Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 197 (2008) 1305–1322
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

Finite element approximation of the hyperbolic wave equation in mixed form

Ramon Codina

Universitat Politècnica de Catalunya, Jordi Girona 1-3, Edifici C1, 08034 Barcelona, Spain

Received 30 July 2007; received in revised form 19 October 2007; accepted 8 November 2007
Available online 22 November 2007

Abstract

The purpose of this paper is to present a finite element approximation of the scalar hyperbolic wave equation written in mixed form, that is, introducing an auxiliary vector field to transform the problem into a first-order problem in space and time. We explain why the standard Galerkin method is inappropriate to solve this problem, and propose as alternative a stabilized finite element method that can be cast in the variational multiscale framework. The unknown is split into its finite element component and a remainder, referred to as subscale. As original features of our approach, we consider the possibility of letting the subscales to be time dependent and orthogonal to the finite element space. The formulation depends on algorithmic parameters whose expression is proposed from a heuristic Fourier analysis.

Keywords: Hyperbolic wave equation; Stabilized finite element methods; Orthogonal subscales

1. Introduction

In most engineering problems dealing with waves, the wave equation is written in irreducible form, that is, with a single scalar unknown η depending on the spatial variable $\boldsymbol{x}$ and time t, so that if c is the wave speed this equation reads

$$\frac{1}{c^2}\partial_{tt}^2\eta - \Delta\eta = f, \tag{1}$$

where $\partial_{tt}^2 \equiv \partial_t\partial_t$ is the second order time derivative, Δ is the Laplacian operator and f is a given forcing term. This equation needs to be solved in a spatial domain $\Omega \subset \mathbb{R}^d$ ($d = 1$, 2 or 3) with appropriate boundary conditions and in a time interval $[0, T]$, giving $\eta(\boldsymbol{x}, 0)$ and $\partial_t\eta(\boldsymbol{x}, 0)$ as initial conditions.

However, in some cases it is convenient to consider the mixed form of (1), which consist in solving for η *as well as for a vector function* $\boldsymbol{u}(\boldsymbol{x}, t)$ the problem

$$\mu_\eta\partial_t\eta + \nabla\cdot\boldsymbol{u} = f_\eta, \tag{2}$$

$$\mu_u\partial_t\boldsymbol{u} + \nabla\eta = \boldsymbol{f}_u, \tag{3}$$

where $\mu_\eta > 0$ and $\mu_u > 0$ are coefficients such that $c^2 = (\mu_\eta\mu_u)^{-1}$ and the forcing terms f_η and $\boldsymbol{f}_u$ must be such that $\mu_u\partial_t f_\eta - \nabla\cdot\boldsymbol{f}_u = f$. Another possibility to transform (1) into a first order system is to define $\xi = \partial_t\eta$ and $\boldsymbol{u} = \nabla\eta$. This is the natural option in elastodynamics [4], although in many other physical applications, such as acoustic waves or gravity waves in fluids, the original problem is in fact (2)–(3) and its irreducible form (1). We shall briefly discuss an example of each situation in the following section. It is clear however that the linear problem (2)–(3) is only a model for more involved situations, either in solid mechanics or in nonlinear waves in shallow waters, cases in which the mixed form is mandatory.

The differential operator in (1) is of second order in both space and time, whereas (2)–(3) is a first order evolution problem with first order spatial derivatives. The most popular approach to deal with (1) is to use the Galerkin method for the spatial discretization and then to integrate in time using a finite difference scheme (see for example

E-mail address: ramon.codina@upc.edu

0045-7825/$ - see front matter
doi:10.1016/j.cma.2007.11.006

[15]). Numerical difficulties are considered to be concentrated on the time integration scheme, giving for granted that the Galerkin method is optimal for the spatial discretization of the Laplace operator. Therefore, research has been focused on devising time integration schemes for (1) (or vector counterparts) with given design properties.

Much less attention has been paid to the mixed form (2) and (3) of the wave problem. As in most mixed problems, there is a compatibility condition between the interpolation spaces for η and $\boldsymbol{u}$ which can be expressed as an inf–sup condition (see [5] for background, for example). As mentioned in Section 3, this condition on the interpolation of η and $\boldsymbol{u}$ is in fact similar to the condition for pressure and velocity in the case of the Stokes problem or the Darcy problem, but even if it is satisfied it does not guarantee stability of η in the space where it must belong. Our objective will be to present a formulation allowing equal and continuous interpolation for η and $\boldsymbol{u}$. Apart from simplifying the numerical implementation, this has two additional benefits. On the one hand, we will show that improved stability on η can be obtained with respect to some classical methods that satisfy the inf–sup condition. On the other hand, the classical Lagrange interpolation naturally allows for mass lumping through the use of special quadrature rules, a requirement to design explicit time integration schemes and not always possible using some interpolations satisfying the inf–sup condition (see, e.g., [4] and references therein).

Our formulation is based on the variational multiscale approach in the format introduced in [16,17]. The basic idea is to split the unknowns into a *resolvable* component, which can be reproduced by the discretization method (in our case finite elements) and the remainder, which we will call *sub-grid scale* or *subscale*. Rather than solving exactly for the latter, the formulation results from a closed form approximation for the subscales, which is designed in order to capture their *effect* on the discrete finite element solution. This leads to a formulation that allows the use of equal η–$\boldsymbol{u}$ interpolations. We prove analytically this fact in a particular case, only aiming to explain the stabilization mechanism introduced by the approximation of the subscales.

This paper is organized as follows. In Section 2, we state the initial and boundary value problem to be solved, both in its differential and in its weak form. We also present two examples of wave problems that will serve us to illustrate a discussion on the way to scale the equations. The Galerkin space discretization is presented in Section 3, where the reasons for its failure are explained. The main contribution of this work is presented in Section 4, where a stabilized finite element method is proposed. After presenting the basis of the formulation, its application to the mixed form of the wave equation is studied in detail. The algorithmic parameters on which the formulation depends are designed on the basis of a Fourier analysis of the problem, similar to that proposed already in [8] for the incompressible Navier–Stokes equations, although extended to general first-order systems. A stability estimate is then proved in the particular case in which the space of subcales is orthogonal to the finite element space, a possibility introduced in [10] to stabilize velocity–pressure interpolations in the Stokes problem (see also [9] for a full analysis of the method applied to the linearized Navier–Stokes equations). Section 5 presents the results of some numerical experiments only intended to demonstrate that the stabilized formulation proposed in fact suppresses the instabilities of the Galerkin method. Some concluding remarks close the paper.

2. Problem statement

2.1. Initial and boundary value problem

The differential equations (2) and (3) need to be supplied with adequate initial and boundary condition to define the problem to be solved.

As for the boundary conditions, we consider two possibilities. Let the boundary $\partial\Omega$ be split into two disjoint sets Γ_{I} and Γ_{R}. On the former, we consider prescribed the scalar η and on Γ_{R} the normal component of vector $\boldsymbol{u}$ is assumed to be given. Without loss of generality, we will consider both boundary conditions as homogeneous. The initial conditions to be considered are of the form $\eta(\boldsymbol{x},0)=\eta^0(\boldsymbol{x})$ and $\boldsymbol{u}(\boldsymbol{x},0)=\boldsymbol{u}^0(\boldsymbol{x})$.

The differential form of the initial and boundary value problem to be considered consists therefore in finding $\eta(\boldsymbol{x},t)$ and $\boldsymbol{u}(\boldsymbol{x},t)$ such that

$$\mu_\eta\partial_t\eta+\nabla\cdot\boldsymbol{u}=f_\eta,\quad \text{in } \Omega,\ t>0, \tag{4}$$

$$\mu_u\partial_t\boldsymbol{u}+\nabla\eta=\boldsymbol{f}_u,\quad \text{in } \Omega,\ t>0, \tag{5}$$

$$\eta=0,\quad \text{on } \Gamma_{\mathrm{I}},\ t>0, \tag{6}$$

$$\boldsymbol{n}\cdot\boldsymbol{u}=0,\quad \text{on } \Gamma_{\mathrm{R}},\ t>0, \tag{7}$$

$$\eta(\boldsymbol{x},0)=\eta^0(\boldsymbol{x}),\quad \text{in } \Omega, \tag{8}$$

$$\boldsymbol{u}(\boldsymbol{x},0)=\boldsymbol{u}^0(\boldsymbol{x}),\quad \text{in } \Omega. \tag{9}$$

Eqs. (4) and (5) can be re-written as

$$\begin{bmatrix}\mu_\eta & \mathbf{0}\\ \mathbf{0} & \mu_u\boldsymbol{I}\end{bmatrix}\partial_t\begin{bmatrix}\eta\\ \boldsymbol{u}\end{bmatrix}+\begin{bmatrix}0 & \nabla\cdot(\cdot)\\ \nabla(\cdot) & \mathbf{0}\end{bmatrix}\begin{bmatrix}\eta\\ \boldsymbol{u}\end{bmatrix}=\begin{bmatrix}f_\eta\\ \boldsymbol{f}_u\end{bmatrix},$$

where $\boldsymbol{I}$ is the $d\times d$ identity matrix.

At this point it is convenient to introduce some notation. Given X, a space of functions defined on Ω, its norm will be denoted by $\|\cdot\|_X$, and the space of functions such that their X-norm is C^k continuous in the time interval $[0,T]$ will be denoted by $C^k([0,T];X)$. We will be interested only in the cases $k=0$ and $k=1$. Three particular spaces X will be relevant in the presentation: $L^2(\Omega)$, $H^1(\Omega)$, the space of functions in $L^2(\Omega)$ with derivatives also in $L^2(\Omega)$, and $H(\mathrm{div},\Omega)$, the space of vector functions with components and divergence in $L^2(\Omega)$. A bold character will be used to denote the vector counterpart of the first two spaces.

As it will be explained below, for regular enough data the problem is well posed for $\eta\in C^0([0,T],V_\eta)\cap C^1([0,T],L^2(\Omega))$ and $\boldsymbol{u}\in C^0([0,T],\boldsymbol{V}_u)\cap C^1([0,T],\boldsymbol{L}^2(\Omega))$, where

$$V_\eta = \{\xi \in H^1(\Omega) | \xi = 0 \text{ on } \Gamma_{\rm I}\}, \tag{10}$$
$$\boldsymbol{V}_u = \{\boldsymbol{v} \in H(\text{div}, \Omega) | \boldsymbol{n}\cdot\boldsymbol{v} = 0 \text{ on } \Gamma_{\rm R}\}. \tag{11}$$

The force terms are assumed to be $f_\eta \in C^1([0,T]; L^2(\Omega))$ and $\boldsymbol{f}_u \in C^0([0,T]; H(\text{div}, \Omega))$. If we define the spaces

$$V = V_\eta \times \boldsymbol{V}_u,$$
$$L = L^2(\Omega) \times \boldsymbol{L}^2(\Omega),$$

problem (4)–(9) can be cast into the following abstract framework: find $\boldsymbol{U} \in C^0([0,T]; V) \cap C^1([0,T]; L)$ such that

$$\boldsymbol{\mu}\partial_t \boldsymbol{U} + \mathscr{A}\boldsymbol{U} = \boldsymbol{F}, \tag{12}$$
$$\boldsymbol{U}(0) = \boldsymbol{U}_0, \tag{13}$$

where $\boldsymbol{\mu} = \text{diag}(\mu_\eta, \mu_u \boldsymbol{I})$, $\boldsymbol{F} = (f_\eta, \boldsymbol{f}_u)$, $\boldsymbol{U}_0 = (\eta^0(\boldsymbol{x}), \boldsymbol{u}^0(\boldsymbol{x}))$ and $\mathscr{A} : V \to L$ is defined by

$$\boldsymbol{U} = \begin{bmatrix} \eta \\ \boldsymbol{u} \end{bmatrix} \mapsto \mathscr{A}\boldsymbol{U} = \begin{bmatrix} \nabla\cdot\boldsymbol{u} \\ \nabla\eta \end{bmatrix}.$$

2.2. Variational problem

2.2.1. Scaling

Let us start noting that the way in which we have written Eqs. (2) and (3) allows us to guarantee that the pointwise work $\boldsymbol{U}^{\rm t}\boldsymbol{F}$ is dimensionally well defined. Here and in what follows $\boldsymbol{U}$, possibly with a subscript, denotes an element in the domain of $\mathscr{A}$ as a subspace of L, whereas $\boldsymbol{F}$, perhaps also with a subscript, denotes an element in the range of $\mathscr{A}$, which is also in L. The components of $\boldsymbol{U}$ will be generically denoted as $\boldsymbol{U} = (\eta, \boldsymbol{u})$, and the components of $\boldsymbol{F}$ as $\boldsymbol{F} = (f_\eta, \boldsymbol{f}_u)$.

In most problems, making $\boldsymbol{U}^{\rm t}\boldsymbol{F}$ well defined usually implies a scaling of the original equations and a redefinition of the unknowns of the problem if the temporal derivative is written without any coefficient. In our case, if $[\cdot]$ denotes a dimensional group we have that

$$[\boldsymbol{U}^{\rm t}\boldsymbol{F}] = \frac{1}{\mathtt{L}}[\eta][\boldsymbol{u}] = \frac{1}{\mathtt{T}}[\mu_\eta][\eta]^2 = \frac{1}{\mathtt{T}}[\mu_u][\boldsymbol{u}]^2, \tag{14}$$

where $\mathtt{L}$ is a length dimension and $\mathtt{T}$ a time dimension.

