 0,\ \forall\, \mathbf{v} \in \mathcal{V} \tag{14}$$

which is the variational formulation of the problem.

2.1.2. Semi-discretization in space

We consider as usual $\{\mathcal{V}_h, h > 0\}$ a family of finite dimensional subspaces of $\mathcal{V}$, where h is an approximation parameter destined to tend to 0. We assume the standard approximation property:

$$\forall\, \mathbf{v} \in \mathcal{V}, \quad \lim_{h \to 0} \inf_{\mathbf{v}_h \in \mathcal{V}_h} \|\mathbf{v} - \mathbf{v}_h\| = 0 \tag{15}$$

The most classical example is the approximation with conforming Lagrange finite elements of degree $k \geq 1$, the so-called P_k finite elements, on a family of meshes of Ω (in which case the approximation parameter is nothing but the stepsize of the mesh).

We consider the following semi-discrete problem: find $\mathbf{u}_h : \mathbb{R}^+ \mapsto \mathcal{V}_h$ such that

$$\frac{d^2}{dt^2}\left[\int_\Omega \mathbf{u}_h . \mathbf{v}_h\right] + \int_\Omega \nabla H(\partial_x \mathbf{u}_h) . \partial_x \mathbf{v}_h = 0,\ \forall\, \mathbf{v}_h \in \mathcal{V}_h\,. \tag{16}$$

We can write an algebraic formulation of (16) after having introduced the vector $\mathbf{U}_h \in \mathbb{R}^{N_h}$ (resp. $\mathbf{V}_h \in \mathbb{R}^{N_h}$) of the components of $\mathbf{u}_h$ (resp. $\mathbf{v}_h$) in an appropriate basis of $\mathcal{V}_h$. We first introduce the linear operator in $\mathbb{R}^{N_h}$, M_h defined by:

$$\Big(M_h \mathbf{U}_h, \mathbf{V}_h\Big)_h = \int_\Omega \mathbf{u}_h . \mathbf{v}_h, \quad \forall\, \mathbf{v}_h \in \mathcal{V}_h\,. \tag{17}$$

By analogy with the formula

$$\int_\Omega -\partial_x \nabla H(\partial_x \mathbf{u}) \cdot \mathbf{v} = \int_\Omega \nabla H(\partial_x \mathbf{u}) \cdot \partial_x \mathbf{v}$$

we introduce the nonlinear function in $\mathbb{R}^{N_h}$ (the complicated notation is chosen for convenience to emphasize the analogy with the continuous case - note that $-\partial_x$ is the formal adjoint of ∂_x)

$$\mathbf{D}_h^*\big(\nabla H(\mathbf{D}_h)\big) : \mathbf{U}_h \mapsto \mathbf{D}_h^*\big(\nabla H(\mathbf{D}_h \mathbf{U}_h)\big), \tag{18}$$

defined by

$$\Big(\mathbf{D}_h^*\big(\nabla H(\mathbf{D}_h \mathbf{U}_h)\big), \mathbf{V}_h\Big)_h = \int_\Omega \nabla H(\partial_x \mathbf{u}_h) \cdot \partial_x \mathbf{v}_h, \quad \forall\, \mathbf{v}_h \in \mathcal{V}_h\,. \tag{19}$$

Then, (16) is clearly equivalent to the following nonlinear differential system in $\mathbb{R}^{N_h}$ (where $\mathbf{U}_h(t)$ is the vector of the degrees of freedom of $\mathbf{u}_h(t)$)

$$\mathbf{M}_h \frac{d^2\mathbf{U}_h}{dt^2} + \mathbf{D}_h^*\big(\nabla H(\mathbf{D}_h\mathbf{U}_h)\big) = 0. \tag{20}$$

The effective implementation (after time discretization - see the next paragraph) inevitably requires the computation of the integrals in the right hand sides of (17), (19). For the nonlinear part (19), it is not possible to compute exactly these integrals - except for very particular H. This is the case for instance with the string model of section 3. That is why these integrals will be evaluated approximately, which will lead to the following new definitions for $\mathbf{M}_h$ and $\mathbf{D}_h^*\big(\nabla H(\mathbf{D}_h)\big)$, $\forall\, \mathbf{v}_h \in \mathcal{V}_h$:

$$\begin{cases} \Big(M_h\mathbf{U}_h, \mathbf{V}_h\Big)_h = \oint_\Omega^h \mathbf{u}_h.\mathbf{v}_h, \\ \Big(\mathbf{D}_h^*\big(\nabla H(\mathbf{D}_h\mathbf{U}_h)\big), \mathbf{V}_h\Big)_h = \oint_\Omega^h \nabla H(\partial_x\mathbf{u}_h)\cdot\partial_x\mathbf{v}_h, \end{cases} \tag{21}$$

where the linear form $f \mapsto \oint_\Omega^h f$ is an approximate integral

$$\oint_\Omega^h f \simeq \int_\Omega f \quad \text{for small } h. \tag{22}$$

In practice, this approximate integral will be constructed, in the context of finite elements, by decomposing the global integral as the sum of integrals along the segments of the finite element mesh and using inside each segment a given quadrature rule. As a result, the integral becomes exact as soon as f is piecewise (according to the mesh) polynomial of a certain degree. In particular the calculation of the mass matrix $\mathbf{M}_h$ may be exact in this degree is large enough since, contrary to $\nabla H(\partial_x\mathbf{u}_h)\cdot\partial_x\mathbf{v}_h$, the product $\mathbf{u}_h.\mathbf{v}_h$ is piecewise polynomial.

An important property is required, namely that

$$f \geq 0 \quad \Longrightarrow \quad \oint_\Omega^h f \geq 0. \tag{23}$$

This will be the case in the finite element context as long as quadrature formulas with positive quadrature weights are chosen.

Remark 2.1. *In practice, for appropriate quadrature formulas adapted to the finite element space $\mathcal{V}_h$, the positivity of the quadrature weights induces a stronger property, namely the existence of $\gamma > 0$ such that*

$$\forall \mathbf{v}_h \in \mathcal{V}_h, \quad \oint_\Omega^h |\mathbf{v}_h|^2 \geq \gamma \int_\Omega |\mathbf{v}_h|^2.$$

On the other hand, one can choose a quadrature rule that makes the mass matrix become diagonal. This is called mass lumping and can lead to explicit schemes (see section 10.4 pages 305 to 313 of [22] for a mathematical approach).

For a smooth enough function H, typically $H \in C^2(\mathbb{R})$, the existence and uniqueness of a local (in a maximum time interval $[0, T_h[$) solution $\mathbf{u}_h$ of (16) is a direct and easy consequence of standard theorems from the theory of ordinary differential equations [20], with the regularity

$$\mathbf{u}_h \in C^2(0, T_h; \mathcal{V}_h).$$

Our next result allows us to show that the solution is for each h global in time ($T_h = +\infty$) and provides H^1 stability estimates.

Theorem 2.1. *The scheme (16) preserves a semi discrete energy, i.e. the solution $\mathbf{u}_h$ of the scheme satisfies:*

$$\frac{d}{dt}E_h(t) = 0, \quad \textit{with} \quad E_h(t) = \frac{1}{2}\oint_\Omega^h |\partial_t\mathbf{u}_h|^2 + \oint_\Omega^h H\big(\partial_x\mathbf{u}_h\big).$$

Proof. This property comes directly from the variational formulation, with $\mathbf{v}_h = \partial_t\mathbf{u}_h$ which belongs to $\mathcal{V}_h$. Then we have:

$$\oint_\Omega^h \Big(\partial_t^2\mathbf{u}_h\Big).\Big(\partial_t\mathbf{u}_h\Big) + \oint_\Omega^h \nabla H(\partial_x\mathbf{u}_h).\partial_x\Big(\partial_t\mathbf{u}_h\Big) = 0$$

that can be rewritten

$$\oint_\Omega^h \partial_t\Big(\frac{1}{2}\Big|\partial_t\mathbf{u}_h\Big|^2\Big) + \oint_\Omega^h \partial_t\Big(H(\partial_x\mathbf{u}_h)\Big) = 0,$$

which leads to the result.

□

We easily deduce from theorem 2.1 the following discrete H^1 stability result:

Corollary 2.1. *Let us assume hypothesis ($\mathcal{H}2$). Then, there exists $C > 0$ such that*

$$\oint_\Omega^h |\partial_x\mathbf{u}_h(t)|^2 \leq C\, E_h(0), \qquad \forall t \geq 0. \tag{24}$$

Proof. Theorem 2.1 implies

$$\oint_\Omega^h H(\partial_x\mathbf{u}_h)\, dx = E_h(0) - \frac{1}{2}\oint_\Omega^h \Big|\partial_t\mathbf{u}_h\Big|^2\, dx.$$

Using the hypotheses (23) and ($\mathcal{H}2$), we get

$$K\oint_\Omega^h \Big|\partial_x\mathbf{u}_h\Big|^2 \leq \oint_\Omega^h H(\partial_x\mathbf{u}_h)\, dx \leq E_h(0).$$

□

Remark 2.2. *As in Remark 2.1, with appropriate quadrature formulas adapted to the finite element space $\mathcal{V}_h$, (24) yields uniform H^1-upper bounds for (16) (one uses in particular a discrete form of Poincaré's inequality.)*

2.2. Time discretization : construction of energy preserving schemes

2.2.1. A class of energy preserving schemes

As announced previously, we investigate the question of finding finite difference schemes that preserve rigorously a discrete energy. Such schemes are well known in the linear case, which corresponds to

$$H(\mathbf{v}) = \frac{1}{2}\mathbf{A}\mathbf{v}\cdot\mathbf{v}, \quad (\Longrightarrow \quad \nabla H(\mathbf{v}) = \mathbf{A}\mathbf{v}) \tag{25}$$

that is to say to the linear hyperbolic system

$$\partial_t^2 \mathbf{u} - \mathbf{A}\,\partial_{xx}^2 \mathbf{u} = 0 \tag{26}$$

and its corresponding semi-discrete version, using the notation of the previous paragraph

$$\frac{d^2}{dt^2}\left[\oint_\Omega^h \mathbf{u}_h\cdot\mathbf{v}_h\right] + \oint_\Omega^h \mathbf{A}\,\partial_x\mathbf{u}_h\cdot\partial_x\mathbf{v}_h = 0, \forall\ \mathbf{v}_h\in\mathcal{V}_h\,, \tag{27}$$

which preserves the quadratic discrete energy

$$E_h(t) = \frac{1}{2}\oint_\Omega^h |\partial_t\mathbf{u}_h|^2 + \frac{1}{2}\oint_\Omega^h \mathbf{A}\,\partial_x\mathbf{u}_h\cdot\partial_x\mathbf{u}_h. \tag{28}$$

In this case, there is a natural class of energy preserving schemes, called the θ-schemes, where $\theta\in[0,1/2]$ is an averaging parameter. Those schemes belong to the more general class of Newmark schemes (see chapter XX of [9]) that also contains dissipative schemes. Using a constant time step Δt and denoting by $\mathbf{u}_h^n$ the approximation of $\mathbf{u}_h(t^n)$, this scheme is, in its variational form:

$$\begin{aligned}&\forall\ \mathbf{v}_h\in\mathcal{V}_h,\ \oint_\Omega^h \frac{\mathbf{u}_h^{n+1}-2\mathbf{u}_h^n+\mathbf{u}_h^{n-1}}{\Delta t^2}\cdot\mathbf{v}_h + \\ &\oint_\Omega^h \mathbf{A}\,\partial_x\big(\theta\mathbf{u}_h^{n+1}+(1-2\theta)\mathbf{u}_h^n+\theta\mathbf{u}_h^{n-1}\big)\cdot\partial_x\mathbf{v}_h = 0.\end{aligned} \tag{29}$$

The solution of this scheme satisfies the conservation of a discrete energy

$$E_h^{n+\frac{1}{2}} = E_h^{n-\frac{1}{2}} \tag{30}$$

where the discrete energy $E_h^{n+\frac{1}{2}}$ corresponding to time $t^{n+\frac{1}{2}} = (n+\frac{1}{2})\Delta t$ is given by

$$\begin{aligned}E_h^{n+\frac{1}{2}} &= \frac{1}{2}\oint_\Omega^h \Big|\frac{\mathbf{u}_h^{n+1}-\mathbf{u}_h^n}{\Delta t}\Big|^2 + \frac{1}{2}\oint_\Omega^h \mathbf{A}\,\partial_x\big(\mathbf{u}_h^{n+1/2}\big)\cdot\partial_x\big(\mathbf{u}_h^{n+1/2}\big) \\ &+\frac{1}{2}\left(\theta-\frac{1}{4}\right)\Delta t^2\oint_\Omega^h \mathbf{A}\,\partial_x\Big(\frac{\mathbf{u}_h^{n+1}-\mathbf{u}_h^n}{\Delta t}\Big)\cdot\partial_x\Big(\frac{\mathbf{u}_h^{n+1}-\mathbf{u}_h^n}{\Delta t}\Big)\end{aligned}$$

where we have set $\mathbf{u}_h^{n+1/2} := \dfrac{\mathbf{u}_h^{n+1}+\mathbf{u}_h^n}{2}$.

The identity (30) is easily derived by taking $v_h = \big(\mathbf{u}_h^{n+1} - \mathbf{u}_h^{n-1}\big)/2\Delta t$ as a test function in (29).

The conservation of the energy $E_h^{n+\frac{1}{2}}$ automatically provides the stability of the scheme when $\theta\geq 1/4$ since $E_h^{n+\frac{1}{2}}$ is always positive.

When $\theta < 1/4$, the scheme is stable under the stability condition

$$(1-4\theta)\,\frac{\Delta t^2}{4}\sup_{\mathbf{v}_h\in\mathcal{V}_h}\left[\frac{\oint_\Omega^h \mathbf{A}\,\partial_x\mathbf{v}_h\cdot\partial_x\mathbf{v}_h}{\oint_\Omega^h |\mathbf{v}_h|^2}\right] \leq 1 \tag{31}$$

which is nothing but the condition that ensures the positivity of $E_h^{n+\frac{1}{2}}$.

When $\theta = 0$, one gets the well-known leap-frog scheme - or explicit scheme - which is explicit in practice when one achieves mass lumping (see remark 2.1).

Our objective is in some sense to generalize the θ-scheme to the nonlinear case, with as main objective the preservation of a discrete energy guaranteeing the stability of the scheme.

Of course the problematic of energy preserving schemes is close to the problematic of symplectic schemes [30] for the discretization of Hamiltonian differential equations, whose purpose is to preserve other invariants of the continuous problems. These invariants are of more geometrical nature and linked to the preservation of symmetries of the system. In general, such schemes cannot preserve a discrete energy (see [34]) but can succeed in "almost preserving" such an energy over large times [18, 19, 29]. Some authors have carried out a comparison between symplectic and energy preserving schemes, and found the latter "more accurate" (see [10, 14, 15]).