We will also need to scale properly the inner product *of forces* in L or, more precisely, to introduce a scaling matrix $\boldsymbol{M}$ so that the pointwise product $\boldsymbol{F}_1^{\rm t}\boldsymbol{M}\boldsymbol{F}_2$ is dimensionally well defined, with $\boldsymbol{F}_i$, $i = 1, 2$, two vectors of forces in L. The choice of this scaling is equivalent to choose the way the equations are written in dimensionless form, if this is the option adopted. Let us describe a possibility to do this. If we choose the scaling matrix of the form $\boldsymbol{M} = \text{diag}\ (m_\eta, m_u\boldsymbol{I})$, we will have that

$$[\boldsymbol{F}_1^{\rm t}\boldsymbol{M}\boldsymbol{F}_2] = \frac{1}{\mathtt{L}^2}[m_\eta][\boldsymbol{u}]^2 = \frac{1}{\mathtt{L}^2}[m_u][\eta]^2. \tag{15}$$

Let $C_{\mathtt{L}}$ be a characteristic length of the domain. In view of (14), if we take $m_\eta = C_{\mathtt{L}}\sqrt{\mu_u/\mu_\eta}$, $m_u = C_{\mathtt{L}}\sqrt{\mu_\eta/\mu_u}$, that is, if

$$\boldsymbol{M} = \text{diag}(m_\eta, m_u\boldsymbol{I}), \quad m_\eta := C_{\mathtt{L}}\sqrt{\frac{\mu_u}{\mu_\eta}}, \quad m_u := C_{\mathtt{L}}\sqrt{\frac{\mu_\eta}{\mu_u}}, \tag{16}$$

condition (15) will hold and, moreover, $[\boldsymbol{U}^{\rm t}\boldsymbol{F}] = [\boldsymbol{F}_1^{\rm t}\boldsymbol{M}\boldsymbol{F}_2]$ (although this condition is not mandatory).

Finally, we also need to define a scalar product in L for two elements $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ in the domain of $\mathscr{A}$. It can be easily checked that a way to define this product in a dimensionally consistent manner is to use $\boldsymbol{M}^{-1}$ as scaling matrix, since in this case

$$[\boldsymbol{U}_1^{\rm t}\boldsymbol{M}^{-1}\boldsymbol{U}_2] = \frac{1}{[C_{\mathtt{L}}]}\left[\sqrt{\frac{\mu_\eta}{\mu_u}}\right][\eta]^2 = \frac{1}{[C_{\mathtt{L}}]}\left[\sqrt{\frac{\mu_u}{\mu_\eta}}\right][\boldsymbol{u}]^2$$
$$= \frac{1}{[C_{\mathtt{L}}]}[\eta][\boldsymbol{u}]. \tag{17}$$

It is observed that for $C_{\mathtt{L}}$ a characteristic length we have that $[\boldsymbol{U}_1^{\rm t}\boldsymbol{M}^{-1}\boldsymbol{U}_2] = [\boldsymbol{U}^{\rm t}\boldsymbol{F}]$. However, this is in fact not necessary, and we could take $C_{\mathtt{L}} = 1$. We will see that the final formulation to be proposed does not depend on $C_{\mathtt{L}}$.

The different scalings defined according to the variables being multiplied define different inner products in L, that we denote as follows:

$$(\boldsymbol{U}, \boldsymbol{F})_L := \int_\Omega \boldsymbol{U}^{\rm t}\boldsymbol{F}\,{\rm d}\boldsymbol{x} = \int_\Omega \eta f_\eta\,{\rm d}\boldsymbol{x} + \int_\Omega \boldsymbol{u}\cdot\boldsymbol{f}_u\,{\rm d}\boldsymbol{x}, \tag{18}$$

$$(\boldsymbol{F}_1, \boldsymbol{F}_2)_{L,M} := \int_\Omega \boldsymbol{F}_1^{\rm t}\boldsymbol{M}\boldsymbol{F}_2\,{\rm d}\boldsymbol{x} = C_{\mathtt{L}}\sqrt{\frac{\mu_u}{\mu_\eta}}\int_\Omega f_{\eta,1}f_{\eta,2}\,{\rm d}\boldsymbol{x} + C_{\mathtt{L}}\sqrt{\frac{\mu_\eta}{\mu_u}}\int_\Omega \boldsymbol{f}_{u,1}\cdot\boldsymbol{f}_{u,2}\,{\rm d}\boldsymbol{x}, \tag{19}$$

$$(\boldsymbol{U}_1, \boldsymbol{U}_2)_{L,M^{-1}} := \int_\Omega \boldsymbol{U}_1^{\rm t}\boldsymbol{M}^{-1}\boldsymbol{U}_2\,{\rm d}\boldsymbol{x} = \frac{1}{C_{\mathtt{L}}}\sqrt{\frac{\mu_\eta}{\mu_u}}\int_\Omega \eta_1\eta_2\,{\rm d}\boldsymbol{x} + \frac{1}{C_{\mathtt{L}}}\sqrt{\frac{\mu_u}{\mu_\eta}}\int_\Omega \boldsymbol{u}_1\cdot\boldsymbol{u}_2\,{\rm d}\boldsymbol{x}, \tag{20}$$

where $\boldsymbol{M}$ is given by (16). It is observed that if the coordinates were curvilinear, $\boldsymbol{M}$ would play the role of the metric that transforms the "natural" covariant character of forces $\boldsymbol{F}$ into contravariant, and likewise $\boldsymbol{M}^{-1}$ transforms the "natural" contravariant character of $\boldsymbol{U}$ into covariant.

It remains to define the way to scale the inner product in V (and its associated norm). We do this as follows:

$$(\boldsymbol{U}_1, \boldsymbol{U}_2)_V := \frac{1}{C_{\mathtt{L}}}\sqrt{\frac{\mu_\eta}{\mu_u}}\int_\Omega \eta_1\eta_2\,{\rm d}\boldsymbol{x} + C_{\mathtt{L}}\sqrt{\frac{\mu_\eta}{\mu_u}}\int_\Omega \nabla\eta_1\cdot\nabla\eta_2\,{\rm d}\boldsymbol{x} + \frac{1}{C_{\mathtt{L}}}\sqrt{\frac{\mu_u}{\mu_\eta}}\int_\Omega \boldsymbol{u}_1\cdot\boldsymbol{u}_2\,{\rm d}\boldsymbol{x} + C_{\mathtt{L}}\sqrt{\frac{\mu_u}{\mu_\eta}}\int_\Omega (\nabla\cdot\boldsymbol{u}_1)(\nabla\cdot\boldsymbol{u}_2)\,{\rm d}\boldsymbol{x}. \tag{21}$$

We finally set $\|\boldsymbol{U}\|_V := (\boldsymbol{U}, \boldsymbol{U})_V^{1/2}$, $\|\boldsymbol{F}\|_{L,M} := (\boldsymbol{F}, \boldsymbol{F})_{L,M}^{1/2}$, $\|\boldsymbol{U}\|_{L,M^{-1}} := (\boldsymbol{U}, \boldsymbol{U})_{L,M^{-1}}^{1/2}$.

2.2.2. First variational formulation

Problem (12)–(13) is equivalent to find $\boldsymbol{U} \in C^0([0,T]; V) \cap C^1([0,T]; L)$ such that

$$(\boldsymbol{\mu}\partial_t \boldsymbol{U}, \boldsymbol{V})_L + (\mathscr{A}\boldsymbol{U}, \boldsymbol{V})_L = (\boldsymbol{F}, \boldsymbol{V})_L \quad \forall \boldsymbol{V} = (\xi, \boldsymbol{v}) \in V, \quad (22)$$

$$(\boldsymbol{U}(0), \boldsymbol{V})_L = (\boldsymbol{U}_0, \boldsymbol{V})_L \quad \forall \boldsymbol{V} = (\xi, \boldsymbol{v}) \in L, \quad (23)$$

which is *the first* variational form of the problem we will consider. Note that, because of the density of V in L, the test function in (22) can be taken directly in L. In expanded form, this equation reads:

$$\mu_\eta(\partial_t \eta, \xi) + (\nabla \cdot \boldsymbol{u}, \xi) = (f_\eta, \xi) \quad \forall \xi \in V_\eta, \quad (24)$$

$$\mu_u(\partial_t \boldsymbol{u}, \boldsymbol{v}) + (\nabla \eta, \boldsymbol{v}) = (\boldsymbol{f}_u, \boldsymbol{v}) \quad \forall \boldsymbol{v} \in V_u. \quad (25)$$

The well-posedness of problem (22)–(23) relies on the following properties:

1. $\mathscr{A}$ is monotone: $(\mathscr{A}\boldsymbol{U}, \boldsymbol{U})_L \geqslant 0$.
2. $\mathscr{A}$ is maximal. In the case $\mathscr{A}$ is monotone and L reflexive, as in our case, this is implied by the existence of constants C_1 and C_2 such that

$$\|\mathscr{A}\boldsymbol{U}\|_{L,M} \geqslant C_1\|\boldsymbol{U}\|_V - C_2\|\boldsymbol{U}\|_{L,M^{-1}} \quad \forall \boldsymbol{U} \in V.$$

In our case, monotonicity and maximality of $\mathscr{A}$ are trivially checked. Since $\mathscr{A}\boldsymbol{V} \in L$ for all $\boldsymbol{V} \in V$, with the scalings introduced (see (18)–(21)) we have that

$$(\mathscr{A}\boldsymbol{U}, \boldsymbol{U})_L = \int_\Omega \eta \nabla \cdot \boldsymbol{u}\, \mathrm{d}\boldsymbol{x} + \int_\Omega \nabla \eta \cdot \boldsymbol{u}\, \mathrm{d}\boldsymbol{x} = 0,$$

$$\|\mathscr{A}\boldsymbol{U}\|^2_{L,M} = C_{\mathrm{L}}\sqrt{\frac{\mu_u}{\mu_\eta}} \int_\Omega |\nabla \cdot \boldsymbol{u}|^2\, \mathrm{d}\boldsymbol{x} + C_{\mathrm{L}}\sqrt{\frac{\mu_\eta}{\mu_u}} \int_\Omega |\nabla \eta|^2\, \mathrm{d}\boldsymbol{x} = \|\boldsymbol{U}\|^2_V - \|\boldsymbol{U}\|^2_{L,M^{-1}},$$

that is, the constants C_1 and C_2 can be taken both equal to 1.

If these two conditions hold, Hille–Yosida theorem guarantees that there exists a unique solution to the problem that can be bounded as follows (see, for example, [14]):

$$\sup_{t\in[0,T]} \|\boldsymbol{U}\|_{L,M^{-1}} \leqslant C\left(\|\boldsymbol{U}_0\|_{L,M^{-1}} + T \sup_{t\in[0,T]} \|\boldsymbol{F}\|_{L,M}\right), \quad (26)$$

$$\sup_{t\in[0,T]} \|\partial_t \boldsymbol{U}\|_{L,M^{-1}} \leqslant C\left(\|\boldsymbol{U}_0\|_V + T \sup_{t\in[0,T]} \|\partial_t \boldsymbol{F}\|_{L,M}\right), \quad (27)$$

$$\sup_{t\in[0,T]} \|\boldsymbol{U}\|_V \leqslant C\left(\|\boldsymbol{U}_0\|_V + T \sup_{t\in[0,T]} \|\partial_t \boldsymbol{F}\|_{L,M}\right). \quad (28)$$

Here C denotes a positive constant which may depend on the coefficients μ_η and μ_u.

Let us indicate how to prove these bounds. Bound (26) is obtained by taking $\boldsymbol{V} = \boldsymbol{U}(\cdot, t)$ in (22), using the monotonicity of $\mathscr{A}$ and integrating from $t = 0$ to an arbitrary $t\prime$. Bound (27) follows using a similar argument applied to the equation differentiated with respect to t. The important point is bound (28). It follows taking $\boldsymbol{V} = \boldsymbol{M}\mathscr{A}\boldsymbol{U}$ in the variational equation:

$$\begin{aligned}\|\mathscr{A}\boldsymbol{U}\|^2_{L,M} &= (\mathscr{A}\boldsymbol{U}, \boldsymbol{M}\mathscr{A}\boldsymbol{U})_L \\ &= (\boldsymbol{F}, \boldsymbol{M}\mathscr{A}\boldsymbol{U})_L - (\boldsymbol{\mu}\partial_t \boldsymbol{U}, \boldsymbol{M}\mathscr{A}\boldsymbol{U})_L \\ &\leqslant \|\boldsymbol{F}\|^2_{L,M} + C\|\boldsymbol{\mu}\partial_t \boldsymbol{U}\|^2_{L,M} + \frac{1}{2}\|\mathscr{A}\boldsymbol{U}\|^2_{L,M} \\ &\leqslant \|\boldsymbol{F}\|^2_{L,M} + C\|\partial_t \boldsymbol{U}\|^2_{L,M^{-1}} + \frac{1}{2}\|\mathscr{A}\boldsymbol{U}\|^2_{L,M},\end{aligned}$$

and then using the maximality of $\mathscr{A}$ and bounds (26) and (27).

It is important to note that estimates (26)–(28) are meaningless for the long term time behavior ($T \to \infty$) and other techniques to prove stability are required in this case. However, they are enough for our purpose to explain the failure of the Galerkin method and of designing a stabilized finite element formulation.

2.2.3. Second variational formulation

Similarly to what is done for the mixed approximation of the standard Poisson problem (see [5]), it is possible to prescribe the boundary conditions for η weakly and to reduce the regularity requirements. This can be done integrating by parts the second term in the left-hand-side of (25), using that $\boldsymbol{n} \cdot \boldsymbol{v} = 0$ on Γ_{R} and prescribing $\eta = 0$ on Γ_{I} weakly. Noting that no spatial derivatives of η appear, the resulting variational formulation is, instead of (24) and (25): find $\eta(\cdot, t) \in \bar{V}_\eta = L^2(\Omega)$, $\boldsymbol{u}(\cdot, t) \in V_u$ such that

$$\mu_\eta(\partial_t \eta, \xi) + (\nabla \cdot \boldsymbol{u}, \xi) = (f_\eta, \xi) \quad \forall \xi \in \bar{V}_\eta = L^2(\Omega), \quad (29)$$

$$\mu_u(\partial_t \boldsymbol{u}, \boldsymbol{v}) - (\nabla \cdot \boldsymbol{v}, \eta) = (\boldsymbol{f}_u, \boldsymbol{v}) \quad \forall \boldsymbol{v} \in V_u. \quad (30)$$

The regularity in time required is the same as in the previous formulation.

The advantage of (29)–(30) is that discontinuous approximations can be employed for η. In fact, classical elements used to approximate the Poisson problem in mixed form can be used for the spatial approximation of (29)–(30) (see, for example, [2,4]).

2.3. Two examples

To close this section, let us state two examples of wave problems that will serve to explain the scaling described previously.

2.3.1. Waves in an elastic bar

Let us consider an elastic bar of length ℓ, elastic modulus E, constant section of area A and made of a material of density ρ. If the bar is fixed at $x = 0$, free at $x = \ell$ and loaded with an axial load $p(x)$, $0 < x < \ell$, the initial and boundary value problem to be solved is

$$\rho\partial_t v - \partial_x \sigma = p, \quad 0 < x < \ell, \; t \in (0, T),$$

$$\frac{1}{EA}\partial_t \sigma - \partial_x v = 0, \quad 0 < x < \ell, \; t \in (0, T),$$

$$v = 0, \quad x = 0, \; t \in (0, T),$$

$$\sigma = 0, \quad x = \ell, \ t \in (0, T),$$
$$v = v_0, \quad 0 < x < \ell, \ t = 0,$$
$$\sigma = \sigma_0, \quad 0 < x < \ell, \ t = 0.$$

In these equations, v is the temporal derivative of the axial displacement u ($v = \partial_t u$), $\sigma = EA\partial_x u$ is the axial stress and $v_0(x)$ and $\sigma_0(x)$ are the initial conditions for $v(x,t)$ and $\sigma(x,t)$, respectively. Identifying these equations with problem (4)–(9) it is seen that $\eta \equiv v$, $\boldsymbol{u} \equiv -\sigma$, $\Gamma_{\rm I} = \{0\}$, $\Gamma_{\rm R} = \{\ell\}$. The important point is that in this case $\mu_\eta \equiv \rho$ and $\mu_u \equiv (EA)^{-1}$. As it is well known, the wave speed is $c = \sqrt{EA/\rho}$. The scaling matrix in (16) is now given by

$$\boldsymbol{M} = \mathrm{diag}\left(C_{\rm L}\sqrt{\frac{1}{\rho EA}}, C_{\rm L}\sqrt{\rho EA} \right),$$

where $C_{\rm L}$ may be either taken as $C_{\rm L} = \ell$ or as $C_{\rm L} = 1$, if we allow the three different products (14), (15) and (17) to have different dimensions (which is the natural choice in unbounded domains).

2.3.2. Gravity waves in shallow water flows

In this case the problem to be solved can be written as

$$\frac{1}{H}\partial_t \eta + \nabla \cdot \boldsymbol{u} = f_\eta, \quad \text{in } \Omega \subset \mathbb{R}^2, t > 0,$$
$$\frac{1}{g}\partial_t \boldsymbol{u} + \nabla \eta = \boldsymbol{f}_u, \quad \text{in } \Omega, \ t > 0,$$

with the same form of boundary and initial conditions as in (6)–(9). Thus, the equations to be solved are exactly (4)–(9), now η having the physical meaning of being the water elevation relative to the boundary conditions on $\Gamma_{\rm I}$, $\boldsymbol{u}$ the mean velocity, f_η and $\boldsymbol{f}_u$ come from non-homogeneous boundary conditions typically imposed for the water elevation on $\Gamma_{\rm I}$, H is the water depth (here assumed to be constant for simplicity) and g is the gravity acceleration.

Also in this case, the important point is the form of the scaling. Now $\mu_\eta \equiv H^{-1}$, $\mu_u \equiv g^{-1}$, $c = \sqrt{gH}$ (wave speed) and the scaling matrix is given by

$$\boldsymbol{M} = \mathrm{diag}\left(C_{\rm L}\sqrt{\frac{H}{g}}, C_{\rm L}\sqrt{\frac{g}{H}}\boldsymbol{I} \right).$$

Similarly to the previous case, $C_{\rm L}$ may be either taken as $C_{\rm L} = \mathrm{diam}(\Omega)$ (if Ω is bounded) or as $C_{\rm L} = 1$.

3. Space discretization using the Galerkin method

3.1. Galerkin method

Let us consider the variational formulation (22), which is supplied with the initial condition (23). In order to discretize it using the standard Galerkin method, we simply have to approximate space V by a finite element space $V_h \subset V$ (conforming approximation). For the moment, let us assume that the interpolations of both η and $\boldsymbol{u}$ are continuous, so that they are conforming using the two variational formulations described in the previous section.

The discrete problem using the Galerkin method is: find $\boldsymbol{U}_h \in C^1([0,T]; V_h)$ such that

$$(\boldsymbol{\mu}\partial_t \boldsymbol{U}_h, \boldsymbol{V}_h)_L + (\mathscr{A}\boldsymbol{U}_h, \boldsymbol{V}_h)_L = (\boldsymbol{F}, \boldsymbol{V}_h)_L \quad \forall \boldsymbol{V}_h \in V_h.$$

Note that the test function and trial solution spaces are the same.

Bounds (26) and (27) can be proved for the discrete problem exactly as for the continuous one. However, *bound (28) does not hold for the discrete problem*. The reason is quite simple: *$\mathscr{A}\boldsymbol{U}_h$ cannot be taken as test function*, since for $\boldsymbol{U}_h \in V_h$, $\mathscr{A}\boldsymbol{U}_h \notin V_h$. Observe also that from the numerical point of view, (28) is what *prevents oscillations*, since in our case it gives control on the divergence of $\boldsymbol{u}$ and the gradient of η. As a conclusion, the standard Galerkin method may yield oscillations.

3.2. Compatibility conditions

The previous argument explains a fact known from numerical evidence: the Galerkin formulation fails to solve wave problems in mixed form unless special care is put in the selection of the interpolating spaces for η and $\boldsymbol{u}$. Let us denote them by $V_{\eta,h}$ and $\boldsymbol{V}_{u,h}$, respectively.

Using the first variational form given by (24)–(25), several ways to overcome the instability of the Galerkin method can be found in the literature, particularly in the context of gravity waves in shallow water flows. Referring for example to the Boussinesq model for shallow waters, an early finite difference approximation can be found in [1] and another popular finite difference model in [22]. Finite element approximations were introduced later, see for example [18,20,21,23]. In these references, high frequency oscillations over the grid used to discretize the domain are mentioned, and methods to overcome them by ad-hoc filtering techniques or by the addition of numerical viscosity are reported for example in [21] and references therein. In the finite difference context, different grids can be used for the approximation of velocity and water elevation (see [18], for example). Surprisingly, there seems to be no explicit association between the instability problems encountered and the lack of stability of the Galerkin method. In [19] spurious propagating modes for the one dimensional linear equations are found from a classical dispersion analysis, both using continuous and discontinuous interpolations for the unknowns (for the continuous Galerkin method there are also spurious stationary modes).

The second variational form given by (29)–(30) has been studied for example in [2] (see also [4,3] for several extensions, including elastic waves). The way to obtain a stable formulation is based on choosing $V_{\eta,h}$ and $\boldsymbol{V}_{u,h}$ as for a stable mixed interpolation of the Poisson problem. Classical examples of these type of elements are the Raviart–Thomas (RT) and the Brezzi–Douglas–Marini (BDM) families (see [5] for background). For these elements, the interpolation for η is discontinuous, although this keeps conformity for formulation (29)–(30). In [2] it is shown that stability can be obtained for the wave equation as for the mixed Poisson problem. First, the divergence of $\boldsymbol{u}_h$ (the finite element approximation to $\boldsymbol{u}$) can be bounded in terms of the data by using the commutation of the divergence and the projection onto the appropriate finite element space, and then the L^2 norm of η_h (the approximation to η) can be bounded by using an inf–sup condition known to hold for the elements under consideration. Obviously, this approach does not lead to stability for $\nabla\eta_h$.

3.3. Matrix formulation

To close this section, let us formulate the discrete Galerkin approximation to problem (24)–(25) in matrix form and compare it with the irreducible form that would arise from (1).

Let M_η and M_u be the mass matrices arising from the interpolation of η and $\boldsymbol{u}$, respectively, D the matrix arising from the divergence term in (24) and F_η the resulting right-hand-side (RHS) vector. Let also G be the matrix arising from the gradient term in (25) and F_u the resulting RHS vector. If X_η and X_u are the arrays of nodal unknowns of η and $\boldsymbol{u}$, respectively, the matrix form of the discrete version of (24) and (25) is

$$\begin{bmatrix} \mu_\eta \mathsf{M}_\eta & 0 \\ 0 & \mu_u \mathsf{M}_u \end{bmatrix} \begin{bmatrix} \dot{\mathsf{X}}_\eta \\ \dot{\mathsf{X}}_u \end{bmatrix} + \begin{bmatrix} 0 & \mathsf{D} \\ \mathsf{G} & 0 \end{bmatrix} \begin{bmatrix} \mathsf{X}_\eta \\ \mathsf{X}_u \end{bmatrix} = \begin{bmatrix} \mathsf{F}_\eta \\ \mathsf{F}_u \end{bmatrix},$$

where the dot denotes time derivative. It is understood that boundary conditions are already incorporated in this matrix expression. Taking the time derivative of the first equation and using the second it is found that

$$\frac{1}{c^2}\mathsf{M}_\eta \ddot{\mathsf{X}}_\eta - \mathsf{D}\mathsf{M}_u^{-1}\mathsf{G}\mathsf{X}_\eta = \mu_u \dot{\mathsf{F}}_\eta - \mathsf{D}\mathsf{M}_u^{-1}\mathsf{F}_u, \tag{31}$$

where, as in the Introduction, we have defined $c^2 = (\mu_\eta\mu_u)^{-1}$.

On the other hand, if we had started directly from (1) the discrete algebraic equation would have had the matrix structure

$$\frac{1}{c^2}\mathsf{M}_\eta \ddot{\mathsf{X}}_\eta - \mathsf{L}\mathsf{X}_\eta = \mathsf{F}, \tag{32}$$

where L is the standard matrix arising from a Galerkin discretization of the Laplacian operator Δ (using continuous interpolation functions for η, as we are assuming) and F is the array that results from weighting f in (1) with the test functions corresponding to the interpolation of η.

Comparing (31) and (32) it is observed that the difference in the finite element approximation of the mixed and the irreducible wave equation relies on two different approximations to the Laplacian operator, namely, $\mathsf{D}\mathsf{M}_u^{-1}\mathsf{G}$ in the case of the mixed form and the classical approximation L for the irreducible form (provided an adequate identification of the forcing terms is done).

4. Stabilized finite element method

4.1. General framework

In this section, we present a stabilized finite element method aimed to overcome the instability problems of the standard Galerkin method when using *equal and continuous interpolation for the unknowns* η *and* $\boldsymbol{u}$. We start considering an abstract linear first order partial differential equation of the form

$$\boldsymbol{\mu}\partial_t \boldsymbol{U} + \boldsymbol{A}_i \partial_i \boldsymbol{U} = \boldsymbol{F}, \tag{33}$$

where $\boldsymbol{A}_i$ are $n_{\text{unk}} \times n_{\text{unk}}$ matrices, ∂_i is the partial derivative with respect to the i-th coordinate, $i = 1, 2, \dots, d$, repeated indexes in $\boldsymbol{A}_i\partial_i\boldsymbol{U}$ imply summation up to d, and $\boldsymbol{F}$ is a vector of n_{unk} components, n_{unk} being the number of scalar unknowns in $\boldsymbol{U}$ and d the space dimension.

The wave equation is mixed form can be recast as a problem of type (33) by taking $n_{\text{unk}} = d + 1$ and, for example for $d = 2$,

$$\boldsymbol{U} = \begin{bmatrix} \eta \\ u_1 \\ u_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_\eta & 0 & 0 \\ 0 & \mu_u & 0 \\ 0 & 0 & \mu_u \end{bmatrix}, \quad \boldsymbol{A}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$
$$\boldsymbol{A}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{34}$$

In spite of its "convective" look, the instabilities of the Galerkin method applied to Eq. (33) are not due to convection, but to the mathematical structure dictated by matrices (34) and by the different functional setting of the variables in play. For an application of the present ideas to the simple convection–diffusion equation, see [8].

Appropriate boundary conditions have to be appended to (33). The resulting problem is of the abstract form (12), with $\mathscr{A}\boldsymbol{U} = \boldsymbol{A}_i\partial_i\boldsymbol{U}$, and the Galerkin method applied to it suffers from the instability problems described in the previous section.

The stabilized finite element method we will present has its roots in the variational multiscale decomposition proposed in [17]. Our arguments here are completely heuristic, and are only intended to motivate a numerical formulation. Once it is stated, it will be clear why is it consistent and where does the stabilization mechanism introduced come from.

Let us split the unknown $\boldsymbol{U}$ as $\boldsymbol{U} = \boldsymbol{U}_h + \boldsymbol{U}'$, where $\boldsymbol{U}_h$ is the finite element solution we are looking for and $\boldsymbol{U}'$ the component of $\boldsymbol{U}$ that cannot be captured by the finite element mesh. We will call it *subgrid scale* or, simply, *subscale*. The idea is that an approximation for $\boldsymbol{U}'$ will lead to a problem for $\boldsymbol{U}_h$ with enhanced stability problems with respect to the standard Galerkin method.

Let us consider the problem posed for $\boldsymbol{U}(\cdot, t)$ in an adequate subspace of $\boldsymbol{L}^2(\Omega)$ that depends on the boundary conditions, so that $\boldsymbol{A}_i\partial_i\boldsymbol{U}(\cdot, t) \in \boldsymbol{L}^2(\Omega)$ (which corresponds to L in the previous section). Now we will simply denote the inner product in this space by $(\cdot,\cdot)$.

The weak form of the problem consists in finding $\boldsymbol{U}_h$ and $\boldsymbol{U}'$ such that

$$(\boldsymbol{\mu}\partial_t \boldsymbol{U}_h, \boldsymbol{V}_h) + (\boldsymbol{\mu}\partial_t \boldsymbol{U}', \boldsymbol{V}_h) + (\boldsymbol{A}_i\partial_i \boldsymbol{U}_h, \boldsymbol{V}_h) + (\boldsymbol{A}_i\partial_i \boldsymbol{U}', \boldsymbol{V}_h) = (\boldsymbol{F}, \boldsymbol{V}_h), \tag{35}$$

$$(\boldsymbol{\mu}\partial_t \boldsymbol{U}_h, \boldsymbol{V}') + (\boldsymbol{\mu}\partial_t \boldsymbol{U}', \boldsymbol{V}') + (\boldsymbol{A}_i\partial_i \boldsymbol{U}_h, \boldsymbol{V}') + (\boldsymbol{A}_i\partial_i \boldsymbol{U}', \boldsymbol{V}') = (\boldsymbol{F}, \boldsymbol{V}'), \tag{36}$$

for all $\boldsymbol{V}_h$ in the finite element space and $\boldsymbol{V}'$ in the space of subscales. This set of equations is obviously equivalent to the classical weak form of the problem.