In what follows, we are going to investigate a class of three point schemes for the time discretization of (16). These schemes have the same type of structure as the θ-schemes and include all the linear schemes. They will be based on a function that we shall call the "approximate gradient":

$$\left|\begin{array}{lccl}\mathbb{\nabla} H: & \mathbb{R}^N\times\mathbb{R}^N\times\mathbb{R}^N & \longrightarrow & \mathbb{R}^N \\ & (\mathbf{u},\mathbf{v},\mathbf{w}) & \longrightarrow & \mathbb{\nabla} H(\mathbf{u},\mathbf{v},\mathbf{w})\end{array}\right. \tag{32}$$

that should satisfy the consistency condition

$$\forall\mathbf{v}\in\mathbb{R}^N, \quad \mathbb{\nabla} H(\mathbf{v},\mathbf{v},\mathbf{v}) = \nabla H(\mathbf{v}). \tag{33}$$

Using such an approximate gradient, the fully discrete version of (16) is

$$\begin{aligned}&\forall\ \mathbf{v}_h\in\mathcal{V}_h,\ \oint_\Omega^h \frac{\mathbf{u}_h^{n+1}-2\mathbf{u}_h^n+\mathbf{u}_h^{n-1}}{\Delta t^2}\cdot\mathbf{v}_h + \\ &\oint_\Omega^h \mathbb{\nabla} H(\partial_x\mathbf{u}_h^{n+1},\partial_x\mathbf{u}_h^n,\partial_x\mathbf{u}_h^{n-1})\cdot\partial_x\mathbf{v}_h = 0.\end{aligned} \tag{34}$$

For the sequel, we shall assume that $\mathbb{\nabla} H$ is a "smooth enough" function. Note that

- One obtains an explicit scheme (provided mass lumping) as soon as

$$\mathbb{\nabla} H(\mathbf{u},\mathbf{v},\mathbf{w}) \text{ is independent of } \mathbf{u}. \tag{35}$$

- One obtains a scheme which is reversible in time and second order accurate (see remark 2.3) if and only if

$$\mathbb{\nabla} H(\mathbf{u},\mathbf{v},\mathbf{w}) = \mathbb{\nabla} H(\mathbf{w},\mathbf{v},\mathbf{u}), \quad \forall\, (\mathbf{u},\mathbf{v},\mathbf{w}) \in (\mathbb{R}^N)^3. \tag{36}$$

Remark 2.3. *To check the second order accuracy of the time approximation, we introduce the truncation error defined as the linear form*

$$\mathcal{E}(\mathbf{v}_h) = \oint_\Omega^h \frac{\mathbf{u}_h^{n+1} - 2\mathbf{u}_h^n + \mathbf{u}_h^{n-1}}{\Delta t^2} \cdot \mathbf{v}_h + \oint_\Omega^h \mathbb{\nabla} H(\partial_x \mathbf{u}_h^{n+1}, \partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^{n-1}) \cdot \partial_x \mathbf{v}_h \tag{37}$$

with $\mathbf{u}_h^n := \mathbf{u}_h(t^n)$ *and* $\mathbf{u}_h(\cdot)$ *is the solution of (16), which is a smooth function of time. Using a Taylor expansion we obtain*

$$\frac{\mathbf{u}_h^{n+1} - 2\mathbf{u}_h^n + \mathbf{u}_h^{n-1}}{\Delta t^2} = \frac{d^2 \mathbf{u}_h}{dt^2}(t^n) + O(\Delta t^2). \tag{38}$$

On the other hand, denoting $D_1 \mathbb{\nabla} H(\mathbf{u},\mathbf{v},\mathbf{w})$ *(resp.* $D_3 \mathbb{\nabla} H(\mathbf{u},\mathbf{v},\mathbf{w})$*) the differential of the application* $\mathbf{u} \mapsto \mathbb{\nabla} H(\mathbf{u},\mathbf{v},\mathbf{w})$ *(resp.*$\mathbf{w} \mapsto \mathbb{\nabla} H(\mathbf{u},\mathbf{v},\mathbf{w})$*), we have*

$$\begin{aligned}\mathbb{\nabla} H(\partial_x \mathbf{u}_h^{n+1}, \partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^{n-1}) &= \mathbb{\nabla} H(\partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^n) \\ &+ D_1 \mathbb{\nabla} H(\partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^n)\, \partial_x\big(\mathbf{u}_h^{n+1} - \mathbf{u}_h^n\big) \\ &+ D_3 \mathbb{\nabla} H(\partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^n)\, \partial_x\big(\mathbf{u}_h^{n-1} - \mathbf{u}_h^n\big) \\ &+ O(\Delta t^2)\end{aligned}$$

that is to say, using the consistency condition (33)

$$\begin{aligned}\mathbb{\nabla} H(\partial_x \mathbf{u}_h^{n+1}, \partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^{n-1}) &= \nabla H(\partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^n) \\ &+ D_1 \mathbb{\nabla} H(\partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^n)\, \partial_x\big(\mathbf{u}_h^{n+1} - \mathbf{u}_h^n\big) \\ &+ D_3 \mathbb{\nabla} H(\partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^n)\, \partial_x\big(\mathbf{u}_h^{n-1} - \mathbf{u}_h^n\big) \\ &+ O(\Delta t^2)\end{aligned}$$

Differentiating with respect to $\mathbf{u}$ *the symmetry condition (36), we get*

$$D_1 \mathbb{\nabla} H(\mathbf{u},\mathbf{v},\mathbf{w}) = D_3 \mathbb{\nabla} H(\mathbf{w},\mathbf{v},\mathbf{u}),\ \forall\, (\mathbf{u},\mathbf{v},\mathbf{w}) \in (\mathbb{R}^N)^3$$

so that we can write

$$\begin{aligned}&\mathbb{\nabla} H(\partial_x \mathbf{U}_h^{n+1}, \partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^{n-1}) = \nabla H(\partial_x \mathbf{u}_h^n) + O(\Delta t^2) \\ &+ D_1 \mathbb{\nabla} H(\partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^n)\, \big(\partial_x \mathbf{u}_h^{n+1} - 2\partial_x \mathbf{U}_h^n + \partial_x \mathbf{u}_h^{n-1}\big),\end{aligned}$$

that is to say , since

$$\big(\partial_x \mathbf{u}_h^{n+1} - 2\partial_x \mathbf{u}_h^n + \partial_x \mathbf{u}_h^{n-1}\big) = O(\Delta t^2)\ ,$$

$$\mathbb{\nabla} H(\partial_x \mathbf{u}_h^{n+1}, \partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^{n-1}) = \nabla H(\partial_x \mathbf{u}_h^n) + O(\Delta t^2). \tag{39}$$

It suffices to substitute (38) and (39) into (37) and to use the fact that $\mathbf{u}_h(\cdot)$ *is a solution of (16) to conclude that*

$$\mathcal{E}(\mathbf{v}_h) = O(\Delta t^2).$$

To investigate the preservation of a discrete energy, we choose $\mathbf{v}_h = \big(\mathbf{u}_h^{n+1} - \mathbf{u}_h^{n-1}\big)/2\Delta t$ as a test function in (34) (as in the linear case) and obtain

$$\begin{aligned}0 = \frac{1}{\Delta t}\Big\{ \frac{1}{2}\oint_\Omega^h \Big|\frac{\mathbf{u}_h^{n+1} - \mathbf{u}_h^n}{\Delta t}\Big|^2 - \frac{1}{2}\oint_\Omega^h \Big|\frac{\mathbf{u}_h^n - \mathbf{u}_h^{n-1}}{\Delta t}\Big|^2 \Big\} \\ + \oint_\Omega^h \mathbb{\nabla} H(\partial_x \mathbf{u}_h^{n+1}, \partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^{n-1}) \cdot \partial_x\big(\frac{\mathbf{u}_h^{n+1} - \mathbf{u}_h^{n-1}}{2\Delta t}\big).\end{aligned} \tag{40}$$

This identity leads us to make the following definition

Definition 2.1. *The function* $\mathbb{\nabla} H$ *is called "conservative" (we shall also say that the corresponding scheme is "conservative" or "energy preserving"- cf lemma 2.1) if and only if there exists a scalar function (a discrete potential energy)*

$$\left|\begin{array}{lccc}\mathbb{H}: & \mathbb{R}^N \times \mathbb{R}^N & \longrightarrow & \mathbb{R} \\ & (\mathbf{u},\mathbf{v}) & \longrightarrow & \mathbb{H}(\mathbf{u},\mathbf{v})\end{array}\right. \tag{41}$$

such that $\forall\, (\mathbf{u},\mathbf{v},\mathbf{w}) \in (\mathbb{R}^N)^3$,

$$\mathbb{\nabla} H(\mathbf{u},\mathbf{v},\mathbf{w}) \cdot \frac{(\mathbf{u} - \mathbf{w})}{2} = \mathbb{H}(\mathbf{u},\mathbf{v}) - \mathbb{H}(\mathbf{v},\mathbf{w}). \tag{42}$$

Remark 2.4. *Note that (42) implies in particular the symmetry of* $\mathbb{H}$*:*

$$\forall\, (\mathbf{u},\mathbf{v}) \in \mathbb{R}^N \times \mathbb{R}^N, \quad \mathbb{H}(\mathbf{u},\mathbf{v}) = \mathbb{H}(\mathbf{v},\mathbf{u}). \tag{43}$$

Of course, for consistency reasons, the discrete potential energy $\mathbb{H}$ should also satisfy (in agreement with (33))

$$\forall \mathbf{v} \in \mathbb{R}^N, \quad \mathbb{H}(\mathbf{v},\mathbf{v}) = H(\mathbf{v}). \tag{44}$$

Assuming that $\mathbb{\nabla} H$ is conservative, we deduce from (42) that

$$\left|\begin{array}{l}\mathbb{\nabla} H(\partial_x \mathbf{u}_h^{n+1}, \partial_x \mathbf{u}_h^n, \partial_x \mathbf{u}_h^{n-1}) \cdot \partial_x\big(\frac{\mathbf{u}_h^{n+1} - \mathbf{u}_h^{n-1}}{2\Delta t}\big) = \\ \qquad = \frac{1}{\Delta t}\Big\{ \mathbb{H}(\partial_x \mathbf{u}_h^{n+1}, \partial_x \mathbf{u}_h^n) - \mathbb{H}(\mathbf{u}_h^n, \partial_x \mathbf{u}_h^{n-1}) \Big\}\end{array}\right. \tag{45}$$

which is a discrete equivalent of the derivation rule for composed functions:

$$\frac{\partial}{\partial t} H(\partial_x \mathbf{u}) = \nabla H(\partial_x \mathbf{u}) \cdot \partial_{xt}^2 \mathbf{u}. \tag{46}$$

Joining (45) to (40) leads to the following lemma:

Lemma 2.1. *If* $\mathbf{u}_h^n$ *is a solution of (34), with* $\mathbb{\nabla} H$ *conservative in the sense of definition 2.1, it satisfies the energy conservation property:*

$$E_h^{n+\frac{1}{2}} = E_h^{n-\frac{1}{2}}, \tag{47}$$

where the discrete energy $E_h^{n+\frac{1}{2}}$ *corresponding to time* $t^{n+\frac{1}{2}} = (n+\frac{1}{2})\Delta t$ *is given by*

$$E_h^{n+\frac{1}{2}} = \frac{1}{2}\oint_\Omega^h \Big|\frac{\mathbf{u}_h^{n+1} - \mathbf{u}_h^n}{\Delta t}\Big|^2 + \oint_\Omega^h \mathbb{H}\big(\partial_x \mathbf{u}_h^{n+1}, \partial_x \mathbf{u}_h^n\big) \tag{48}$$

where $\mathbb{H}$ *is the discrete potential energy associated to* $\mathbb{\nabla} H$ *(see (42)).*

An immediate consequence (we omit the details of the proof) of this lemma is the following corollary:

Corollary 2.2. *If the discrete potential energy $\mathbb{H}$ is positive*

$$\forall (\mathbf{u},\mathbf{v}) \in \mathbb{R}^N, \quad \mathbb{H}(\mathbf{u},\mathbf{v}) \geq 0, \tag{49}$$

then the scheme is unconditionally L^2-stable in the sense that the L^2 norm in space of any solution $\mathbf{u}_h^n$ is uniformly (with respect to h and Δt) bounded.

Remark 2.5. *In the linear case, i.e. $H(\mathbf{v}) = \frac{1}{2}\,\mathbf{A}\,\mathbf{v}\cdot\mathbf{v}$, the θ-scheme corresponds to*

$$\left|\begin{array}{l} \mathbb{\nabla} H(\mathbf{u},\mathbf{v},\mathbf{w}) = \mathbf{A}\left(\theta\mathbf{u} + (1-2\theta)\mathbf{v} + \theta\mathbf{w}\right) \\ \mathbb{H}(\mathbf{u},\mathbf{v}) = \dfrac{1}{2}\,\mathbf{A}\,\dfrac{\mathbf{u}+\mathbf{v}}{2}\cdot\dfrac{\mathbf{u}+\mathbf{v}}{2} + \dfrac{1}{2}\,(\theta-\dfrac{1}{4})\mathbf{A}\,(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v}) \end{array}\right.$$

and the positivity property (49) is unconditionally satistied if and only if $\theta \geq \frac{1}{4}$.

Before investigating more elaborate discretizations, let us consider the case of the most naïve scheme to discretize (16), which is the most natural extension to the non linear case of the explicit leap frog scheme ($\theta = 0$) for the linear case : $\forall\, \mathbf{v}_h \in \mathcal{V}_h$,

$$\oint_\Omega^h \frac{\mathbf{u}_h^{n+1} - 2\mathbf{u}_h^n + \mathbf{u}_h^{n-1}}{\Delta t^2}\cdot\mathbf{v}_h + \oint_\Omega^h \nabla H(\partial_x \mathbf{u}_h^n)\cdot\partial_x\mathbf{v}_h = 0. \tag{50}$$

which corresponds to

$$\mathbb{\nabla} H(\mathbf{u},\mathbf{v},\mathbf{w}) = \nabla H(\mathbf{v}). \tag{51}$$

Lemma 2.2. *The explicit scheme (50) is conservative in the sense of the definition 2.1 if and only if the original equation is linear.*

Proof. According to definition 2.1, we look for a discrete potential energy $\mathbb{H}(\mathbf{u}.\mathbf{v})$ such that $\forall\, (\mathbf{u},\mathbf{v},\mathbf{w}) \in (\mathbb{R}^N)^3$

$$\mathbb{H}(\mathbf{u},\mathbf{v}) - \mathbb{H}(\mathbf{v},\mathbf{w}) = F(\mathbf{v})\cdot\frac{(\mathbf{u}-\mathbf{w})}{2}, \quad \text{with } F(\mathbf{v}) := \nabla H(\mathbf{v}). \tag{52}$$

Our objective is to show that if $\mathbb{H}(\mathbf{u},\mathbf{v})$ exists, $F(\mathbf{v}) = \nabla H(\mathbf{v})$ is necessarily linear in $\mathbf{v}$. In what follows we shall denote $\nabla_1\mathbb{H}(\mathbf{u},\mathbf{v})$ (resp $\nabla_2\mathbb{H}(\mathbf{u},\mathbf{v})$) the gradient of the function $\mathbf{u}\mapsto\mathbb{H}(\mathbf{u},\mathbf{v})$ (respectively $\mathbf{v}\mapsto\mathbb{H}(\mathbf{u},\mathbf{v})$).

First, differentiating (52) with respect to $\mathbf{u}$ leads to the identity

$$\nabla_1\mathbb{H}(\mathbf{v},\mathbf{w}) = \frac{1}{2}F(\mathbf{v}), \quad \forall\,(\mathbf{v},\mathbf{w}) \in \mathbb{R}^N\times\mathbb{R}^N. \tag{53}$$

Next, differentiating (52) with respect to $\mathbf{v}$ leads to

$$\nabla_2\mathbb{H}(\mathbf{u},\mathbf{v}) - \nabla_1\mathbb{H}(\mathbf{v},\mathbf{w}) = \frac{1}{2}DF(\mathbf{v})^*(\mathbf{u}-\mathbf{w}), \quad \forall\,(\mathbf{u},\mathbf{v},\mathbf{w}) \in \mathbb{R}^N\times\mathbb{R}^N\times\mathbb{R}^N, \tag{54}$$

where ($DF(\mathbf{u}) \in \mathcal{L}(\mathbb{R}^N)$ denotes the differential of $F(\mathbf{u})$ and $DF(\mathbf{u})^*$ its adjoint with respect the usual inner product in $\mathbb{R}^N$). Using (53) one can write

$$\nabla_2\mathbb{H}(\mathbf{u},\mathbf{v}) = \frac{1}{2}F(\mathbf{w}) + \frac{1}{2}DF(\mathbf{v})^*(\mathbf{u}-\mathbf{w}), \quad \forall\,(\mathbf{u},\mathbf{v},\mathbf{w}) \in \mathbb{R}^N\times\mathbb{R}^N\times\mathbb{R}^N. \tag{55}$$

Finally, we differentiate (55) with respect to $\mathbf{w}$ to obtain

$$0 = \frac{1}{2}DF(\mathbf{w})^* - \frac{1}{2}DF(\mathbf{v})^*, \quad \forall\,(\mathbf{v},\mathbf{w}) \in \mathbb{R}^N\times\mathbb{R}^N. \tag{56}$$

This means that $DF(\mathbf{v})$ is constant in $\mathbf{v}$, i.e. that F is linear, which completes the proof.

□

We have in fact a more general result.

Lemma 2.3. *Let us consider a scheme of the form (34) that is explicit, i.e. $\mathbb{\nabla} H$ is independent of $\mathbf{u}$, and consistent. It is conservative in the sense of the definition 2.1 if and only if ∇H is linear and $\mathbb{\nabla} H(\mathbf{u},\mathbf{v},\mathbf{w}) = \nabla H(\mathbf{v})$.*

Proof. Assume that there exists a discrete potential energy $\mathbb{H}(\mathbf{u}.\mathbf{v})$ such that $\forall\,(\mathbf{u},\mathbf{v},\mathbf{w}) \in (\mathbb{R}^N)^3$

$$\mathbb{H}(\mathbf{u},\mathbf{v}) - \mathbb{H}(\mathbf{v},\mathbf{w}) = F(\mathbf{v},\mathbf{w})\cdot\frac{(\mathbf{u}-\mathbf{w})}{2}, \tag{57}$$

with $F(\mathbf{v}) := \mathbb{\nabla} H(\mathbf{v},\mathbf{w})$.
Differentiating twice (57) once with respect to u the other with respect to $\mathbf{w}$, we get

$$D_2F(\mathbf{v},\mathbf{w}) = 0,$$

where $D_2F(\mathbf{v},\mathbf{w})$ is the differential of the function $\mathbf{w}\mapsto F(\mathbf{v},\mathbf{w})$. This means that $F(\mathbf{v},\mathbf{w})$ does not depend on $\mathbf{w}$, i.e.

$$F(\mathbf{v},\mathbf{w}) \equiv F(\mathbf{v}).$$

The consistency condition then implies $F(\mathbf{v}) = \nabla H(\mathbf{v})$ and we can then use lemma 2.2 to conclude.

□

This last lemma shows that, except in the linear case, conservativity in the sense of definition 2.1 implies that the scheme is **implicit**.

2.2.2. The case of the scalar nonlinear wave equation

We consider the scalar wave equation (N=1), in which case we can write without any ambiguity (we omit the index h for simplicity)

$$\mathbf{u} \equiv u_1, \tag{58}$$

and (9) simply becomes

$$\partial_t^2 u_1 - \partial_x\left[F(\partial_x u_1)\right] = 0, \quad F := H' \quad (\equiv \nabla H), \tag{59}$$

and the scheme (34) may be written $\forall\, v_1 \in \mathcal{V}_h$,

$$\oint_\Omega^h \frac{u_1^{n+1} - 2u_1^n + u_h^{n-1}}{\Delta t^2}\cdot v_1 + \oint_\Omega^h \mathbb{\nabla} H(\partial_x u_1^{n+1}, \partial_x u_1^n, \partial_x u_1^{n-1})\cdot\partial_x v_1 = 0. \tag{60}$$

The conservativity condition (42) is simply

$$\mathbb{\nabla} H(u_1,v_1,w_1)\cdot\frac{(u_1-w_1)}{2}=\mathbb{H}(u_1,v_1)-\mathbb{H}(v_1,w_1),\quad\forall\,(u_1,v_1,w_1)\in(\mathbb{R})^3. \tag{61}$$

We notice that, given a discrete energy function $\mathbb{H}(u_1,v_1):\mathbb{R}^2\mapsto\mathbb{R}$, satisfying the symmetry condition (that we know to be necessary - cf remark 2.4):

$$\forall\,(u_1,v_1)\in\mathbb{R}\times\mathbb{R},\quad \mathbb{H}(u_1,v_1)=\mathbb{H}(v_1,u_1). \tag{62}$$

Condition (61) determines completely $\mathbb{\nabla} H(u_1,v_1,w_1)$ (this is due to the fact that $\mathbb{\nabla} H(u_1,v_1,w_1)$ is real valued, which holds only when $N=1$) as (note that we use the symmetry of $\mathbb{H}$ and that $\mathbb{H}$ is smooth):

$$\mathbb{\nabla} H(u_1,v_1,w_1)=\begin{cases}\dfrac{\mathbb{H}(u_1,v_1)-\mathbb{H}(w_1,v_1)}{u_1-w_1}, & \text{if}\quad u_1\neq w_1,\\ \dfrac{\partial\mathbb{H}}{\partial u_1}(u_1,v_1)\equiv\dfrac{\partial\mathbb{H}}{\partial v_1}(u_1,v_1), & \text{if}\quad u_1=w_1.\end{cases} \tag{63}$$

In fact, given any positive symmetric function $\mathbb{H}(u_1,v_1)$ satisfying the consistency condition

$$\forall v_1\in\mathbb{R},\quad \mathbb{H}(v_1,v_1)=H(v_1), \tag{64}$$

the choice of $\mathbb{\nabla} H(u_1,v_1,w_1)$ given by (63) provides a consistent, energy preserving numerical scheme.

The rest is simply a question of the choice of a positive function $\mathbb{H}$ satisfying both symmetry (62) and consistency (64). The two simplest choices (in our opinion) are

$$\mathbb{H}(u_1,v_1)=\frac{1}{2}\Big\{H(u_1)+H(v_1)\Big\}, \tag{65}$$

$$\mathbb{H}(u_1,v_1)=H\Big(\frac{u_1+v_1}{2}\Big). \tag{66}$$

The choice (65) leads to the scheme

$$\forall\,v_1\in\mathcal{V}_h,\ \oint_\Omega^h\frac{u_1^{n+1}-2u_1^n+u_h^{n-1}}{\Delta t^2}\cdot v_1+\oint_\Omega^h\frac{H(\partial_x u_1^{n+1})-H(\partial_x u_1^{n-1})}{\partial_x u_1^{n+1}-\partial_x u_1^{n-1}}\cdot\partial_x v_1=0, \tag{67}$$

while the choice (66) leads to the scheme

$$\forall\,v_1\in\mathcal{V}_h,\ \oint_\Omega^h\frac{u_1^{n+1}-2u_1^n+u_h^{n-1}}{\Delta t^2}\cdot v_1+\oint_\Omega^h\frac{H(\partial_x u_1^{n+1/2})-H(\partial_x u_1^{n-1/2})}{\partial_x u_1^{n+1/2}-\partial_x u_1^{n-1/2}}\cdot\partial_x v_1=0. \tag{68}$$

where by convention $u_1^{n+1/2}=\dfrac{u_1^{n+1}+u_1^n}{2}$.

Remark 2.6. *The schemes are not rigorously defined because of the presence of the denominators. To be more rigorous, we should introduce, for any function of one variable $\Phi:\mathbb{R}\mapsto\mathbb{R}$, the function of 2 variables:*

$$\delta\Phi(u_1,w_1)=\begin{cases}\dfrac{\Phi(u_1)-\Phi(w_1)}{u_1-w_1}, & \text{if}\quad u_1\neq w_1,\\ \Phi'(w_1), & \text{if}\quad u_1=w_1.\end{cases} \tag{69}$$

and rewrite (67) and (68) as respectively

$$\forall\,v_1\in\mathcal{V}_h,\ \oint_\Omega^h\frac{u_1^{n+1}-2u_1^n+u_h^{n-1}}{\Delta t^2}\cdot v_1+\oint_\Omega^h\delta H(\partial_x u_1^{n+1},\partial_x u_1^{n-1})\cdot\partial_x v_1=0, \tag{70}$$

$$\forall\,v_1\in\mathcal{V}_h,\ \oint_\Omega^h\frac{u_1^{n+1}-2u_1^n+u_h^{n-1}}{\Delta t^2}\cdot v_1+\oint_\Omega^h\delta H(\partial_x u_1^{n+1/2},\partial_x u_1^{n-1/2})\cdot\partial_x v_1=0. \tag{71}$$

Remark 2.7. *The scheme (67) is found in several publications in the scalar case ([12, 16, 23, 25, 31]).*

The reader will easily check that in the linear case, the scheme (67) gives the θ-scheme with $\theta=1/2$ while the scheme (68) gives the θ-scheme with $\theta=1/4$. Other θ-schemes can be recovered in the linear case by choosing for the discrete energy $\mathbb{H}$ an appropriate linear combination of the two functions (67) and (68).

2.2.3. A class of partially decoupled implicit schemes

We come back to the general case of systems, i.e. $N\geq 1$ and set $\mathbf{u}_h=(u_1,u_2,\cdots,u_N)$. If we look at the equation number ℓ of the system (9), it may be written:

$$\partial_t^2 u_\ell-\partial_x\left[\partial_\ell H(\partial_x u_\ell,\partial_x u_{j\neq\ell})\right]=0, \tag{72}$$

where $\partial_\ell H$ denotes the derivative of H with respect to its ℓ^{th} variable and by convention, for any $\mathbf{v}=(v_j)\in\mathbb{R}^N$,

$$v_{j\neq\ell}:=(v_1,\cdots,v_{\ell-1},v_{\ell-1},\cdots,v_N) \tag{73}$$

so that

$$\mathbf{v}=(v_j)\equiv(v_\ell,v_{j\neq\ell}).$$

Assuming that $u_{j\neq\ell}$ is known, this is for u_ℓ a 1D equation very similar to (59). Thus, the most natural generalization of the scheme (70) is (with $\mathcal{V}_h=\prod\mathcal{V}_\ell$)

$$\oint_\Omega^h\frac{u_\ell^{n+1}-2u_\ell^n+u_\ell^{n-1}}{\Delta t^2}\cdot v_\ell+\oint_\Omega^h\delta_\ell H(\partial_x u_\ell^{n+1},\partial_x u_\ell^{n-1};\partial_x u_{j\neq\ell}^n)\cdot\partial_x v_\ell=0,\quad\forall\,v_\ell\in\mathcal{V}_\ell,\quad\ell=1,\cdots,N, \tag{74}$$

where we have introduced a new notation for the multidimensional generalization of (69) : to any scalar function of N variables $\Phi(v_1,\cdots,v_N)$, we associate the function of $N+1$ variables (with notations similar to the previous ones - we omit the details):

$$\delta_\ell\Phi(u_\ell,w_\ell;v_{j\neq\ell})=\begin{cases}\dfrac{\Phi(u_\ell,v_{j\neq\ell})-\Phi(w_\ell,v_{j\neq\ell})}{u_\ell-w_\ell}, & \text{if}\quad u_\ell\neq w_\ell,\\ \partial_\ell\Phi(w_\ell,v_{j\neq\ell}), & \text{if}\quad u_\ell=w_\ell.\end{cases}$$

which satisfies in particular

$$\delta_\ell\Phi(u_\ell,w_\ell;v_{j\neq\ell})\,(u_\ell-w_\ell)=\Phi(u_\ell,v_{j\neq\ell})-\Phi(w_\ell,v_{j\neq\ell}). \tag{75}$$

The scheme (74) is clearly of the form (34) with

$$\mathbb{V}H(\mathbf{u},\mathbf{v},\mathbf{w}) := \big(\ \delta_\ell H(u_\ell, w_\ell; v_{j\neq\ell})\ \big)_{1\leq\ell\leq N}. \tag{76}$$

The question is : is this scheme energy preserving in the sense of definition 2.1 ?

In other words, we wish to know if it is always possible to find a function $\mathbb{H} : (\mathbb{R}^N)^2 \to \mathbb{R}$ such that for any $(\mathbf{u}, \mathbf{v}, \mathbf{w})$,

$$\mathbb{H}(\mathbf{u},\mathbf{v}) - \mathbb{H}(\mathbf{v},\mathbf{w}) = \frac{1}{2}\sum_{k=1}^{N}\left[H(u_k; v_{l\neq k}) - H(w_k; v_{l\neq k})\right] \tag{77}$$

We will answer this question gradually in N.

Lemma 2.4. *Assume $N = 1$. The approximate gradient defined by (76) is conservative in the sense of definition 2.1, and the only discrete potential energy $\mathbb{H}$ that satisfies the consistency property (44) is written*

$$\mathbb{H}(u_1, v_1) = \frac{1}{2}\left[H(u_1) + H(v_1)\right] \quad . \tag{78}$$

Proof. When $N = 1$, the problem (77) becomes : find a function $\mathbb{H} : \mathbb{R}^2 \to \mathbb{R}$ such that for any $(u_1, v_1, w_1) \in \mathbb{R}^3$:

$$\mathbb{H}(u_1, v_1) - \mathbb{H}(v_1, w_1) = \frac{H(u_1) - H(w_1)}{2}. \tag{79}$$

Let us assume that such a function exists and differentiate (79) with respect to u_1. We get:

$$\partial_1\mathbb{H}(u_1, v_1) = \frac{1}{2}H'(u_1),$$

which implies the existence of $\phi : \mathbb{R} \to \mathbb{R}$ such that

$$\mathbb{H}(u_1, v_1) = \frac{1}{2}H(u_1) + \phi(v_1) \quad . \tag{80}$$

Substituting (2.2.3) into (79), we get : $\forall\,(u_1, v_1, w_1) \in \mathbb{R}^3$,

$$\left[\frac{1}{2}H(u_1) + \phi(v_1)\right] - \left[\frac{1}{2}H(v_1) + \phi(w_1)\right] = \frac{H(u_1) - H(w_1)}{2}.$$

After differentiation in v_1, we get

$$\phi'(v_1) = \frac{1}{2}H'(v_1) \quad \Rightarrow \quad \phi(y) = \frac{1}{2}H(y) + c, \quad c \in \mathbb{R},$$

which yields

$$\mathbb{H}(u_1, v_1) = \frac{1}{2}\left[H(u_1) + H(v_1)\right] + c \quad .$$

It is easy to verify that (79) is satisfied, and (44) is satisfied if and only if $c = 0$.

□

Lemma 2.5. *Assume $N = 2$. The approximate gradient defined by (76) is conservative in the sense of definition 2.1, and the only discrete potential energy $\mathbb{H}$ that satisfies the consistency property (44) is written*

$$\mathbb{H}(u_1, u_2; v_1, v_2) = \frac{1}{2}\left[H(u_1, v_2) + H(v_1, u_2)\right]. \tag{81}$$

Proof. The problem (77) becomes : find a function $\mathbb{H} : \mathbb{R}^4 \to \mathbb{R}$ such that, for any $(u_1, u_2, v_1, v_2, w_1, w_2) \in \mathbb{R}^6$,

$$\mathbb{H}(u_1, u_2; v_1, v_2) - \mathbb{H}(v_1, v_2; w_1, w_2) = \frac{1}{2}\Big\{H(u_1, v_2) - H(w_1, v_2) + H(v_1, u_2) - H(v_1, w_2)\Big\}.$$

It is clear that $\mathbb{H}$ given by (81) is a solution. Let us show that $\mathbb{H}$ is necessarily of this form.