The objective is to introduce some approximations to compute $\boldsymbol{U}'$ and, as a consequence, leading to the stabilized finite element problem for $\boldsymbol{U}_h$. First of all, we consider that both $\boldsymbol{U}_h$ and $\boldsymbol{U}'$ are continuous across interelement boundaries. Problem (35)–(36) can be written as

$$(\boldsymbol{\mu}\partial_t \boldsymbol{U}_h, \boldsymbol{V}_h) + (\boldsymbol{A}_i\partial_i \boldsymbol{U}_h, \boldsymbol{V}_h) + (\boldsymbol{\mu}\partial_t \boldsymbol{U}', \boldsymbol{V}_h) - (\boldsymbol{U}', \boldsymbol{A}_i^{\mathrm{t}}\partial_i \boldsymbol{V}_h) = (\boldsymbol{F}, \boldsymbol{V}_h), \tag{37}$$

$$\boldsymbol{\mu}\partial_t \boldsymbol{U}' + P'(\boldsymbol{A}_i\partial_i \boldsymbol{U}') = P'(\boldsymbol{F} - (\boldsymbol{\mu}\partial_t \boldsymbol{U}_h + \boldsymbol{A}_i\partial_i \boldsymbol{U}_h)) =: \boldsymbol{R}_h, \tag{38}$$

where P' is the L^2-projection onto the space of subscales. The time derivative of $\boldsymbol{U}'$ could be neglected assuming that $\boldsymbol{U}'$ varies in time much more slowly than $\boldsymbol{U}_h$. This is defined in [8] as the assumption of *quasi-static* subscales. However, here we will keep this time derivative, leading to what is termed in [13,12] in the context of the incompressible Navier–Stokes equations as *dynamic subscales*. It is shown in the references indicated that it improves considerably the time stability of the resulting formulation. The assumption of quasi-static subscales will be used in this section when explaining the stabilization *in space* introduced by the formulation proposed.

It could also be possible to approximate $\boldsymbol{U}$ or some of its components using discontinuous finite element functions, and therefore $\boldsymbol{V}_h$ would also be discontinuous. In this case, it would be necessary to approximate also the subscales on the element boundaries.

Eq. (38) needs to be approximated to obtain a closed-form expression for $\boldsymbol{U}'$ which, once inserted into (37), will lead to the stabilized finite element problem for $\boldsymbol{U}_h$. Observe that $\boldsymbol{R}_h$ in (38) can be considered as the residual of the finite element approximation projected onto the space of subscales.

The basic heuristic idea is to consider that since $\boldsymbol{U}'$ is the component of the unknown unresolved by the finite element space, its Fourier transform in space must be dominated by wave numbers of the form $\frac{1}{h}\boldsymbol{k}$, where $\boldsymbol{k} = (k_1,\ldots,k_d)$ is dimensionless and of order $\mathcal{O}(1)$ and h is the mesh size.

As done in the previous section for the wave equation, let $\boldsymbol{M}$ be a symmetric and positive matrix that defines an inner product in the space of forcing terms, and let $|\cdot|_M$ be the norm with respect to this matrix, that is, $|\boldsymbol{F}|_M^2 = \boldsymbol{F}^{\mathrm{t}}\boldsymbol{M}\boldsymbol{F}$ (recall that $\boldsymbol{F}^{\mathrm{t}}\boldsymbol{F}$ in general is not even dimensionally meaningful, see (15)). Likewise, let $\|\cdot\|_{K,M}$ be the $L^2(K)$-norm of $|\cdot|_M$, where K is the domain of a generic element of the finite element mesh. Similar definitions apply for the inner product in the space of unknowns weighted with $\boldsymbol{M}^{-1}$.

The objective is to approximate the spatial operator $P'(\boldsymbol{A}_i\partial_i \boldsymbol{U}')$ in (38). Neglecting boundary values of $\boldsymbol{U}'$ on the element boundaries, its Fourier transform can be approximated by

$$P'(\widehat{\boldsymbol{A}_i\partial_i \boldsymbol{U}'}) \approx \mathscr{S}(\boldsymbol{k})\widehat{\boldsymbol{U}'} \equiv -\mathrm{i}\frac{1}{h}k_j\boldsymbol{A}_j\widehat{\boldsymbol{U}'},$$

where the Fourier transform of a function f has been denoted by $\hat{f}$ and $\mathrm{i}=\sqrt{-1}$. Taking the M-norm of $\mathscr{S}(\boldsymbol{k})\widehat{\boldsymbol{U}'}$ yields

$$\widehat{\boldsymbol{U}'}^{\mathrm{t}}\overline{\mathscr{S}(\boldsymbol{k})}^{\mathrm{t}}\boldsymbol{M}\mathscr{S}(\boldsymbol{k})\widehat{\boldsymbol{U}'} = \widehat{\boldsymbol{U}'}^{\mathrm{t}}\left(\frac{1}{h^2}k_ik_j\boldsymbol{A}_i^{\mathrm{t}}\boldsymbol{M}\boldsymbol{A}_j\right)\widehat{\boldsymbol{U}'},$$

where $\overline{\mathscr{S}(\boldsymbol{k})}$ is the complex conjugate of $\mathscr{S}(\boldsymbol{k})$. Neglecting again boundary values of $\boldsymbol{U}'$ on the element boundaries, Plancherel's formula yields

$$\begin{aligned}
\|P'(\boldsymbol{A}_i\partial_i \boldsymbol{U}')\|_{K,M} &\approx \|P'(\widehat{\boldsymbol{A}_i\partial_i \boldsymbol{U}'})\|_M \\
&\approx \int |\mathscr{S}(\boldsymbol{k})\widehat{\boldsymbol{U}'}|_M^2\mathrm{d}\boldsymbol{k} \\
&\leqslant \int |\mathscr{S}(\boldsymbol{k})|_M^2|\widehat{\boldsymbol{U}'}|_{M^{-1}}^2\mathrm{d}\boldsymbol{k} \\
&= \int |\mathscr{S}(\boldsymbol{k}^0)|_M^2|\widehat{\boldsymbol{U}'}|_{M^{-1}}^2\mathrm{d}\boldsymbol{k} \\
&= |\mathscr{S}(\boldsymbol{k}^0)|_M^2\|\widehat{\boldsymbol{U}'}\|_{M^{-1}}^2 \\
&\approx |\mathscr{S}(\boldsymbol{k}^0)|_M^2\|\boldsymbol{U}'\|_{K,M^{-1}}^2,
\end{aligned} \tag{39}$$

where $\boldsymbol{k}^0$ is a wave number whose existence is guaranteed by the mean value theorem and the integrals extend over the wave number space.

Let us consider now the approximation $P'(\boldsymbol{A}_i\partial_i \boldsymbol{U}') \approx \boldsymbol{\tau}^{-1}\boldsymbol{U}'$, where $\boldsymbol{\tau}$ is a symmetric and positive definite matrix to be determined. Clearly, a sufficient condition for (39) to hold also for this approximation is that $|\boldsymbol{\tau}^{-1}|_M^2 \leqslant |\mathscr{S}(\boldsymbol{k}^0)|_M^2$. In particular, *our proposal is to design* $\boldsymbol{\tau}$ *so that the equality holds*, that is to say,

$$\sup_{|\boldsymbol{X}|_{M^{-1}}=1} \boldsymbol{X}^{\mathrm{t}}\boldsymbol{\tau}^{-1}\boldsymbol{M}\boldsymbol{\tau}^{-1}\boldsymbol{X} = \sup_{|\boldsymbol{X}|_{M^{-1}}=1} \boldsymbol{X}^{\mathrm{t}}\frac{1}{h^2}(k_i^0k_j^0\boldsymbol{A}_i^{\mathrm{t}}\boldsymbol{M}\boldsymbol{A}_j)\boldsymbol{X}. \tag{40}$$

Of course $\boldsymbol{k}^0$ is unknown, and their components have to be understood as algorithmic constants. The hope is that the approximated subscale will bound the residual of the finite element solution as the exact subscale bounds the residual of the finite element component of the exact solution. This bound will not be exactly the same, but will have the same asymptotic behavior in terms of h and the coefficients of the equation to be solved.

A practical way to impose condition (40) is to compute the spectrum of matrices $\boldsymbol{\tau}^{-1}\boldsymbol{M}\boldsymbol{\tau}^{-1}$ and $h^{-2}(k_i^0k_j^0\boldsymbol{A}_i^{\mathrm{t}}\boldsymbol{M}\boldsymbol{A}_j)$

with respect to matrix $\boldsymbol{M}^{-1}$ and impose that at least the spectral radius be the same.

The final problem for the finite element component of the unknown is obtained by inserting the approximation $P'(\boldsymbol{A}_i\partial_i\boldsymbol{U}') \approx \boldsymbol{\tau}^{-1}\boldsymbol{U}'$ into (38), which leads to the *ordinary* differential equation

$$\boldsymbol{\mu}\partial_t\boldsymbol{U}' + \boldsymbol{\tau}^{-1}\boldsymbol{U}' = \boldsymbol{R}_h, \quad (41)$$

which needs to be solved at each integration point of each element. The most natural way to do this is to use the same time integration scheme for the subscale $\boldsymbol{U}'$ as for the finite element unknown $\boldsymbol{U}_h$. Note that $\boldsymbol{U}'$ needs to be stored at the integration points. An efficient way to do this is explained in [12]. The expression for $\boldsymbol{U}'$ at each time step inserted into (37) yields the stabilized finite element problem for $\boldsymbol{U}_h$.

In the particular case of quasi-static subscales, that is, when $\boldsymbol{\mu}\partial_t\boldsymbol{U}' \approx 0$, the problem can be simplified. Noting that the approximation for the subscale is *local* to each element and using the notation $\langle\cdot,\cdot\rangle_K$ for the integral of the product of two functions on a generic element K, the final discrete variational equation is

$$(\partial_t\boldsymbol{U}_h, \boldsymbol{V}_h) + (\boldsymbol{A}_i\partial_i\boldsymbol{U}_h, \boldsymbol{V}_h) - (\boldsymbol{F}, \boldsymbol{V}_h) + \sum_K \langle \boldsymbol{A}_i^{\mathrm{t}}\partial_i\boldsymbol{V}_h, \boldsymbol{\tau}P'(\partial_t\boldsymbol{U}_h + \boldsymbol{A}_i\partial_i\boldsymbol{U}_h - \boldsymbol{F})\rangle_K = 0, \quad (42)$$

where the sum extends to all elements of the finite element partition. We will apply now this general framework to the

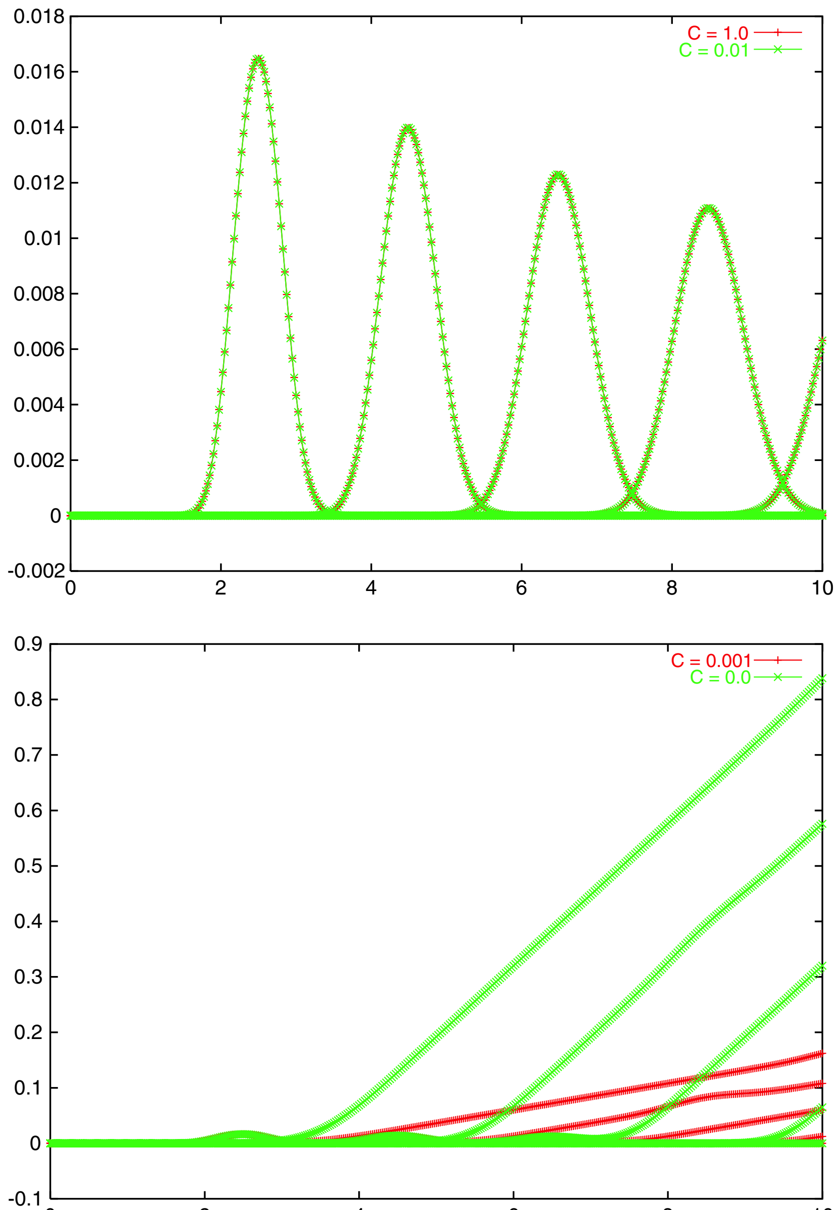

Fig. 1. Propagation of the solitary wave using the backward Euler scheme for different values of the constant in the stabilization parameter. Backward Euler time integration.

problem we are considering. The first step is to design matrix $\boldsymbol{\tau}$.

4.2. Application to the mixed form of the wave equation

4.2.1. Stabilization parameters

Our objective now is to design the matrix of stabilization parameters $\boldsymbol{\tau}$ so that (40) is satisfied. In fact, the optimal situation would be to choose $\boldsymbol{\tau}$ satisfying $\boldsymbol{\tau}^{-1}\boldsymbol{M}\boldsymbol{\tau}^{-1} = h^{-2}(k_i^0 k_j^0 \boldsymbol{A}_i^{\mathrm{t}}\boldsymbol{M}\boldsymbol{A}_j)$. However, we will also try to choose $\boldsymbol{\tau}$ as simple as possible. In particular, we shall see that it is possible to take $\boldsymbol{\tau}$ *diagonal* and satisfy condition (40), although the equality just mentioned *will not hold*.

For the sake of simplicity in the notation, we will consider the case $d = 2$, although the final result applies also to $d = 3$. For the matrices given by (34) it is found that

$$\frac{1}{h^2} k_i^0 k_j^0 \boldsymbol{A}_i^{\mathrm{t}}\boldsymbol{M}\boldsymbol{A}_j = \frac{1}{h^2}\begin{bmatrix} m_u|\boldsymbol{k}^0|^2 & 0 & 0 \\ 0 & m_\eta(k_1^0)^2 & m_\eta k_1^0 k_2^0 \\ 0 & m_\eta k_1^0 k_2^0 & m_\eta(k_2^0)^2 \end{bmatrix},$$

where m_η and m_u are defined in (16). This matrix is singular. This is due to the fact that it does not contain the information on the boundary conditions that allows to invert the differential operator from where it comes.

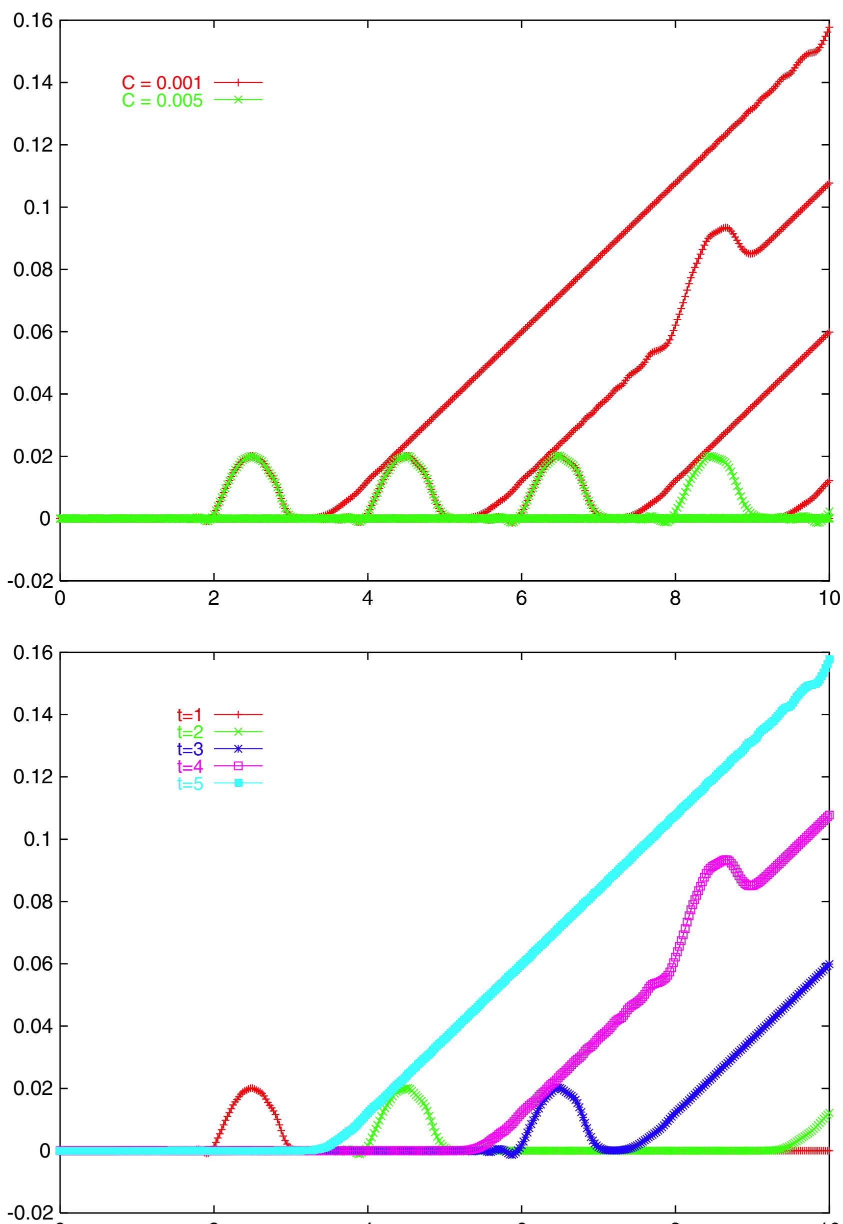

Fig. 2. Propagation of the boundary effect for a one dimensional wave. Top: numerical results for $C = 0.005$ (stable) and $C = 0.001$ (unstable). Bottom: evolution of the profile in the unstable case. Crank–Nicolson time integration in both cases.

Let us denote by $\mathrm{Spec}_{M^{-1}}(\boldsymbol{B})$ the spectrum of a matrix $\boldsymbol{B}$ with respect to the inner product $\boldsymbol{M}^{-1}$, that is, the set of eigenvalues λ of the generalized eigenvalue problem $\boldsymbol{BU} = \lambda \boldsymbol{M}^{-1}\boldsymbol{U}$. It turns out that

$$\mathrm{Spec}_{M^{-1}}(h^{-2}k_i^0k_j^0\boldsymbol{A}_i^{\mathrm{t}}\boldsymbol{MA}_j) = \left\{\frac{C_{\mathrm{L}}^2|\boldsymbol{k}^0|^2}{h^2}, 0, \frac{C_{\mathrm{L}}^2|\boldsymbol{k}^0|^2}{h^2}\right\}.$$

Obviously, $\boldsymbol{k}^0$ is unknown, so that we may consider its norm as an algorithmic constant.

Let us assume now that $\boldsymbol{\tau}$ is diagonal. Since the two scalar equations for the components of $\boldsymbol{u}$ have the same form, we may take it as $\boldsymbol{\tau} = \mathrm{diag}(\tau_\eta, \tau_u, \tau_u)$. Clearly, $\boldsymbol{\tau}^{-1}\boldsymbol{M}\boldsymbol{\tau}^{-1}$ will also be diagonal, and it is impossible that it behaves as $h^{-2}(k_i^0k_j^0\boldsymbol{A}_i^{\mathrm{t}}\boldsymbol{MA}_j)$. However, since $\mathrm{Spec}_{M^{-1}}(\boldsymbol{\tau}^{-1}\boldsymbol{M}\boldsymbol{\tau}^{-1}) = \{\tau_\eta^{-2}m_\eta, \tau_u^{-2}m_u, \tau_u^{-2}m_u\}$, a way to choose $\boldsymbol{\tau}$ satisfying (40) is to take τ_η and τ_u satisfying

$$\tau_\eta^{-2}m_\eta = \frac{C_{\mathrm{L}}^2|\boldsymbol{k}^0|^2}{h^2}, \quad \tau_u^{-2}m_u = \frac{C_{\mathrm{L}}^2|\boldsymbol{k}^0|^2}{h^2}.$$

Calling the algorithmic constant $C_\tau \equiv |\boldsymbol{k}^0|^{-1}$, it follows that

$$\boldsymbol{\tau} = (\tau_\eta, \tau_u\boldsymbol{I}), \quad \tau_\eta = C_\tau h\sqrt{\frac{\mu_u}{\mu_\eta}}, \quad \tau_u = C_\tau h\sqrt{\frac{\mu_\eta}{\mu_u}}. \tag{43}$$

This is the expression we were looking for. Note that it does not depend on the choice of the lengthscale C_{L}, as it has been anticipated before.

It can be readily checked that, when $d = 1$, $\boldsymbol{\tau}^{-1}\boldsymbol{M}\boldsymbol{\tau}^{-1} = h^{-2}(k_i^0k_j^0\boldsymbol{A}_i^{\mathrm{t}}\boldsymbol{MA}_j)$, which is the optimal situation, whereas in the case $d = 2$ matrix $h^{-2}(k_i^0k_j^0\boldsymbol{A}_i^{\mathrm{t}}\boldsymbol{MA}_j)$ has two eigenvalues equal to the diagonal entries of $\boldsymbol{\tau}^{-1}\boldsymbol{M}\boldsymbol{\tau}^{-1}$, and the third one is zero. Condition (40), however, is satisfied.

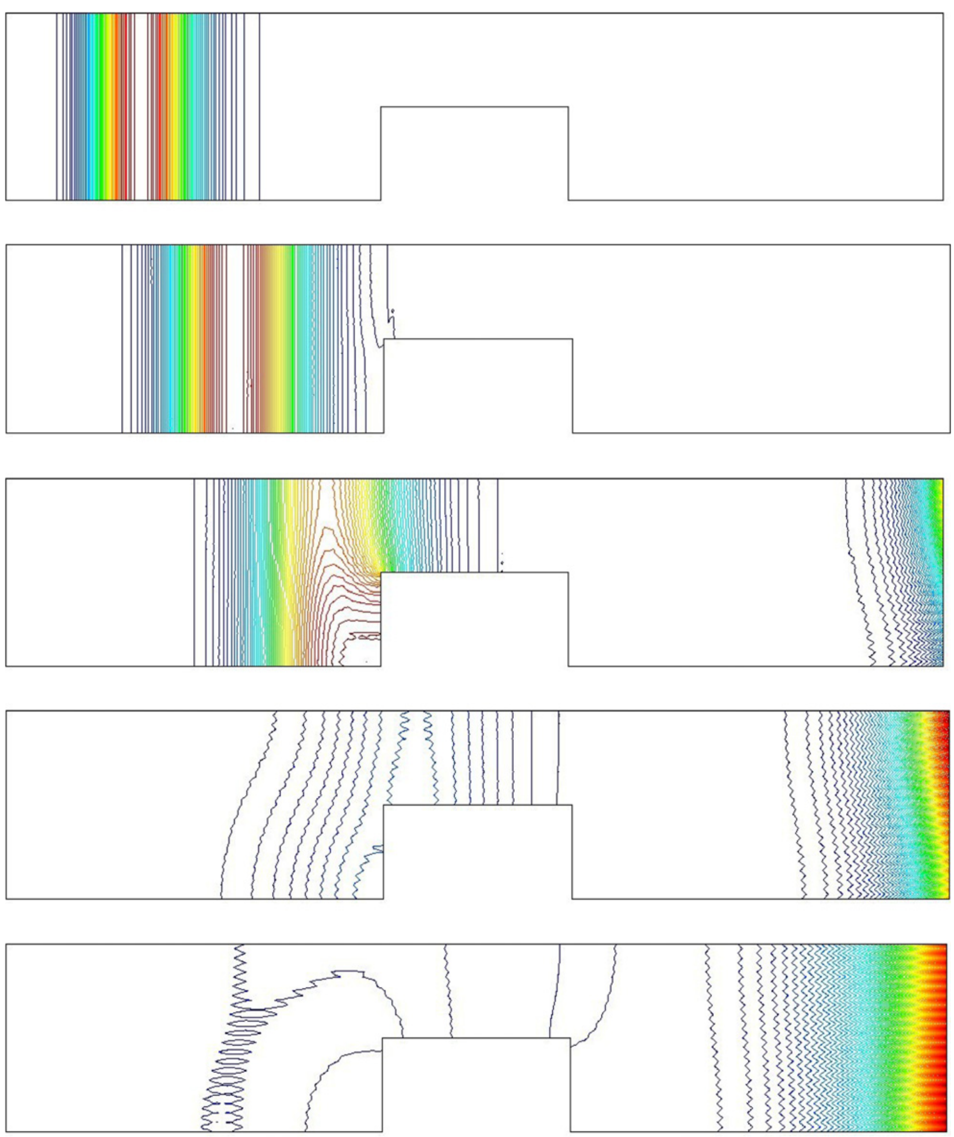

Fig. 3. Galerkin solution of the propagation of a wave over an obstacle. From the top to the bottom: $t = 1, 2, 3, 4$ and 5.

4.2.2. Stabilized formulation

Once the matrix of stabilization parameters has been designed, the stabilized formulation is now complete. Let us see which is the final discrete variational form for the wave problem in mixed form. Assuming for simplicity that the subscales are quasi-static (that is to say, using (42)) and taking the mesh size uniform, so that the stabilization parameters in (43) are the same for all the elements of the finite element mesh, the equations for $(\eta_h, \boldsymbol{u}_h)$ are

$$\begin{aligned} 0 = {} & \mu_\eta(\partial_t\eta_h, \xi_h) - (\boldsymbol{u}_h, \nabla\xi_h) - (f_\eta, \xi_h) \\ & + \mu_u(\partial_t\boldsymbol{u}_h, \boldsymbol{v}_h) + (\nabla\eta_h, \boldsymbol{v}_h) - (\boldsymbol{f}_u, \boldsymbol{v}_h) \\ & + \tau_\eta(P'(\mu_\eta\partial_t\eta_h + \nabla\cdot\boldsymbol{u}_h - f_\eta), \nabla\cdot\boldsymbol{v}_h) \\ & + \tau_u(P'(\mu_u\partial_t\boldsymbol{u}_h + \nabla\eta_h - \boldsymbol{f}_u), \nabla\xi_h), \end{aligned} \tag{44}$$

which must hold for all test functions ξ_h and $\boldsymbol{v}_h$ in the appropriate spaces. The terms in the first row of this variational equation correspond to the Galerkin contribution, whereas those multiplied by τ_η and τ_u should provide stabilization.