We first take $u_2 = v_2 = w_2 = \lambda$:

$$\mathbb{H}(u_1, \lambda; v_1, \lambda) - \mathbb{H}(v_1, \lambda; w_1, \lambda) = \frac{H(u_1, \lambda) - H(w_1, \lambda)}{2}.$$

The problem is reduced to a problem similar to (79) where λ plays the role of a parameter. By Lemma 2.4, we know that there exists a scalar 1D function $c_1(\lambda)$ such that

$$\mathbb{H}(u_1, \lambda; v_1, \lambda) = \frac{H(u_1, \lambda) + H(v_1, \lambda)}{2} + c_1(\lambda). \tag{82}$$

In the same way, taking $u_1 = v_1 = w_1 = \mu$ in (82), we deduce that

$$\mathbb{H}(\mu, u_2; \mu, v_2) = \frac{H(\mu, u_2) + H(\mu, v_2)}{2} + c_2(\mu). \tag{83}$$

Thus taking $u_1 = v_1 = \mu$ in (82) and $u_2 = v_2 = \lambda$ in (83), we find that $c_1(\lambda) = c_2(\mu)$, $\forall\ (\lambda, \mu) \in \mathbb{R}^2$, that is to say

$$c_1(\lambda) = c_2(\mu) = c \in \mathbb{R}, \tag{84}$$

The identity (82), combined with (84), implies that, (u_1, v_1) being fixed, we know (up to an additive constant) the function $(u_2, v_2) \mapsto \mathbb{H}(u_1, u_2; v_1, v_2)$ along the diagonal $u_2 = v_2$. To determine completely this function, it suffices a priori to establish a first order ODE satisfied by $\mathbb{H}(u_1, u_2; v_1, v_2)$. This can be done by differentiating (82) with respect to u_2:

$$\frac{\partial\mathbb{H}}{\partial u_2}(u_1, u_2; v_1, v_2) = \frac{1}{2}\,\partial_2 H(v_1, u_2). \tag{85}$$

Reintegrating (85) between u_2 and v_2 leads to

$$\mathbb{H}(u_1, u_2; v_1, v_2) = \mathbb{H}(u_1, v_2; v_1, v_2) + \frac{1}{2}\left[H(v_1, u_2) - H(v_1, v_2)\right].$$

Using (82) and (84), we get

$$\mathbb{H}(u_1, u_2; v_1, v_2) = \frac{1}{2}\left[H(u_1, v_2) + H(v_1, u_2)\right] + c. \tag{86}$$

The consistency condition (44) yields $c = 0$, i.e. (81).

□

Unfortunately, the extension of these results to dimensions greater than 2 is not possible, except for very special potential energies H (the sums of functions of two variables). Let us state a precise result:

Lemma 2.6. *Assume $N \geq 2$. The approximate gradient defined by (76) is conservative in the sense of definition 2.1, i.e. there exists a discrete potential energy $\mathbb{H}$ such that (77) is satisfied, if and only if*

$$\exists \left\{ H_{ij} : \mathbb{R}^2 \mapsto \mathbb{R},\ i<j \right\} \quad / \quad H(\mathbf{v}) = \sum_{i<j} H_{ij}(v_i, v_j) \tag{87}$$

or equivalently

$$\forall\, p<q<r, \quad \partial^3_{pqr}\, H(\mathbf{v}) = 0. \tag{88}$$

In this case the only discrete potential energy $\mathbb{H}$ that satisfies the consistency property (44) is given by

$$\mathbb{H}(\mathbf{u},\mathbf{v}) = \frac{1}{2} \sum_{i<j} \{H_{ij}(u_i,v_j) + H_{ij}(v_i,u_j)\} \tag{89}$$

Proof. The easy part of the proof is to check that if H is given by (87), (77) is satisfied with the discrete potential energy (89). The reader can refer to [7] for this point of the proof as well as a more detailed version of what follows.

It is less immediate that the existence of $\mathbb{H}$ satisfying (77) implies (87) and (89). We shall show this in three steps.

Step 1 : we prove that for any N we have

$$\big(\mathcal{P}_N\big) \begin{cases} \text{The existence of } \mathbb{H} \text{ satisfying (77) implies that} \\ \mathbb{H}\big(\mathbf{u};\mathbf{v}\big) = \frac{1}{2}\Big[\sum_{k=1}^{N} H(u_k,(v_l)_{l\neq k}) - (N-2)H\big(\mathbf{v}\big)\Big] + c. \end{cases}$$

The proof is by induction on N. First note that for $N=1$ (resp. $N=2$), $\big(\mathcal{P}_N\big)$ is nothing but Lemma 2.4 (resp. Lemma 2.5).

Let us assume that $\big(\mathcal{P}_N\big)$ is true and let us prove $\big(\mathcal{P}_{N+1}\big)$.

We start from (77) written with $N+1$ variables, i.e.

$$\mathbb{H}(\mathbf{u},\mathbf{v}) - \mathbb{H}(\mathbf{v},\mathbf{w}) = \frac{1}{2}\sum_{k=1}^{N+1} \left[H(u_k;v_{l\neq k}) - H(w_k;v_{l\neq k})\right]. \tag{90}$$

We follow the proof of lemma 2.5. Let us choose in (90) $(\mathbf{u},\mathbf{v},\mathbf{w})$ such that $u_{N+1} = v_{N+1} = w_{N+1} = \lambda$. In this case, the last term in the sum of the right hand side vanishes and we have, (setting $\mathbf{u}^N = (u_1,\cdots,u_N) \in \mathbb{R}^N$ and $\mathbf{v}^N = (v_1,\cdots,v_N) \in \mathbb{R}^N$)

$$\begin{aligned} &\mathbb{H}_\lambda\big(\mathbf{u}^N,\mathbf{v}^N\big) - \mathbb{H}_\lambda\big(\mathbf{u}^N,\mathbf{v}^N\big) \ = \\ &\qquad = \sum_{k=1}^{N} \Big[H_\lambda(u_k,(v_l)_{l\neq k}) - H_\lambda(w_k,(v_l)_{l\neq k})\Big], \end{aligned} \tag{91}$$

where we have set for any $(\mathbf{u}^N,\mathbf{v}^N) \in \mathbb{R}^N$ and $\lambda \in \mathbb{R}$

$$\mathbb{H}_\lambda\big(\mathbf{u}^N,\mathbf{v}^N\big) := \mathbb{H}\big((\mathbf{u}^N,\lambda),(\mathbf{v}^N,\lambda)\big), \quad H_\lambda\big(\mathbf{u}^N\big) = H_\lambda\big(\mathbf{u}^N,\lambda\big).$$

In (91) we recognize (77) for $(\mathbb{H}_\lambda, H_\lambda)$. Thus applying $\big(\mathcal{P}_N\big)$ we deduce that, for some $c_{N+1} : \mathbb{R} \to \mathbb{R}$,

$$\begin{aligned} &\mathbb{H}\big((u_k)_{1\leq k\leq N},\lambda;\,(v_k)_{1\leq k\leq N},\lambda\big) = c_{N+1}(\lambda) + \\ &\frac{1}{2}\Big[\sum_{k=1}^{N} H(u_k,(v_l)_{l\neq k},\lambda) - (N-2)H\big((v_k)_{1\leq k\leq N},\lambda\big)\Big]. \end{aligned} \tag{92}$$

Using the same argument after having chosen $x_i = y_i = z_i = \lambda_i$ for $i = 1,\dots,N$, we deduce the existence of $c_i : \mathbb{R} \to \mathbb{R}$ such that, with obvious notations

$$\begin{aligned} &\mathbb{H}\big((u_k)_{k\neq i},\lambda\,;\,(v_k)_{k\neq i},\lambda_i\big) = c_i(\lambda) + \\ &\frac{1}{2}\Big[\sum_{k=1}^{N} H(u_k,(v_l)_{l\notin\{i,k\}},\lambda_i) - (N-2)H\big((v_k)_{k\neq i},\lambda\big)\Big]. \end{aligned} \tag{93}$$

Taking $x_k = y_k = z_k = \lambda_k$ and $\lambda = \lambda_i$ in (93), we obtain

$$\forall \lambda = (\lambda_\ell)_\ell \in \mathbb{R}^{N+1}, \quad c_i(\lambda_i) = c_j(\lambda_j), \qquad \text{for } i \neq j,$$

which implies the existence of a constant c such that

$$c_i(\lambda_i) = c, \qquad \text{for any } 1 \leq i \leq N+1.$$

Thus, going back to (92) with $\lambda = v_{N+1}$, we can write

$$\begin{aligned} &\mathbb{H}\big(\mathbf{u}^N, v_{N+1};\mathbf{v}\big) = c + \\ &\qquad \frac{1}{2}\Big[\sum_{k=1}^{N} H(u_k,(v_l)_{l\neq k}) - (N-2)H\big(\mathbf{v}\big)\Big]. \end{aligned} \tag{94}$$

We now differentiate (90) with respect to u_{N+1} to get

$$\frac{\partial \mathbb{H}}{\partial u_{N+1}}(\mathbf{u},\mathbf{v}) = \frac{1}{2}\,\partial_{N+1} H\big(\mathbf{v}^N, u_{N+1}\big) \tag{95}$$

that we reintegrate (in u_{N+1} again) between v_{N+1} and u_{N+1} to obtain

$$\begin{aligned} \mathbb{H}\big(\mathbf{u}^N, u_{N+1};\mathbf{v}\big) &= \mathbb{H}\big(\mathbf{u}^N, v_{N+1};\mathbf{v}\big) \\ &\quad + \frac{1}{2}\Big[H\big(\mathbf{v}^N,u_{N+1}\big) - H\big(\mathbf{v}^N,v_{N+1}\big)\Big]. \end{aligned}$$

We can use (94) for replacing $\mathbb{H}\big(\mathbf{u}^N, v_{N+1};\mathbf{v}\big)$ and obtain

$$\begin{aligned} \mathbb{H}(\mathbf{u};\mathbf{v}) = c \ &+ \ \frac{1}{2}\Big[H\big(\mathbf{v}^N,u_{N+1}\big) - H\big(\mathbf{v}\big)\Big] \\ &+ \frac{1}{2}\Big[\sum_{k=1}^{N} H(u_k,(v_l)_{l\neq k},\lambda) - (N-2)H\big(\mathbf{v}\big)\Big] \end{aligned}$$

which is nothing but $\big(\mathcal{P}_{N+1}\big)$.

Step 2 : we show that $\big(\mathcal{P}_N\big)$ and (77) imply (87).

We substitute the expression for $\mathbb{H}$ given by $\big(\mathcal{P}_N\big)$ into (77):

$$\begin{aligned} &\frac{1}{2}\Big[\sum_{k=1}^{N} H(u_k,(v_l)_{l\neq k}) - (N-2)H(\mathbf{v})\Big] \\ &\quad -\frac{1}{2}\Big[\sum_{k=1}^{N} H\big(v_k,(w_l)_{l\neq k}\big) - (N-2)H(\mathbf{w})\Big] \\ &\quad = \frac{1}{2}\sum_{k=1}^{N}\left[H\big(u_k;(v_l)_{l\neq k}\big) - H\big(w_k;(v_l)_{l\neq k}\big)\right] \end{aligned}$$

which leads to

$$\begin{aligned} &(N-2)\Big[H(\mathbf{v}) - H(\mathbf{w})\Big] = \\ &\qquad \sum_{k=1}^{N}\Big[H\big(w_k,(v_l)_{l\neq k}\big) - H\big(v_k,(w_l)_{l\neq k}\big)\Big]. \end{aligned}$$

Let us differentiate the above equality with respect to v_p:

$$(N-2)\,\partial_p H(\mathbf{v}) = \sum_{k\neq p} \partial_p H\big(w_k, (v_l)_{l\neq k}\big) - \partial_p H\big(v_p, (w_l)_{l\neq p}\big).$$

Differentiating again with respect to w_q with $q \neq p$, we get

$$\partial^2_{pq} H(w_q, (v_l)_{l\neq p}) = \partial^2_{pq} H(v_p, (w_l)_{l\neq q}).$$

Finally, differentiating in v_r with $r \notin \{p, q\}$ we get

$$\partial^3_{pqr} H(w_q, (v_l)_{l\neq p}) = 0,$$

which yields (88), and thus (87), since $\mathbf{v}$ and $\mathbf{w}$ are arbitrary.

Step 3: we show that (87) implies (89).

The reader will easily check that, using (87)

$$\begin{aligned}\sum_{k=1}^{N} H\big(u_k, (v_l)_{l\neq k}\big) \quad &= \sum_{i<j} \Big\{ H_{ij}(u_i, v_j) + H_{ij}(v_i, u_j) \Big\} \\ &+ \sum_{i<j} \sum_{k\notin\{i,j\}} H_{ij}\big(v_i, v_j\big).\end{aligned}$$

Substituting this formula into the expression for $\mathbb{H}$ given by $(\mathcal{P}_N)$, we get

$$\begin{aligned}\mathbb{H}(\mathbf{u}, \mathbf{v}) \quad &= \frac{1}{2} \sum_{i<j} \Big\{ H_{ij}(u_i, v_j) + H_{ij}(v_i, u_j) \Big\} \\ &+ \frac{1}{2} \sum_{i<j} \sum_{k\notin\{i,j\}} H_{ij}\big(v_i, v_j\big) \\ &- \frac{N-2}{2} \sum_{i<j} H_{ij}\big(v_i, v_j\big) + c.\end{aligned}$$

Then it suffices to remark that, by permutation of sums

$$\sum_{i<j} \sum_{k\notin\{i,j\}} H_{ij}\big(v_i, v_j\big) = (N-2) \sum_{i<j} H_{ij}\big(v_i, v_j\big)$$

and to use the consistency condition (44) to get $c = 0$. □

This lemma shows that, except for very particular cases, energy preservation in the sense of definition 2.1 is only possible for **fully implicit schemes**.

2.2.4. Construction of fully implicit preserving schemes

Following the same idea as in the previous section, we can first look for schemes of the form

$$\begin{aligned}&\oint_\Omega^h \frac{u_\ell^{n+1} - 2u_\ell^n + u_\ell^{n-1}}{\Delta t^2} \cdot v_\ell + \\ &\oint_\Omega^h \delta_\ell H(\partial_x u_\ell^{n+1}, \partial_x u_\ell^{n-1}; \partial_x [\![u_j^n]\!]^\ell_{j\neq\ell}) \cdot \partial_x v_\ell = 0, \qquad (96) \\ &\qquad\qquad \forall\, v_\ell \in \mathcal{V}_\ell, \quad \ell = 1, \cdots, N,\end{aligned}$$

where $[\![u_j^n]\!]^\ell$ represents some approximation of $u_j(t^n)$, to be determined, using u_j^{n+1}, u_j^n and u_j^{n+1} and that we authorize to depend on ℓ. The decoupled schemes of section 2.2.3 correspond to $[\![u_j^n]\!] = u_j^n$. We need another choice to ensure conservativity for any n. Let us consider

$$[\![u_j^n]\!]^\ell = u_j^{n+sg(\ell-j)}, \qquad (97)$$

where $sg(\cdot)$ is the usual sign function.