Let us comment two possible projections P' that are also used in other applications of stabilized finite element methods:

- $P' = I$ (identity). This could be considered the most common option in stabilized finite element methods (see [6] and references therein).
- $P' = P_h^\perp$, where P_h is the projection onto the adequate finite element space. Stabilized formulations arising from subscales orthogonal to the finite element space were first proposed in [7].

In this paper we will concentrate in the second option, although *similar stability properties hold for the first*. We will see in the next subsection that the terms added to the Galerkin formulation indeed provide additional stability,

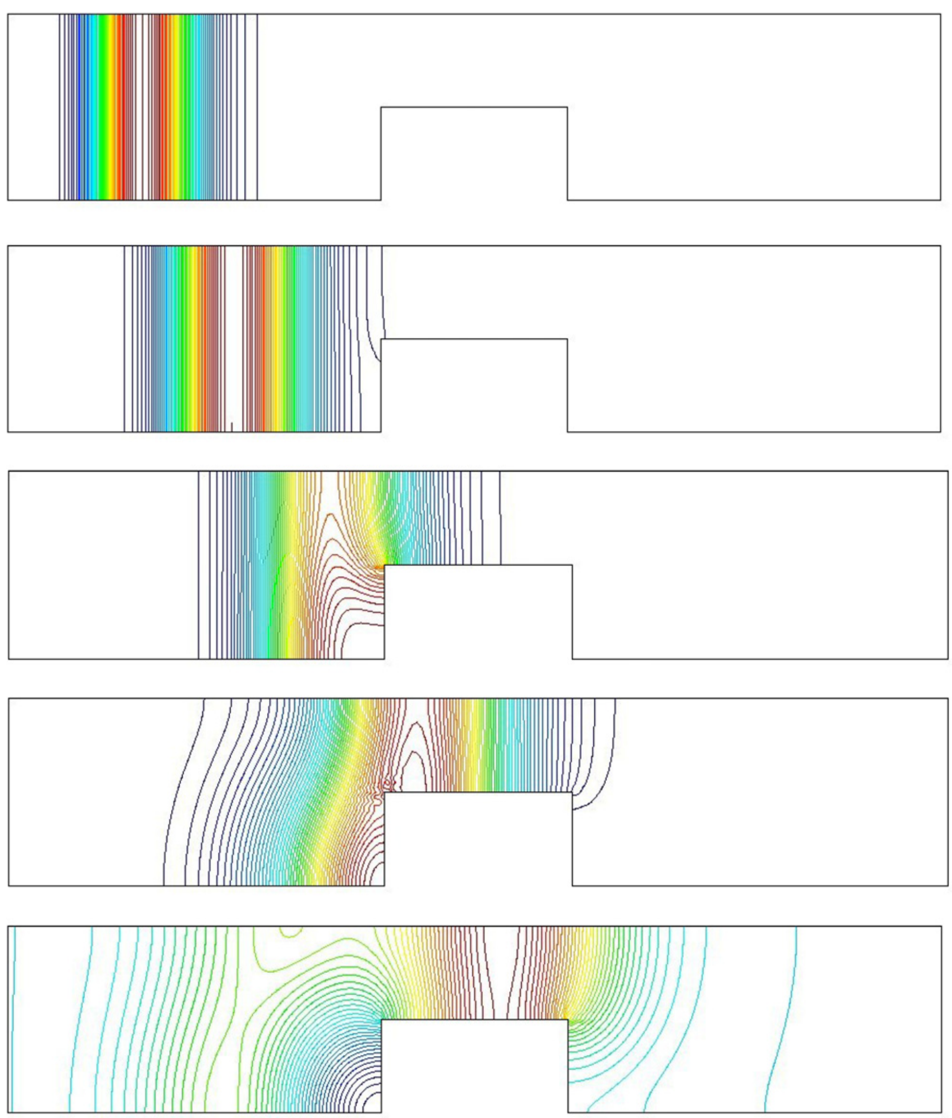

Fig. 4. Stabilized solution ($C = 0.01$) of the propagation of a wave over an obstacle. From the top to the bottom: $t = 1, 2, 3, 4$ and 5.

and in Section 5 this enhanced stability will be numerically demonstrated.

4.2.3. Stability estimate

Let us consider the case $P' = P_h^{\perp}$, that is to say, the space of subscales is orthogonal to the finite element space and, as before, that the subscales are quasi-static. It has to be noted that in this case $P'(\partial_t \eta_h) = 0$, $P'(\partial_t \boldsymbol{u}_h) = 0$. The mass matrix of the linear system will *not* be modified with respect to the Galerkin method.

We assume that η and $\boldsymbol{u}$ are interpolated using equal continuous functions. We will obtain in this situation a stability estimate for the finite element unknowns $\eta_h(\boldsymbol{x}, t)$ and $\boldsymbol{u}_h(\boldsymbol{x}, t)$, solution of the semidiscrete problem (continuous in time). For that purpose, it is enough to consider the case without forcing terms in (44). The problem to be considered is then

$$0 = \mu_\eta(\partial_t \eta_h, \xi_h) - (\boldsymbol{u}_h, \nabla \xi_h) + \mu_u(\partial_t \boldsymbol{u}_h, \boldsymbol{v}_h) + (\nabla \eta_h, \boldsymbol{v}_h) + \tau_\eta(P_h^{\perp}(\nabla \cdot \boldsymbol{u}_h), \nabla \cdot \boldsymbol{v}_h) + \tau_u(P_h^{\perp}(\nabla \eta_h), \nabla \xi_h). \tag{45}$$

If at each time t we take $\xi_h = \eta_h$, $\boldsymbol{v}_h = \boldsymbol{u}_h$, it is found that

$$\frac{1}{2}\mu_\eta \frac{\mathrm{d}}{\mathrm{d}t}\|\eta_h\|^2 + \frac{1}{2}\mu_u \frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{u}_h\|^2 + \tau_\eta\|P_h^{\perp}(\nabla \cdot \boldsymbol{u}_h)\|^2 + \tau_u\|P_h^{\perp}(\nabla \eta_h)\|^2 = 0.$$

Here and in what follows, we use the abbreviation $\|\cdot\| \equiv \|\cdot\|_{L^2(\Omega)}$. Integrating from $t = 0$ to any time t' one gets

$$\frac{1}{2}\mu_\eta\|\eta_h(t')\|^2 + \frac{1}{2}\mu_u\|\boldsymbol{u}_h(t')\|^2 + \tau_\eta \int_0^{t'} \|P_h^{\perp}(\nabla \cdot \boldsymbol{u}_h(t))\|^2 \mathrm{d}t + \tau_u \int_0^{t'} \|P_h^{\perp}(\nabla \eta_h(t))\|^2 \mathrm{d}t = \frac{1}{2}\mu_\eta\|\eta_h(0)\|^2 + \frac{1}{2}\mu_u\|\boldsymbol{u}_h(0)\|^2. \tag{46}$$

If now we differentiate (45) with respect to time and at each time t we take $\xi_h = \partial_t \eta_h$, $\boldsymbol{v}_h = \partial_t \boldsymbol{u}_h$ we get

$$\frac{1}{2}\mu_\eta \frac{\mathrm{d}}{\mathrm{d}t}\|\partial_t \eta_h\|^2 + \tau_\eta\|P_h^{\perp}(\nabla \cdot \partial_t \boldsymbol{u}_h)\|^2 + \frac{1}{2}\mu_u \frac{\mathrm{d}}{\mathrm{d}t}\|\partial_t \boldsymbol{u}_h\|^2 + \tau_u\|P_h^{\perp}(\nabla \partial_t \eta_h)\|^2 = 0,$$

which after integration from $t = 0$ to any time t' yields

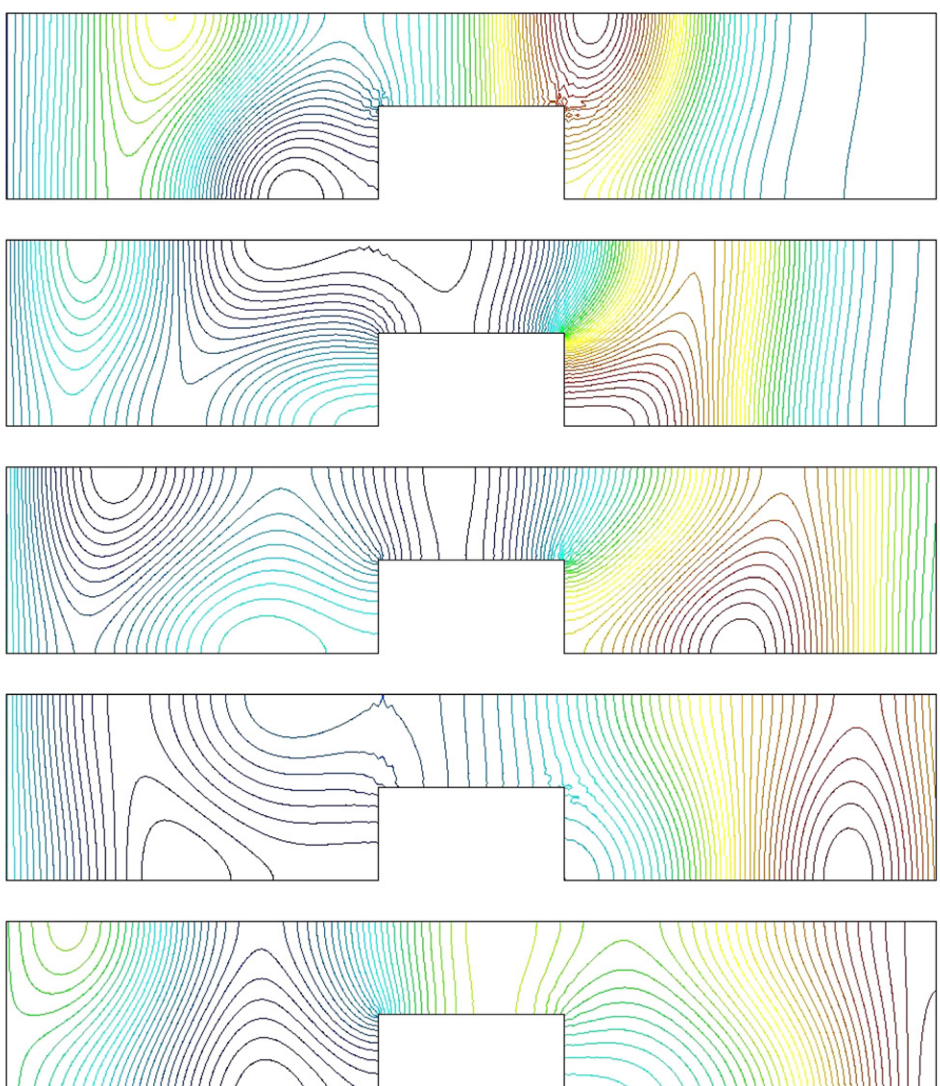

Fig. 5. Stabilized solution ($C = 0.01$) of the propagation of a wave over an obstacle. From the top to the bottom: $t = 6$, 7, 8, 9 and 10.

$$\frac{1}{2}\mu_\eta\|\partial_t\eta_h(t')\|^2+\frac{1}{2}\mu_u\|\partial_t\boldsymbol{u}_h(t')\|^2 + \tau_\eta\int_0^{t'}\|P_h^\perp(\nabla\cdot\partial_t\boldsymbol{u}_h(t))\|^2\mathrm{d}t+\tau_u\int_0^{t'}\|P_h^\perp(\nabla\partial_t\eta_h(t))\|^2\mathrm{d}t \leqslant \frac{1}{2}\mu_\eta\|\partial_t\eta_h(0)\|^2+\frac{1}{2}\mu_u\|\partial_t\boldsymbol{u}_h(0)\|^2. \quad (47)$$

We assume now that the family of finite element meshes is quasi-uniform, so that there exists a constant C_{inv} such that

$$\|\nabla v_h\| \leqslant \frac{C_{\text{inv}}}{h}\|v_h\| \quad (48)$$

for all functions v_h in the finite element spaces, either of η or of $\boldsymbol{u}$.

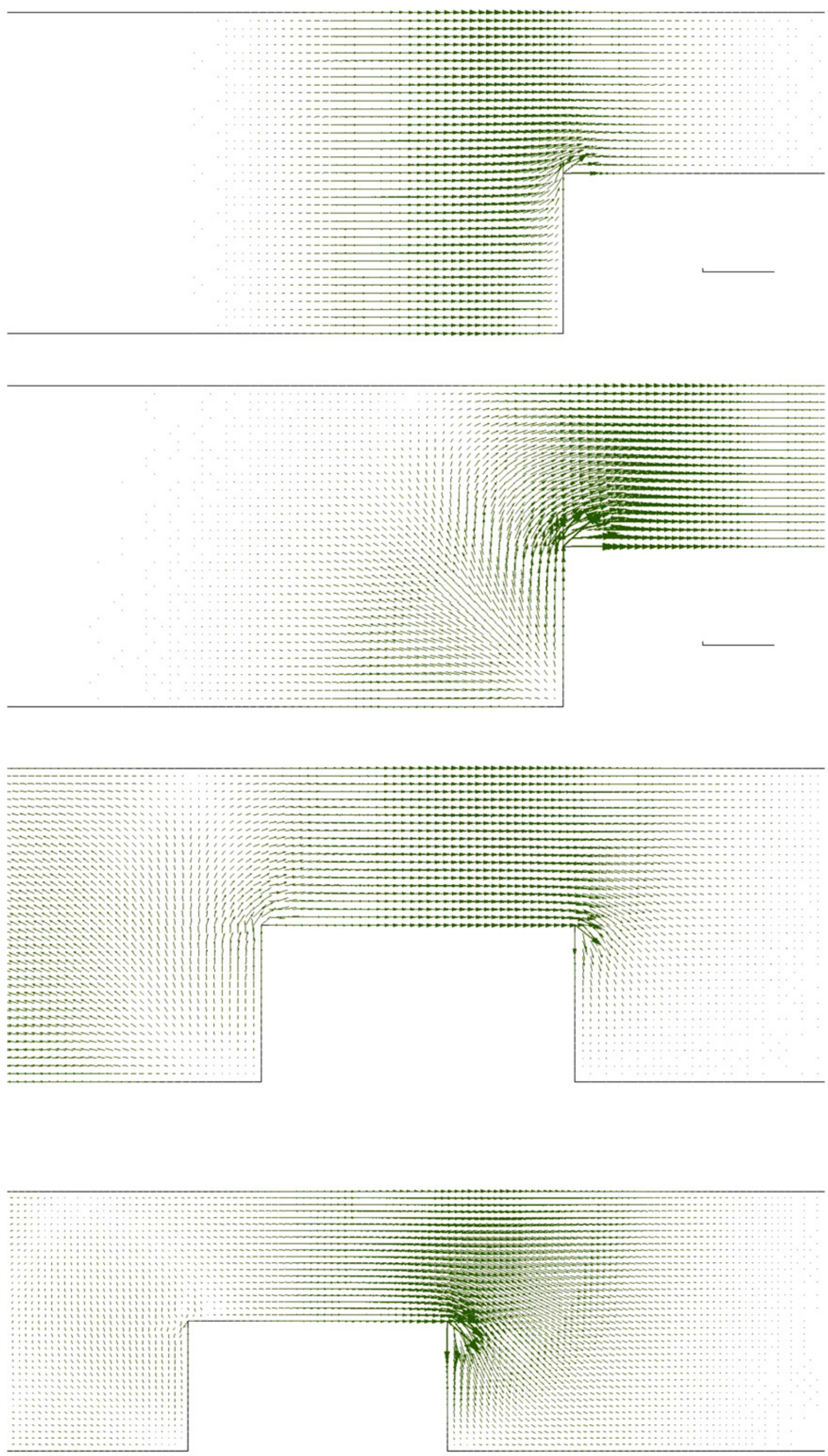

Fig. 6. Some details of the vector field u for the stabilized solution of the propagation of a wave over an obstacle. From the top to the bottom: $t = 3, 4, 5$ and 6.