The interpretation of this choice is the following (the principle is similar to the one of the Gauss-Seidel algorithm for linear systems) : in the ℓ^{th} equation of (96), associated to the component u_ℓ, the other components u_j are evaluated at time t^{n-1} if $j < \ell$ and at time t^{n+1} if $j > \ell$. As an illustration, for $N = 2$ our scheme may be written

$$\left\{\begin{aligned}&\oint_\Omega^h \frac{u_1^{n+1} - 2u_1^n + u_1^{n-1}}{\Delta t^2} \cdot v_1 + \\ &\oint_\Omega^h \frac{H(\partial_x u_1^{n+1}, \partial_x u_2^{n-1}) - H(\partial_x u_1^{n-1}, \partial_x u_2^{n-1})}{\partial_x u_1^{n+1} - \partial_x u_1^{n-1}} \cdot \partial_x v_1 = 0, \\ &\qquad\qquad \forall\, v_1 \in \mathcal{V}_1, \\ &\oint_\Omega^h \frac{u_2^{n+1} - 2u_2^n + u_2^{n-1}}{\Delta t^2} \cdot v_2 + \\ &\oint_\Omega^h \frac{H(\partial_x u_1^{n+1}, \partial_x u_2^{n+1}) - H(\partial_x u_1^{n+1}, \partial_x u_2^{n-1})}{\partial_x u_2^{n+1} - \partial_x u_2^{n-1}} \cdot \partial_x v_2 = 0, \\ &\qquad\qquad \forall\, v_2 \in \mathcal{V}_2.\end{aligned}\right. \qquad (98)$$

With the choice (97), (96) enters the class (34) with

$$\Big[\mathbb{\nabla} H(\mathbf{u}, \mathbf{v}, \mathbf{w})\Big]_\ell := \delta_\ell H\big(u_\ell, w_\ell; \big[\beta_{j\ell} u_j + (1-\beta_{j\ell}) w_j\big]_{j\neq\ell}\big) \qquad (99)$$

where by definition $\beta_{j\ell} = 1$ if $j < \ell$, $\quad 0$ if $j > \ell$.

Lemma 2.7. *The approximate gradient (99) is conservative in the sense of definition 2.1 and the associate discrete potential energy is given by:*

$$\mathbb{H}(\mathbf{u}, \mathbf{v}) = \frac{1}{2} \Big\{ H(\mathbf{u}) + H(\mathbf{v}) \Big\}. \qquad (100)$$

Proof. Using (99), we have

$$\begin{aligned}&\mathbb{\nabla} H(\mathbf{u}, \mathbf{v}, \mathbf{w}) \cdot \frac{(\mathbf{u} - \mathbf{w})}{2} = \\ &\frac{1}{2} \sum_\ell \big(\delta_\ell H(u_\ell, w_\ell; \big[\beta_{j\ell} u_j + (1-\beta_{j\ell}) w_j \big]_{j\neq\ell}) \big) (u_\ell - w_\ell).\end{aligned}$$

Using (75) and the definition of $\beta_{j\ell}$, we can write

$$\begin{aligned}\mathbb{\nabla} H(\mathbf{u}, \mathbf{v}, \mathbf{w}) \cdot \frac{(\mathbf{u} - \mathbf{w})}{2} = \frac{1}{2} \Big[\sum_\ell &H(u_1, \cdots, u_\ell, w_{\ell+1}, \cdots w_N) \\ &- H(u_1, \cdots, u_{\ell-1}, w_\ell, \cdots w_N) \Big]\end{aligned}$$

As this is a telescopic sum, this results in

$$\begin{aligned}&\mathbb{\nabla} H(\mathbf{u}, \mathbf{v}, \mathbf{w}) \cdot \frac{(\mathbf{u} - \mathbf{w})}{2} = \\ &\frac{1}{2} H(u_1, \cdots, u_N) - \frac{1}{2} H(w_1, \cdots, w_N) \equiv \mathbb{H}(\mathbf{u}, \mathbf{v}) - \mathbb{H}(\mathbf{v}, \mathbf{w}).\end{aligned}$$

which completes the proof. □

Remark 2.8. *The reader will notice that, at a given time t^{n+1}, if the equations are solved one by one when increasing the value of ℓ, the calculation of each component can be done in a decoupled way as in the scheme (74) : one has to solve a triangular non linear system instead of a diagonal non linear system with (74).*

As emphasized by remark 2.8, the numbering of the equations of (9) (or equivalently the ranking of the components of $\mathbf{u}$) has an influence on the resulting scheme constructed with the above procedure. In other words, to any permutation $p \in \mathcal{S}_N$, where $\mathcal{S}_N$ is the group of permutations of $\{1, \cdots, N\}$, we can associate a scheme of the same nature as (96,74) namely

$$\oint_\Omega^h \frac{u_\ell^{n+1} - 2u_\ell^n + u_\ell^{n-1}}{\Delta t^2} \cdot v_\ell + \oint_\Omega^h \delta_\ell H(\partial_x u_\ell^{n+1}, \partial_x u_\ell^{n-1}; \partial_x u_{j\neq\ell}^{n+sg(p(\ell)-p(j))}) \cdot \partial_x v_\ell = 0, \quad (101)$$
$$\forall\, v_\ell \in \mathcal{V}_\ell, \quad \ell = 1, \cdots, N,$$

which corresponds to the discrete gradient

$$\mathbb{V}(\mathbf{u}, \mathbf{v}, \mathbf{w}) \equiv \mathbb{V}^{(p)}(\mathbf{u}, \mathbf{v}, \mathbf{w}) \ , \quad (102)$$

$$\Big[\mathbb{V}^{(p)}(\mathbf{u}, \mathbf{v}, \mathbf{w})\Big]_\ell := \delta_\ell H(u_\ell, w_\ell; \big[\, \beta_{j\ell}^{(p)} u_j + (1 - \beta_{j\ell}^{(p)}) w_j \,\big]_{j\neq\ell}),$$

where by definition

$$\beta_{j\ell}^{(p)} = 1 \text{ if } p(j) < p(\ell), \quad 0 \text{ if } p(j) > p(\ell). \quad (103)$$

Any of these schemes will preserve the same energy (48) with $\mathbb{H}$ given by (100).

As an illustration, for $N = 2$, we obtain a scheme that differs from (98) by exchanging the roles of the indices 1 and 2:

$$\begin{cases} \displaystyle\oint_\Omega^h \frac{u_1^{n+1} - 2u_1^n + u_1^{n-1}}{\Delta t^2} \cdot v_1 + \oint_\Omega^h \frac{H(\partial_x u_1^{n+1}, \partial_x u_2^{n+1}) - H(\partial_x u_1^{n-1}, \partial_x u_2^{n+1})}{\partial_x u_1^{n+1} - \partial_x u_1^{n-1}} \cdot \partial_x v_1 = 0, \\ \qquad\qquad \forall\, v_1 \in \mathcal{V}_1, \\ \displaystyle\oint_\Omega^h \frac{u_2^{n+1} - 2u_2^n + u_2^{n-1}}{\Delta t^2} \cdot v_2 + \oint_\Omega^h \frac{H(\partial_x u_1^{n-1}, \partial_x u_2^{n+1}) - H(\partial_x u_1^{n-1}, \partial_x u_2^{n-1})}{\partial_x u_2^{n+1} - \partial_x u_2^{n-1}} \cdot \partial_x v_2 = 0, \\ \qquad\qquad \forall\, v_2 \in \mathcal{V}_2. \end{cases} \quad (104)$$

A major drawback to these schemes is that neither of them is centered in time (the property (36) is not satisfied) and they are consequently only first order accurate in time. From this point of view, they can not be considered as a generalization of the θ-schemes.

To restore the second order accuracy which is valid for the scalar case, the idea is to take the "average" of these schemes, by choosing the average approximate gradient. For instance, in the case $N = 2$, the "average" of the schemes (98) and (104) is

$$\begin{cases} \displaystyle\oint_\Omega^h \frac{u_1^{n+1} - 2u_1^n + u_1^{n-1}}{\Delta t^2} \cdot v_1 + \frac{1}{2}\oint_\Omega^h \frac{H(\partial_x u_1^{n+1}, \partial_x u_2^{n+1}) - H(\partial_x u_1^{n-1}, \partial_x u_2^{n+1})}{\partial_x u_1^{n+1} - \partial_x u_1^{n-1}} \cdot \partial_x v_1 \\ \displaystyle\quad + \frac{1}{2}\oint_\Omega^h \frac{H(\partial_x u_1^{n+1}, \partial_x u_2^{n-1}) - H(\partial_x u_1^{n-1}, \partial_x u_2^{n-1})}{\partial_x u_1^{n+1} - \partial_x u_1^{n-1}} \cdot \partial_x v_1 = 0, \\ \qquad\qquad \forall\, v_1 \in \mathcal{V}_1, \\ \displaystyle\oint_\Omega^h \frac{u_2^{n+1} - 2u_2^n + u_2^{n-1}}{\Delta t^2} \cdot v_2 + \frac{1}{2}\oint_\Omega^h \frac{H(\partial_x u_1^{n+1}, \partial_x u_2^{n+1}) - H(\partial_x u_1^{n+1}, \partial_x u_2^{n-1})}{\partial_x u_2^{n+1} - \partial_x u_2^{n-1}} \cdot \partial_x v_2 \\ \displaystyle\quad + \frac{1}{2}\oint_\Omega^h \frac{H(\partial_x u_1^{n-1}, \partial_x u_2^{n+1}) - H(\partial_x u_1^{n-1}, \partial_x u_2^{n-1})}{\partial_x u_2^{n+1} - \partial_x u_2^{n-1}} \cdot \partial_x v_2 = 0, \\ \qquad\qquad \forall\, v_2 \in \mathcal{V}_2. \end{cases}$$

The generalization to any N consists in taking

$$\mathbb{V}H(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \frac{1}{N!} \sum_{p \in \mathcal{S}_N} \mathbb{V}H^{(p)}(\mathbf{u}, \mathbf{v}, \mathbf{w}). \quad (105)$$

Clearly, this discrete gradient is still conservative with the discrete energy given by (100). Moreover, it is centered in time (and then second order accurate) because (36) is satisfied. To check (36) easily we introduce the bijection $\mathcal{I}$ from $\mathcal{S}_N$ into itself

$$p \mapsto q = \mathcal{I}(p) \quad \text{such that} \quad q(j) = p(N + 1 - j). \quad (106)$$

Then, according to (105) and (102), we write

$$\Big[\mathbb{V}H(\mathbf{w}, \mathbf{v}, \mathbf{u})\Big]_\ell = \frac{1}{N!} \sum_{p \in \mathcal{S}_N} \big(\, \delta_\ell H(w_\ell, u_\ell; \big[\, \beta_{j\ell}^{(p)} w_j + (1 - \beta_{j\ell}^{(p)}) u_j \,\big]_{j\neq\ell}) \,\big).$$

We write $p = \mathcal{I}(q)$ in the sum (so that q describes $\mathcal{S}_N$ when p describes $\mathcal{S}_N$) and notice that

$$p = \mathcal{I}(q) \Longrightarrow \beta_{j\ell}^{(p)} = 1 - \beta_{j\ell}^{(q)}$$

to write

$$\Big[\mathbb{V}H(\mathbf{w}, \mathbf{v}, \mathbf{u})\Big]_\ell = \frac{1}{N!} \sum_{q \in \mathcal{S}_N} \big(\, \delta_\ell H(w_\ell, u_\ell; \big[\, + (1 - \beta_{j\ell}^{(q)}) w_j + \beta_{j\ell}^{(p)} u_j \,\big]_{j\neq\ell}) \,\big).$$

Since $\delta_\ell H(u_\ell, w_\ell; v_{j\neq\ell})$ is symmetric in (u_ℓ, w_ℓ) we have

$$\Big[\mathbb{V}H(\mathbf{w}, \mathbf{v}, \mathbf{u})\Big]_\ell = = \frac{1}{N!} \sum_{q \in \mathcal{S}_N} \big(\, \delta_\ell H(u_\ell, w_\ell; \big[\, \beta_{j\ell}^{(q)} u_j + (1 - \beta_{j\ell}^{(q)}) w_j + \,\big]_{j\neq\ell}) \,\big) = \Big[\mathbb{V}H(\mathbf{u}, \mathbf{v}, \mathbf{w})\Big]_\ell.$$

In formula (105), $\mathbb{V}H(\mathbf{u},\mathbf{v},\mathbf{w})$ appears as a sum of $N!$ terms. However, we are going to see that, component by component, it can be rewritten as the sum of only 2^{N-1} terms, which will be useful for the practical implementation of the scheme.

To state the result, it is useful to introduce the sets

$$J_\ell = \{1,\dots,N\}\setminus\{\ell\},\quad \ell = 1,\cdots,N,$$

and for each $1\le\ell\le N$, $\Sigma_\ell = \{\sigma : J_\ell \longrightarrow \{+1,-1\}\}$, the set of applications from J_ℓ into $\{+1,-1\}$ (that contains 2^{N-1} elements). Finally to each $\sigma\in\Sigma_\ell$, we associate the integer $\mu(\sigma)$ defined by

$$\mu(\sigma) = \#\left\{l\in J_k,\ \sigma(l)=+1\right\} = \#\sigma^{-1}(+1).$$

Lemma 2.8. *The approximate gradient defined by (105) is also given by*

$$\Big[\mathbb{V}H(\mathbf{u},\mathbf{v},\mathbf{w})\Big]_\ell = \sum_{\sigma\in\Sigma_\ell}\zeta(\sigma)\ \delta_\ell H\big(u_\ell,w_\ell;\langle u_j,w_j\rangle_\sigma\big), \tag{107}$$

where $\zeta(\sigma) = \dfrac{\mu(\sigma)!\,\big(N-1-\mu(\sigma)\big)!}{N!}$ *and*

$$\langle u_j,w_j\rangle_\sigma := \Big(\frac{1+\sigma(j)}{2}\Big)\,u_j + \Big(\frac{1-\sigma(j)}{2}\Big)\,w_j. \tag{108}$$

Proof. Let us introduce the map

$$\begin{array}{ccccc}\Phi_\ell & : & \mathcal{S}_N & \longrightarrow & \Sigma_\ell\\ & & p & \mapsto & \Phi_\ell(p)=\sigma_p^\ell\end{array} \tag{109}$$

where σ_p^ℓ is defined by

$$\forall\, j\in\Sigma_\ell,\quad \sigma_p^\ell(j) = sg\big(p(\ell)-p(j)\big). \tag{110}$$

We can rearrange the sum (105) as:

$$\Big[\ \mathbb{V}H(\mathbf{u},\mathbf{v},\mathbf{w})\ \Big]_\ell = \frac{1}{N!}\sum_{\sigma\in\mathcal{S}_\ell}\ \sum_{p\in\Phi_\ell^{-1}(\sigma)}\Big[\ \mathbb{V}H^{(p)}(\mathbf{u},\mathbf{v},\mathbf{w})\ \Big]_\ell.$$

Next, we remark that, inside each level set $\Phi_\ell^{-1}(\sigma)$ of Φ_ℓ ,

$$\Big[\ \mathbb{V}H^{(p)}(\mathbf{u},\mathbf{v},\mathbf{w})\ \Big]_\ell$$

is independent of p. Indeed, from (103) and (110), we have

$$\beta_{j\ell}^{(p)} = \frac{1+\sigma_p^\ell(j)}{2}\quad\Big(\equiv\frac{1+\sigma(j)}{2}\quad\text{if } p \text{ describes } \Phi_\ell^{-1}(\sigma)\Big).$$

Therefore, using (108), we deduce that for any $p\in\Phi_\ell^{-1}(\sigma)$

$$\Big[\ \mathbb{V}H^{(p)}(\mathbf{u},\mathbf{v},\mathbf{w})\ \Big]_\ell = \delta_\ell H\big(u_\ell,w_\ell;\langle u_j,w_j\rangle_\sigma\big),$$

which yields

$$\Big[\ \mathbb{V}H(\mathbf{u},\mathbf{v},\mathbf{w})\ \Big]_\ell = \sum_{\sigma\in\mathcal{S}_\ell}\Big[\#\Phi_\ell^{-1}(\sigma)\Big]\ \delta_\ell H\big(u_\ell,w_\ell;\langle u_j,w_j\rangle_\sigma\big).$$

To conclude it suffices to show that

$$\#\Phi_\ell^{-1}(\sigma) = \mu(\sigma)!\big(N-1-\mu(\sigma)\big)!\,. \tag{111}$$

Given $\sigma\in\Sigma_\ell$, we set $m=\mu(\sigma)\in\{1,\cdots,N-1\}$ and

$$\mathcal{I}_+ = \{j\in J_\ell\,|\,\sigma(j)=+1\}\quad(\#\mathcal{I}_+ = m)$$
$$\mathcal{I}_- = \{j\in J_\ell\,|\,\sigma(j)=-1\}\quad(\#\mathcal{I}_- = N-1-m).$$

Next, it suffices to remark that, with $\mathcal{B}(E<F)$ denoting the set of bijections between E and F,

$$(i)\ \Phi_\ell(p)=\sigma \iff (ii)\ \begin{cases} p(\ell) = m+1\\ p|_{\mathcal{I}_+}\in\ \mathcal{B}(\mathcal{I}_+;\{1,\cdots,m\})\\ p|_{\mathcal{I}_-}\mathcal{B}(\mathcal{I}_-;\{m+2,\cdots,N\})\end{cases} \tag{112}$$

Indeed, $\Phi_\ell(p)=\sigma$ means that when j describes $\mathcal{I}_+$ (resp. j describes $\mathcal{I}_-$), $p(j)$ takes m values strictly smaller than $p(\ell)$ (resp. $N-1-m$ values strictly greater than $p(\ell)$). As a consequence, the only possibility is $p(\mathcal{I}_+)=\{1,\cdots,m\}$, $p(\ell)=m+1$ and $p(\mathcal{I}_+)=\{m+2,\cdots,N\}$. This proves $(i)\ \Rightarrow\ (ii)$. The inverse statement is obvious.