Evaluating (45) at $t = 0$, taking $\xi_h = \partial_t \eta_h(0)$, $\boldsymbol{v}_h = \partial_t \boldsymbol{u}_h(0)$, using Young's inequality $ab \leqslant \frac{1}{2\alpha}a^2 + \frac{\alpha}{2}b^2$ for all constants $\alpha > 0$, and using the inverse estimate (48) and the expression of τ_η and τ_u in (43), we obtain

$$\begin{aligned}
&\mu_\eta \|\partial_t \eta_h(0)\|^2 + \mu_u \|\partial_t \boldsymbol{u}_h(0)\|^2 \\
&\quad \leqslant \|\nabla \cdot \boldsymbol{u}_h(0)\| \|\partial_t \eta_h(0)\| + \tau_\eta \|\nabla \cdot \boldsymbol{u}_h(0)\| \|\nabla \cdot \partial_t \boldsymbol{u}_h(0)\| \\
&\qquad + \|\nabla \eta_h(0)\| \|\partial_t \boldsymbol{u}_h(0)\| + \tau_u \|\nabla \eta_h(0)\| \|\nabla \partial_t \eta_h(0)\| \\
&\quad \leqslant \frac{1}{2\alpha_1} \mu_\eta \|\partial_t \eta_h(0)\|^2 + \frac{\alpha_1}{2} \frac{1}{\mu_\eta} \|\nabla \cdot \boldsymbol{u}_h(0)\|^2 \\
&\qquad + \frac{1}{2\alpha_2} \mu_u \|\partial_t \boldsymbol{u}_h(0)\|^2 + \frac{\alpha_2}{2} \frac{(C_\tau C_{\mathrm{inv}})^2}{\mu_\eta} \|\nabla \cdot \boldsymbol{u}_h(0)\|^2 \\
&\qquad + \frac{1}{2\alpha_3} \mu_u \|\partial_t \boldsymbol{u}_h(0)\|^2 + \frac{\alpha_3}{2} \frac{1}{\mu_u} \|\nabla \eta_h(0)\|^2 \\
&\qquad + \frac{1}{2\alpha_4} \mu_\eta \|\partial_t \eta_h(0)\|^2 + \frac{\alpha_4}{2} \frac{(C_\tau C_{\mathrm{inv}})^2}{\mu_u} \|\nabla \eta_h(0)\|^2.
\end{aligned}$$

Taking the constants α_i, $i = 1, \ldots, 4$, sufficiently large, it follows that there exists a constant C for which

$$\mu_\eta \|\partial_t \eta_h(0)\|^2 + \mu_u \|\partial_t \boldsymbol{u}_h(0)\|^2 \leqslant C\left(\frac{1}{\mu_\eta} \|\nabla \cdot \boldsymbol{u}_h(0)\|^2 + \frac{1}{\mu_u} \|\nabla \eta_h(0)\|^2 \right). \quad (49)$$

Here and in what follows, C will denote a generic constant, not necessarily the same in different appearances, but now independent of the coefficients μ_η and μ_u.

To obtain stability for $\nabla \cdot \boldsymbol{u}_h$ and $\nabla \eta_h$ it is seen from (46) that we need to bound only their component in the appropriate finite element space. For that, we can take as tests functions in (45)

$$\xi_h = \tau_\eta P_h(\bar{\xi}_h), \quad \bar{\xi}_h = \nabla \cdot \boldsymbol{u}_h, \boldsymbol{v}_h = \tau_u P_h(\bar{\boldsymbol{v}}_h), \quad \bar{\boldsymbol{v}}_h = \nabla \eta_h,$$

which yields

$$\begin{aligned}
&\tau_\eta \|P_h(\bar{\xi}_h)\|^2 + \tau_u \|P_h(\bar{\boldsymbol{v}}_h)\|^2 \\
&\quad \leqslant \tau_\eta \mu_\eta \|\partial_t \eta_h\| \|P_h(\bar{\xi}_h)\| + \tau_u \mu_u \|\partial_t \boldsymbol{u}_h\| \|P_h(\bar{\boldsymbol{v}}_h)\| \\
&\qquad + \tau_\eta \tau_u \|P_h^\perp(\bar{\xi}_h)\| \|P_h^\perp(\nabla \cdot P_h(\bar{\boldsymbol{v}}_h))\| \\
&\qquad + \tau_\eta \tau_u \|P_h^\perp(\bar{\boldsymbol{v}}_h)\| \|P_h^\perp(\nabla P_h(\bar{\xi}_h))\| \\
&\quad \leqslant \frac{1}{2\alpha_1} \tau_\eta \|P_h(\bar{\xi}_h)\|^2 + \frac{\alpha_1}{2} \tau_\eta \mu_\eta^2 \|\partial_t \eta_h\|^2 \\
&\qquad + \frac{1}{2\alpha_2} \tau_u \|P_h(\bar{\boldsymbol{v}}_h)\|^2 + \frac{\alpha_2}{2} \tau_u \mu_u^2 \|\partial_t \boldsymbol{u}_h\|^2 \\
&\qquad + \frac{1}{2\alpha_3} \tau_u \|P_h(\bar{\boldsymbol{v}}_h)\|^2 + \frac{\alpha_3}{2} (C_{\mathrm{inv}} C_\tau)^2 \tau_\eta \|P_h^\perp(\xi_h)\|^2 \\
&\qquad + \frac{1}{2\alpha_4} \tau_\eta \|P_h(\bar{\xi}_h)\|^2 + \frac{\alpha_4}{2} (C_{\mathrm{inv}} C_\tau)^2 \tau_u \|P_h^\perp(\bar{\boldsymbol{v}}_h)\|^2.
\end{aligned}$$

Once again, taking the constants α_i, $i = 1, \ldots, 4$, sufficiently large, it follows that there exists a constant C for which

$$\begin{aligned}
\tau_\eta \|P_h(\bar{\xi}_h)\|^2 + \tau_u \|P_h(\bar{\boldsymbol{v}}_h)\|^2 &\leqslant C(\tau_\eta \mu_\eta^2 \|\partial_t \eta_h\|^2 + \tau_u \mu_u^2 \|\partial_t \boldsymbol{u}_h\|^2 \\
&\quad + \tau_\eta \|P_h^\perp(\bar{\xi}_h)\|^2 + \tau_u \|P_h^\perp(\bar{\boldsymbol{v}}_h)\|^2)
\end{aligned}$$

which combined with (46), (47) and (49), and noting that $\tau_\eta \mu_\eta = \tau_u \mu_u$, yields

$$\begin{aligned}
&\tau_\eta \int_0^{t'} \|P_h(\bar{\xi}_h)\|^2 \mathrm{d}t + \tau_u \int_0^{t'} \|P_h(\bar{\boldsymbol{v}}_h)\|^2 \mathrm{d}t \\
&\quad \leqslant C(\mu_\eta \|\eta_h(0)\|^2 + \mu_u \|\boldsymbol{u}_h(0)\|^2 + \tau_\eta \|\nabla \cdot \boldsymbol{u}_h(0)\|^2 t' \\
&\qquad + \tau_u \|\nabla \eta_h(0)\|^2 t').
\end{aligned}$$

This, together with (46), implies the stability estimates we were looking for:

$$\mu_\eta \max_{0<s<t} \|\eta_h(s)\|^2 + \mu_u \max_{0<s<t} \|\boldsymbol{u}_h(s)\|^2 \leqslant \mu_\eta \|\eta_h^0\|^2 + \mu_u \|\boldsymbol{u}_h^0\|^2, \quad (50)$$

$$\begin{aligned}
&\tau_\eta \int_0^t \|\bar{\xi}_h(s)\|^2 \mathrm{d}s + \tau_u \int_0^t \|\bar{\boldsymbol{v}}_h(s)\|^2 \mathrm{d}s \\
&\quad \leqslant C(\mu_\eta \|\eta_h^0\|^2 + \mu_u \|\boldsymbol{u}_h^0\|^2 + \tau_\eta \|\nabla \cdot \boldsymbol{u}_h^0\|^2 t + \tau_u \|\nabla \eta_h^0\|^2 t).
\end{aligned} \quad (51)$$

From the numerical point of view, estimate (50), which bounds the $C^0(0, T; L^2(\Omega))$ norm of the unknowns, is weaker that (51), since in this last case we have some control on the divergence of $\boldsymbol{u}_h$ and the gradient of η_h.

5. Numerical examples

In this section we present the results of two very simple numerical examples. Our intention is only to show that the Galerkin formulation applied to the mixed form of the wave equation is unstable and that this instability can be overcome by using the stabilized formulation proposed here. For two more elaborate examples, with application to nonlinear oscillations of the water elevation in harbors, the reader is referred to [11] where, in particular, an accuracy test is presented.

Since our objective is to point out a instability *in space* of the Galerkin method, we will also use very simple time integration schemes. In particular, to make sure that the source of instability is not the integration in time, we will use in both examples the backward Euler scheme, in spite of its poor accuracy. Our conclusions apply to all the schemes we have tested, in particular the accurate fourth order scheme described in [11]. In the first example we will also present results for the second order Crank–Nicolson scheme. Likewise, we have tried different element types (triangles and quadrilaterals, linear and quadratic), although here we will only present results for linear elements in 1D and bilinear quadrilateral elements in 2D. Again, for numerical results obtained in unstructured triangular meshes, see [11]. It will be seen that the stabilization simply removes the oscillations and the unphysical behavior of the Galerkin method, without changing the mean amplitudes and the mean phases of the propagated profiles.

Quasi-static orthogonal subscales have been used in both examples. In order to deal with the orthogonal projection of a discrete function f_h at time step $n+1$ we have computed it as $P_h^\perp(f_h)^{n+1} \approx f_h^{n+1} - P_h(f_h^n)$. This explicit treatment of the second term makes the formulation inexpensive and does not spoil stability, as we have observed from numerical experiments. This is in accordance with the stability analysis presented in [10] for the pressure stabilization of the incompressible Navier–Stokes equations.

5.1. Solitary wave

In this first example we have solved the apparently trivial problem of propagating the profile $\eta_0(x) = 0.02\sin(\pi x)$ for $0 \leqslant x \leqslant 1$, $\eta_0(x) = 0$ for $x > 1$, in a one-dimensional domain $[0, 10]$. We have taken the equation parameters $\mu_\eta = \mu_u = 1$ and the forcing terms $f_\eta = f_u = 0$, so that the profile $\eta_0(x)$ has to propagate in time with a speed $c = 1$.

In this one-dimensional problem we have used 40 linear elements to discretize the interval $[0, 10]$ and we have taken the time step size $\delta t = 0.02$, so that $c\delta t < h = 0.025$. For the values of the physical properties chosen, the stabilization parameters in (43) simply become $\tau_\eta = \tau_u = Ch$. Obviously, the case $C = 0$ corresponds to the Galerkin method.

We have experimentally observed that for $C \geqslant C_0$ results are virtually identical for a wide range of values of the constant C. In this one-dimensional example, the solution is perfectly stable for $C \geqslant C_0 = 0.005$. Fig. 1 (top) shows the results obtained for $C = 0.01$ and $C = 1$ using the backward Euler method. They are almost identical and perfectly smooth and stable. Note that no special

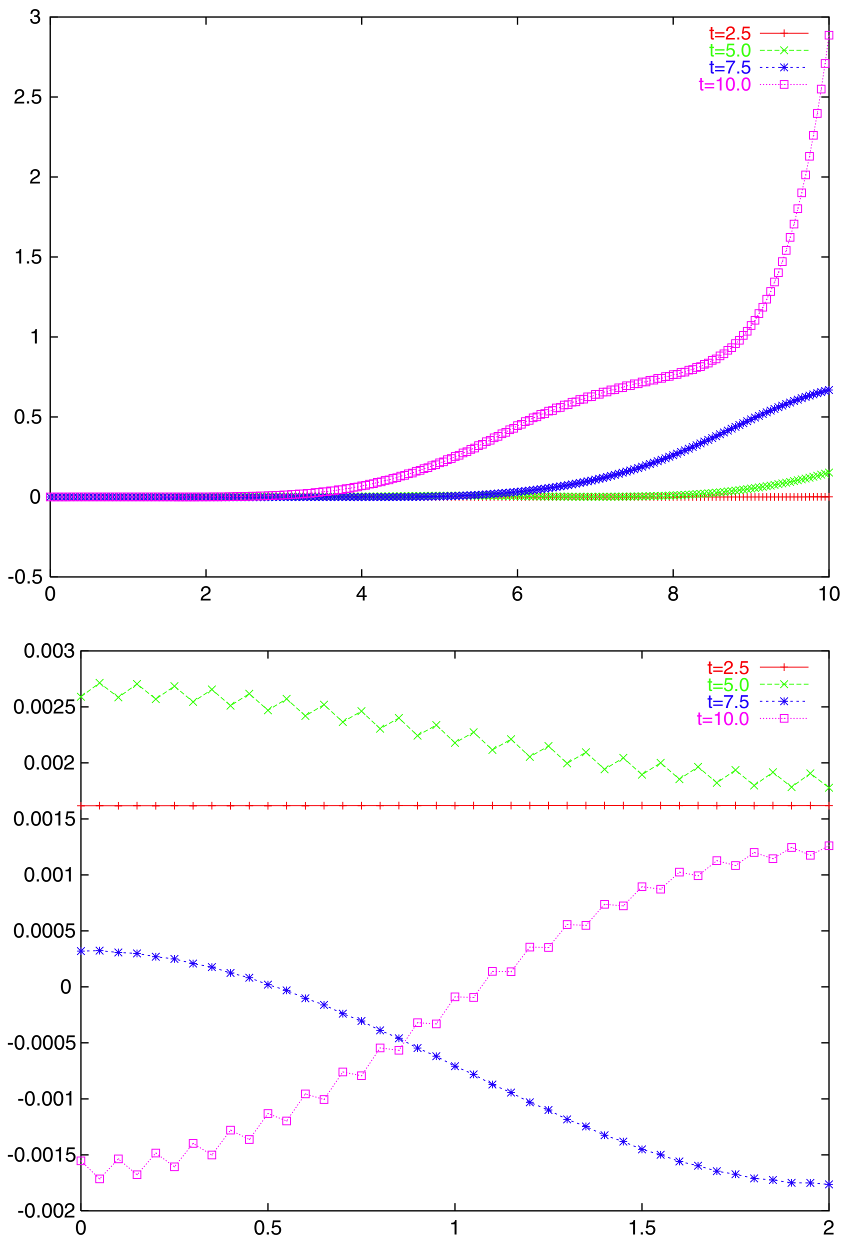

Fig. 7. Evolution of the $y = 1.5$ section (top) and the $x = 2$ section (bottom) of η for the Galerkin solution of the wave propagation over an obstacle.

treatment of the boundary $x = 10$ has been used. Spurious boundary effects begin to be noticeable for $C = 10^3$.

For $C < 0.005$ results are unstable. In this case, the instability is manifested by a spurious effect of the boundary that is more pronounced the smaller the value of C. Fig. 1 (bottom) shows the results for $C = 0.001$ and $C = 0$ (Galerkin method), also using the backward Euler scheme.

In order to see in more detail which is the temporal evolution of the solution we have plotted in Fig. 2 the solution for $C = 0.001$ and $C = 0.005$ (top) and shown how the former is increasingly affected in time by the above mentioned spurious effect of the boundary layer (bottom). In this case we have used the second order Crank–Nicolson scheme to integrate in time. As expected, the stable case ($C = 0.005$) has some overshoots and undershoots, typical of the Crank–Nicolson method.

5.2. Wave propagation over an obstacle

In this second example we have computed the propagation of the same profile as before, namely $\eta_0(x) = 0.02\sin(\pi x)$ for $0 \leqslant x \leqslant 1$, $\eta_0(x) = 0$ for $x > 1$, in the rectangle $[0, 10] \times [0, 2]$ with an obstacle represented by the rectangle $[4, 6] \times [0, 1]$. This domain has been meshed with a uniform mesh of 200×40 bilinear elements ($h = 0.05$). Once the obstacle is excluded, the final number of elements is 7200.

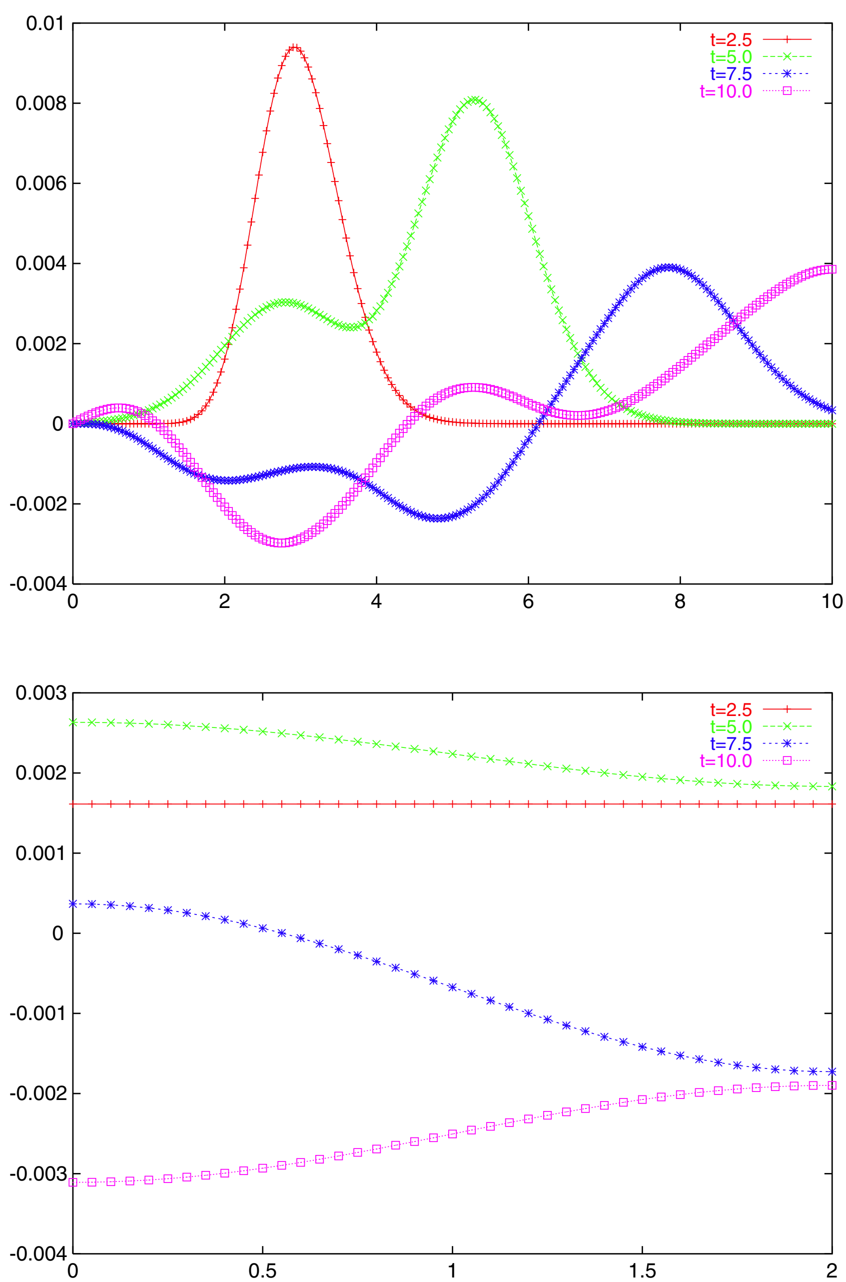

Fig. 8. Evolution of the $y = 1.5$ section (top) and the $x = 2$ section (bottom) of η for the stabilized solution of the wave propagation over an obstacle.

Similar conclusions to those drawn in the previous example apply. In particular, the Galerkin solution turns out to be unstable, instabilities being generated mainly on the right boundary $x = 10$. However, some oscillations in the solution can also be observed in the interior of the computational domain. Contour lines of the scalar field η obtained using the backward Euler method are shown in Fig. 3.

As in the one-dimensional example, the solution obtained for values of the stability constant larger than 0.005 are perfectly stable. The evolution of the scalar field η is plotted in Fig. 4 and Fig. 5. Results are completely smooth and free of any spurious boundary effect. This is also so for the vector field $\boldsymbol{u}$, some details of which are plotted in Fig. 6. Note that once again no special treatment of the boundary $x = 10$ has been used.

Finally, in order to have a better understanding of the evolution of η in time we have plotted it at different time instants along sections $y = 1.5$ and $x = 2$ in Fig. 7 for the Galerkin method and in Fig. 8 using the stabilized formulation proposed. It is clearly observed how the instabilities of the Galerkin case appear mainly on the boundary, although results are also oscillating in the interior of the computational domain.

6. Conclusions

As for all mixed problems, the wave equation requires a compatibility condition for the interpolation of the unknowns when it is written in mixed form. In this paper we have presented a particular and original point of view to describe the need for such a compatibility condition and we have proposed a stabilized finite element method to avoid the need to comply with it.

Our formulation is based on the variational multiscale formalism. Even though we have used a space of subscales orthogonal to the finite element space, our conclusions also apply to the more classical approach of taking the subscales directly proportional to the residual of the finite element solution. This alternative displays similar numerical properties, yielding fully stable numerical schemes. However, the stability analysis presented in this paper for orthogonal subscales is slightly more involved when the orthogonality does not hold.

The method depends on a matrix of stabilization parameters. A Fourier analysis has allowed us to formally justify that it is possible to take this matrix diagonal, and also to obtain an expression for the stabilization parameters entering the formulation.

Even though the main ideas of the formulation rely on previous work, several aspects are novel contributions of this paper. Apart from the extension of the orthogonal-subscale and dynamic-subscale concepts to the mixed wave equation, the heuristic Fourier analysis applied to systems is new, as well as the discussion on the proper scaling it requires. The validity of the formulation is justified from the numerical examples and its stability analysis.

The resulting formulation certainly possesses the stabilization properties is was designed for. This can be concluded both from the stability analysis performed in a simple setting and also from the numerical examples presented here. In other much more complex situations the formulation also behaves perfectly well, from the point of view of stability. We believe that the objective of developing a stable finite element method for the wave equation written in mixed form has been fully accomplished.

References

[1] M.B. Abbot, H.M. Petersen, O. Skovgaard, On the numerical modelling of short waves in shallow water, J. Hydraul. Res. 16 (1978) 173–203.
[2] E. Bécache, P. Joly, C. Tsogka, An analysis of new mixed finite elements for the approximation of wave propagation problems, SIAM J. Numer. Anal. 37 (4) (2000) 1053–1084.
[3] E. Bécache, P. Joly, C. Tsogka, Fictitious domains, mixed finite elements and perfectly matched layers for 2-D elastic wave propagation, J. Comput. Acoust. 9 (3) (2001) 1175–1201.
[4] E. Bécache, P. Joly, C. Tsogka, A new family of mixed finite elements for the linear elastodynamic problem, SIAM J. Numer. Anal. 39 (6) (2002) 2109–2132.
[5] F. Brezzi, M. Fortin, Mixed and Hybrid Finite Element Methods, Springer Verlag, 1991.
[6] R. Codina, Comparison of some finite element methods for solving the diffusion–convection-reaction equation, Comput. Methods Appl. Mech. Engrg. 156 (1998) 185–210.
[7] R. Codina, Stabilization of incompressibility and convection through orthogonal sub-scales in finite element methods, Comput. Methods Appl. Mech. Engrg. 190 (2000) 1579–1599.
[8] R. Codina, Stabilized finite element approximation of transient incompressible flows using orthogonal subscales, Comput. Methods Appl. Mech. Engrg. 191 (2002) 4295–4321.
[9] R. Codina, Analysis of a stabilized finite element approximation of the Oseen equations using orthogonal subscales, Applied Numerical Mathematics, to appear, doi:10.1016/j.apnum.2006.11.011.
[10] R. Codina, J. Blasco, Stabilized finite element method for the transient Navier–Stokes equations based on a pressure gradient projection, Comput. Methods Appl. Mech. Engrg. 182 (2000) 277–300.
[11] R. Codina, J.M. González-Ondina, G. Díaz-Hernández, J. Principe, Finite element approximation of the modified Boussinesq equations using a stabilized formulation. Int. J. Numer. Methods Fluid, to appear, doi:10.1002/fld.1718.
[12] R. Codina, J. Principe, Dynamic subscales in the finite element approximation of thermally coupled incompressible flows, Int. J. Numer. Methods Fluid 54 (2007) 707–730.
[13] R. Codina, J. Principe, O. Guasch, S. Badia, Time dependent subscales in the stabilized finite element approximation of incompressible flow problems, Comput. Methods Appl. Mech. Engrg. 196 (2007) 2413–2430.
[14] A. Ern, J.-L. Guermond, Theory and Practice of Finite Element, Springer-Verlag, 2004.
[15] T.J.R. Hughes, The Finite Element Method, Prentice-Hall, 1987.
[16] T.J.R. Hughes, Multiscale phenomena: Green's function, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized formulations, Comput. Methods Appl. Mech. Engrg. 127 (1995) 387–401.
[17] T.J.R. Hughes, G.R. Feijóo, L. Mazzei, J.B. Quincy, The variational multiscale method – a paradigm for computational mechanics, Comput. Methods Appl. Mech. Engrg. 166 (1998) 3–24.
[18] H.P. Langtangen, G. Pedersen, Computational models for weakly dispersive nonlinear water waves, Comput. Methods Appl. Mech. Engrg. 160 (1998) 337–358.

[19] D.Y. Le Roux, G.F. Carey, Stability/dispersion analysis of the discontinuous Galerkin linearized shallow water system, Int. J. Numer. Methods Fluid 48 (2005) 325–347.
[20] Y.S. Li, S.-X. Liu, Y.-X. Yu, G.-Z. Lai, Numerical modeling of the Boussinesq equations by finite element method, Coastal Engrg. 37 (1999) 97–122.
[21] M. Walkley, M. Berzins, A finite element method for the two-dimensional extended Boussinesq equations, Int. J. Numer. Methods Fluid 39 (2002) 865–885.
[22] G. Wei, J.T. Kirby, Time dependent numerical code for extended Boussinesq equations, J. Waterway, Port, Coastal Ocean Engrg. 121 (1995) 251–261.
[23] S.-B. Woo, P.L.-F. Liu, Finite element model for modified Boussinesq equations. I: Model development, J. Waterway, Port, Coastal, Ocean Engrg. 130 (2004) 1–16.