Then, to count the number of antecedents of σ via Φ_ℓ, from (112), it suffices to multiply the numbers of bijections in a set with m elements by the numbers of bijections in a set with $N-1-m$ elements, which leads to (111).

□

Finally, the equations of the scheme associated to (105) (or (107)) are

$$\begin{aligned}0 = &\oint_\Omega^h \frac{u_\ell^{n+1}-2u_\ell^n+u_\ell^{n-1}}{\Delta t^2}\cdot v_\ell \\ &+\sum_{\sigma\in\Sigma_\ell}\zeta(\sigma)\oint_\Omega^h \delta_\ell H(\partial_x u_\ell^{n+1},\partial_x u_\ell^{n-1};\partial_x u_{\tilde\ell\neq\ell}^{n+\sigma(\tilde\ell)})\cdot\partial_x v_\ell,\\ &\qquad\forall\ v_\ell\in\mathcal{V}_\ell,\quad \ell=1,\dots,N.\end{aligned} \tag{113}$$

The starting procedure must be second order accurate in order to preserve the global accuracy of the scheme. We propose the following formulas, based on the classical starting procedures of finite elements:

$$\begin{cases} u_\ell^0 = \mathcal{I}_h\,u_{0,\ell}\\ \\ u_\ell^1 = \mathcal{I}_h\,u_{0,\ell} + \Delta t\,\mathcal{I}_h\,u_{1,\ell} - \dfrac{\Delta t}{2}\mathcal{I}_h\left(M^{-1}F^\ell(\mathbf{u}_0)\right)\end{cases}$$

where $\mathcal{I}_h$ is the interpolator operator on $\mathcal{V}_\ell$, and for any basis function φ_j of $\mathcal{V}_\ell$, $M_{i,j}=\oint_\Omega^h\varphi_i\varphi_j$ and

$$F_j^\ell(\mathbf{u}) = \sum_{\sigma\in\Sigma_\ell}\zeta(\sigma)\oint_\Omega^h\delta_\ell H(\partial_x u_\ell^{n+1},\partial_x u_\ell^{n-1};\partial_x u_{\tilde\ell\neq\ell}^{n+\sigma(\tilde\ell)})\cdot\partial_x\varphi_j.$$

To give a concrete example of this scheme, it becomes in the particular case $N=3$ the following scheme : Find $(u_1,u_2,u_3)\in\mathcal{V}_h$ such that for any $(v_1,v_2,v_3)\in\mathcal{V}_h$,

$$\begin{cases}\oint_\Omega^h\dfrac{u_1^{n+1}-2u_1^n+u_1^{n-1}}{\Delta t^2}\cdot v_1\\ \quad+\dfrac{1}{6}\Big[\,2\,\delta_1H(\partial_x u_1^{n+1},\partial_x u_1^{n-1};\partial_x u_2^{n+1},\partial_x u_3^{n+1})+\\ \qquad 2\,\delta_1H(\partial_x u_1^{n+1},\partial_x u_1^{n-1};\partial_x u_2^{n-1},\partial_x u_3^{n-1})+\\ \qquad\ \delta_1H(\partial_x u_1^{n+1},\partial_x u_1^{n-1};\partial_x u_2^{n+1},\partial_x u_3^{n-1})+\\ \qquad\ \delta_1H(\partial_x u_1^{n+1},\partial_x u_1^{n-1};\partial_x u_2^{n-1},\partial_x u_3^{n+1})\Big]\,\partial_x v_1 = 0,\end{cases}$$

$$\left\{\begin{array}{l}\displaystyle\oint_\Omega^h \frac{u_2^{n+1}-2u_2^n+u_2^{n-1}}{\Delta t^2}\cdot v_2\\ \quad+\frac{1}{6}\Big[\,2\,\delta_2 H(\partial_x u_2^{n+1},\partial_x u_2^{n-1};\partial_x u_1^{n+1},\partial_x u_3^{n+1})+\\ \qquad 2\,\delta_2 H(\partial_x u_2^{n+1},\partial_x u_2^{n-1};\partial_x u_1^{n-1},\partial_x u_3^{n-1})+\\ \qquad \delta_2 H(\partial_x u_2^{n+1},\partial_x u_2^{n-1};\partial_x u_1^{n+1},\partial_x u_3^{n-1})+\\ \qquad \delta_2 H(\partial_x u_2^{n+1},\partial_x u_2^{n-1};\partial_x u_1^{n-1},\partial_x u_3^{n+1})\Big]\,\partial_x v_2=0,\end{array}\right.$$

$$\left\{\begin{array}{l}\displaystyle\oint_\Omega^h \frac{u_3^{n+1}-2u_3^n+u_3^{n-1}}{\Delta t^2}\cdot v_3\\ \quad+\frac{1}{6}\Big[\,2\,\delta_3 H(\partial_x u_3^{n+1},\partial_x u_3^{n-1};\partial_x u_1^{n+1},\partial_x u_2^{n+1})+\\ \qquad 2\,\delta_3 H(\partial_x u_3^{n+1},\partial_x u_3^{n-1};\partial_x u_1^{n-1},\partial_x u_2^{n-1})+\\ \qquad \delta_3 H(\partial_x u_3^{n+1},\partial_x u_3^{n-1};\partial_x u_1^{n+1},\partial_x u_2^{n-1})+\\ \qquad \delta_3 H(\partial_x u_3^{n+1},\partial_x u_3^{n-1};\partial_x u_1^{n-1},\partial_x u_2^{n+1})\Big]\,\partial_x v_3=0.\end{array}\right.$$

3. Application to the nonlinear string model

3.1. The geometrically exact model and its variants

3.1.1. Establishment of the geometrically exact model

We are interested in the string vibration, for instance a piano string. The problem has been formulated in its nonlinear version in [28], then used and modified by several authors. The geometrically exact model uses an exact geometric description of the movement of the string : this introduces geometric nonlinearity. The most complete model takes into account the three components of the points of the string (which corresponds to $N=3$ with the notation of the previous section, see section 3.1.2) but, for simplicity, we present the model where it is assumed that the movement of the string remains in a plane (x,y) ($N=2$). Figure 1, inspired by [2], presents the unknowns of the problem. What follows is widely based on [33].

We will use the following notation : Ω is a segment of $\mathbb{R}$ or

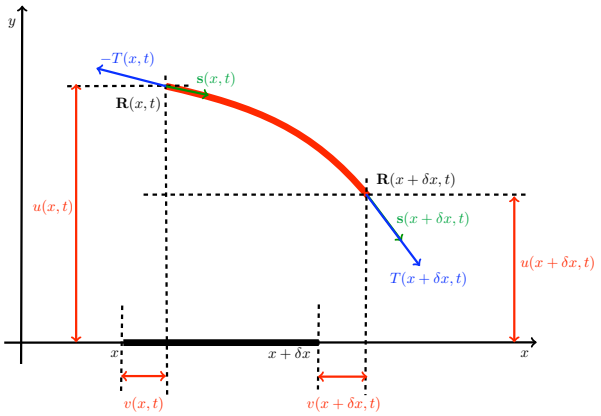

Figure 1: Transversal and longitudinal motions of a string.

$\mathbb{R}$ itself (infinite string), $x\in\Omega$ is the coordinate along the string, and $t>0$ is time. The function $u(x,t)$ indicates the transversal component of the string motion (along $\mathbf{e}_y$), and $v(x,t)$ indicates the longitudinal component of the string motion (along $\mathbf{e}_x$). E is Young's modulus of the string, A is the section area, μ is the lineic mass of the string, and T_0 is its tension at rest. They are supposed to be independent of x and t. The position vector for a point marked by x is

$$\mathbf{R}(x,t)=\big(x+v(x,t)\big)\,\mathbf{e_x}+u(x,t)\,\mathbf{e_y}=x\,\mathbf{e_x}+\mathbf{U}(x,t),$$

where $\mathbf{U}(x,t)$ is the vector of unknowns $\big(v(x,t),u(x,t)\big)$. We apply Newton's second law to a string element $[x,x+\delta x]$. We make the assumptions that the only forces acting on the string come from the tensions $T(x,t)$ and $T(x+\delta x,t)$ of the string at the extreme points of the segment and that these are tangent to the string. If we denote $T(x,t)\in\mathbb{R}$ the tension of the string and $\mathbf{s}(x,t)\in\mathbb{R}^2$ the (oriented) tangent unit vector to the string at point x and time t, we have

$$\frac{1}{\delta x}\Big[\mu\frac{d^2}{dt^2}\Big(\int_x^{x+\delta x}\mathbf{R}(y,\delta x,t)\,dy\Big)\Big]=\frac{1}{\delta x}\big[T(x+\delta x,t)\mathbf{s}(x+\delta x,t)-T(x,t)\mathbf{s}(x,t)\big],$$

hence, taking the limit when $\delta x\to 0$, $\mu\dfrac{\partial^2\mathbf{R}}{\partial t^2}=\dfrac{\partial}{\partial x}\big[T\ \mathbf{s}\big]$. The physical stress-strain relation gives us an expression of the tension T, varying along the string, according to the deformation of the string, namely the relative extension $\delta a(x,t)$. The length of the element at rest is δx and becomes

$$\Big|\mathbf{R}(x+\delta x,t)-\mathbf{R}(x,t)\Big|,$$

hence the relative extension, after a Taylor expansion and neglecting $O(|\delta x|^2)$ is

$$\delta a(x,\delta x,t):=\frac{\Big|\mathbf{R}(x+\delta x,t)-\mathbf{R}(x,t)\Big|-\delta x}{\delta x}\simeq\Big|\frac{\partial\mathbf{R}}{\partial x}\Big|-1.$$

In the general case, the constitutive stress-strain law can be written

$$T(x,t)=\phi\big(\delta a\,(x,t)\,\big)\tag{114}$$

for some function $\phi:\mathbb{R}\to\mathbb{R}$. Defining $d:\mathbb{R}^N\to\mathbb{R}$ and $\mathbf{F}:\mathbb{R}^N\to\mathbb{R}^N$ as:

$$d(\mathbf{v})=\big|\mathbf{e_x}+\mathbf{v}\big|-1\qquad\text{and}\qquad\mathbf{F}\big(\mathbf{v}\big)=\phi\circ d\big(\mathbf{v}\big)\,\frac{\mathbf{e_x}+\mathbf{v}}{\big|\mathbf{e_x}+\mathbf{v}\big|},$$

the mechanical PDE system takes the same form, which can lead to a Hamiltonian form

$$\mu\,\partial_t^2\mathbf{U}-\partial_x\nabla\Big[\Phi\circ d\Big(\partial_x\mathbf{U}\Big)\Big]=0.\tag{115}$$

Setting $H=\Phi\circ d$, we have the nonlinear Hamiltonian systems of wave equations class:

$$\mu\,\partial_t^2\mathbf{U}-\partial_x\Big[\nabla H\big(\partial_x\mathbf{U}\big)\Big]=0.\tag{116}$$

In the case where the stress-strain law is affine, it is called Hooke's law. The constant T_0 is called the prestress of the string, and the law can be written as

$$\phi(\tau)=T_0+EA\,\tau.\tag{117}$$

Projecting the system on the axes, we obtain

$$\begin{cases} \mu\,\partial_t^2 u = \partial_x \left[EA\,\partial_x u - (EA - T_0)\,\dfrac{\partial_x u}{\sqrt{(\partial_x u)^2 + (1+\partial_x v)^2}} \right], \\ \mu\,\partial_t^2 v = \partial_x \left[EA\,\partial_x v - (EA - T_0)\,\dfrac{(1+\partial_x v)}{\sqrt{(\partial_x u)^2 + (1+\partial_x v)^2}} \right]. \end{cases}$$

With appropriate space and time scaling, we can write the following equivalent system, depending on a unique parameter $0 < \alpha < 1$ given by the formula $\alpha = \frac{EA - T_0}{EA}$.

$$\begin{cases} \partial_t^2 u = \partial_x \left[\partial_x u - \alpha\,\dfrac{\partial_x u}{\sqrt{(\partial_x u)^2 + (1+\partial_x v)^2}} \right], \\ \partial_t^2 v = \partial_x \left[\partial_x v - \alpha\,\dfrac{(1+\partial_x v)}{\sqrt{(\partial_x u)^2 + (1+\partial_x v)^2}} \right]. \end{cases} \tag{118}$$

Using the notation $\mathbf{u} = (u, v)$ and $\forall\ (u_x, v_x) \in \mathbb{R}^2$,

$$H_{ex}(u_x, v_x) = \frac{1}{2}\,u_x^2 + \frac{1}{2}\,v_x^2 - \alpha\left[\sqrt{u_x^2 + (1+v_x)^2} - (1+v_x)\right],$$

we can write the string system in the form of a Hamiltonian system of wave equations (NLSWE):

$$\boxed{\partial_t^2 \mathbf{u} - \partial_x\left[\nabla H_{ex}(\partial_x \mathbf{u})\right] = 0} \tag{119}$$

3.1.2. The geometrically exact model with three unknowns

We can generalize the geometrically exact model to the non planar motion of a string, considering two transversal displacements u_1 and u_2, and the longitudinal displacement v. The system of three equations can be derived in the same way as the previous system:

$$\begin{cases} \partial_t^2 u_1 - \partial_x \left[\partial_x u_1 - \alpha\,\dfrac{\partial_x u_1}{\sqrt{(\partial_x u_1)^2 + (\partial_x u_2)^2 + (1+\partial_x v)^2}} \right] = 0, \\ \partial_t^2 u_2 - \partial_x \left[\partial_x u_2 - \alpha\,\dfrac{\partial_x u_2}{\sqrt{(\partial_x u_1)^2 + (\partial_x u_2)^2 + (1+\partial_x v)^2}} \right] = 0, \\ \partial_t^2 v - \partial_x \left[\partial_x v - \alpha\,\dfrac{1+\partial_x v}{\sqrt{(\partial_x u_1)^2 + (\partial_x u_2)^2 + (1+\partial_x v)^2}} \right] = 0. \end{cases}$$

Using the notation $\mathbf{u} = (u_1, u_2, v)$ and

$$H_{ex}(u_1, u_2, v) = \frac{1}{2}\,u_1^2 + \frac{1}{2}\,u_2^2 + \frac{1}{2}\,v^2 - \alpha\left[\sqrt{u_1^2 + u_2^2 + (1+v)^2} - (1+v)\right],$$

the previous system can also be written in a NLSWE form.

3.1.3. Approximations of the geometrically exact model

Under the assumption of small deformations, it is natural to look for approximate models by considering various Taylor expansions of H_{ex} for small values of (u_x, v_x). Let us note that, near the origin,

$$\begin{aligned} H_{ex}(u_x, v_x) = {} & \frac{1-\alpha}{2}\,u_x^2 + \frac{1}{2}\,v_x^2 \\ & + \frac{\alpha}{2}\left[u_x^2 v_x - u_x^2 v_x^2 + \frac{1}{4}\,u_x^4\right] + O(|\mathbf{u}_x|^5). \end{aligned} \tag{120}$$

The linear model. The linear model is obtained by considering only the quadratic terms in (120)

$$H_{ex}(u_x, v_x) \simeq H_{DL2}(u_x, v_x) = \frac{1-\alpha}{2}u_x^2 + \frac{1}{2}\,v_x^2,$$

and we write the classical linear model of two uncoupled wave equations (since $\alpha < 1$):

$$\begin{cases} \partial_t^2 u - (1-\alpha)\,\partial_x^2 u = 0, & x \in \Omega, \quad t > 0, \\ \partial_t^2 v - \partial_x^2 v = 0, & x \in \Omega, \quad t > 0. \end{cases}$$

From this system we can deduce an approximate propagation speed for each direction : 1 for the longitudinal direction, and $\sqrt{1-\alpha}$ for the transversal direction. Since $\alpha \simeq 1$ in real applications, we obtain the well known result in mechanics saying that the longitudinal waves have a propagation speed much higher than the transversal waves. However, with this model, we cannot take into account the coupling between the longitudinal and transverse components of the displacement field, which is an essential characteristic of the behavior of a piano string.

Higher order models. As in [5], it is natural to consider the third and fourth order approximations of H_{ex}, namely

$$\begin{aligned} H_{DL3} &= \frac{1-\alpha}{2}\,u_x^2 + \frac{1}{2}\,v_x^2 + \frac{\alpha}{2}\,u_x^2 v_x\ , \\ H_{DL4} &= \frac{1-\alpha}{2}\,u_x^2 + \frac{1}{2}\,v_x^2 + \frac{\alpha}{2}\left[u_x^2 v_x - u_x^2 v_x^2 + \frac{1}{4}\,u_x^4\right]\ . \end{aligned}$$

However, the corresponding models should be rejected because the fundamental assumption ($\mathcal{H}2$) is not satisfied (the energy is neither positive nor bounded from below). In [1, 2, 5], the authors proposed a less natural model, consisting in neglecting the quartic term $-\frac{\alpha}{2}\,u_x^2 v_x^2$ in (120). This can be fully justified (see [7]) when the string is submitted to transverse solicitations only. This gives the following expression

$$H_{BS} = \frac{1-\alpha}{2}\,u_x^2 + \frac{1}{2}\,v_x^2 + \frac{\alpha}{2}\left[u_x^2 v_x + \frac{1}{4}\,u_x^4\right]$$

leading to the system

$$\begin{cases} \partial_t^2 u = \partial_x\left[(1-\alpha)\,\partial_x u + \alpha\,(\partial_x u\,\partial_x v + \frac{1}{2}(\partial_x u)^3)\right], \\ \partial_t^2 v = \partial_x\left[\partial_x v - \frac{\alpha}{2}(\partial_x u)^2\right]. \end{cases} \tag{121}$$

which will be referred to as the "Bank and Sujbert" model in the following.

3.2. Properties of the string models

3.2.1. Assumptions ($\mathcal{H}1$) to ($\mathcal{H}5$)

In this section, we investigate which of the assumptions ($\mathcal{H}1$) to ($\mathcal{H}5$) are satisfied by the nonlinear models H_{ex} and H_{BS} .

The regularity assumption ($\mathcal{H}1$) is obviously satisfied with H_{BS} and locally around the origin with H_{ex}. One can

show (see [7]) that the coercivity assumption ($\mathcal{H}2$) is satisfied with $K = \frac{1-\alpha}{2}$. Because $\alpha < 1$, both H_{BS} and H_{ex} are locally convex : ($\mathcal{H}3$), which ensures the local hyperbolicity of (118) and (121). However, neither H_{ex} nor H_{BS} is globally convex. It can be shown (see [7] again) that assumption ($\mathcal{H}4$), that ensures a global bound for the propagation velocity of the solution, is satisfied by H_{ex} with $c_+ = 1$. However, it is not satisfied by H_{BS}. For instance, one computes that

$$|\nabla H_{BS}|^2(u_x, 0) = \frac{\alpha^2}{4} u_x^6 + \alpha \left(1 - \frac{3\,\alpha}{4}\right) u_x^4 + (1-\alpha)^2 u_x^2$$

which cannot be bounded by a constant times $H_{BS}(u_x, 0)$, which is a polynomial of degree 4. This difference is due to the fact that while H_{BS} and H_{ex} are close to each other around the origin, they are very different at infinity : H_{BS} grows superlinearly while H_{ex} grows linearly. For the same reason, H_{ex} satisfies ($\mathcal{H}5$) but H_{BS} does not.

3.2.2. Existence of a classical solution

Up to now, we have not spoken of existence or uniqueness results. The results of this paragraph rely on the first order system form of NLSWE (1) as given in section 1.3 and more specifically on its non conservative form:

$$\begin{cases} \text{Find } \mathbf{U} : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}^n, \\ \partial_t \mathbf{U} + A(\mathbf{U})\, \partial_x \mathbf{U} = 0, \\ \mathbf{U}(x, 0) = \mathbf{U}_0(x) \end{cases} \tag{122}$$

where $A(\mathbf{U})$ is given by (4) or (7) in the case of NLSWE.

In what follows, we shall refer to the notation of section 1.3. Let us first recall some classical definitions:

Definition 3.1. LINEARLY DEGENERATE FIELD
The couple $(\mu_k, \mathbf{r}_k)$ is said to be linearly degenerate on $\mathcal{D}$ (L.D.) if

$$\nabla \mu_k(\mathbf{u}) \cdot \mathbf{r}_k(\mathbf{u}) = 0, \qquad \forall\, \mathbf{u} \in \mathcal{D}.$$

Definition 3.2. GENUINELY NONLINEAR FIELD
The couple $(\mu_k, \mathbf{r}_k)$ is said to be genuinely nonlinear on $\mathcal{D}$ (G.N.L.) if

$$\nabla \mu_k(\mathbf{u}) \cdot \mathbf{r}_k(\mathbf{u}) \neq 0, \qquad \forall\, \mathbf{u} \in \mathcal{D}.$$

The theory of hyperbolic systems shows that in general, even if initial conditions are very smooth, classical solutions do not exist beyond some finite time interval. However, we can state a global existence result for (122), using the results of Li Ta-tsien page 89 of [32], who considered $A(\mathbf{u})$ locally C^2, under the following assumptions:

Assumption 1: The system (122) is **locally hyperbolic** (cf. definition 1.1) and strictly hyperbolic at the origin:

$$\mu_1(0) < \mu_2(0) < \cdots < \mu_n(0).$$

Assumption 2: The system (122) is locally **linearly degenerate**, i.e. all couples $(\mu_k, \mathbf{r}_k)$ are L.D. on a neighborhood of 0.

Theorem 3.1 (Li Ta-tsien). *Suppose that $A(\mathbf{U})$ is $\mathcal{C}^2$ in a neighborhood of $\mathbf{U} = 0$ and satisfies assumptions 1 and 2. Suppose that $\mathbf{U}_0$ is $\mathcal{C}^1$ and has compact support such that*

$$Supp(\mathbf{U}_0) \subseteq [\alpha_0, \beta_0].$$

Then, there exists $\theta_0 > 0$ such that if

$$\theta := (\beta_0 - \alpha_0)\, \sup_{x \in \mathbb{R}} |\mathbf{U}_0'(x)| < \theta_0,$$

(122) admits a unique global solution

$$\mathbf{U} \equiv \mathbf{U}(x, t) \in \mathcal{C}^1(\mathbb{R} \times \mathbb{R}^+).$$

In the case of nonlinear Hamiltonian systems of wave equations, the unknown vector $\mathbf{U}$ is $(\partial_t \mathbf{u}, \partial_x \mathbf{u})$ and the matrix $A(\mathbf{U})$ is given by (7). The regularity needed on $A(\mathbf{U})$ is then deferred on $D^2 H(\mathbf{U}_x)$, so that we can write:

Corollary 3.1. *Suppose in (1) that H is $\mathcal{C}^4$ and strictly convex in a neighborhood of $\mathbf{u}_x = 0$ (local form of ($\mathcal{H}3$)). Assume also that $\mathbf{u}_0$ is $\mathcal{C}^2$, $\mathbf{u}_1$ is $\mathcal{C}^1$ and both have compact support included in $[\alpha_0, \beta_0]$. Suppose furthermore that the eigenpairs $\big(\lambda_i(\mathbf{v}), \mathbf{r}_i(\mathbf{v})\big)$, i $\in$ $[1, N]$ of $D^2 H(\mathbf{v})$ satisfy*

$$\nabla \lambda_i(\mathbf{v}) \cdot \mathbf{v}_i(\mathbf{v}) = 0, \qquad \forall\, \mathbf{v} \in \mathcal{D} \tag{123}$$

$\mathcal{D}$ being the neighborhood of 0. Then there exists $\theta_0 > 0$ such that if

$$\theta := (\beta_0 - \alpha_0) \sup_{x \in \mathbb{R}} \big[\, |\mathbf{u}_0''(x)|\,,\, |\mathbf{u}_1'(x)| \,\big] < \theta_0,$$

Cauchy problem (1) admits a unique global $\mathcal{C}^2$ solution $\mathbf{u} \equiv \mathbf{u}(x, t) \in \mathcal{C}^2(\mathbb{R} \times \mathbb{R}^+)$.

PROOF. Wefirst remark that $A(\mathbf{U})$ depends on $\mathbf{U}_x$ only via $D^2 H$ (see (7)). The eigenpairs of $A(\mathbf{U})$ are given by:

$$\mu_i^{\pm}(\mathbf{U}) = \pm\sqrt{\lambda_i(\mathbf{U}_x)},$$

$$\mathbf{r}_i^{\pm}(\mathbf{U}) = \Big(\pm \sqrt{\lambda_i(\mathbf{U}_x)}\mathbf{v}_i(\mathbf{U}_x), \mathbf{v}_i(\mathbf{U}_x) \Big), \quad \forall\, 1 \leq i \leq N.$$

One then computes

$$\nabla \mu_i^{\pm}(\mathbf{U}) \cdot \mathbf{r}_i^{\pm}(\mathbf{U}) = \pm \frac{\nabla \lambda_i(\mathbf{U}_x)}{2\sqrt{\lambda_i(\mathbf{U}_x)}} \cdot \mathbf{v}_i(\mathbf{U}_x) = 0$$

using (123). The corollary is then a rephrasing of theorem 3.1 adapted to the NLSWE. □

We can show that corollary 3.1 can be applied to the geometrically exact model. Indeed, $D^2 H_{ex}$ has the following eigenvalues:

$$\lambda_1(u_x, v_x) = 1 \qquad \text{and} \qquad \lambda_2(u_x, v_x) = 1 - \frac{\alpha}{\sqrt{u_x^2 + (1 + v_x)^2}}$$

associated with the eigenvectors:

$$\mathbf{v_1}(u_x, v_x) = \begin{pmatrix} u_x \\ 1 + v_x \end{pmatrix} \quad \text{and} \quad \mathbf{v_2}(u_x, v_x) = \begin{pmatrix} -(1 + v_x) \\ u_x \end{pmatrix}$$

We can also easily show that for $i \in [1,2]$

$$\nabla\lambda_i(u_x, v_x) \cdot v_i(u_x, v_x) = 0, \ \forall u_x \neq 0 \text{ and } v_x \neq -1.$$

The first order form of (118) is then **linearly degenerate**. Under assumptions on initial data, we can then apply theorem (3.1) and conclude that there exists a global $\mathcal{C}^2$ solution of the equation. All the computations of the first part of the article are valid since solutions are smooth enough.

In the case of the Bank-Sujbert model (121), the calculation of $\nabla\lambda.v$ leads to:

$$\begin{cases} \nabla\lambda_1.v_1 = & \frac{\alpha}{8}\Big[9u_x^2 + 6\sqrt{\Delta} - 12\frac{1}{\sqrt{\Delta}} + 27\frac{1}{\sqrt{\Delta}}u_x^4 \\ & +36\frac{1}{\sqrt{\Delta}}u_x^2 v_x + 9u_x^2 + 12\frac{1}{\sqrt{\Delta}}v_x^2 + 6v_x\Big] \\ \nabla\lambda_2.v_2 = & \alpha\Big[\frac{9}{4}u_x^2 + \frac{3}{4}v_x - \frac{3}{8}\sqrt{\Delta} + \frac{3}{2}\frac{1}{\sqrt{\Delta}} - \frac{27}{8}\frac{1}{\sqrt{\Delta}}u_x^4 \\ & -\frac{9}{2}\frac{1}{\sqrt{\Delta}}u_x^2 v_x\Big], \end{cases}$$

where $\Delta = 4\Big[\big(1 - v_x - \frac{3}{2}u_x^2\big)^2 + 4u_x^2\Big]$.

This calculation shows that the first order form of (121) is **genuinely nonlinear** unless $\alpha = 0$. Consequently, theorem 3.1 cannot be applied in this case. In fact, if the system is G.N.L. (all couples $(\lambda, \mathbf{r})$ are G.N.L., see definition 3.2), there exists a **blow up** result in [32] which states that the $\mathcal{C}^2$ norm of the solution of (121) must blow up in a finite time depending on the size of the initial data.

Remark 3.1. *This existence result is true for the geometrically exact model for $N \geqslant 2$ but is false when the equation is scalar ($N = 1$) for the nonlinear case. As for the approximate Bank-Sujbert model mentioned earlier, the nonlinear scalar case is genuinely nonlinear, hence the blow up theorem can be applied, meaning that the $\mathcal{C}^2$ norm of the solution must blow up in a finite time depending on the size of the initial data. This result has been mentioned before by John [21] and Kleinerman and Majda [24].*

The nonlinear scalar wave equation often found in the literature is the following:

$$u_{tt} - \big(K(u_x)\big)_x = 0.$$

It is possible to write this second order scalar equation as a first order system having two opposite eigenvalues $\lambda^{\pm} = \pm\sqrt{K'(u_x)}$ associated with the two eigenvectors

$$v^{\pm} = \begin{pmatrix} 1 \\ \mp\sqrt{K'(u_x)} \end{pmatrix} \quad \Rightarrow \quad \nabla\lambda^{\pm}.v^{\pm} = \pm\frac{K''(u_x)}{2\sqrt{K'(u_x)}}$$

The classical "nonlinear string model" is to set

$$K(v) = \frac{v}{\sqrt{1+v^2}} \ ,$$

which leads indeed to a genuinely nonlinear field.

3.3. Computational algorithm

Finite element discretization. Concretely, in the case of the nonlinear string, Ω is a segment, noted $[0, L]$. For each direction, we discretize $H_0^1\big([0,L]\big)$ with Lagrange P_k elements. We can evaluate the dimension of $\mathcal{V}_h$, that is to say the number of degrees of freedom, which depends on the degree k of the chosen basis functions, on the number of elements $N_x - 1$ in the mesh, and on the size N of the system. The amount leads to N_h degrees of freedom, with $N_h = N[(N_x - 1)k + 1]$.

Nonlinear resolution. Programming the scheme (113) amounts to nullifying, at each time step, a function $F : \mathbb{R}^{N_h} \longrightarrow \mathbb{R}^{N_h}$ which is a priori highly non linear. The method we use to find a zero of F is Newton's method, which consists in going from an initial point, then inverting the jacobian matrix of F until we find a point $\mathbf{U}^*$ such that $F(\mathbf{U}^*)$ is "sufficiently close" to zero. We have to calculate this Jacobian matrix, which depends on the point at which we estimate it, since the problem is nonlinear.
We can write the scheme's solution $\mathbf{u}_h$ as its decomposition on basis functions:

$$\mathbf{u}_h = \sum_{\ell=1}^{N}\sum_{j=1}^{N_d} u_{\ell,j}\psi_{\ell,j}$$

where $\psi_{\ell,j}$ is a vector having only one non zero component, directed in the direction ℓ, which is ϕ_j ie the basis function of P_k associated with the degree of freedom j, and N_d is the number of degrees of freedom for each direction. The unknown of the problem, at each time step, is the vector $\mathbf{U}_h = \big(u_{\ell,j}^{n+1}\big)_{\ell,j}$, the values $\big(u_{\ell,j}^{n}\big)_{\ell,j}$ and $\big(u_{\ell,j}^{n-1}\big)_{\ell,j}$ being considered known and constant. Then, the scheme consists in N_h lines, and the line corresponding to the direction ℓ and the degree of freedom j can be written

$$\begin{aligned} F_{\ell,j}(\mathbf{U}_h) = & \oint_{[0\ L]}^{h} \frac{u_\ell^{n+1} - 2u_\ell^n + u_\ell^{n-1}}{\Delta t^2}\phi_j \\ & + \oint_{[0\ L]}^{h} \sum_{\sigma\in\Sigma_\ell} \zeta(\sigma)\, \delta_\ell H(\partial_x u_\ell^{n+1}, \partial_x u_\ell^{n-1}; \partial_x u_{\tilde{\ell}\neq\ell}^{n+\sigma(\tilde{\ell})})\partial_x\phi_j, \end{aligned}$$

where $u_\ell = \sum_{p=1}^{N_d} u_{\ell,p}\phi_p$.
The Jacobian of this scheme is then a matrix of the applications from $\mathbb{R}^{N_h}$ to $\mathbb{R}$:

$$\begin{cases} \dfrac{\partial F_{\ell,j}}{\partial u_{\ell,n}} = \dfrac{1}{\Delta t^2}\displaystyle\oint_{[0\ L]}^{h} \phi_n\phi_j \\ \qquad + \displaystyle\sum_{\sigma\in\Sigma_\ell} \zeta(\sigma) \oint_{[0\ L]}^{h} \frac{\partial\delta_\ell H}{\partial\ell}(\partial_x u_\ell^{n+1}, \partial_x u_\ell^{n-1}; \partial_x u_{\tilde{\ell}\neq\ell}^{n+\sigma(\tilde{\ell})})\partial_x\phi_n\partial_x\phi_j \\ \dfrac{\partial F_{\ell,j}}{\partial u_{k,n}} = \displaystyle\sum_{\substack{\sigma\in\Sigma_\ell \\ \sigma(k)=+1}} \zeta(\sigma)\oint_{[0\ L]}^{h} \frac{\partial\delta_\ell H}{\partial k}(\partial_x u_\ell^{n+1}, \partial_x u_\ell^{n-1}; \partial_x u_{\tilde{\ell}\neq\ell}^{n+\sigma(\tilde{\ell})})\partial_x\phi_n\partial_x\phi_j. \end{cases}$$

3.4. Tests of numerical results for the string model

Using our preserving scheme (113), we have implemented the nonlinear string model described in section 3.1 (geometrically exact model), and the approximate Bank-Sujbert model. The numerical results were interesting for showing the influence of nonlinearity on the behavior of a vibrating string, and to show the comparison between exact and developed models. We remind that the system (118) is equivalent to the geometrically exact model, and has been obtained by a space and time scaling, introducing $\alpha = \frac{EA-T_0}{EA}$. Increasing α amounts to decreasing T_0 (all other parameters equal) ie slackening the string. The following numerical experiments have been made on scaled systems, using linear shape functions ($\mathbb{P}_1$) for the finite elements space approximation, a space step $\Delta x = 0.01$ and a time step $\Delta t = 0.0033$.

3.4.1. Influence of the nonlinearity

Numerical experiments have been carried out on very simple problems to show the influence of the nonlinearity on the string vibration. Nonlinear behavior can come from two different factors : either the nonlinear factor α comes closer to 1, or the initial amplitude grows. Figure 2 shows the deformation of a string with a sine function as initial condition in the transversal direction. The time is represented as the third space variable, and the different snapshots are displayed in different colors from blue to red. The nonlinear factor is changed between the lines, and the initial amplitude of the sine is changed between the columns.

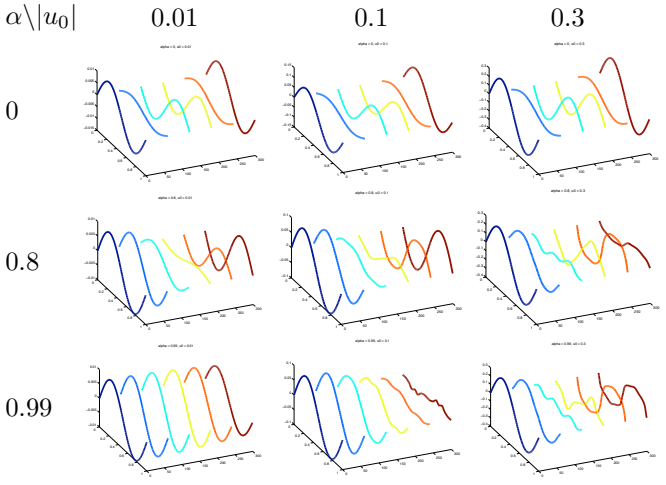

Figure 2: Evolution of the string deformation (time evolution in each subfigure from blue (left) to red (right)), for different values of α and initial amplitude.

The first line has been made with a nonlinear factor $\alpha = 0$, which leads to two linear uncoupled wave equations. Indeed, the initial amplitude of the data has just a scaling influence on the vibration of the string, expressing the linear behavior of the solution.

The first column shows, for small initial data, the influence of the nonlinear factor α. We can notice that the vibration slows when α increases, which is in agreement with the second order Taylor expansion of the potential energy H_{ex}: we indeed get as approximated system two uncoupled linear wave equations, with a celerity of 1 for the longitudinal wave and $\sqrt{1-\alpha}$ for the transversal wave, which is the one that we can observe. For $\alpha = 0$ the celerity is $c = 1$ and for $\alpha = 0.99$ the celerity is $c = 0.1$, ten times less.

Finally, if we look at the last two lines, we can see that increasing the initial amplitude leads to unusual behaviors of the string, pointing out the nonlinear influence of the equation, and especially the stretching of the string due to the presence of longitudinal waves.

We add that the simulations presented here have shown a very good energy preservation (about 10^{-13} relative error on the energy preservation for a Newton tolerance of 10^{-13} on the ℓ^2 norm of $F(\mathbf{U}_h)$).

3.4.2. Comparison with approximate Bank-Sujbert model

Another interesting point was to compare the string deformation when we use the geometrically exact model and when we use the Taylor expansion used in [2] and [5]. Our scheme makes it easy to switch from one model to another, and the result of simulation can be seen in figure 3.

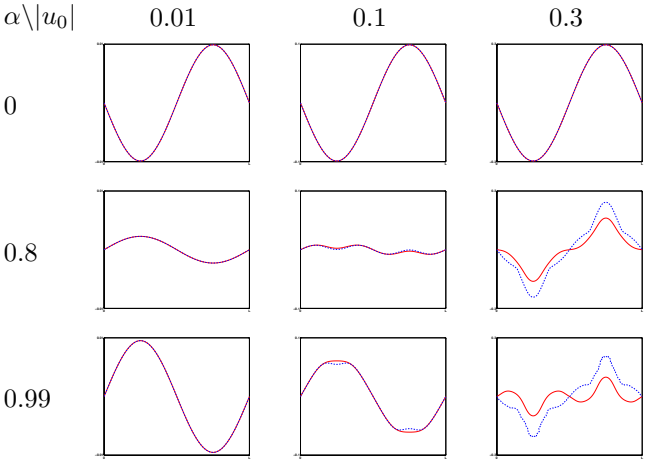

Figure 3: Comparison of a snapshot of the string deformation for exact (red, solid) and approximate (blue, dashed) model, for different values of α and initial amplitude.

The system of lines and columns is exactly the same as in the previous figure : different nonlinear factors α in lines, different initial amplitudes in columns. Each subfigure shows the string deformation by the exact model in red (solid line) and by the expanded model in blue (dashed line).

For $\alpha = 0$ (the first line) the two models coincide, and

we can see that the simulations give the same result (red (solid) curve and blue (dashed) curve are the same). But for more realistic values of α (for real piano strings it can be more than 0.999) the two curves are about the same for a very small initial amplitude, but are slightly different for $|u_0| = 0.1$ and very different for $|u_0| = 0.3$. This illustrates the fact that the expanded model is a good approximation of the exact model for small amplitudes, but becomes rather bad when the amplitudes grow.

Conclusions & Perspectives

In this paper we have proposed one solution for achieving our goal : constructing energy preserving schemes for nonlinear Hamiltonian systems of wave equations. Work is currently in progress on an energy preserving discretization of a more general class of systems. The future plan for this study is to develop a mathematical and a numerical model for the full grand piano. One of the main difficulties is to consider the coupling of the strings' movement with the vibrations of the soundboard and with the sound radiation, and we believe that the method that we developed, using an energy technique, will make it possible to achieve stability for the whole coupled problem.

References

[1] GV Anand. Large-amplitude damped free vibration of a stretched string. *Journal of the Acoustical Society of America*, 45(5):1089–1096, 1969.

[2] B Bank and L Sujbert. Generation of longitudinal vibrations in piano strings: From physics to sound synthesis. *Journal of the Acoustical Society of America*, 117:2268–2278, 2005.

[3] P Betsch and P Steinmann. Conservation properties of a time fe method. part i: time-stepping schemes for n-body problems. *Int J Numer Meth Eng*, 49(5):599–638, Jan 2000.

[4] P Betsch and P Steinmann. Conservation properties of a time fe method - part ii: Time-stepping schemes for non-linear elastodynamics. *Int J Numer Meth Eng*, 50(8):1931–1955, Jan 2001.

[5] S Bilbao. Conservative numerical methods for nonlinear strings. *Journal of the Acoustical Society of America*, 118(5):3316–3327, 2005.

[6] B Cano. Conserved quantities of some hamiltonian wave equations after full discretization. *Numerische Mathematik*, 103:197–223, 2006.

[7] J Chabassier and P Joly. Energy preserving schemes for nonlinear hamiltonian systems of wave equations: Application to the vibrating piano string. *Research Report*, (RR-7168):70, 2010.

[8] YS Chin and C Qint. Explicit energy-conserving schemes for the three-body problem. *Journal of Computational Physics*, 83(2):485–493, 1989.

[9] R Dautray, JL Lions, C Bardos, M Cessenat, P Lascaux, A Kavenoky, B Mercier, O Pironneau, B Scheurer, and R Sentis. Mathematical analysis and numerical methods for science and technology - vol. 6. *Springer*, 2000.

[10] J de Frutos and JM Sanz-Serna. Accuracy and conservation properties in numerical integration: the case of the korteweg-de vries equation. *Numerische Mathematik*, 75(4):421–445, 1997.

[11] S Dmitriev, P Kevrekidis, and N Yoshikawa. Standard nearesneighbour discretizations of klein gordon models cannot preserve both energy and linear momentum. *J Phys A-Math Gen*, 39:7217–7226, 2006.

[12] D Furihata. Finite-difference schemes for nonlinear wave equation that inherit energy conservation property. *Journal of Computational and Applied Mathematics*, 134(1-2):37–57, 2001.

[13] E Godlewski and PA Raviart. *Hyperbolic systems of conservation laws*. 1991.

[14] O Gonzalez. Exact energy and momentum conserving algorithms for general models in nonlinear elasticity. *Computer Methods in Applied Mechanics and Engineering*, 190:1763–1783, 2000.

[15] O Gonzalez and J Simo. On the stability of symplectic and energy-momentum algorithms for non-linear hamiltonian systems with symmetry. *Computer Methods in Applied Mechanics and Engineering*, 134:197–222, 1996.

[16] D Greenspan. Conservative numerical methods for ẍ= f(x). *Journal of Computational Physics*, 56:28–41, 1984.

[17] M Gross, P Betsch, and P Steinmann. Conservation properties of a time fe method. part iv: Higher order energy and momentum conserving schemes. *Int J Numer Meth Eng*, 63(13):1849–1897, Jan 2005.

[18] E Hairer. Backward analysis of numerical integrators and symplectic methods. *Ann. Numer. Math.*, 1(1-4):107–132, 1994.

[19] E Hairer and C Lubich. The life-span of backward error analysis for numerical integrators. *Numerische Mathematik*, 76:441–462, 1997.

[20] MW Hirsch, S Smale, and LR Devaney. Differential equations, dynamical systems, and an introduction to chaos. page 417, 2004.

[21] F John. Formation of singularities in one-dimensional nonlinear wave propagation. *Comm. Pure Appl. Math.*, 27(3):377–405, 1974.

[22] NA Kampanis, VA Dougalis, and JA Ekaterinakis. Effective computational methods for wave propagation. *Chapman & Hall / CRC*, 2008.

[23] P Kevrekidis. On a class of discretizations of hamiltonian nonlinear partial differential equations. *Physica D: Nonlinear Phenomena*, 183(1-2):68–86, Sep 2003.

[24] S Klainerman and A Majda. Formation of singularities for wave equations including the nonlinear vibrating string. *Comm. Pure Appl. Math.*, 33(3):241–263, 1980.

[25] S Li and L Vu-Quoc. Finite difference calculus invariant structure of a class of algorithms for the nonlinear klein-gordon equation. *SIAM Journal on Numerical Analysis*, 32(6):1839–1875, 1995.

[26] RE Mickens. A non-standard finite-difference scheme for conservative oscillators. *Journal of Sound and Vibration*, 240(3):587–591, 2001.

[27] RE Mickens. A numerical integration technique for conservative oscillators combining nonstandard finite-difference methods with a hamilton's principle. *Journal of Sound and Vibration*, 285(1-2):477–482, 2005.

[28] PM Morse and KU Ingard. *Theoretical Acoustics*. 1968.

[29] S Reich. Backward error analysis for numerical integrators. *SIAM Journal on Numerical Analysis*, 36(5):1549–1570, 1999.

[30] J Sanz-Serna. Symplectic integrators for hamiltonian problems: an overview. *Acta Numerica*, 1991.

[31] W Strauss and L Vazquez. Numerical solution of a nonlinear klein-gordon equation. *Journal of Computational Physics*, 28(2):271–278, 1978.

[32] Li Ta-Tsien. *Global classical solutions for quasilinear hyperbolic systems*. 1994.

[33] C Valette and C Cuesta. *Mécanique de la corde vibrante*. 1993.

[34] G Zhong and JE Marsden. Lie-poisson hamilton-jacobi theory and lie-poisson integrators. *Physics Letters A*, 133(3):134–139, Nov 1988.